EXPLORE REGIONAL VARIATION IN THE EFFECTS OF BUILT ENVIRONMENT ON DRIVING WITH HIGH RESOLUTION U.S. NATION-WIDE DATA

Liming Wang
Portland State University
Toulan School of Urban Studies and Planning
Portland, OR, 97222
lmwang@pdx.edu

Word Count: 4275 words + 9 table(s)/figure(s) $\times$ 250 = 6525 words

Submission Date: August 1, 2022

ABSTRCT

There have now been numerous studies on the relationship between travel behavior and built environment over the last a few decades. Prior studies have mostly focused on producing point estimates of model coefficients and ended up with a wide range of estimates for the built environment elasticity of travel behavior, including household VMT. With few exceptions, previous studies use data from a single region or a small number of regions, and thus are not able to sufficiently investigate the regional variation in built environment elasticity. A few papers have addressed the heterogeneity of elasticity among different population groups and neighborhood types (for example, *1*, *2*), but so far have paid little attention to regional variation of elasticity. In this paper, I use the 2009 U.S. National Household Travel Survey and high resolution built environment measures in the Smart Location Database to investigate the regional variation in the effect of built environment.

Keywords: Built Environment, VMT, Regional Variation of Elasticity

INTRODUCTION

There have now been numerous studies on the relationship between travel behavior and built environment over the last a few decades. Prior studies have predominately focused on producing point estimates of model coefficients, but they ended up with a wide range of estimates for built environment's effects on travel behavior, including household VMT. While more recent studies started to present the heterogeneity in this relationship, few have investigated the regional variation of the relationship.

One of the reasons for this omission is that, with very few exceptions, previous studies use data from a single region or a small number of regions, and thus are not able to sufficiently investigate the regional variation in the effects of built environment, because, until recently, it requires major effort to harmonize data, for example, travel surveys, from different regions (see, for example (*3*)). Even then, questions remain whether the measures from in data pieced together from multiple regions are comparable as they were collected at different times by different agencies with different survey instruments.

In this paper, I explore the regional variation in the effects of built environment with a unique high resolution U.S. nationwide dataset that was created by joining the 2009 U.S. National Household Travel Survey and the high resolution built environment measures from the EPA Smart Location Database. I apply fixed effects and hierarchical mixed effects models and compare the results to investigate the regional variation in the effects of built environment. I find that there is indeed substantial regional variation in the effects, so much so that there is almost no fixed effect of most built environment measures after considering the random effects. I conclude the paper with a discussion of the implications and limitations of this research.

LITERATURE REVIEW

Over the last a few decades, there has been extensive research on the elasticity of travel to built environment. Ewing and Cervero (*4*) identify and synthesize more than a hundred papers on the topic between 1990s and 2010. Aston et al. (*5*) collected 187 studies on the topic of built environment and transit use.

A primary focus of this research is the direction and magnitude of the elasticity. For example, the elasticity of VMT to density (various measures) in the research reviewed by Ewing and Cervero (*4*) ranges from 0.03 to -1.05 (Table A-1, page 282), with a synthesized elasticity of -0.04. Based the low built environment elasticity of driving, Stevens (*6*) argues that densification has limited potential as a policy tool to reduce VMT. Research of Aston et al. (*5*) shows that research design has a substantial impact on the size of estimated effect.

Most prior studies uses various measures of built environment with data from a single region or a few regions. A few studies use nationwide data but with coarse measures of built environment. For example, Cervero and Murakami (*7*) use the 2001 US National Household Travel Survey to investigate the relationship between built environments on vehicle miles traveled using the UZA (Urbanized Area) as the unit of analysis and find population and job density at the UZA level is negatively associated with the average VMT per capita of a UZA. A few more recent papers use the 2009 US National Household Travel Survey to examine the heterogeneity of travel behavior across different neighborhood types and population groups (*1*, *2*), but they ignore regional variations and/or heterogeneity. Bento et al (*8*) use 1990 NPTS to estimate pooled models of travel outcomes including driving and their models do not allow regional variation of coefficients.

Ewing et al (*3*) argue that hierarchical multi-level models would be a more appropriate model structure when the data include multiple regions, as the households from the same region likely shared many of the regional characteristics. They study the varying influences of the built environment on household travel with pooled data from 15 regions in the US. Even though they apply hierarchical linear modelling (HLM), they do not investigate the random effects of built environment due the small number of regions in their data. Ewing et al (*9*) uses hierarchical models to study travel outcomes including VMT using data from 6 regions in the US and their models only consider random intercept (no random slope considered).

Other research compares models estimated with data from a few different places. Zhang (*10*) compares mode choice models estimated with data from Boston and Hong Kong and finds that elasticity of mode choice probabilities are largely consistent between Boston and Hong Kong even though there are variations in the elasticity for some of the modes. He does not look into driving distance (VMT).

To my knowledge, there is little research that systematically investigates the regional variations in built environment elasticity and this paper aims to start to fill that gap with a high resolution dataset including all large UZAs in the U.S.

DATA

I utilize a unique data set created by joining two US nation-wide data sources that provides information including travel behavior, household social-economic characteristics, and high resolution built environment measures.

National Household Travel Survey

The 2009 National Household Travel Survey (NHTS) (*11*) is a nation-wide travel survey conducted by the Federal Highway Administration of US Department of Transportation that surveyed more than 150,000 households between 2008 and 2009. Travel diaries over 24 hours on the day of the survey, as well as survey participants' socio-demographic characteristics, are captured in the survey. For this study, I focus on total household vehicle miles traveled (VMT). Figure 1 shows a histogram of survey day household VMT. I applied the household weights in the NHTS dataset in descriptive statistics and all model estimation.

I was able to access the confidential residence Census Block Group (2010 geography) for all households in the 2009 NHTS. I used the residence block group to join the household characteristics and travel outcomes in NHTS with the Smart Location Database to create a unique nationwide dataset with rich social-economic characteristics and built environment information.

Smart Location Database

The Smart Location Database (SLD) is a US nationwide database with extensive built environment variables organized around the 5D categorization: D1 - Density, D2 - Diversity, D3 - Design, D4 - Distance to transit and D5 - Destination accessibility. It includes more than 90 built environment measures such as population density, diversity of land use, neighborhood design, destination accessibility, transit service, employment, and demographics. Most measures are available for every Census block group covering the whole United States (*12*) except for the D4 transit variables, most of which are missing for a substantial portion of the dataset. For this reason, I am limited in my choice of D4 variables in my model specifications.

Variables within the same 5D category in the SLD are generally highly correlated. For example, D1B (block group level population density) has a Pearson's correlation of 1, 0.923, 0.923,

