VEHICLE MILES TRAVELLED AND THE BUILT ENVIRONMENT: EVIDENCE FROM VEHICLE SAFETY INSPECTION DATA IN THE BOSTON METROPOLITAN AREA

Mi Diao. diaomi@mit.edu, Massachusetts Institute of Technology[1]

Joseph Ferreira, Jr. jf@mit.edu, Massachusetts Institute of Technology

[1]Corresponding Author

ABSTRACT

This study examines the relationship between the built environment and vehicle miles travelled (VMT) by taking advantage of a unique dataset of millions of odometer readings from annual vehicle safety inspections for all private passenger vehicles registered in the Boston Metropolitan Area, Massachusetts, U.S.A. With advanced GIS techniques and database management tools, VMT and a set of comprehensive built-environment measures are computed for a statewide 250m*250m grid cell layer developed by MassGIS (the State's GIS Office). We use factor analysis to extract built-environment and demographic factors that may affect VMT, and integrate the factors into the regression models. Spatial regression techniques are applied to correct spatial autocorrelation. The empirical results suggest that there are significant associations between built environment factors and household vehicle usage. In particular, distance to non-work destination, connectivity, accessibility to transit and jobs play significant roles in explaining the VMT variations across grid cells. The research findings can facilitate the dialogue among regional planning agencies, local government and the public regarding growth management and sustainable regional development strategies and scenarios.

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

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 causing increasing concerns about the sustainability of our world. 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 about two thirds of the total transportation energy use (IEA 2004). As an important source of GHG emissions, transportation plays a critical role in the global efforts to achieve sustainable development. Multiple strategies to reduce transportation energy use and emissions are currently explored, such as fuel-efficient vehicles, financial (dis)incentives, and various smart growth policies. Among these policy options, smart growth policies invite special interests due to their financial and political feasibility.

Central to smart growth strategies is leveraging the interconnection between the built environment and travel behavior to reduce travel demand. The built environment comprises urban design, land use, and the transportation system, and encompasses patterns of human activity within the physical environment (Handy et al. 2002). Smart growth policies try to reshape household travel behavior by changing the built environment via such mechanisms as regional planning, zoning, provisions of alternative transportation nodes, and so forth.

The relationship between transportation and the built environment has long been studied and is recognized as complex, as reviewed in Handy (1996), Boarnet and Crane (2000), Crane (2000), Ewing and Cervero (2001), and Frank and Engelke (2001). There continues to be debates about whether the relationship is “strong” or “weak” (Krizek 2005). Trip-based survey data (for sampled individuals and households) are the preferred instrument for empirical analysis of travel behavior since the unit of analysis, an individual trip, can be readily associated with the mode availability, travel cost, demographic factors, and built environment measures. However, the high expense of individual travel surveys tends to limit sample sizes 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 neighborhood level.

Another line of research characterizes both the built environment and travel using aggregate measures. Newman and Kenworthy (1999) analyze the relationship between density and energy use for an international sample of cities and find significant negative correlation between density and energy use. However, besides the fundamental problem of comparing places with different cultural, political and historical contexts, this study is also criticized by its use of simple measures of urban form and travel (Handy 1996). Holtzclaw (1994) uses odometer reading data from biennial auto emission inspections to derive estimates of total travel for twenty eight zip code zones in California and relate it to built environment measures. The result shows that annual vehicle miles travelled is significantly associated with neighborhood density. 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. The study finds zonal VMT per worker to increase with increasing distance from the CBD, and/or other major employment zones within the urban area. Holtzclaw et. al. (2002) use socio-demographic variables to control for population difference across different zones and find that auto ownership and mileage per car are functions of neighborhood urban design and socio-economic characteristics in the Chicago, Los Angles, and San Francisco regions.

The aggregate approach has provided promising evidence of the potential effectiveness of smart growth policies in reducing travel demand (Handy 1996). However as many researchers have suggested, this approach also has significant shortcomings: (1) It does not allow for an exploration of underlying factors and the mechanisms by which the built environment influences individual decisions; (2) The zones used in previous aggregate studies are usually very large in size. For example, Holzclaw et al. (2002) use zip code as their unit of analysis. At such aggregated level, the intra-zone variations of the built environment and demographic measure could be too large to ignore; (3) Previous studies either omit or include very few demographic variables in their statistical analyses, thus make limited effort to control the residential self-selection problem and construct causal relationships (Brownstone 2008); and (4) spatial autocorrelation may affect significantly but is neglected.

In this study, we take advantage of a newly-available unique dataset, the odometer readings from annual safety inspections for all private passenger vehicles registered in Metro Boston to develop an extensive and spatial detailed analysis of the built environment and household vehicle usage. Vehicle Miles Travelled (VMT) is taken as the primary variable of interest, which is a convenient measure that reduces the multi-dimensional travel demand (number of trips, the spatial distribution of these trips, the modes and routes chosen to execute these trips) to a single variable (Miller and Ibrahim 1998). The basic spatial unit for our analysis is a statewide 250m by 250m grid cell layer developed by MassGIS, the State's Office for Geographic and Environmental information. We perform multivariate regression analyses at the grid cell level to identify built-environment and demographic factors that are significantly associated with household vehicle usage. Spatial econometric techniques are applied to address the potential spatial autocorrelation.

Given the nature of our analysis, we raise two cautions at the outset. First, the objective of this study is not to project the impact of a given policy on vehicle usage, which requires a dynamic model of land use-transportation interaction (Miller and Ibrahim 1998). Our more modest objectives are to examine the spatial distribution of travel behavior within a metropolitan area, which can be seen as the outcome of this dynamic land use-transportation process, and to clarify the irreducible spatial components of household travel behavior. The second issue concerns the ecological fallacy. In particular, we focus on the aggregate spatial patterns of the relationship between the built environment and household vehicle usage, while the underlying factors and the behavior mechanisms by which the built environment influences individual decisions cannot be revealed by this study.

2. STUDY AREA AND DATA

The Boston Metropolitan Area is selected as the study area of our empirical analyses. Metro Boston exhibits a variety of built-environment characteristics, which makes it a compelling case for our study. Figure 1 maps Metro Boston and the City of Boston.

