5-4-2023

Explore Regional Variation in the Effects of Built Environment on Driving with High Resolution U.S. Nationwide Data

Liming Wang
Portland State University, limwang@pdx.edu

Regional Variation in the Effects of Built Environment on Driving (VMT): A Reproducible Research Project

Liming Wang, Portland State University

Outline

- Rants of existing research on built environment and travel (driving)
- An exploration of regional variation of the connection using US nation wide data
  - Data: 2009 NHTS + SLD
  - Methods: fixed effect regression, mixed effect regression
  - Results: existence of regional variation
  - Reproducible research

Rant #1

Existing research predominantly focuses on a single city/region
- Elasticity of Vehicle Miles Traveled with respect to population density range from -1.05 to +0.03 (with weighted average=-0.04) (Ewing and Cervero, 2010)

Few exceptions: Ewing et al (2015), Tian et al (2020), Sabouri (2021);

The few studies using U.S. nation-wide data use only very coarse resolution (census tract, or even UZA level) data: e.g. Cervero and Murakami (2010);

Rant #2

Existing research predominantly focuses on producing point estimates of the “effects”, or association, between built environment and travel

Rant #3

Published articles of previous research is all we got – no data, data processing or modeling code available, i.e. none* of the research is reproducible by others beyond the authors

* There may be exceptions, but there is no way to search for those

Rant #4

Synthesis (meta analysis) of previous research then tries to distill results from these single region studies into another single estimate

Ewing and Cervero (2001, 2010); Stevens (2017)

Demo project of a solution?

A reproducible research project estimating mixed effect models on the effect of built environment on driving using high resolution U.S. nation-wide data

- An explanation of large spread in point estimate: existence of regional variation
- A demo project following reproducible research practice

Data

- The 2009 U.S. National Household Travel Survey (NHTS)
  - Socio-demographic characteristics
  - Travel behavior outcomes, including household VMT on survey day

- EPA's Smart Location Database (SLD)
  - SLD Diverse set of built environment measures organized by 5D variables

==> 44,000 households across 100 UZAs (filtered by surveyed households >= 100) with detailed travel, socio-demographic and built environment measures

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)

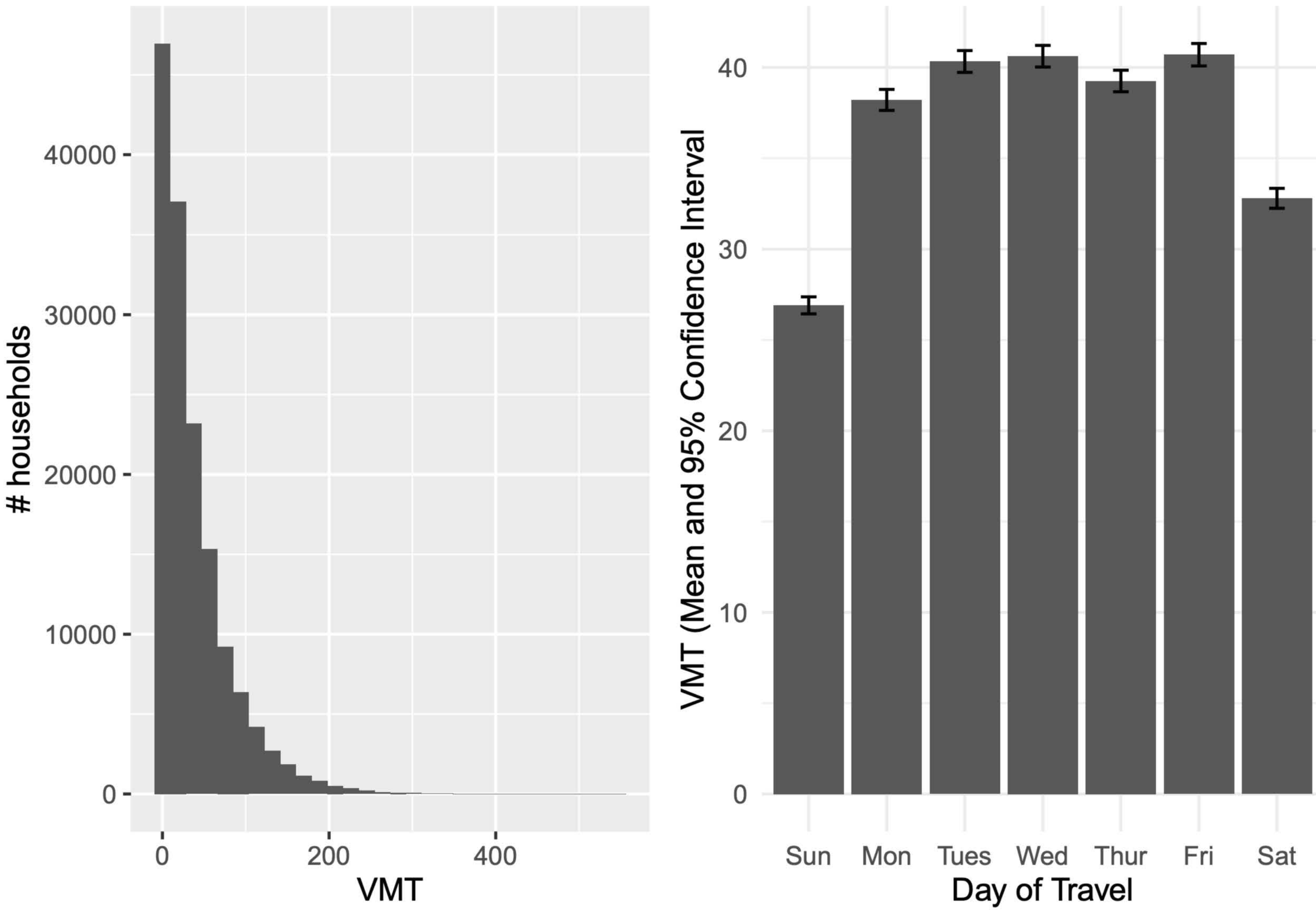
BE Measures in SLD: 5D

- D1 Density: e.g. D1B population density
- D2 Diversity: e.g. D2A_WRKEMP Household workers per Job
- D3 Design: D3a Total road network density
- D4 Distance to Transit: D4b050 Proportion of jobs within 0 5 mile of transit stop
- D5 Destination Accessibility: D5ar1k Jobs within 45 minutes auto travel time 1000