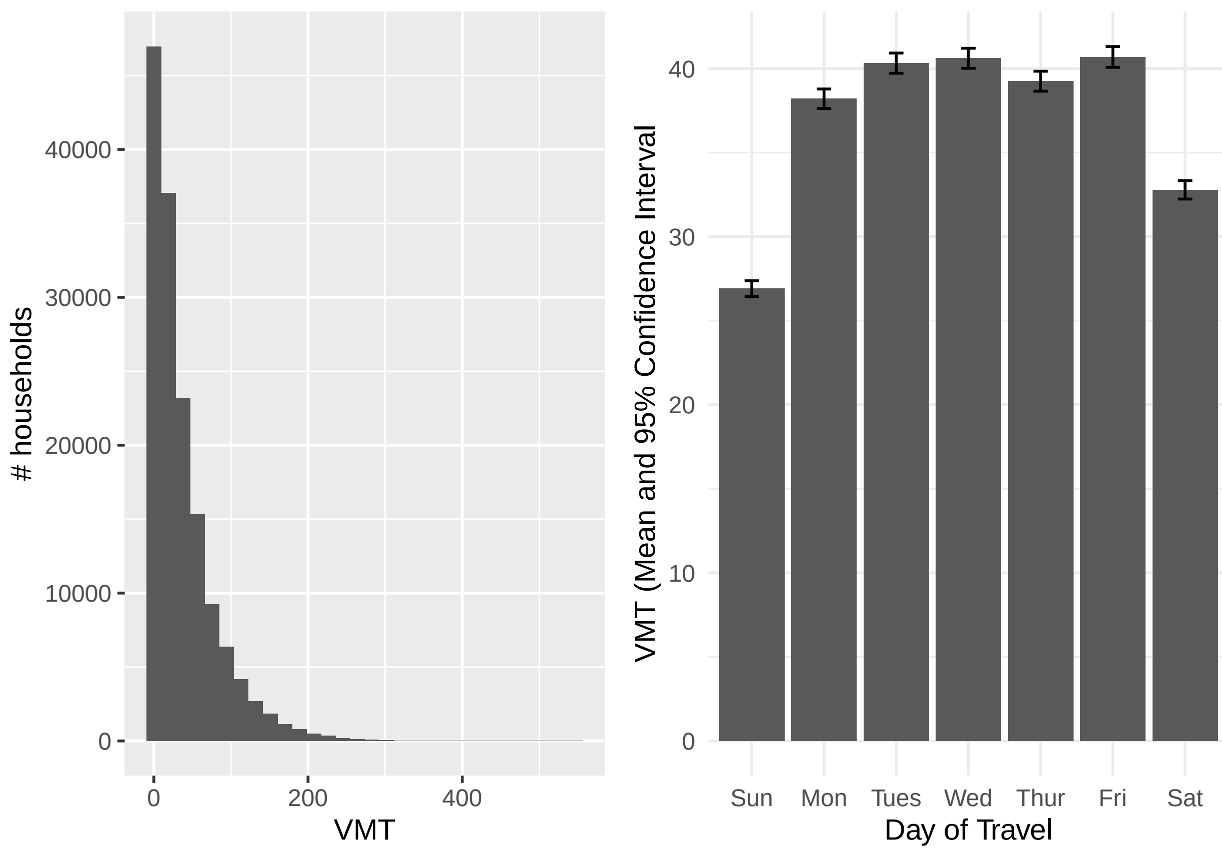


FIGURE 1: Histogram of household VMT and Mean and 95% CI of household VMT by day of travel

1 with D1A (block group level housing density) and 1, 0.565, 0.565, 1 with D1D (block group level activity density). With a few exceptions involving the D3 Design and D4 Transit variables (between D4d - Aggregate frequency of transit service per square mile and D3a - Total road network density, D3aao - Network density in terms of facility miles, D3bao - Intersection density), the correlations between variables across 5D categories are low to moderate (< 0.4). Throughout the paper I provide brief descriptions of built environment variables used in the paper, all of which are at the block group level unless noted otherwise. A complete definition of them can be found in the Smart Location Database documentation (*12*).

Some previous research has utilized the Smart Location Database. For example, Voulgaris et al(*2*) apply factor analysis to reduce the dimensionality of the many variables and cluster analysis to classify locations to a small number of types - loosely labeled as neighborhood types. The approach has the advantage of utilizing more information in the data, but it makes the interpretation, comparision, and application of the research results more difficult. For this reason, I use the original variables in the database. From each of the 5D categories, I choose at least one variable that have the best improvement to model fit while avoiding including variables that are highly correlated.

I exclude a portion of the joined dataset from this study. Households in NHTS were surveyed for a 24-hour period either on weekdays and in weekends, but since households surveyed during weekends have substantial lower VMT (Figure 1), I exclude those households in my sample (n=42979. It is unknown whether households were surveyed on non-weekend holidays as the survey date is not provided. Since I focus on total VMT of all household members, I also exclude households with incomplete member information (FLAG100=02, n=19638). Finally, to use the same sample throughout all models for comparability, I limit my sample to households living in urbanized areas with at least 100 valid household observations because my hierarchical mixed effect models use Urbanized Area (UZA) as a level in the hierarchical models (n=64576. Because there is overlap among these three criteria, I end up with 48122 households out of 150145 in this study. Table 1 shows the descriptive statistics of the variables for observations in (Included) and out of (Excluded) my sample. Finally, 3974 of these 48122 household observations contain missing values in at least one of the variables and are dropped in the model estimation process.

The households in my final sample are distributed among 100 UZAs. Figure 2 shows histograms of UZA-level household VMT and number of household observations by UZA in the final sample used in following analysis.

METHODS

There is a number of model structures applied in household VMT models. In Ewing and Cervero's review, they identify about a dozen model structures used in the literature, including linear regression, Tobit regression, hierarchical linear modeling, non-linear regression, logistic regression, probit regression, propensity score matching, copula-based switching model, simultaneous linear equations (*4*) . In this paper, I apply two-step models for household VMT: 1.) a binomial logistic regression (logit) model of whether households making zero VMT; 2.) a log-linear regression of VMT for households with non-zero VMT. This model structure is the same as what Ewing et al (*3*) use for their household VMT models and has the best predictive performance (*13*).

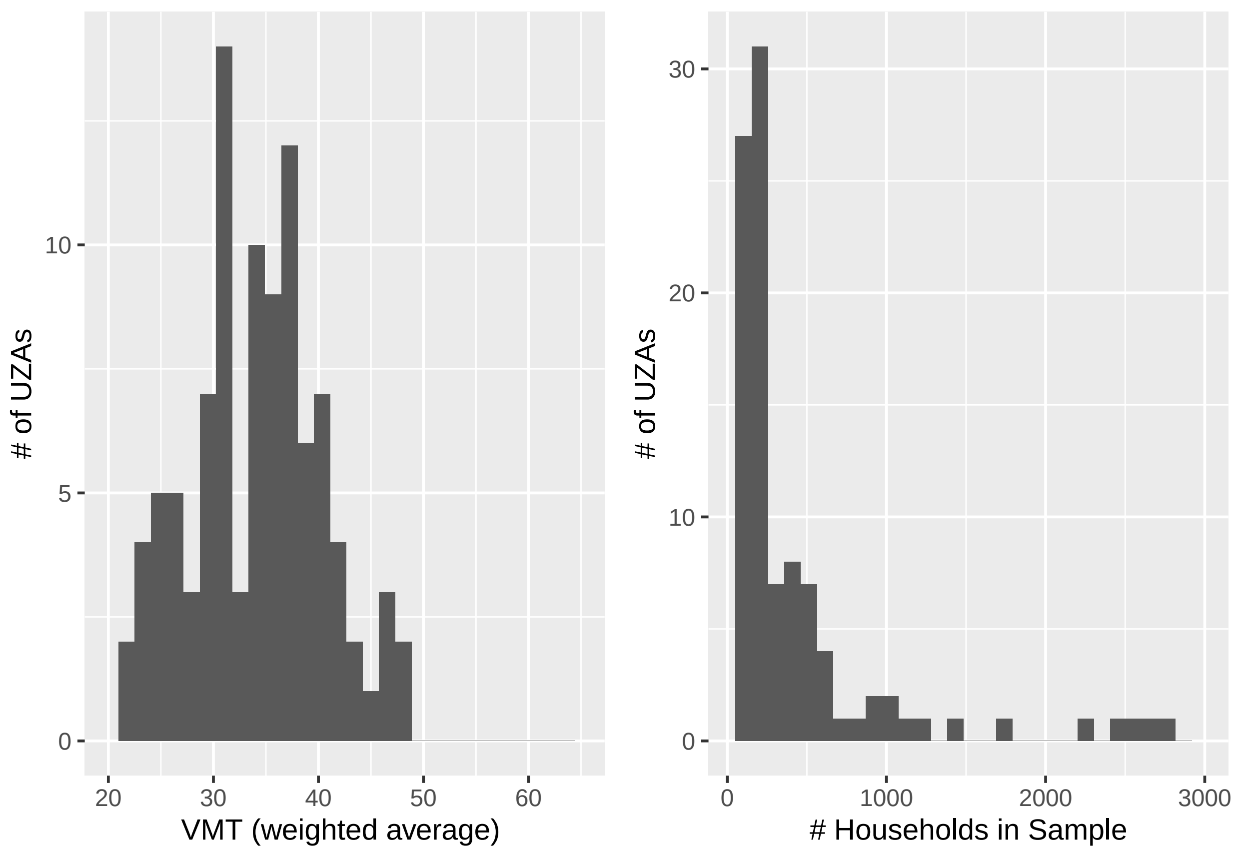


FIGURE 2: Histograms of UZA-level weighted average VMT and Number of household observations by UZA