Boston Metropolitan Area

City of Boston

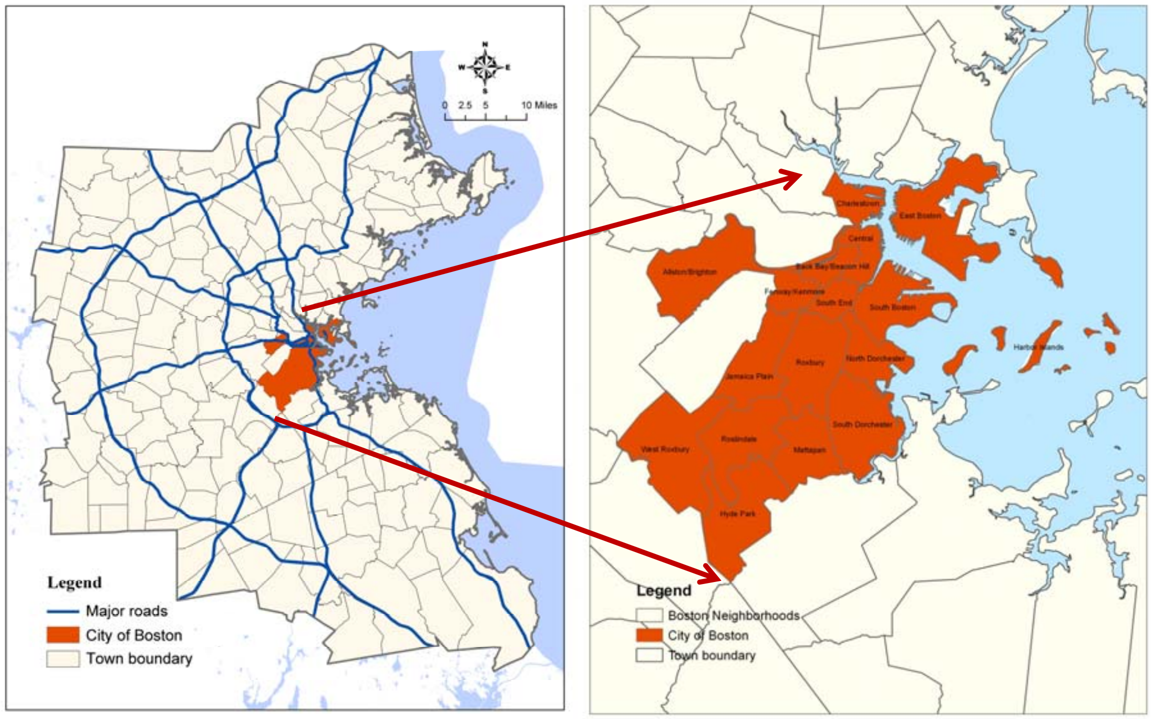


Figure 1- Study Area

This study uses a unique VMT dataset, the annual vehicle safety inspection records from the Registry of Motor Vehicles (RMV) to estimate annual mileage for every private passenger vehicle registered in Metro Boston. Safety inspection is mandated annually beginning within one week of registering a new or used vehicle. The safety inspection utilizes computing equipment that records vehicle identification number (VIN) and odometer reading and transmits these data electronically to the RMV where they can be associated with the street address of the place of residence of the vehicle owner. MassGIS gets access to the safety inspection records from the RMV for a “Climate Roadmap” project that details possible plans for significant reductions in GHG emissions by the 2020-2050 period in Massachusetts. MassGIS compares the two recent vehicle inspection records for all private passenger vehicles, calculates the odometer reading difference, and pro-rates it based upon the time period between inspection records so as to reflect the estimated annual mileage travelled. Each vehicle is then geocoded to the owner's address using GIS tools. MassGIS grants us permission to use this dataset in our study through the research collaboration between MIT and MassGIS. 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 places of garaging but don’t have reliable odometer readings, either because the reported reading was determined to be in error or because two readings sufficiently far apart were not available.

While this dataset lacks individual trip details, it does provide a very high percentage sample of total passenger vehicle miles travelled. Furthermore, 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 percent greater than odometer-based VMT in urban areas. EIA suggests that odometer-based VMT should be obtained if possible (Schipper and Moorhead 2000). Holtzclaw et al. (2002) use a similar dataset in their study, odometer readings from auto emission inspections (smog check), but since California exempts new vehicles from smog checks for the first two years, their measure systematically biases VMT downwards for zones with large numbers of new vehicles (Brownstone 2008).

Our study also benefits from built-environment data with exceptional spatial detail, which are mainly from MassGIS. MassGIS utilized Dun & Bradstreet business location database to locate household non-work destinations, and geocoded these businesses to a point layer. Institutional destinations such as schools, hospitals and parks exist as independent data layers developed and maintained by MassGIS. Road inventory database with detailed information on road networks in the region is from the Massachusetts Department of Transportation. Population and household data are generated from the 2000 Census and constrained by MassGIS to those areas identified as residential by the 2000 land use dataset. Population and households are further assigned to 250m*250m grid cells.

The spatial unit used in this study is a 250m*250m grid cell layer developed by MassGIS. The grid cell approach has often been identified as a possible means of dealing with the Modifiable Areal Unit Problem (Robsen 1969). In this study, a grid cell contains an area just over 15.4 acres, which is very small to capture spatial details and neighborhood effects. Meanwhile, using the grid cell as a basic study unit, one can take advantage of powerful raster analysis tools in GIS software. For each grid cell, we define a catchment area (neighborhood) as the 3*3 nearest grid cells, compute the variable of interest for the catchment area, and assign the value to the grid cell in the middle. The 750m*750m catchment area has a size that is close to the “transportation impact area”, which is conventionally defined as a circle with a 1⁄4-mile radius, a figure that has been backed by behavioral and empirical research (Untermann 1984). The employment of catchment area also helps create a smooth surface, reducing noises in the raw data. Compared with previous research, our study is performed at a much more fine-grained scale. Table 1 compares the grid cells we use in our study and some spatial units that are widely used in transportation research for Metro Boston.