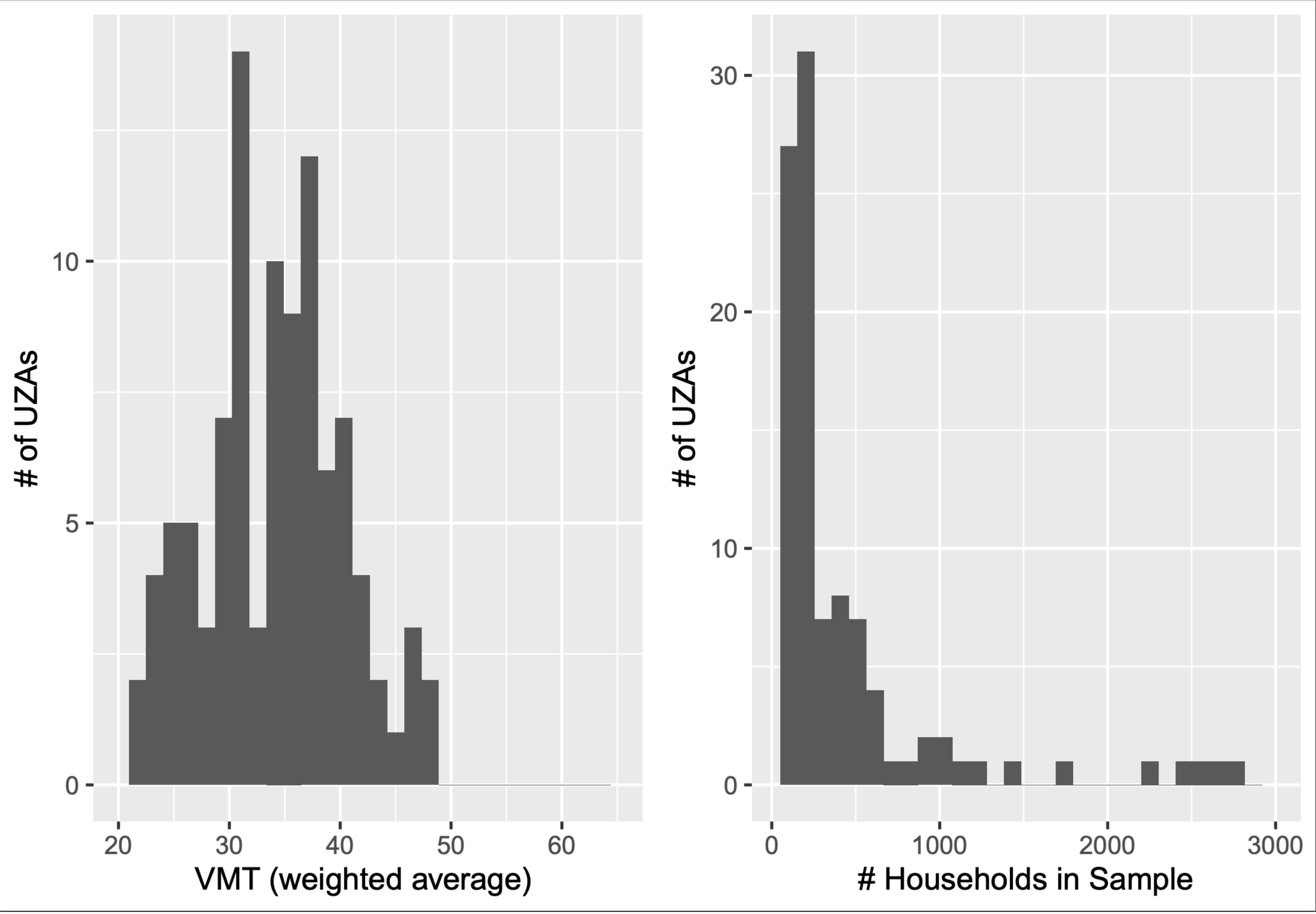
[redacted]

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)

Descriptive Stats of BE variables



households
40000
30000
20000
10000
0
0
200
400
VMT
VMT (Mean and 95% Confidence Interval
40
30
20
10
0
Sun
Mon
Tues
Wed
Thur
Fri
Sat
Day of Travel



Weighted Average Household VMT by UZA

Surveyed Household by UZA

Fixed Effect Models

- Base/pooled Fixed Effect Model

$$\Pr(VMT_{iu}=0)=\frac{\exp(V_{iu})}{1+\exp(V_{iu})}, \qquad V_{iu}=\alpha_u+\beta X_{iu}^{SES}+\gamma X_{iu}^{BE}$$