TABLE 1: Descriptive Statistics of Main Variables

	Sample	
	Included n = 44158	Excluded n = 93778
Household VMT		
	36.7 (37.2)	38.7 (46.3)
Household Size		
	2.3 (1.2)	2.4 (1.3)
Life Cycle		
Couple w/o children	8,709 (19.7%)	20,678 (22%)
Empty Nester	18,297 (41.4%)	37,864 (40.4%)
Parents w/ children	11,608 (26.3%)	26,904 (28.7%)
Single	5,544 (12.6%)	8,332 (8.9%)
Workers		
	1.0 (0.9)	0.9 (0.9)
Family Income		
$10-30k	9,222 (20.9%)	22,410 (23.9%)
$30-50k	8,531 (19.3%)	19,318 (20.6%)
$50-70k	6,668 (15.1%)	14,687 (15.7%)
$70-100k	7,216 (16.3%)	14,576 (15.5%)
<$10k	2,318 (5.2%)	5,632 (6%)
>$100k	10,203 (23.1%)	17,155 (18.3%)
Poverty Status		
0	39,150 (88.7%)	80,805 (86.2%)
1	5,008 (11.3%)	12,973 (13.8%)
Zero Vehicle		
0	41,651 (94.3%)	90,219 (96.2%)
1	2,507 (5.7%)	3,559 (3.8%)
Vehicles per Driver		
	1.1 (0.5)	1.2 (0.6)
D1B Population density		
	9.8 (21.3)	4.4 (12.1)
D2A_WRKEMP Household workers per Job		
	11.5 (37.3)	9.4 (28.8)
D3a Total road network density		
	16.4 (7.2)	10.1 (8.1)
D4b050 Proportion of jobs within 0 5 mile of transit stop		
	0.0 (0.2)	0.0 (0.1)
D5ar1k Jobs within 45 minutes auto travel time 1000		
	138.6 (146.7)	58.9 (100.4)

Base Fixed Effects Models

For each of the modeling steps, the base specification is fixed effect models with UZA specific intercepts, in which I assume that each UZA has a different but constant effect on the response variables ($\Pr(VMT_{hu} = 0)$ and VMT_{hu}), but the effects of built environment factors on driving are the same across all UZAs.

Specifically, in the base fixed effects specification, the $\Pr(VMT_{iu} = 0)$ logit model:

$$\begin{aligned}\Pr(VMT_{iu} = 0) &= \frac{\exp(V_{iu})}{1+\exp(V_{iu})}, \text{where} \\ V_{iu} &= \alpha_u + \beta X_{iu}^{SES} + \gamma X_{iu}^{BE}\end{aligned} \tag{1}$$

and the VMT_{iu} log-linear regression model:

$$log(VMT_{iu}) \sim N(a_u + bX_{iu}^{SES} + cX_{iu}^{BE}, \sigma) \text{for} VMT_{iu} > 0 \tag{2}$$

VMT_{iu} is the vehicle miles traveled by household i living in UZA u. α_u, β, and γ in the $\Pr(VMT_{iu} = 0)$ logit model (Equation (1)) are model coefficients to be estimated for UZA specific fixed effects, social economic status variables (X_{iu}^{SES}), and the built environment variables (X_{iu}^{BE}), respectively; while a_u, b, and c are their counterparts in the VMT_{iu} log-linear regression model (Equation (2)).

Full Fixed Effect Models

In addition to the base fixed effect models with UZA specific intercepts, I estimate another pair of fixed effect models in which I allow the slopes for the built environment variables (X_{iu}^{BE}) to vary by UZA. Mathematically, in the full models, Equation (1) becomes

$$\begin{aligned}\Pr(VMT_{iu} = 0) &= \frac{\exp(V_{iu})}{1+\exp(V_{iu})}, \text{where} \\ V_{iu} &= \alpha_u + \beta X_{iu}^{SES} + \gamma_u X_{iu}^{BE}\end{aligned} \tag{3}$$

and, similarly, Equation (2) turns into

$$\log(VMT_{iu}) \sim N(a_u + bX_{iu}^{SES} + c_u X_{iu}^{BE}, \sigma) \text{for } VMT_{iu} > 0 \tag{4}$$

In Equations (3) and (4), I estimate UZA specific coefficients for built environment variables X_{iu}^{BE}. That is, I have a coefficient for each built environment variable for each UZA, representing possible varying effects of built environment factors across UZA.

While the large sample size of my data allows me to estimate the full fixed effects models with hundreds of parameters, the specification segments the sample by UZA for the built environment variables and leaves a small number of observations usable to estimate the fixed effects for many of the UZAs (Figure 2). This leads to coefficient estimates with large standard errors and statistically insignificant results.

Another problem with the fixed effects models is that the error term is assumed to be independent from each other, which is likely violated as observations from the same group are likely sharing some unobserved factors contributing to their error term, and thus have more correlated errors than those from different groups. In the case of modeling household VMT using the NHTS data, households from the same Urbanized Areas (UZA) share the same regional factors, such as the weather, transportation infrastructure, terrain, etc, that can only be partially captured in the observed data. This likely leads to the violation of model assumptions for fixed effects models.

Mixed Effect Models

Mixed effects models address both limitations of fixed effect models by partially pooling observations from different groups and by including random coefficients that absorb the unobserved group characteristics shared by observations from a group (*14*).

Similar to the fixed effects models, I specify a pair of base and full mixed effects models that allow random effects for the intercepts and built environment factors, respectively. I assume the coefficients for household social-economics characteristics, β and b, are fixed across UZAs, while the built environment measures have random effects that vary from UZA to UZA in addition to mean fixed effects. Comparing the fixed effect models and the mixed effect models allow me to test the hypothesis that there is regional variations in the effects of built environment on driving.

Base mixed effects models

The base mixed effects models only incorporate random effects for the intercept, while all other independent variables have only fixed effects. Mathematically,

$$\begin{aligned} \Pr(VMT_{iu}=0) &= \frac{\exp(V_{iu})}{1+\exp(V_{iu})}, \text{where} \\ V_{iu} &= \alpha_{iu} + \beta X_{iu}^{SES} + \gamma X_{iu}^{BE}, \\ \alpha_{iu} &\sim N(\bar{\alpha}, \sigma_\alpha) \end{aligned} \tag{5}$$

$$VMT_{iu} \sim N(a_{iu} + bX_{iu}^{SES} + cX_{iu}^{BE}, \sigma^2) a_{iu} \sim N(\bar{a}_{iu}, \sigma_a^2) \tag{6}$$

Full mixed effects models:

In the full mixed effects models, in addition to the intercept, random effects are also included for built environment variables:

$$\begin{aligned} V_{iu} &= \alpha_{iu} + \beta X_{iu}^{SES} + \gamma_{iu} X_{iu}^{BE}, \text{where} \\ \alpha_{iu} &\sim N(\bar{\alpha}, \sigma_\alpha), \text{and} \\ \gamma_{iu} &\sim N(\bar{\gamma}, \Sigma_\gamma) \end{aligned} \tag{7}$$

$$\begin{aligned} VMT_{iu} &\sim N(a_{iu} + bX_{iu}^{SES} + cX_{iu}^{BE}, \sigma), \text{where} \\ a_{iu} &\sim N(\bar{a}, \sigma_a), \text{and} \\ c_{iu} &\sim N(\bar{c}, \Sigma_c) \end{aligned} \tag{8}$$

Model Specification

The `glm` function in the base R (*15*) and the `glmer` function from the `lme4` package for R (*16*) is used to estimate the fixed effect models and the mixed effect models, respectively.