Table 1 - Comparison of Spatial Units for Metro Boston

	Grid Cell	TAZ	Block Group	Census Tract
No. of observations	119834	2727	3323	894
No. of observations with population	73714	2606	3319	894
Vehicle count for populated units				
Min	0	0	1	1
Max	3117	3022	11593	13631
Mean	32	941	744	2764
Std. Dev.	49	603	514	1514
Household count for populated units				
Min	0	0	0	0
Max	1624	2318	2211	4260
Mean	22	631	495	1839
Std. Dev.	48	391	246	713
Individual count for populated units				
Min	1	1	2	70
Max	3673	4969	6131	12051
Mean	58	1654	1297	4817
Std. Dev.	112	992	626	1825

3. METHODOLOGY

3.1 Model Specifications

In our base model, VMT is specified as a function of built-environment and demographic factors.

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

where VMT_i is the zonal average VMT per vehicle, per household or per capita for the catchment area of grid cell i; BE_i is a vector of built-environment variables of grid cell i, and DEM_i is a vector of demographic variables of the block group that grid cell i falls in.

Many previous studies suggest that built environment can influence travel behavior. This effect can be partitioned into direct influences associated with the characteristics of the neighborhood where the household locates and indirect influences associated with the travel behavior and built environment characteristics of neighboring areas. We estimate both spatial lag model and spatial error models (Anselin 1993) to capture this spatial effect. Spatial lag suggests a possible diffusion process -- VMT of one place is affected by the independent variables of this place as well as neighboring areas. With spatial lag in OLS regression, the estimation result will be biased and inefficient. Spatial error is indicative of omitted independent variables that are spatially corrected. With spatial error in 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:

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

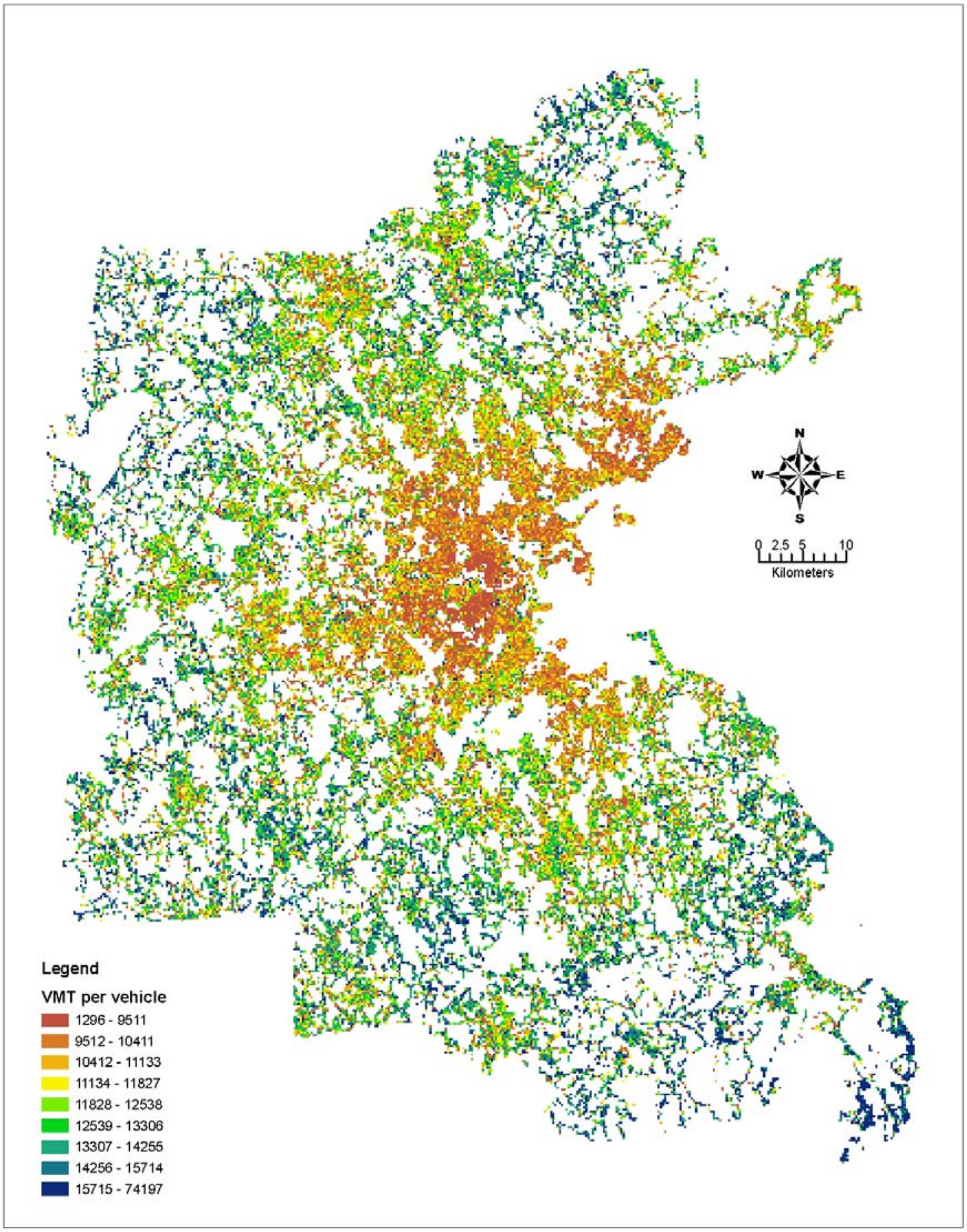
where λ is a spatial error correlation parameter, and μ is an Nx1 vector of i.i.d. standard normal errors.

In Equations (2) and (3), W is the NxN matrix of spatial weight, which is developed assuming constant spatial dependence among grid cells up to a maximum distance. The maximum Euclidean distance used is 750m. Both models can be estimated by maximum likelihood.

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. VMT per vehicle is a single indicator of individual car usage, while VMT per household and VMT per capita are also influenced by auto ownership. VMT per vehicle for each grid cell is first computed based on safety inspection records. Some grid cells have very few vehicles. To address issues related to sparse cells, we apply the spatial interpolation function of GIS software. For grid cells that have at least 12 vehicles with credible odometer readings (denoted as "good" cars), the zonal average annual mileage of all "good" cars is assigned to the grid cell. For grid cells with 1-11 "good" cars, the inverse distance weighted average of 12 closest "good" annual mileages are assigned to the grid cell. VMT per household (VMT per capita) for each grid cell is computed by multiplying VMT per vehicle by total number of vehicles then dividing by number of households (individuals). These odometer-readings-based VMT estimates provide a more accurate and reliable picture of household vehicle usage than survey-based self-report VMT estimates, establishing 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 method. The spatial pattern is what we can expect: VMT are lower in grid cells near urban centers and sub-centers, but higher in suburban areas.