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

- Full/segemented Fixed Effect Model

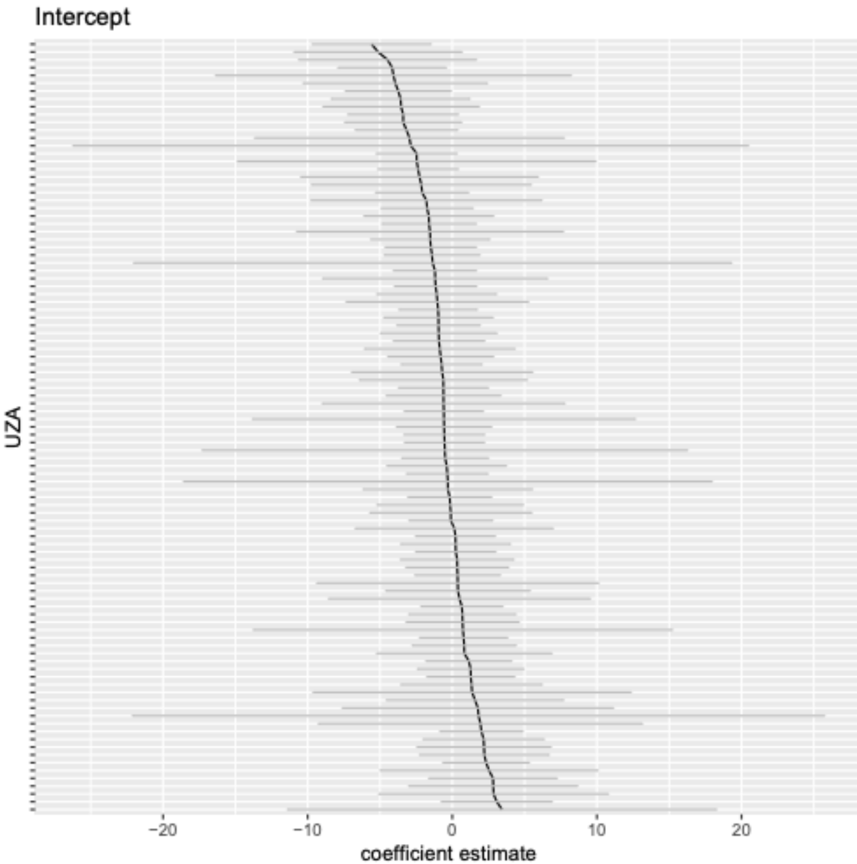
$$\Pr(VMT_{iu}=0)=\frac{\exp(V_{iu})}{1+\exp(V_{iu})}, \qquad V_{iu}=\alpha_u+\beta X_{iu}^{SES}+\gamma_u X_{iu}^{BE}$$

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

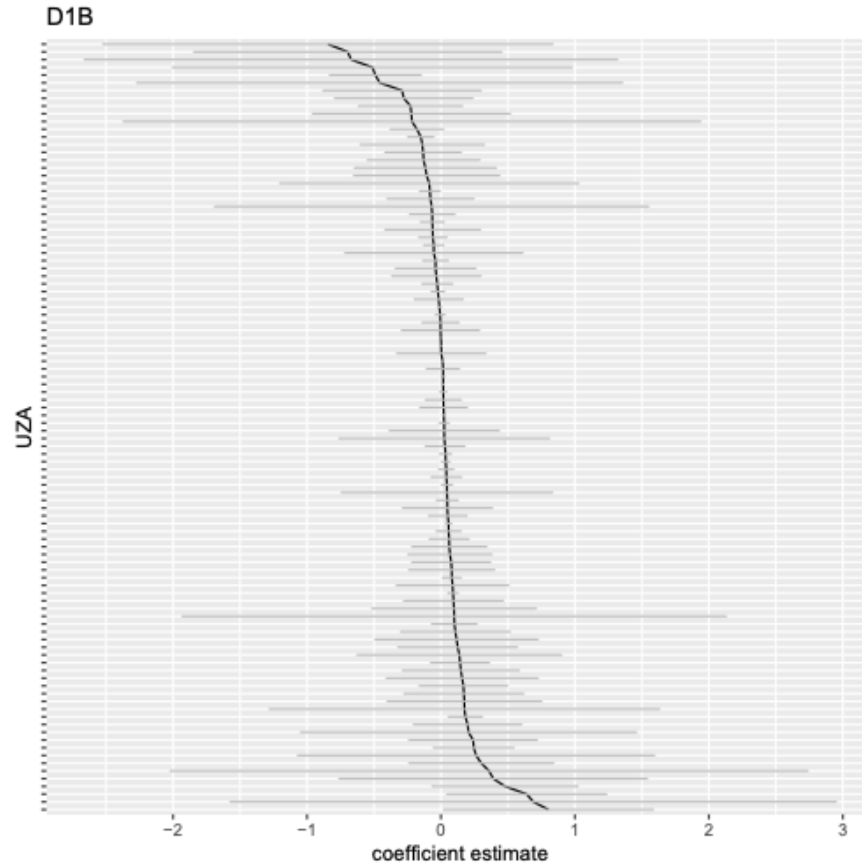
TABLE 2: Estimation Results of Fixed Effects Models

	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.