In the model selection process, I first control for social-economics status of the household. Most of the "usual suspects" that have been reported in the literature to affect travel behavior are included: household size, life cycle, income, number of workers, vehicles in the household. I also included household's poverty status that are derived from income, household size, and Department of Health and Human Services (HHS)'s poverty guidelines.

I then use a forward model selection process to select at least one variables from each of the 5D categories, while avoiding including new variables that are highly correlated with variables already in the model. I keep the same specification throughout all models to maintain comparability of results.

RESULTS

Fixed Effects Models

Table 2 shows the estimation results of the base and full fixed effects models for the ZeroDVMT logit step and DVMT log-linear step. All coefficients in these models have expected signs and, as expected, the built environment measures have significant coefficients that are small in magnitude in comparison with the social economic status variables.

UZA specific intercepts and coefficients in the full fixed effects models are not shown in Table 2 for space reasons. Most of these UZA specific coefficients are not significant, due to the small number of observations for many UZAs.

Figure 3 shows the point estimates and confidence intervals for each UZA specific coefficient in the full fixed effects logit model. The solid line represents the point estimate, while the horizontal grey bar shows the confidential interval. It is apparent in Figure 3 that the point estimates cover a wide range and these estimates for most UZAs are not statistically significant (as showing by their confidence interval intersecting value 0).

Figure 4 shows the point estimates and confidence intervals for each UZA specific coefficient in the full fixed effect log-linear model. Similar to the results for the logit model, the point estimates in the full fixed effects log-linear model also cover a wide range, with the estimates for most UZAs are statistically insignificant.

I then compare the full models against the base models with log-likelihood ratio tests as they are nested specifications (*17*). For the ZeroDVMT logit model, a log-likelihood ratio test indicates that the full model performs better than the base model (p=0, 0, χ^2=1038.188 with 430 degrees of freedom); likewise, for the DVMT log-linear model, the full model performs better than the base model (p=0, 0, χ^2=1620.883 with 430 degrees of freedom). These tests imply that there are indeed regional variations in the effects of built environment.

Mixed Effects Models

For the mixed effect models, the fixed effects coefficients and the variance-covariance matrices for the random effects coefficients are reported in Table 3. For the base models, the model coefficients in the mixed effects model is almost identical with those in the fixed effects models.

However, there is substantial difference in the results for the full models. Note that, except for the coefficient for D2A_WRKEMP and D4b050 in the Zero VMT logit model (even for them, the significance drops), all other built environment measures no longer have significant fixed effects coefficients. That is, after accounting for slopes that vary from UZA to UZA, there is no average effects to speak of for most built environment variables. This is another indication of substantial regional variation in the effects of built environment.

TABLE 3: Estimation Results of Mixed Effects Models

	Base logit	Full logit	Base log-linear	Full log-linear
Intercept	-0.2429	-0.0853	3.4296^{***}	3.4860^{*}
	(0.1250)	(0.1844)	(0.0439)	(1.3636)
Income				

	Base logit	Full logit	Base log-linear	Full log-linear
$10-30k	-0.2953^{***}	-0.3009^{***}	-0.0518	-0.0259
	(0.0814)	(0.0840)	(0.0383)	(0.0380)
$30-50k	-0.7830^{***}	-0.7877^{***}	0.0362	0.0503
	(0.1015)	(0.1047)	(0.0425)	(0.0421)
$50-70k	-1.0311^{***}	-1.1141^{***}	0.1149^{**}	0.1280^{**}
	(0.1084)	(0.1125)	(0.0425)	(0.0422)
$70-100k	-1.0488^{***}	-1.0659^{***}	0.2199^{***}	0.2370^{***}
	(0.1127)	(0.1160)	(0.0423)	(0.0420)
>$100k	-1.0216^{***}	-1.0813^{***}	0.2416^{***}	0.2523^{***}
	(0.1112)	(0.1150)	(0.0421)	(0.0419)
poverty	0.3608^{***}	0.3677^{***}	-0.1971^{***}	-0.1867^{***}
	(0.0683)	(0.0699)	(0.0257)	(0.0251)
HhSize	-0.1933^{***}	-0.2030^{***}	0.0555^{***}	0.0564^{***}
	(0.0182)	(0.0188)	(0.0039)	(0.0039)
Workers	-0.6664^{***}	-0.6559^{***}	0.1929^{***}	0.1918^{***}
	(0.0307)	(0.0315)	(0.0049)	(0.0049)
ZeroVeh	3.5207^{***}	3.7586^{***}		
	(0.1055)	(0.1093)		
VehPerDriver	-0.7323^{***}	-0.6267^{***}	0.1011^{***}	0.0995^{***}
	(0.0602)	(0.0604)	(0.0078)	(0.0079)
Life Cycle				
Single			-0.3374^{***}	-0.3228^{***}
			(0.0164)	(0.0162)
Parents w/ children			0.0170	0.0220^{*}
			(0.0109)	(0.0108)
Empty Nester			-0.1245^{***}	-0.1361^{***}
			(0.0138)	(0.0136)
D1B	0.0037^{***}	0.0007	-0.0026^{***}	-0.0109
	(0.0007)	(0.0202)	(0.0004)	(0.1908)
D2A_WRKEMP	0.0014^{***}	-0.0052	0.0006^{***}	0.0008
	(0.0004)	(0.0034)	(0.0001)	(0.0343)
D3a	0.0126^{***}	0.0053	-0.0099^{***}	-0.0029
	(0.0028)	(0.0123)	(0.0006)	(0.0989)
D4b050	0.9702^{***}	0.6684^{**}	-0.3174^{***}	-0.2360
	(0.0727)	(0.2318)	(0.0243)	(2.5493)
D5ar1k	0.0011^{***}	0.0004	-0.0001^{***}	-0.0073
	(0.0002)	(0.0030)	(0.0000)	(0.0509)
AIC	15984.2240	15972.9718	402503.4709	405449.4779
BIC	16132.0441	16294.6980	402674.0840	405790.7041
Log Likelihood	-7975.1120	-7949.4859	-201231.7355	-202684.7390
Num. obs.	44148	44148	37446	37446
Num. groups: UZA	100	100	100	100
Var: Intercept	0.0594	1.6532	0.0000	185.4059
Var: D1B		0.0323		3.6338

	Base logit	Full logit	Base log-linear	Full log-linear
Var: D2A				
$_{W}RKEMP$		0.0009		0.1174
Var: D3a		0.0128		0.9770
Var: D4b050		0.7558		320.0218
Var: D5ar1k		0.0009		0.2582
Cov: Intercept D1B		0.1963		5.6211
Cov: Intercept D2A				
$_{W}..$		−0.0323		−2.0298
Cov: Intercept D3a		0.1292		−3.7409
Cov: Intercept D4b050		0.0177		170.0113
Cov: Intercept D5ar1k		−0.0356		−5.0077
Cov: D1B D2A				
$_{W}..$		−0.0047		0.0656
Cov: D1B D3a		0.0179		−1.4431
Cov: D1B D4b050		0.0015		4.2281
Cov: D1B D5ar1k		−0.0049		−0.0489
Cov: D2A				
$_{W}..D3a$		−0.0034		−0.0628
Cov: D2A				
$_{W}..D4b050$		−0.0016		−1.8284
Cov: D2A				
$_{W}..D5ar1k$		0.0008		0.0695
Cov: D3a D4b050		0.0075		−3.4491
Cov: D3a D5ar1k		−0.0033		−0.0605
Cov: D4b050 D5ar1k		−0.0013		−5.2450
Var: Residual			737.5981	692.9681

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.