2-(a) VMT per vehicle



2-(b) VMT per household

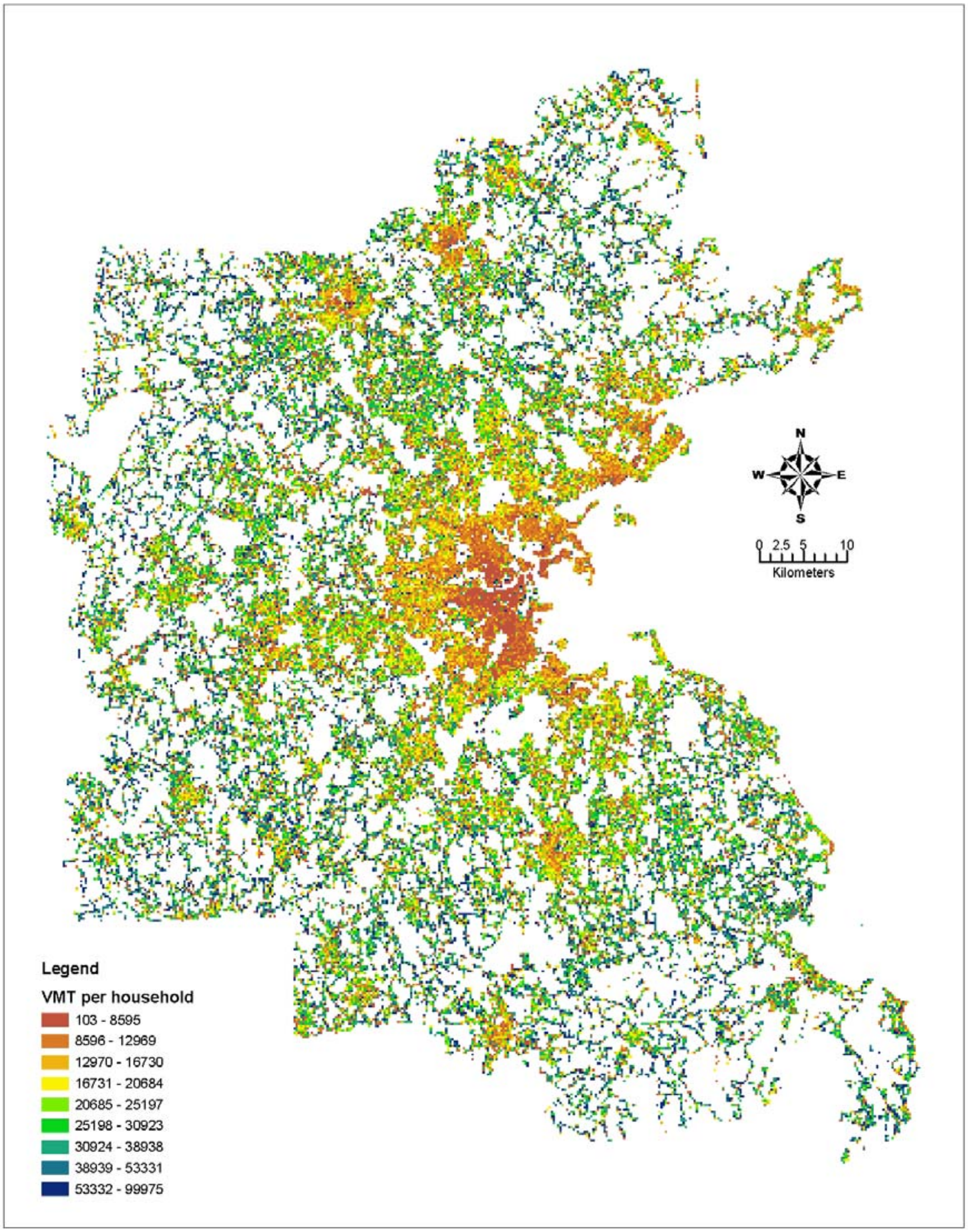


Figure 2 - Grid Cell Level VMT in Metro Boston

The dependent variables of the regression models are VMT per vehicle, VMT per household, and VMT per capita computed for the 9-grid-cell catchment area of each grid cell respectively. After excluding grid cells with extreme values, the total number of grid cells in our empirical analysis is 52929.

3.3 Built-Environment Variables

Twenty seven built-environment variables are computed in this study. Because spatial distribution of destinations can significantly influence travel costs, accessibility to common destinations is an important determinant of housing price. One gravity-type measure of job accessibility is computed at the TAZ level to represent work distance, which takes the following form known as the Hansen accessibility model (1959). Each grid cell is assigned the value of the TAZ that it belongs to.

- Job accessibility: $A_j = \sum_j O_j f(C_{ij})$, where $f(C_{ij}) = \exp(-\beta * C_{ij})$; O_j is the number of jobs in TAZ *j*; *f*(C_{ij}) is an impedance function; C_{ij} is the network distance between TAZ *i* and *j*; *β* is set to 0.1, based on Zhang's calibration using an Activity–Travel Survey conducted by the Central Transportation Planning Staff for the Boston region (2005).

MassGIS computed distances to a variety of non-work destinations at 250m*250m grid cell level using GIS tools. We select eight types of most important non-work destinations based on average trip rate from the 2001 National Household Transportation Survey, including:

- Distance to shopping mall: Euclidian distance to the nearest shopping mall
- Distance to grocery store: Euclidian distance to the nearest grocery store
- Distance to school: Euclidian distance to the nearest school
- Distance to hardware store: Euclidian distance to the nearest hardware store
- Distance to restaurant: Euclidian distance to reach at least 4 restaurants
- Distance to church: Euclidian distance to reach at least 4 churches
- Distance to dentist: Euclidian distance to reach at least 4 dentists
- Distance to gym: Euclidian distance to reach at least 4 gyms

Other built-environment variables describe density, land use mix, road networks, transit proximity, and pedestrian environment respectively. They also have the potential to affect travel costs for different travel modes. Among them, we computed distance-related variables directly for the target grid cell. We computed other measures for the 9-grid-cell catchment area and then assigned the value to the target grid cell.