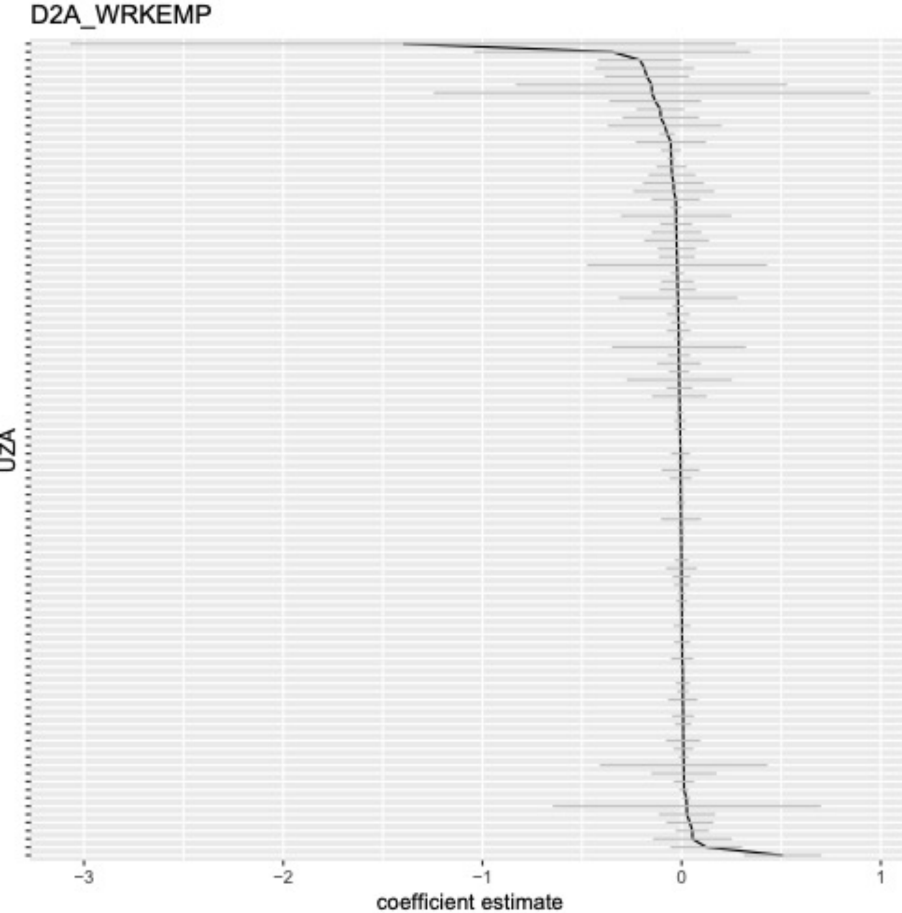


(a) Intercept

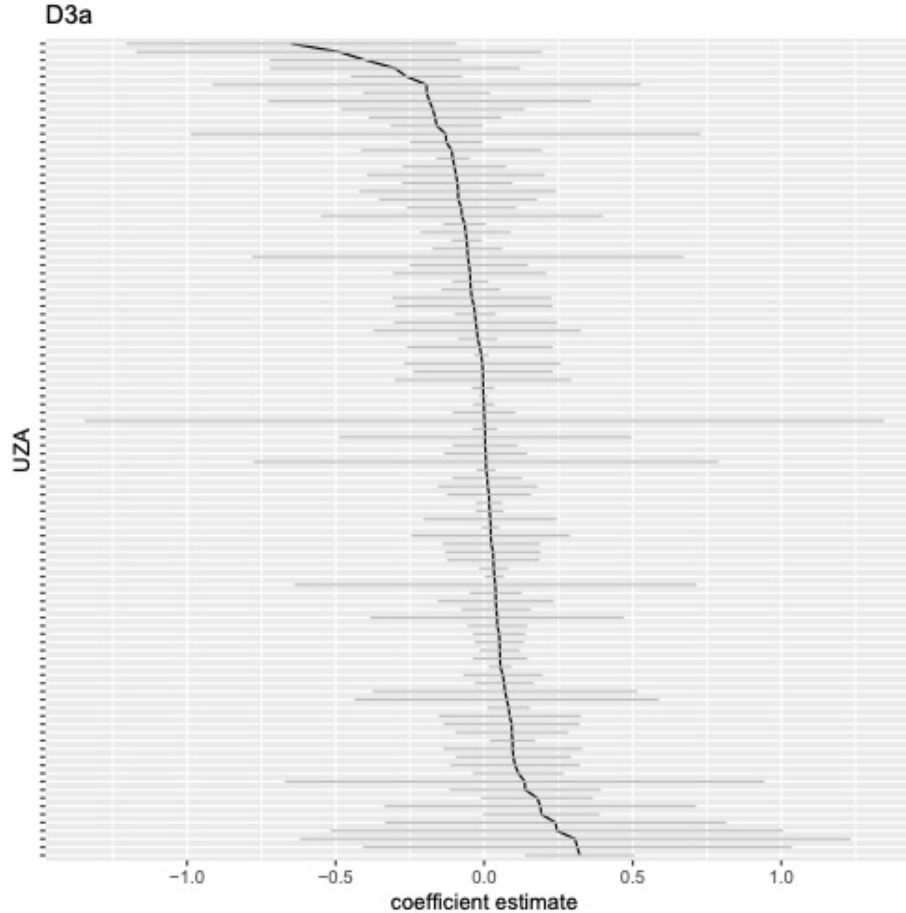


(b) D1B

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

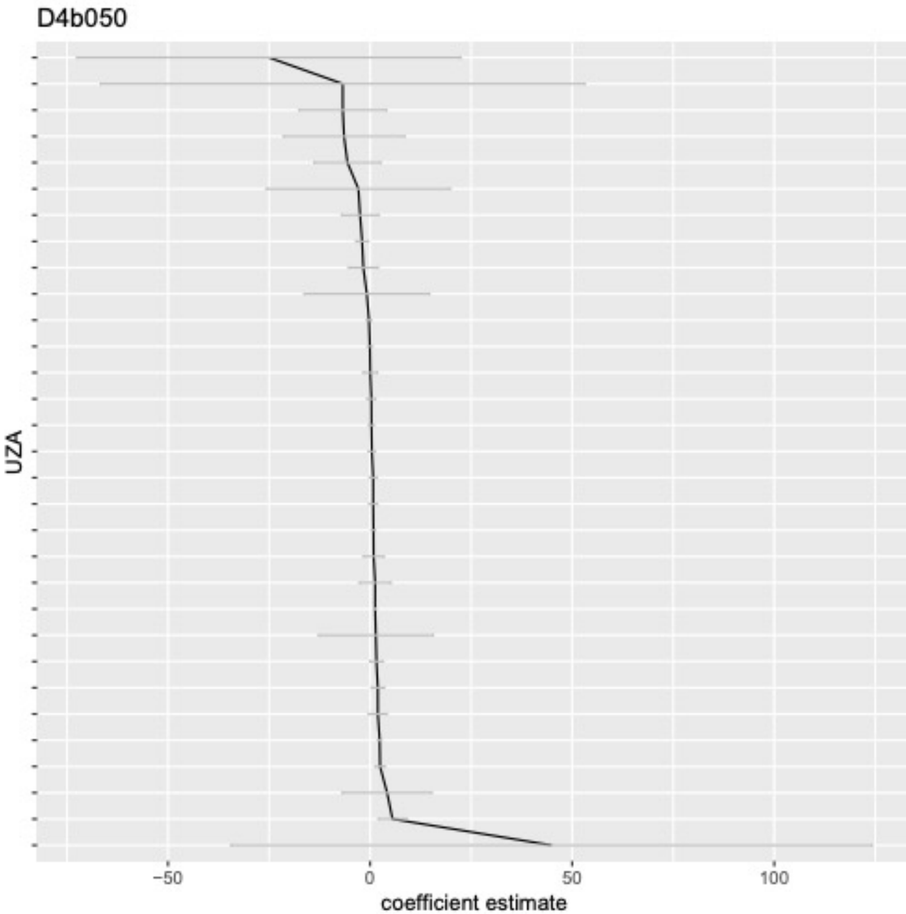


(c) D2A_WRKEMP

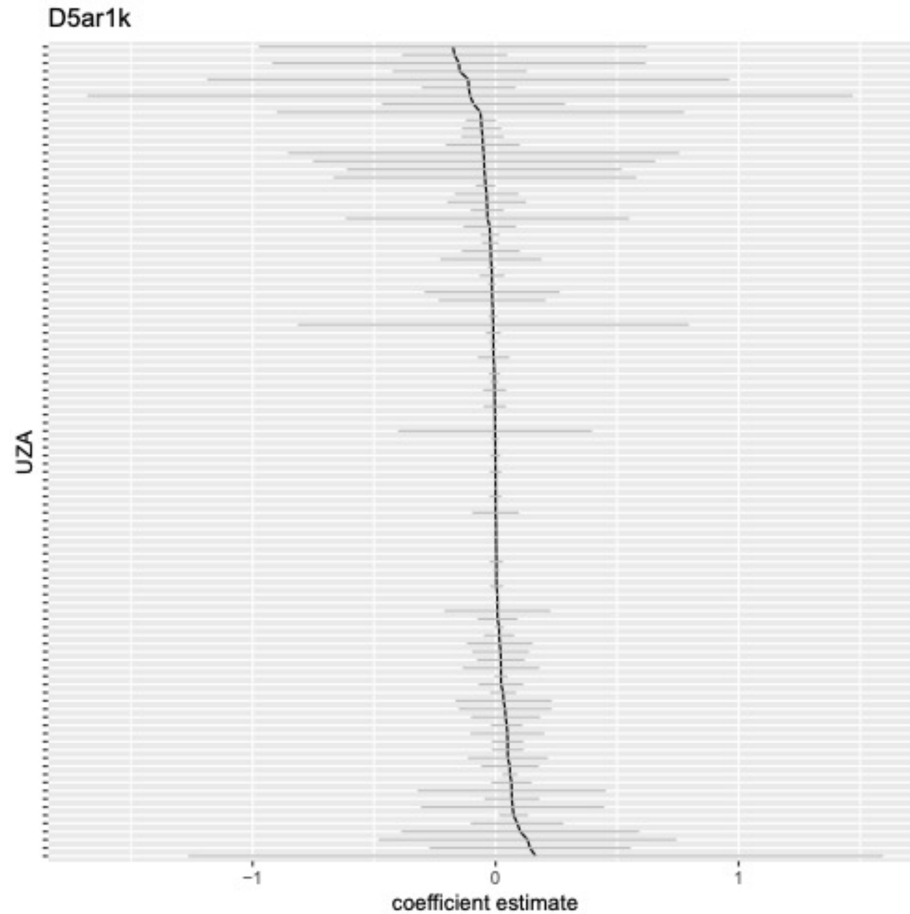


(d) D3a

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



(e) D4b050



(f) D5ar1k

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

Test for Regional Variation

- For the elasticity estimate of VMT wrt population density (D1B) = **-0.086** in pooled fixed effect model, in line with Ewing & Cervero’s estimate of -0.04
- Likelihood ratio test rejects the fixed effect model (no UZA-specific slopes) and favors the full fixed effect model with UZA-specific coefficient (x^2=1620.883, p < 0.000), although most UZA-specific coefficients are not significant.