Random Effects and Regional Variation of Coefficients

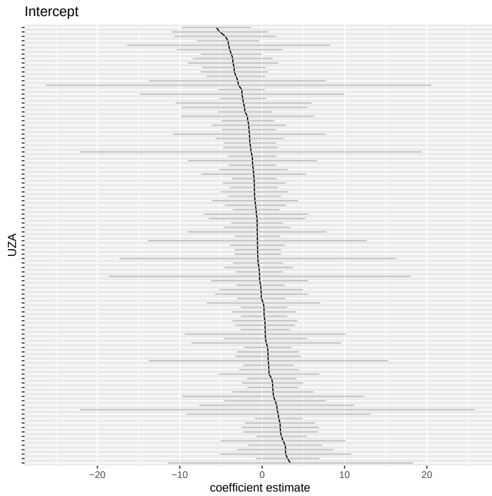
I again use log-likelihood ratio tests between the base and full mixed effects models to test the hypothesis that there is no regional variations in model coefficients (random slopes) for the built environment measures, as they can be seen as pairs of nested models with the fixed effect models being a restricted version of the mixed effect models (*17*). For both the zero VMT logit and DVMT log-linear regression model, the full mixed effect models perform significantly better than the base mixed effect models (p<0.0000001 for both tests), which leads me to reject the hypothesis of no regional variations of model coefficients for built environment measures.

Mixed effect models also provide UZA-specific coefficients for those built environment variables that I assume to have random coefficients (slopes). As can be seen in Figure 5 and 6, the random model coefficients for UZA cover a range. They can be compared with the UZA-specific coefficients from the full fixed effects models in Figure 3 and 4, respectively. Compared with the results from the full fixed effects models, those from the mixed effects have a narrower range, attesting to the advantage of mixed effects models in partially pooling observations. The random

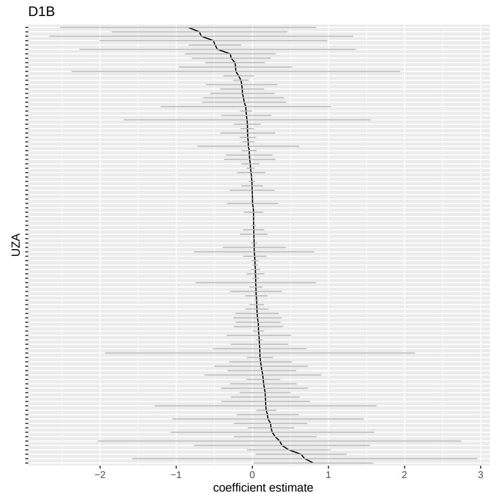
TABLE 2: Estimation Results of Fixed Effects Models

	Base logit	Full logit	Base log-linear	Full log-linear
Intercept	−0.4125 (0.3206)	0.5429 (1.3824)	3.3024 (0.0709)***	3.4381 (0.2357)***
Income				
$10-30k	−0.3054 (0.0822)***	−0.2724 (0.0872)**	−0.0536 (0.0379)	−0.0259 (0.0383)
$30-50k	−0.7953 (0.1024)***	−0.8133 (0.1084)***	0.0342 (0.0421)	0.0503 (0.0424)
$50-70k	−1.0447 (0.1093)***	−1.1197 (0.1160)***	0.1048 (0.0422)*	0.1280 (0.0425)**
$70-100k	−1.0630 (0.1136)***	−1.0585 (0.1194)***	0.2208 (0.0420)***	0.2370 (0.0423)***
>$100k	−1.0358 (0.1121)***	−1.0661 (0.1181)***	0.2315 (0.0419)***	0.2523 (0.0422)***
poverty	0.3593 (0.0687)***	0.3469 (0.0726)***	−0.1983 (0.0255)***	−0.1868 (0.0253)***
HhSize	−0.1928 (0.0183)***	−0.2107 (0.0194)***	0.0522 (0.0039)***	0.0564 (0.0039)***
Workers	−0.6702 (0.0308)***	−0.6742 (0.0324)***	0.1960 (0.0049)***	0.1918 (0.0049)***
ZeroVeh	3.5408 (0.1060)***	3.9653 (0.1159)***		
VehPerDriver	−0.7357 (0.0605)***	−0.6293 (0.0613)***	0.1010 (0.0079)***	0.0995 (0.0079)***
Life Cycle				
Single			−0.3459 (0.0163)***	−0.3228 (0.0163)***
Parents w/ children			0.0171 (0.0108)	0.0220 (0.0108)*
Empty Nester			−0.1293 (0.0137)***	−0.1361 (0.0137)***
D1B	0.0035 (0.0007)***		−0.0014 (0.0004)***	
D2A_WRKEMP	0.0015 (0.0004)***		0.0006 (0.0001)***	
D3a	0.0131 (0.0029)***		−0.0082 (0.0007)***	
D4b050	0.9741 (0.0742)***		−0.2613 (0.0252)***	
D5ar1k	0.0013 (0.0002)***		−0.0008 (0.0001)***	
AIC	17017.8008	16839.6128	401980.2927	401219.4096
BIC	18017.7606	21578.5529	402986.9100	405894.2086
Log Likelihood	−8393.9004	−7874.8064	−200872.1464	−200061.7048
Deviance	18862.6682	17861.4163	27096144.1284	25948285.8808
Num. obs.	44148	44148	37446	37446

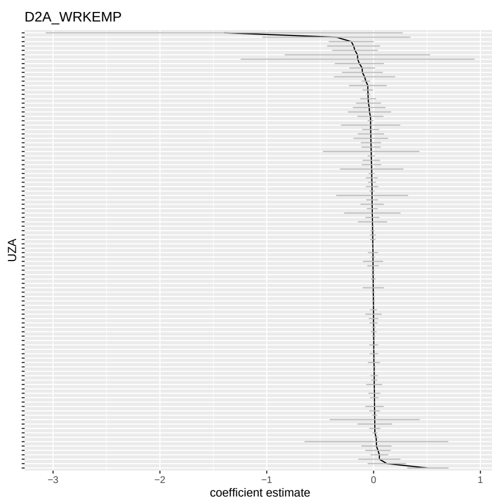
$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Standard errors in parentheses. UZA specific intercepts and coefficients are not shown for space reasons.