- Population density: population / residential area
- Land use mix: the land use mix measure is based on the concept of entropy — a measure of variation, dispersion or diversity (Turner, Gardner and O'Neill, 2001). In the first step, it is computed for each grid cell using $-\sum_j P_j * \ln(P_j) / \ln(J)$, where P_j is the proportion of land in the *jth* land-use category and *J* is the total number of land-use categories considered. In this study, *J*=5: single family, multi-family, commercial, industrial, and recreation and open space. A value of 0 means

the land in the grid cell is exclusively dedicated to a single use, while a value of 1 suggests perfect mixing of the five land uses. Then each grid cell is assigned the average value of the nine grid cells in the catchment area.

- Intersection density: number of intersections / area
- Density of 3-way intersections: number of 3-way intersections / area
- Density of 4-way intersections: number of 4-way intersections / area
- Road density: total length of road / area
- Percent of 4-way intersections: number of 4-way intersections / number of intersections
- Percent of roads with access control: total length of road with access control / total road length
- Average road width: ∑(width of road segment * length of road segment) / total road length
- Percent of roads with over 30-mph speed limit: total length of road segment with over 30-mph speed limit / total road length
- Distance to highway exit: Euclidian distance to the nearest highway exit
- Percent of roads with curbs: total length of road segment with curbs / total road length
- Percent of roads with sidewalks: total length of road segment with sidewalks / total road length
- Average sidewalk width: ∑(sidewalk width of road segment * length of road segment) / total road length
- Distance to subway station: Euclidian distance to the nearest subway station
- Distance to commuter rail station: Euclidian distance to the nearest commuter rail station
- Distance to MBTA bus stop: Euclidian distance to the nearest MBTA bus stop
- Distance to MBTA parking lot: Euclidian distance to the nearest MBTA parking lot

GIS techniques and database management tools are extensively used in the computation of these built-environment variables.

3.4 Demographic Variables

Based on literature, thirteen demographic variables at the block group level are selected to control the zonal difference of populations. The variables included in the analysis can be found in Table 3. Ideally, we should compute demographic variables at grid cell level. Because of data limitation, each grid cell is assigned the value of the block group it belongs to. For population and household counts, block group counts were distributed among only those grid cells in residential area.

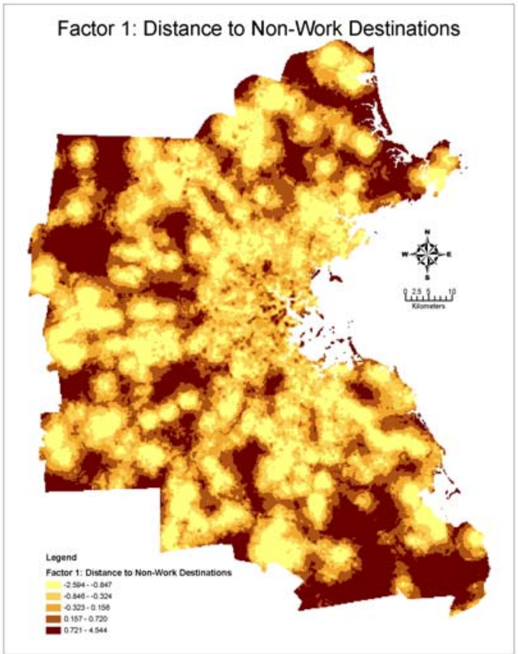
4. EMPIRICAL ANALYSIS

4.1 Factors Analysis

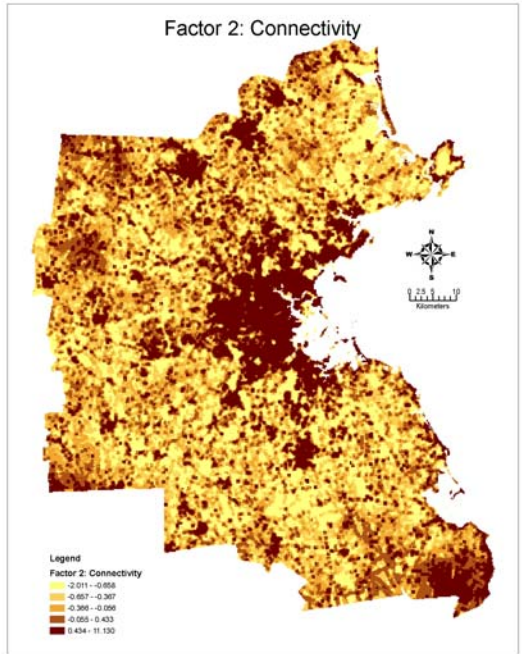
Due to the multi-dimensional nature of the built environment, one central issue in studies of the built environment is the selection of relevant variables from a large set of potentially important variables. Furthermore, many built-environment variables tend to be closely correlated. For example, relatively 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 use factor analysis to reduce the total number of built-environment variables to a small set of factors, and include factor scores in regression models. Principle component analysis with Varimax rotation is performed. Top five factors with initial eigenvalues greater than 1 explain 69.84% 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.

Factor 1 has high loadings on variables for 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, 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”. Good connectivity can improve the connection of people and places and shorten local trips (Crane 1996), thereby reducing household 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. Factor 3 could be positively associated with VMT. Factor 4 leans to the traffic management side, representing the degree of auto dominance. It could decrease travel costs of the auto mode, thus increase vehicle usage. The fifth factor “walkability” describes the pedestrian environment, which can reduce the travel costs of walking, thus decreasing VMT. Figure 3 shows the spatial patterns of built-environment factors. Compared with grid cells in the suburban, grid cells in urban centers have better accessibility to non-work destinations, jobs, and transit systems, better connectivity, and better pedestrian environment as we would expect. Grid cells with higher scores in the “auto dominance” factor tend to be located along major transportation corridors.