Mixed Effect Models

- Base Mixed Effect Model

$$\Pr(VMT_{iu} = 0) = \frac{\exp(V_{iu})}{1 + \exp(V_{iu})}, \text{where}$$
$$V_{iu} = \alpha_{iu} + \beta X_{iu}^{SES} + \gamma X_{iu}^{BE},$$
$$\alpha_{iu} \sim N(\bar{\alpha}, \sigma_\alpha)$$
$$VMT_{iu} \sim N(a_{iu} + bX_{iu}^{SES} + cX_{iu}^{BE}, \sigma^2) a_{iu} \sim N(\bar{a}_{iu}, \sigma_a^2)$$

- Full Mixed Effect Model

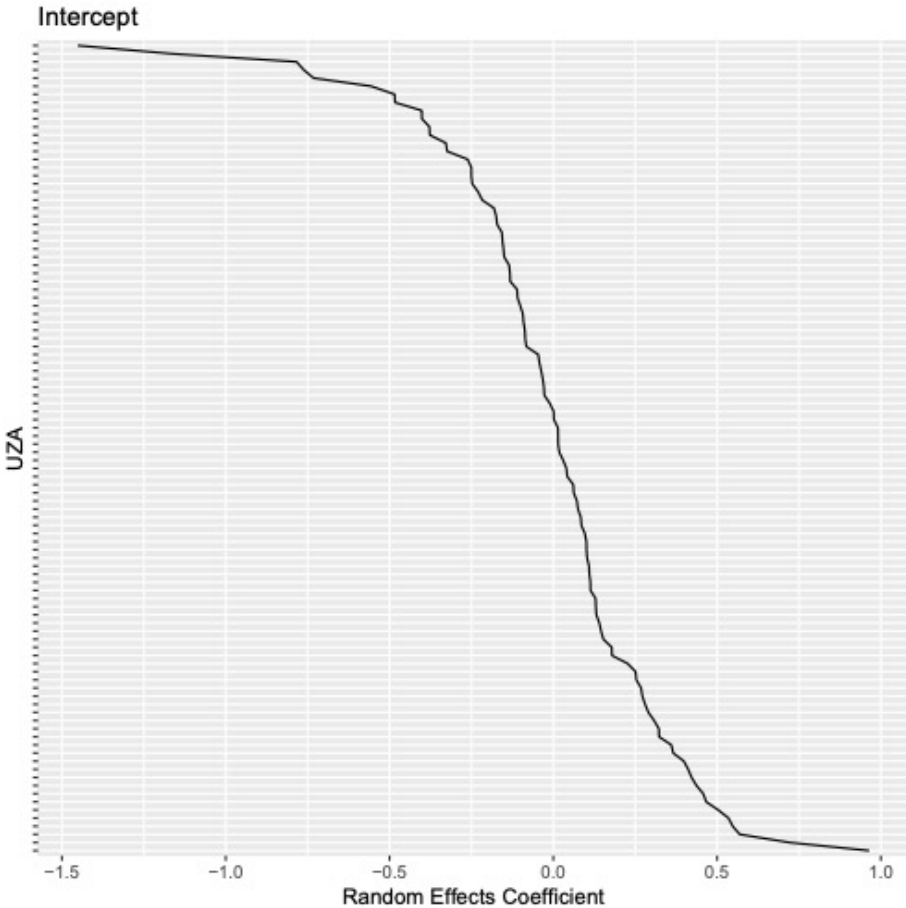
$$V_{iu} = \alpha_{iu} + \beta X_{iu}^{SES} + \gamma_{iu} X_{iu}^{BE}, \text{where}$$
$$\alpha_{iu} \sim N(\bar{\alpha}, \sigma_\alpha), \text{and}$$
$$\gamma_{iu} \sim N(\bar{\gamma}, \Sigma_\gamma)$$
$$VMT_{iu} \sim N(a_{iu} + bX_{iu}^{SES} + cX_{iu}^{BE}, \sigma), \text{where}$$
$$a_{iu} \sim N(\bar{a}, \sigma_a), \text{and}$$
$$c_{iu} \sim N(\bar{c}, \Sigma_c)$$