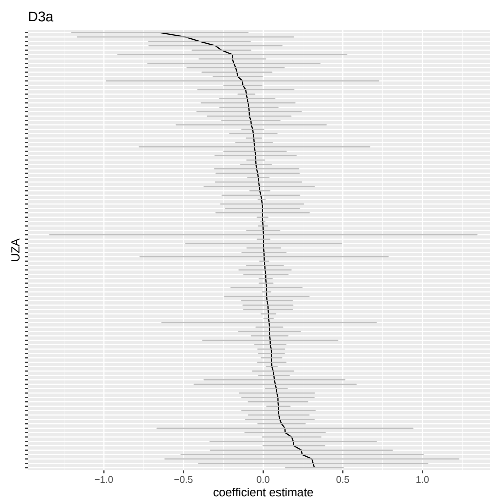
(a) Intercept



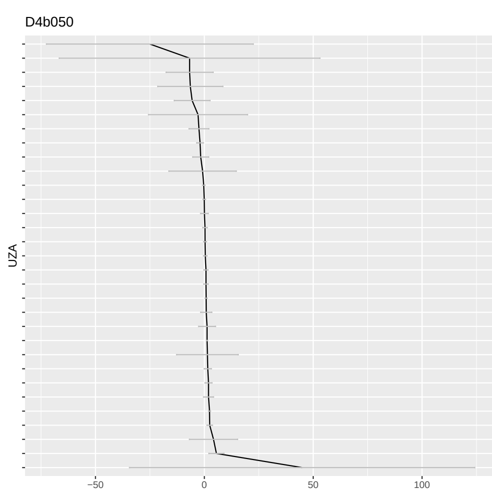
(b) D1B



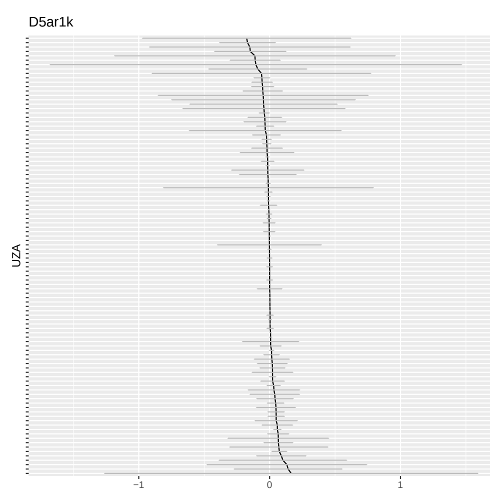
(c) D2A_WRKEMP



(d) D3a

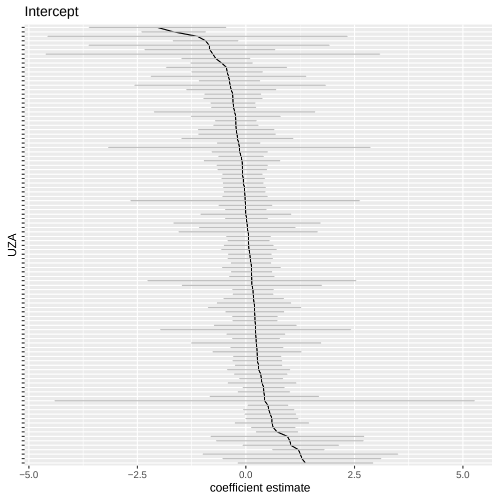


(e) D4b050

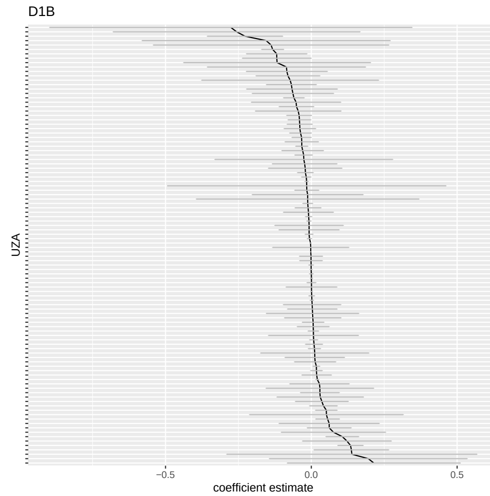


(f) D5ar1k

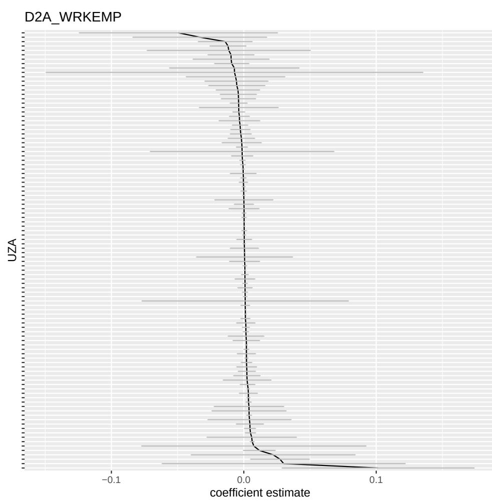
FIGURE 3: Point estimates and confidence intervals from the full ZeroDVMT logit model for each built environment variable by UZA



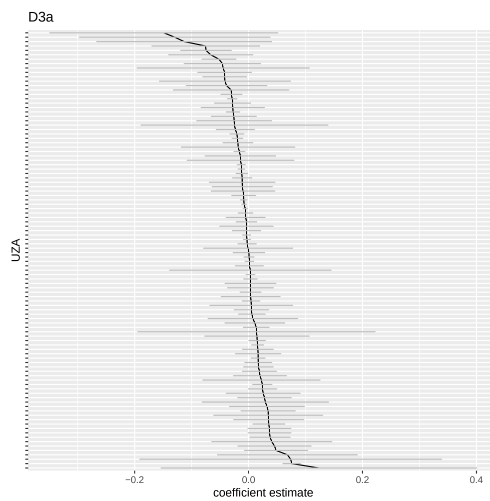
(a) Intercept



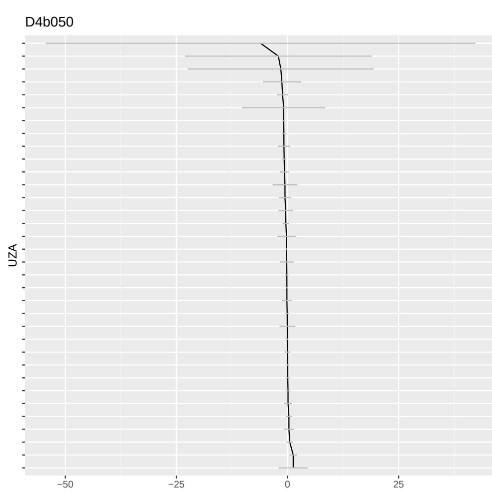
(b) D1B



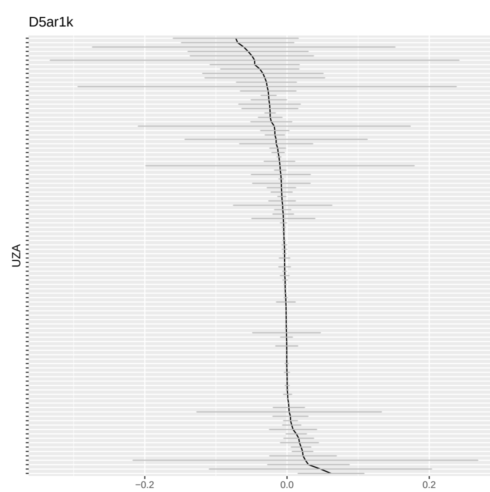
(c) D2A_WRKEMP



(d) D3a



(e) D4b050



(f) D5ar1k

FIGURE 4: Point estimates and confidence intervals from the full DVMT log-linear model for each built environment variable by UZA

effects are also asymmetrical. For example, for the D1B coefficient in the VMT log-linear model (Figure 6), most of the UZAs have expected negative value, while a small percentage of UZAs have positive value.

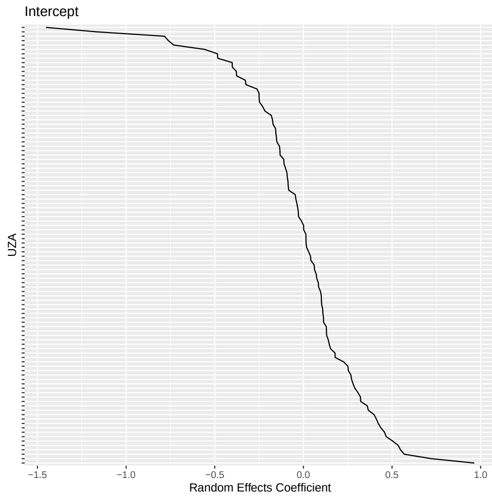
CONCLUSION AND DISCUSSION

In this paper, I start to explore the regional variation in the effects of built environment on driving. With a nationwide dataset and high resolution built environment measures, I evaluate the regional variation in the effect on VMT across the UZAs in the U.S with fixed and mixed effects models. My results indicate there are substantial regional variations in the random coefficients for built environment measures. The long-lasting fixation in the literature on estimating fixed effects of built environment variables is problematic. Particularly, there are no fixed effects for most built environment variables that apply to all UZAs after considering random effects. However, disregarding the effects of built environment on driving because of low average elasticity miss the point too - for certain regions, built environment can and do have large effects on driving.

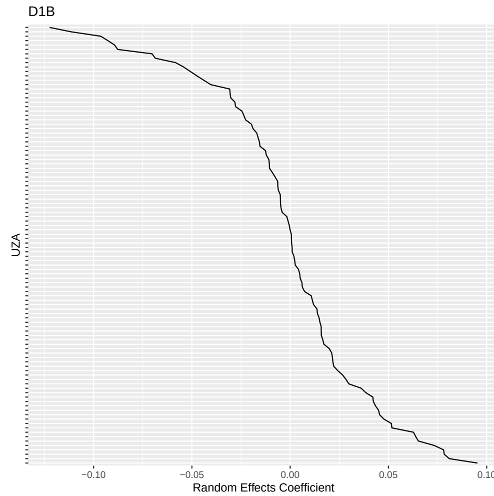
Previous research largely rely on data from a single region or a few regions and thus is not able to sufficiently explore such regional variations. Several research efforts have been dedicated to synthesize numerous previous research and come up with a single point estimate of the effect My research indicates that, while a point estimate may be more straightforward, the effects of built environment may vary substantial from region to region. A few recent papers (*18*) examine the heterogeneity of elasticity across neighborhood types and population groups, but they still assume households living in the same neighborhood types have the same elasticity across different regions. My finding has implications for applications of point estimates of elasticity in projecting effects various policies involving built environment. While point estimates may be much easier to apply, it fails to present the uncertainty in the possible effects. It may also discourage actions since the average effects are small. Indeed, Stevens argues based on the average elasticity that densification policies have minimal effects on driving (*6*).