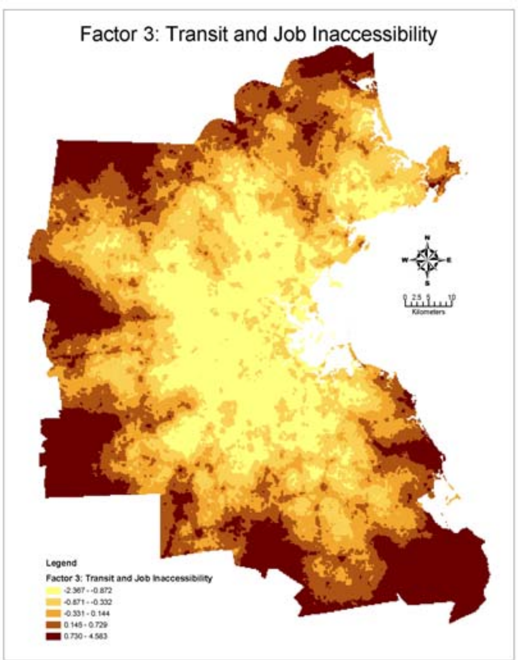
3-(a) - Distance to non-work destinations



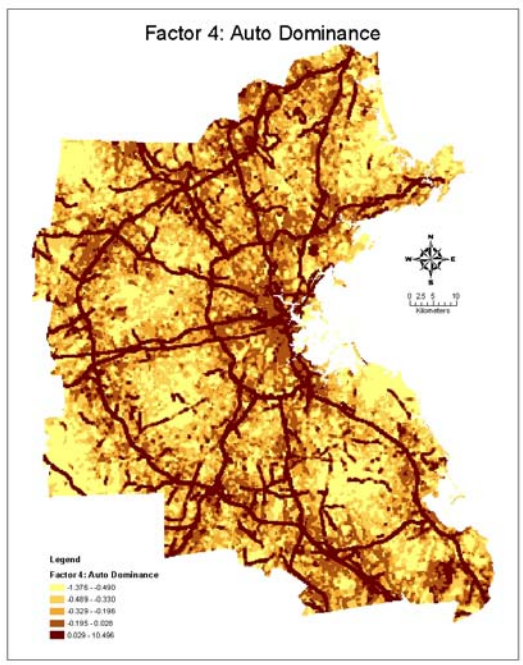
3-(b) - Connectivity



3-(c) - Accessibility to transit and jobs



3-(d) - Auto dominance



3-(e) - Walkability

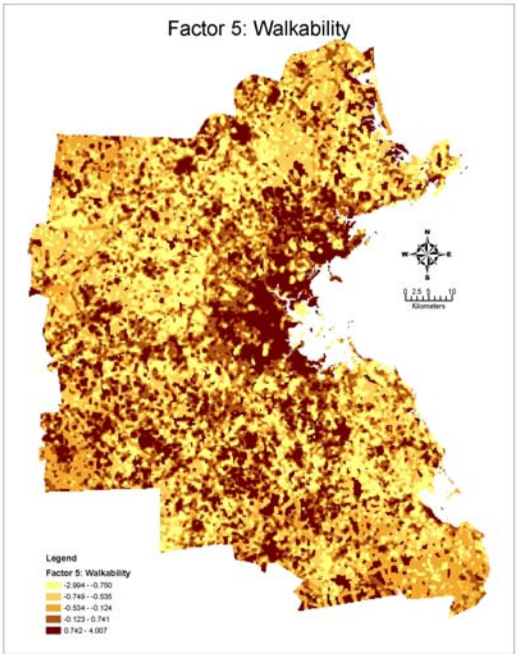


Figure 3 - Maps of Built-Environment Factors

Similarly, we also apply factor analysis to the thirteen demographic variables at block group level and extract from them three demographic factors: wealth, children, and working status. Factor 1 can be seen as an indicator of income level. 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 70.65% of variance in the original variables.

Factors loadings for each built-environment and demographic variable can be found in Table 2 and 3. Factor loadings with absolute value less than 0.35 are suppressed for interpretation convenience. Table 4 presents the descriptive statistics of variables in regression models.

TABLE 2 - Factor Loadings for Built Environment Factors

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

TABLE 3 - Factor Loadings for Demographic Factors

		Factor 1 Wealth	Factor 2 Children	Factor 3 Working Status
1	Pct. of population below poverty level	-0.858		
2	Pct. of owner-occupied housing units	0.826	0.390	
3	Pct. of population with at least 13 years of schooling	0.811		
4	Median household income	0.807		
5	Pct. of population that is white	0.802		
6	Per capita income	0.687		
7	Pct. of occupied housing units with over 2 cars	0.655	0.491	
8	Unemployment rate	-0.602		
9	Pct. of households with less than 3 members		-0.920	
10	Pct. of population 3+ yrs that are enrolled in elementary/high school		0.855	
11	Pct. of population under 5		0.694	
12	Pct. of population 65 years old and over		-0.336	-0.851
13	Pct. of population 16 years old and over in labor force	0.427		0.790

Table 4 - Descriptive Statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
VMT per vehicle	52929	12056.9	1770.8	5219.7	23843.7
VMT per household	52929	27120.6	13315.4	625.3	98954.6
VMT per capita	52929	9372.2	4204.0	85.0	50158.2
BE fac. 1: dis. to non-work destinations	52929	-0.245	0.865	-2.594	3.983
BE fac. 2: connectivity	52929	0.425	1.172	-1.644	11.130
BE fac. 3: inaccessibility to transit & jobs	52929	-0.108	0.973	-2.271	4.583
BE fac. 4: auto dominance	52929	-0.082	0.610	-1.210	6.409
BE fac. 5: walkability	52929	0.080	0.921	-2.664	4.007
DEM fac. 1: wealth	52929	0.594	0.659	-4.084	2.610
DEM fac. 2: children	52929	0.443	0.778	-3.487	3.432
DEM fac. 3: worker	52929	0.082	0.858	-6.902	3.928

4.2 Regression Results

Depending upon the selection of dependent variable and model specification, nine models are estimated in this study:

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.

The spatial lag and spatial error models are estimated with GeoDa 0.9.5 software. Table 5 summarizes statistics for the regression models.

Table 5 - 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
Obs.	52929	52929	52929	52929	52929	52929	52929	52929	52929
R-squared	0.528	0.789	0.810	0.419	0.626	0.631	0.342	0.566	0.573
Log Likelihood	-451118	-432070	-429928	-563404	-553577	-553490	-505662	-496458	-496290

Test	Statistic	p-value	Statistic	p-value	Statistic	p-value
LM-Lag	86282.7	0	43696.6	0	41125.0	0
LM-Error	115263.9	0	46204.5	0	43206.0	0
Robust LM-Lag	633.4	0	605.6	0	293.6	0
Robust LM-Error	29614.7	0	3113.5	0	2374.6	0

The R-squared of the OLS models range from 34.2% to 52.8%. Test of residues indicated that the error term of the OLS models exhibit significant spatial autocorrelation. The likely reasons are the omission of spatially correlated exploratory variables, and the effects of travel behavior in surrounding areas. 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 existence of spatial autocorrelation.