TABLE 3: Estimation Results of Mixed Effects Models

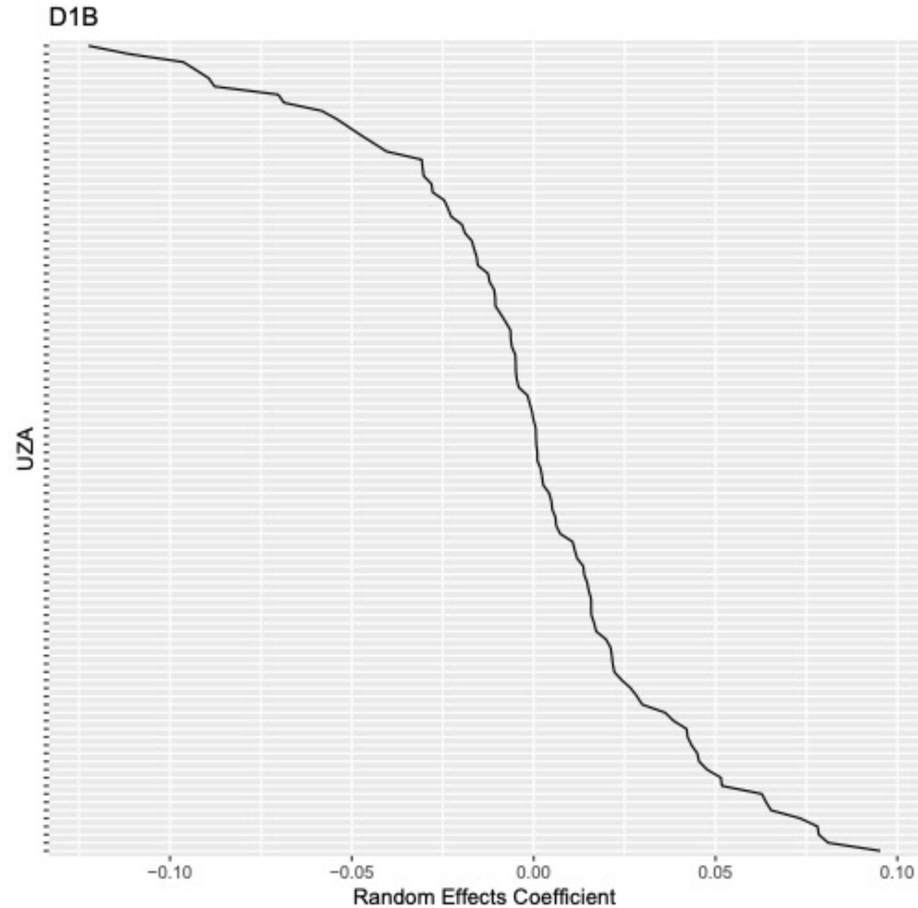
	Base logit	Full logit	Base log-linear	Full log-linear
Intercept	−0.2429	−0.0853	3.4296***	3.4860*
	(0.1250)	(0.1844)	(0.0439)	(1.3636)
Income				
\$10-30k	−0.2953***	−0.3009***	−0.0518	−0.0259
	(0.0814)	(0.0840)	(0.0383)	(0.0380)
\$30-50k	−0.7830***	−0.7877***	0.0362	0.0503
	(0.1015)	(0.1047)	(0.0425)	(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)

TABLE 3: Estimation Results of Mixed Effects Models

	Base logit	Full logit	Base log-linear	Full log-linear
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