I would like acknowledge a number of limitations and future work of this research. Although I went through a variable selection process and my model specification include all the “usual suspects” in a model for household VMT, my model specifications are likely not the best possible. For example, I didn’t consider variables at different geographical resolution, such as census tract- or UZA-level variables, as being considered and tested in other research (*19*). It is also likely that the best model specification varies from model to model, but for my purpose to investigate regional variation, I keep the same model specification across different models. I didn’t consider the potential mixed effects of social-economic status variables or their interactions with built environment variables at the region level. I also didn’t address the potential residential self-selection bias in my models. There has been extensive research documenting the prevalence and magnitude of such bias (*20*). There is research suggesting including SES variables in model specifications helps control potential self-selection bias (*21*). Finally, the 2017 NHTS dataset and a new version of the Smart Location Database have been released by FHWA and EPA, respectively, and it is now possible to update this research with these latest datasets and, furthermore, compare the results across these two cross sectional datasets. However, I haven’t been able to get access to the confidential residence information of survey participants in 2017 NHTS and will have to update this work after I do.

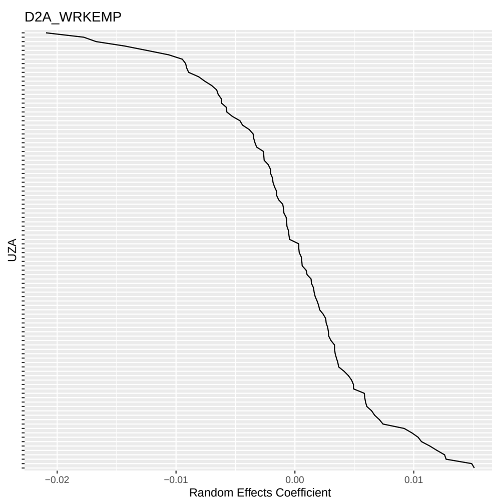
This paper follows the practice of reproducible research: it is written using `rmarkdown/bookdown` formatted into a TRB article using the `rticle` R package and the TRB Latex template (`https:`



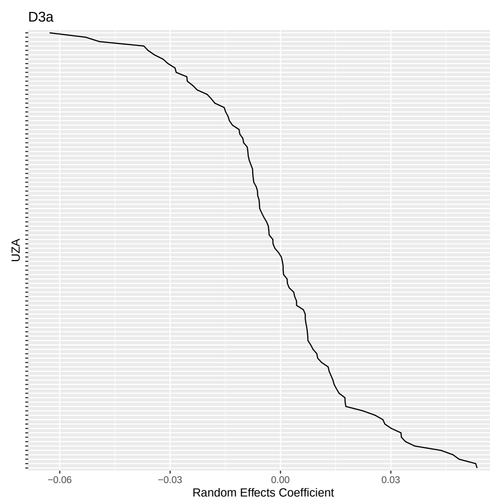
(a) Intercept



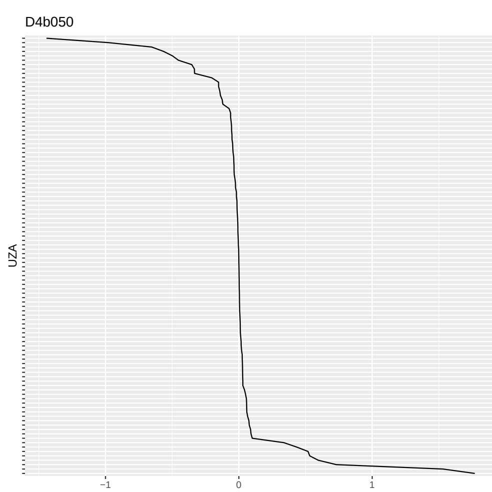
(b) D1B



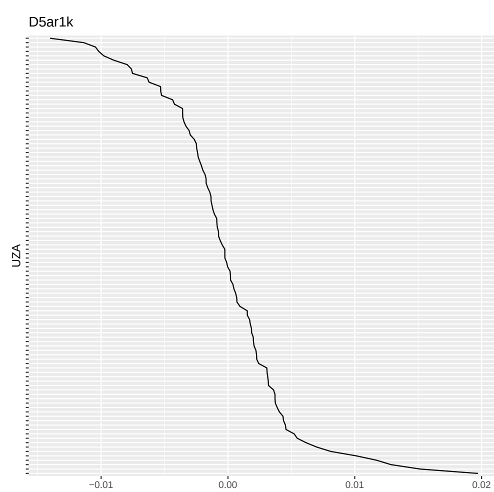
(c) D2A_WRKEMP



(d) D3a

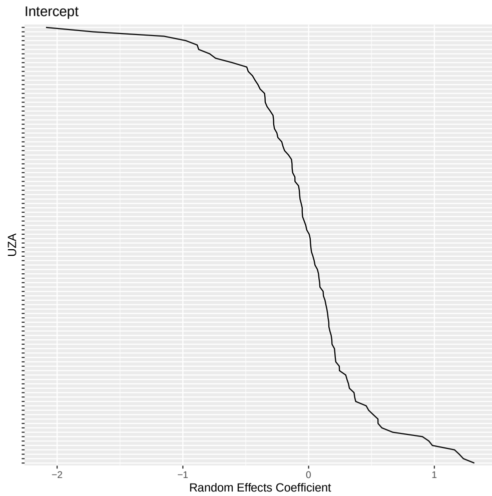


(e) D4b050

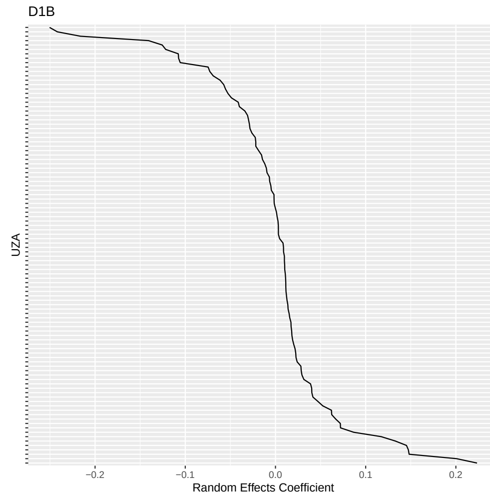


(f) D5ar1k

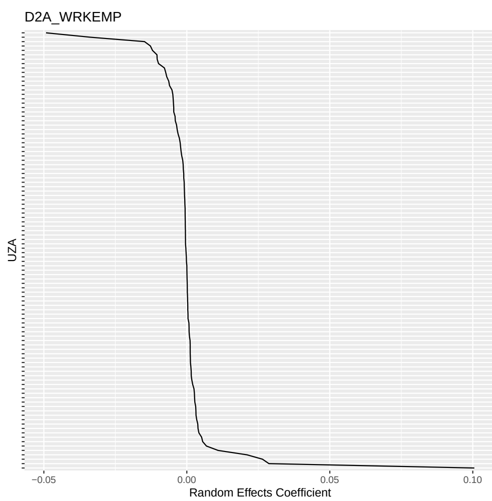
FIGURE 5: Random coefficients from the full ZeroDVMT logit model for each built environment variable by UZA