To capture the spatial effects, we estimate both spatial lag and spatial error models. Anselin et al.'s (1996) Lagrange multiplier tests of spatial lag and spatial error specifications being mutually contaminated by each other are employed to compare the two models. 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. But the robust LM-error test sees rejection of the null at the higher level of significance, favoring the spatial error model. The log-likelihood statistics also support this conclusion, indicating that the spatial error model has better fit to the data than the corresponding spatial lag model and OLS model. The spatial error models can explain 57.3-81.0% of variation in the three VMT variables across grid cells, which is higher than the roughly comparable measures for the OLS models and pretty impressive for a study at such fine-grained scale. The goodness-of-fit statistics for VMT per vehicle models are higher than those for VMT per household and VMT per capita.

Table 6 - Estimation Results of the Spatial Error Models

	VMT per Vehicle			VMT per Household			VMT per Capita		
	Coef.	t		Coef.	t		Coef.	t	
BE Fac.1: dis. to non-work destinations	446.4	21.3	**	3785.0	22.9	**	856.3	15.7	**
BE Fac.2: connectivity	-251.8	-23.4	**	-2944.9	-34.3	**	-831.3	-29.1	**
BE Fac.3: inaccessibility to transit& jobs	1005.8	32.3	**	5862.7	29.9	**	1949.8	30.8	**
BE Fac.4: auto dominance	-10.0	-1.0		587.7	6.1	**	272.0	8.3	**
BE Fac.5: walkability	14.1	1.6		-1551.1	-19.3	**	-588.7	-21.8	**
DEM Fac.1: wealth	-30.1	-2.3	*	793.5	6.0	**	297.4	6.7	**
DEM Fac.2: children	-12.3	-1.3		600.0	6.4	**	-45.5	-1.4	
DEM Fac.3: working status	31.5	4.7	**	136.0	2.0	*	60.0	2.6	**
Lambda	0.91	397.1	**	0.84	231.1	**	0.83	218.8	**
Constant	12415.1	312.6	**	30705.4	127.3	**	10447.7	133.8	**

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

Table 6 presents the estimation results of the three models using the spatial error specification. As is shown in Table 6, most coefficients for demographic factors are statistically significant. One interesting finding is that higher income are associated with lower VMT per vehicle, but higher VMT per household and VMT per capita. It suggests that high-income households tend to own more cars but drive each car less compared to low-income households. Household structure also influences vehicle usage. The number of children in the household tends to increase VMT per household, presumably because of child-related non-work trips. But its effects 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 percentage of population that are working. This factor is positively associated with all three VMT variables, presumably due to the commuting trips.

After controlling for the influence of demographic factors, we find that built-environment factors are indeed important predicators of vehicle usage at 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. It suggests that the spatial distribution of non-work activities is significantly associated with vehicle usage. As the distance to non-work destinations increase, 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 neighborhood tend to reduce household vehicle usage. The coefficients of the auto dominance factor are positive and significant in the VMT per household and VMT per capita models, while its coefficient in the VMT per vehicle model is insignificant. It suggests that an auto-friendly environment influences VMT by increasing the number of cars owned by households rather than increasing the usage of each vehicle. As revealed by the estimated coefficients of the walkability factor, good pedestrian environment are associated with lower VMT per household and VMT per capita, while its effect on VMT per vehicle is insignificant. 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 higher predication power on VMT than demographic factors. Figure 4 plots the change of annual VMT per household due to one standard deviation increase of individual factor. As we can see, accessibility to work and non-work destinations, connectivity, and transit accessibility make much higher contribution to the model than other factors.

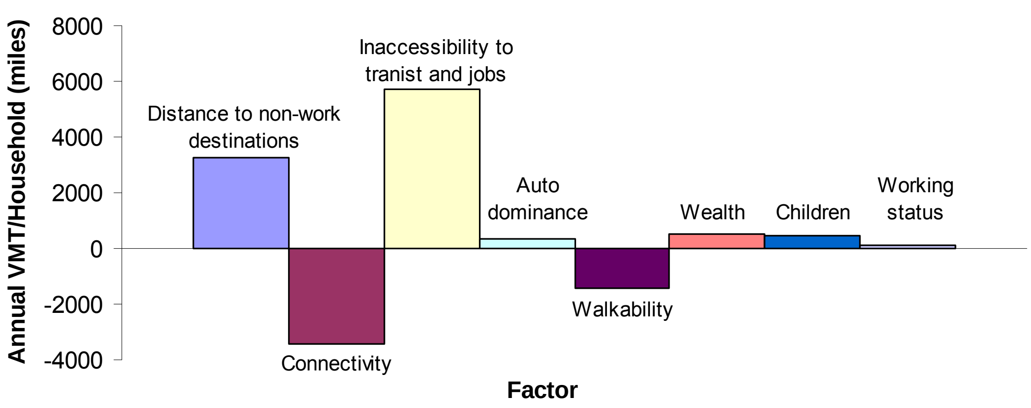


Figure 4 - Contribution of Factors to the Model

5. CONCLUSIONS

This study examines the relationship between the built environment and household vehicle miles travelled in the Boston Metropolitan Area. The VMT measures are derived using annual safety inspection records for all private passenger vehicles registered in Metro Boston. We compute a set of built-environment variables at 250mX250m grid cell level using GIS techniques, apply factor analysis to mitigate multicollinearity, and integrate the built-environment and demographic factors into regression models to explain VMT variations. Spatial regression techniques are applied to correct spatial autocorrelation.