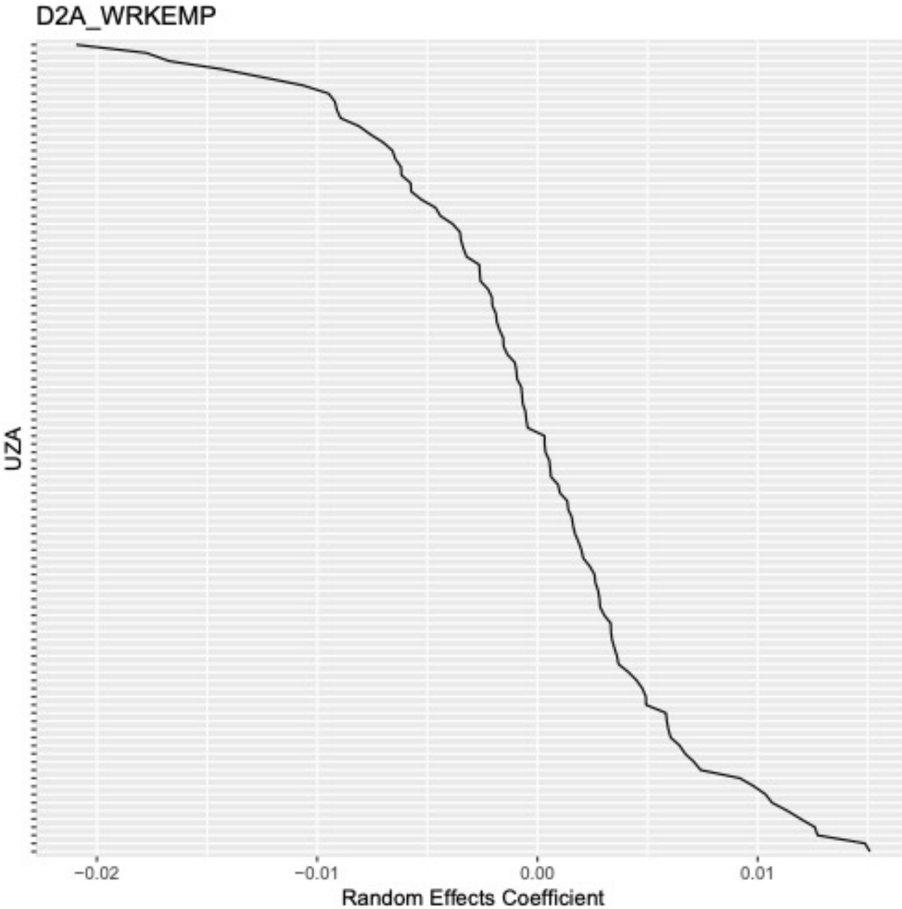


(a) Intercept

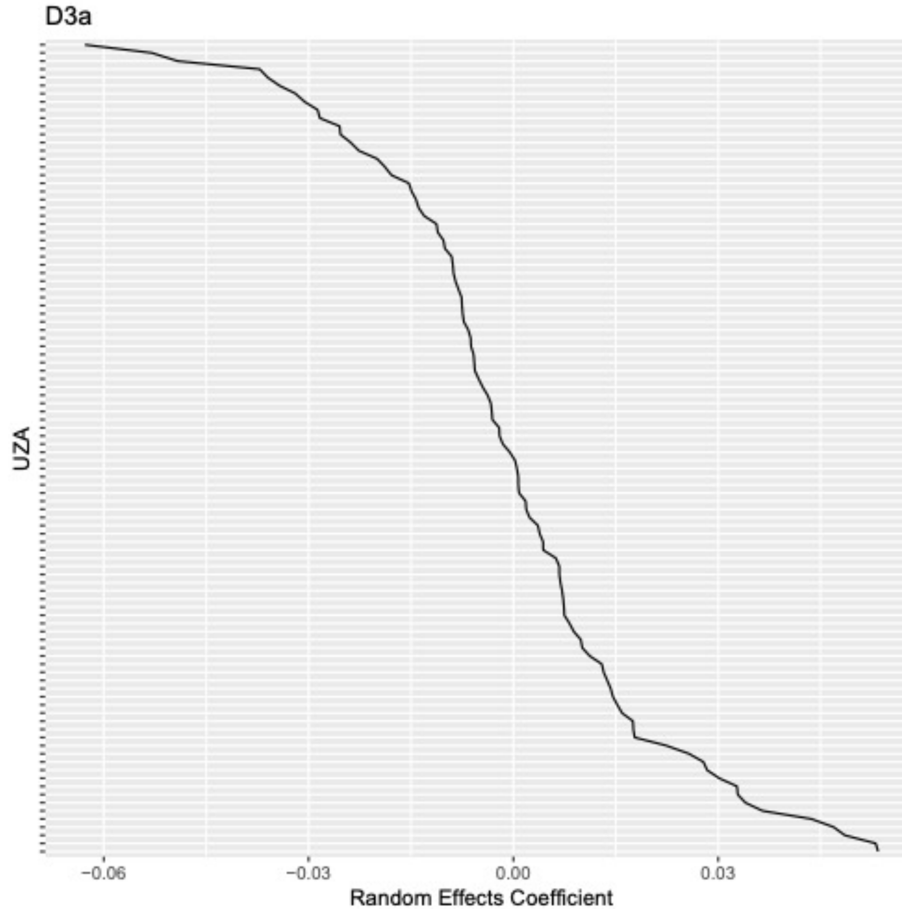


(b) D1B

Mean random coefficients from the full ZeroDVMT logit model for each built environment variable by UZA