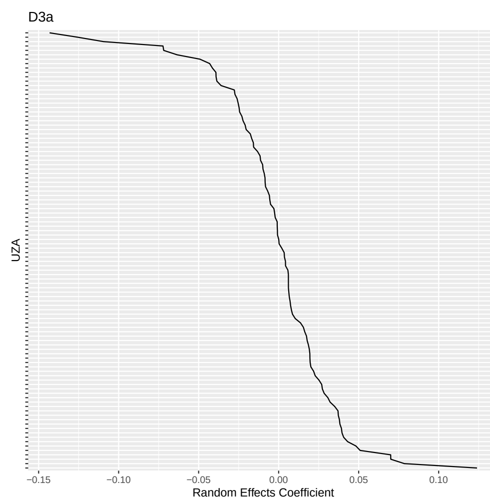
(a) Intercept



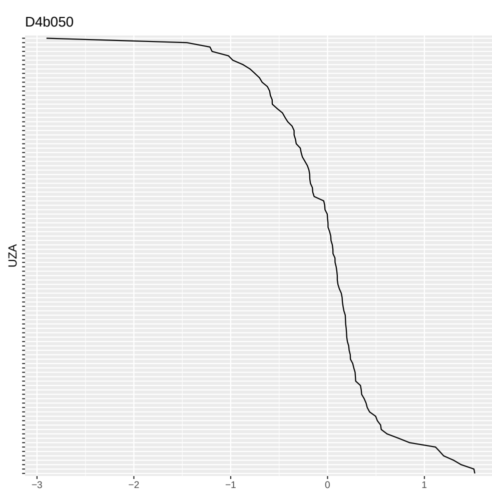
(b) D1B



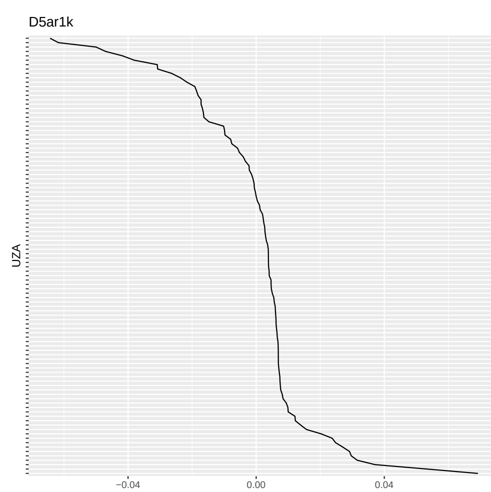
(c) D2A_WRKEMP



(d) D3a



(e) D4b050



(f) D5ar1k

FIGURE 6: Random coefficients from the full DVMT log-linear model for each built environment variable by UZA

//github.com/chiehrosswang/TRB_LaTeX_tex). All the results and figures in the paper are rendered dynamically from R scripts made publicly available on the author's github repository (https://github.com/cities/nhts-mxlm). However, because I used the confidential NHTS data source, I cannot make the complete dataset publicly available and one cannot fully replicate my work.

ACKNOWLEDGEMENTS

The author acknowledges partial financial support from the National Institute for Transportation and Communities (NITC) under grant number NITC-1433 and from Oregon DOT under grant SPR-788.

REFERENCES

00 1. Salon, D. Heterogeneity in the Relationship Between the Built Environment and Driving: Focus on Neighborhood Type and Travel Purpose. *Research in Transportation Economics*, Vol. 52, 2015, pp. 34–45. https://doi.org/10.1016/j.retrec.2015.10.008.

2. Voulgaris, C. T., B. D. Taylor, E. Blumenberg, A. Brown, and K. Ralph. Synergistic Neighborhood Relationships with Travel Behavior: An Analysis of Travel in 30,000 US Neighborhoods. *Journal of Transport and Land Use*, Vol. 10, No. 1, 2016. https://doi.org/10.5198/jtlu.2016.840.

3. Ewing, R., G. Tian, J. Goates, M. Zhang, M. J. Greenwald, A. Joyce, J. Kircher, and W. Greene. Varying Influences of the Built Environment on Household Travel in 15 Diverse Regions of the United States. *Urban Studies*, Vol. 52, No. 13, 2015, pp. 2330–2348. https://doi.org/10.1177/0042098014560991.

4. Ewing, R., and R. Cervero. Travel and the Built Environment. *Journal of the American Planning Association*, Vol. 76, No. 3, 2010, pp. 265–294. https://doi.org/10.1080/01944361003766766.

5. Aston, L., G. Currie, A. Delbosc, Md. Kamruzzaman, and D. Teller. Exploring Built Environment Impacts on Transit Use – an Updated Meta-Analysis. *Transport Reviews*, Vol. 41, No. 1, 2021, pp. 73–96. https://doi.org/10.1080/01441647.2020.1806941.

6. Stevens, M. R. Does Compact Development Make People Drive Less? *Journal of the American Planning Association*, Vol. 83, No. 1, 2017, pp. 7–18. https://doi.org/10.1080/01944363.2016.1240044.

7. Cervero, R., and J. Murakami. Effects of Built Environments on Vehicle Miles Traveled: Evidence from 370 US Urbanized Areas. *Environment and planning A*, Vol. 42, No. 2, 2010, pp. 400–418.

8. Bento, A. M., M. L. Cropper, A. M. Mobarak, and K. Vinha. The Effects of Urban Spatial Structure on Travel Demand in the United States. *Review of Economics and Statistics*, Vol. 87, No. 3, 2005, pp. 466–478.

9. Ewing, R., M. Greenwald, M. Zhang, J. Walters, M. Feldman, R. Cervero, L. Frank, and J. Thomas. Traffic Generated by Mixed-Use Developments—Six-Region Study Using Consistent Built Environmental Measures. *Journal of Urban Planning and Development*, Vol. 137, No. 3, 2011, pp. 248–261. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000068.

10. Zhang, M. The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association*, Vol. 70, No. 3, 2004, pp. 344–360. https://doi.org/10.1080/01944360408976383.

11. U.S. Department of Transportation, Federal Highway Administration. *2009 National Household Travel Survey*. 2009.

12. Ramsey, K., and A. Bell. *Smart Location Database: Version 2.0 User Guide*. US Environment Protection Agency, Washington, DC 20460, 2014.

13. Wang, L., B. Gregor, H. Yang, T. Weidner, and T. Knudson. Capturing the Built Environment-Travel Interaction for Strategic Planning: Development of a Multimodal Travel Module for the Regional Strategic Planning Model (RSPM). *Journal of Transport and Land Use*, Vol. 11, No. 1, 2018. https://doi.org/10.5198/jtlu.2018.1191.

14. Gelman, A., and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge ; New York, 2006.

15. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.

16. Bates, D., M. Mächler, B. Bolker, and S. Walker. Fitting Linear Mixed-Effects Models Using Lme4. *Journal of Statistical Software*, Vol. 67, No. 1, 2015. https://doi.org/10.18637/jss.v067.i01.

17. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, Massachusetts, 1987.

18. Ralph, K., C. T. Voulgaris, B. D. Taylor, E. Blumenberg, and A. E. Brown. Millennials, Built Form, and Travel Insights from a Nationwide Typology of U.S. Neighborhoods. *Journal of Transport Geography*, Vol. 57, 2016, pp. 218–226. https://doi.org/10.1016/j.jtrangeo.2016.10.007.

19. Gehrke, S. R., and L. Wang. Operationalizing the Neighborhood Effects of the Built Environment on Travel Behavior. *Journal of Transport Geography*, Vol. 82, 2020, p. 102561. https://doi.org/10.1016/j.jtrangeo.2019.102561.

20. Cao, X. (Jason)., P. L. Mokhtarian, and S. L. Handy. Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings. *Transport Reviews*, Vol. 29, No. 3, 2009, pp. 359–395. https://doi.org/10.1080/01441640802539195.

21. Naess, P. Tempest in a Teapot: The Exaggerated Problem of Transport-Related Residential Self-Selection as a Source of Error in Empirical Studies. *Journal of transport and land use*, Vol. 7, No. 3, 2014, pp. 57–79.