This study provides some clues to the relationships between the built environment and vehicle usage within the Boston Metro area. The spatial error models explain 57.3-81.0% of VMT variation across 250x250m grid cells, which is pretty impressive for a study at such disaggregate scale. The regression results reveal that both the built environment and demographic factors are significantly associated with VMT. On the demographic side, this study finds that income is negatively associated with VMT per vehicles, but positively associated with VMT per household, suggesting that high-income households tend to own more cars but use each car less. Due to data limitation, the demographic variables are computed at block group level, which is more aggregate than built environment variables. Thus the results may be influenced by the Modifiable Areal Unit Problem. The built environment factors have higher impacts on VMT than demographic factors. In particular, accessibility to work and non-work destinations, connectivity, and transit accessibility are negatively associated with VMT, and their impacts are greater than other factors. Although finding a strong relationship 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), these results still provide some support for those smart growth policies that advocate for increasing accessibility to destinations, creating traditional type high density, mixed use neighborhoods, and improving transit accessibility. The research findings can facilitate the dialogue among regional planning agencies, local government and the public regarding growth management and sustainable regional development strategies and scenarios.

This study also has implications for urban modeling by revealing the opportunities brought about by new spatial data infrastructure. With the development of information technology, the amount of administrative data with location information is rapidly increasing. For example, standardized GIS data layers are becoming more common for data about road networks, parcels, and building footprint, and for transaction information, such as housing transactions, vehicle safety inspections, transit fare cards, utility records, and cell phone use. These administrative datasets are collected regularly by various agencies. Calibrating urban models using administrative data can save the high expense of surveys and enable improved monitoring and modeling of metropolitan areas.

In the future, this study can be extended along multiple directions, for example

1) examine temporal trends in land use-transportation interconnection using time series safety inspection data;

2) construct profiles of fuel economy so that the built environment can be directly linked to energy consumptions and GHG emissions.

3) employ structure equations models to investigate the causal relationships among key variables, such as the built environment, automobile ownership, and travel behavior; and

4) extend the analysis to other North American metropolitan areas.

ACKNOWLEDMENT

We acknowledge the foresight and geoprocessing skill of MassGIS in developing some of the unique datasets used in this research: the Massachusetts grid cell layers of non-work destinations and the geocoded vehicle locations, with VMT estimates, that MassGIS developed from the safety inspection records obtained from the Registry of Motor Vehicles. This research describes a portion of the PhD dissertation work of the first author. We acknowledge the partial support of the US Department of Transportation Region One University Transportation Center through grant MITR21-4 and the computational assistance of the staff and facilities of MIT and the Urban Studies and Planning Department.

REFERENCES

Anselin, L. (1993). Discrete space autoregressive models. In: Goodchild, M., B. Parks and L. Steyaert (Eds.) Environmental Modeling with GIS. Oxford University Press, New York, pp. 454-469.

Anselin, L. A. Bear, R. Florax, and M. Yoon. (1996). Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 26, pp. 77-104.

Badoe, D.A. and E.J. Miller. Transportation-land-use interaction: empirical findings in North America, and their Implications for modeling. Transportation Research Part D, 5, 2000, pp. 235-263.

Boarnet, M.G. and Crane, R. (2000). Travel by Design: The Influence of Urban Form on Travel. Oxford University Press, New York.

Brownstone, D. (2008). Key relationships between the built environment and VMT. Paper prepared for the Committee on the Relationships Among Development Patterns, Vehicle Miles Travelled, and Energy Consumption, Transportation Research Board and the Division on Engineering and Physical Sciences

Crane, R. (1996). On form versus function: will the new urbanism reduce traffic, or increase it? Journal of Planning Education and Research, 15, pp.117-126.

Crane, R. (2000). The influence of urban form on travel: an interpretive review. Journal of Planning Literature, 15(1), pp. 3-23.

Ewing, R. and R. Cervero. (2001). Travel and the built environment: a synthesis. Transportation Research Record, 1780, pp. 87-113.

Frank, L.D. and P.O. Engelke. (2001). The built environment and human activity patterns: exploring the impacts of urban form on public health. Journal of Planning Literature, 16(2), pp. 202-218.

Handy, S. (1996). Methodologies for exploring the link between urban form and travel behavior. Transportation Research D, 1(2), pp. 151-165.

Handy, S., M. Boarnet, R. Ewing R, and R.E. Killingsworth. (2002). How the Built Environment Affects Physical Activity: Views from Urban Planning. American Journal of Preventive Medicine, 23, pp.64-73.

Hansen, W. (1959). How accessibility shapes land use. J. Am. Inst. Plan. 25, pp.73–76.

Holzclaw, J. (1994). Using residential patterns and transit to decrease auto dependence and costs. Natural Resources Defense Council for California Home Energy Efficiency Rating Systems, June 1994.

Holtzclaw, J., R. Clear, H. Dittmar, D. Goldstein and P. Hass. (2002). Location efficiency: neighborhood and socio-economic characteristics determine auto ownership and use – studies in Chicago, Los Angeles and San Francisco. Transportation Planning and Technology, 25, pp. 1-27.

International Energy Agency (IEA). (2004). The IEA/SMP transport spreadsheet model, developed for the World Business Council for Sustainable Development Sustainable Mobility Project.

Krizek, K. (2005). Perspectives on accessibility and travel. In K. Krizek and D. Levinson (Ed.), Access to Destinations. Elsevier Ltd, pp. 171-193.

Miller, E.J. and A. Ibrahim. (1998). Urban form and vehicular travel: some empirical findings. Transportation Research Record: Journal of the Transportation Research Board, 1617, pp. 18-27.

Newman, P. and J. Kenworthy. (1999). Sustainability and Cities: Overcoming Automobile Dependence. Washington, DC: Island Press.

Price, L., S. de la Rue du Can, J. Sinton, E. Worrell, Z. Nan, J. Sathaye, and M. Levine. (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.

Robsen, B.T. (1969). Urban analysis: a study of city structure with special reference to Sunderland. Cambridge University Press, Cambridge.

Schipper, M. and V. Moorhead. (2000). Odometer versus self-reported estimates of vehicle miles traveled. http://www.eia.doe.gov/emeu/consumptionbriefs/transportation/vmt/vmt.html.

Turner, M., R. Gardner and R. O'Neill. (2001). Landscape ecology in theory and practice: pattern and process. Springer-Verlag, New York.

Untermann, R. (1984). Accommodating the pedestrian: adapting towns and neighborhoods for walking and bicycling. Van Nostrand Reinhold, New York.

Zhang, M. (2005). Exploring the relationship between urban form and nonwork travel through time use analysis. Landscape and Urban Planning, 73, pp.244–261.