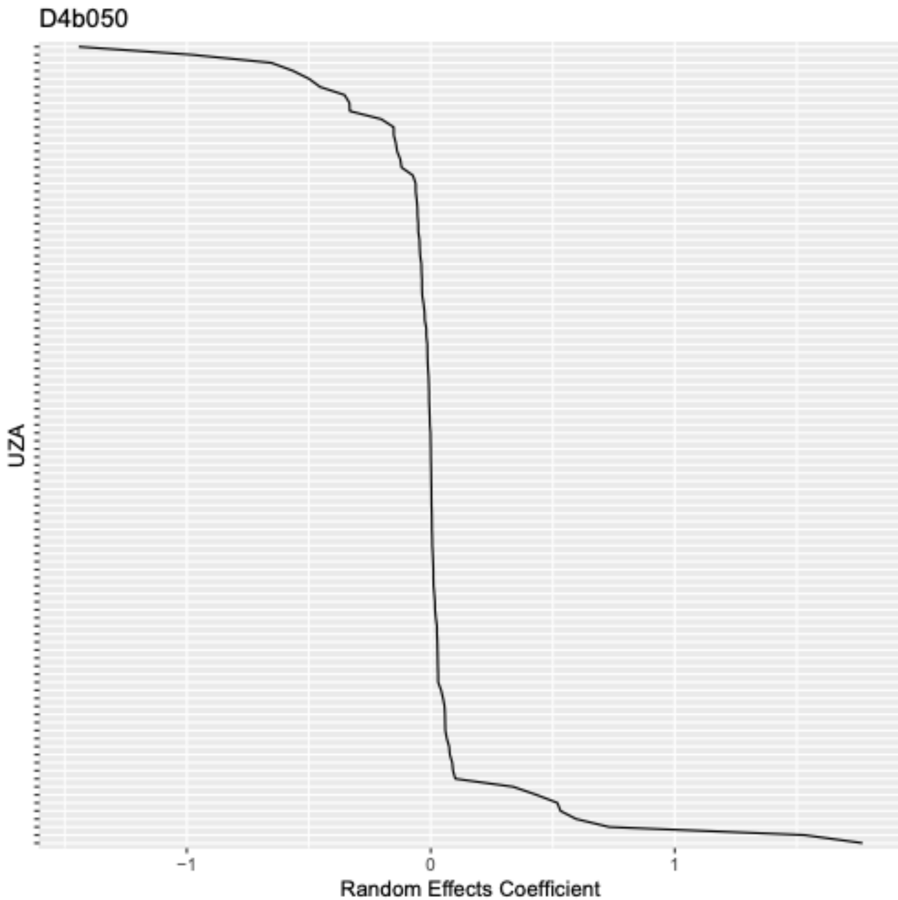


(c) D2A_WRKEMP

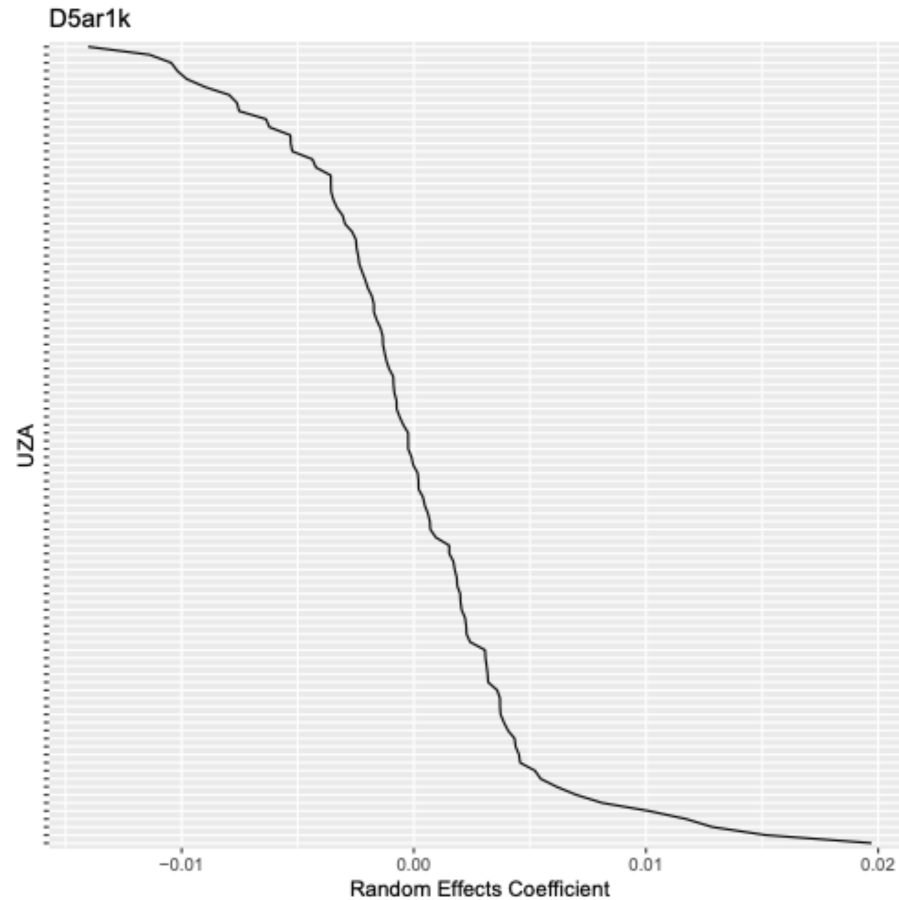


(d) D3a

Mean random coefficients from the full ZeroDVMT logit model for each built environment variable by UZA



(e) D4b050



(f) D5ar1k

Mean random coefficients from the full ZeroDVMT logit model for each built environment variable by UZA

Mixed Effect Models

- Likelihood ratio test also favors full mixed effect model with both random slopes and intercepts ($p<0.0000$).
- Notably, in the mixed effect model, all fixed effect coefficients for the 5D variables are not significant.

Conclusions and Discussion

- Significant regional variation in the estimates of built-environment effects on VMT, even as the mean estimate in line with synthesized weighted average
- With NHTS and SLD data, estimates of UZA-specific effect are mostly unusable (highly variable and statistically insignificant)
- Mixed-effect model provides a solution to the problem by allowing partial pooling
- After considering random effects, there is no statistically significant fixed effects.
- Implications
  - Maybe we should not have been so fixated with producing a single estimate after all?
  - Both sides of the debate miss some part of the big picture

Reproducible Research Practice

The work conducted for the project using R and the `tidyverse` suite of packages, reproducible from raw data to final products

Project reports and papers written using RMarkdown
  - including a TRB paper and the RMarkdown is available at https://github.com/cities/nhts-mxlm

Product of the project including an open source R package `VETravelDemandMM` (https://github.com/cities-lab/VETravelDemandMM) contributed to the VisionEval project (https://visioneval.org), a strategic planning model under active development led by FHWA.

Reproducible Research Practice

The work conducted for the project using R and the `tidyverse` suite of packages, reproducible from raw data to final products

Project reports and papers written using RMarkdown
  - including a TRB paper and the RMarkdown is available at https://github.com/cities/nhts-mxlm

Product of the project including an open source R package `VETravelDemandMM` (https://github.com/cities-lab/VETravelDemandMM) contributed to the VisionEval project (https://visioneval.org), a strategic planning model under active development led by FHWA.

Reproducible Research Practice

Personally, I cannot imagine (myself) doing research any other way; "You have at least one collaborator for any project - at the very least, yourself in 3 months".

Making code available can only go so far; live documentation and literal programming is an essential, but often neglected, step in reproducible research.

Tools (in R, Python) supporting reproducible research have improved tremendously over the last a few years, but the learning curve is still steep and most students and researchers have not been trained or exposed in reproducible research.

Combining reproducible research with open data facilitates dissemination of research and speeds up scientific investigation

Limitations/Future Work

- The confidential part of NHTS data cannot be made public
- Possible to make an anonymized version available?
- Model Selection is rudimental to facilitate model comparison at the cost of model fit
- Explore interaction effect and threshold effect

Acknowledgements

Partial financial support from NITC-1433 and ODOT SPR-788 is appreciated and acknowledged.

NITC
NATIONAL INSTITUTE for TRANSPORTATION and COMMUNITIES

Oregon Department of Transportation