TRANSPORT FINDINGS

Starting Points Matter: Spatial Variation in Marginal Effects for Negative Binomial Trip Models

Clemens A. Pilgram[1], Marlon G. Boarnet[1]

[1] Sol Price School of Public Policy, University of Southern California

Keywords: travel behavior, negative binomial regression, land use, transportation, micromobility

https://doi.org/10.32866/001c.35459

Findings

Several publications studying associations between land use and shared micromobility (such as dockless scooters and bicycles) rely on Negative Binomial regression models or similar, reporting the models' untransformed coefficients. We demonstrate a new way of reporting the associations identified by such models: By reporting marginal effects for several different starting points resembling real-world locations rather than just coefficients, these models' implications can be made more approachable to a lay audience. At the same time, we draw attention to the models' limitations when applied to locations that are outliers in terms of density.

1. Questions

Following the emergence of shared micromobility via scooters and dockless bicycles, several published studies have documented associations between land uses or demographic characteristics and trip generation using these modes. The field has somewhat converged on applying Negative Binomial ("NB") regression models[1] – a type of count model – or variations of them as the model of choice in studying what factors are associated with trip generation. These models possess several convenient properties for studying trip generation, since trips are countable and often highly concentrated, exhibiting generally nonlinear relationships with urban form.

Unlike in linear regression models, coefficeints of NB regressions and related nonlinear models cannot be interpreted directly as the association between an independent variable and trip counts as would be the case for an Ordinary Least Squares regression: Instead, it is necessary to calculate those associations at specific scenarios (commonly referred to as "marginal effects"; the expected change in trip generation associated with some change in an independent variable from a particular starting point in terms of all variables). Nonetheless, it is common for studies reporting the findings of these models to present model coefficients rather than marginal effects (Bai and Jiao 2020; Gehrke et al. 2021; Huo et al. 2021; Jiao and Bai 2020). To the extent marginal effects are reported at all, they are reported for a single, supposedly representative point in the independent variables' distributions (Merlin et al. 2021). These forms of

1 Some studies also use Zero-inflated Negative Binomial ("ZINB") regression, a variation of NB that accounts for large numbers of zero outcomes by combining an NB second stage with a logit to model whether or not to expect nonzero outcomes.

reporting mask great heterogeneity across space in the absolute magnitudes of associations, and may not be the most easily interpretable presentation of those relationships, especially in light of actual density distributions across cities.

Urban structure is, of course, highly concentrated and nonlinear: While most of a city's areal units tend to be of relatively modest densities (in terms of population, or jobs, or amenities), a small handful of tracts – such as central business districts - exhibit supranormal densities. The same nonlinearity and concentration is also true for micromobility trip origination.

Given this, we ask: Do the magnitudes of relationships between land uses and dockless trip generation vary across space within cities? Does this pose a challenge for how to report the findings of an NB regression trip generation model, in particular when trying to make models' outputs accessible to practitioners?

2. Methods

We demonstrate this effect using a simple land use model of scooter trip generation. Our data consist of 1,142,228 scooter trips undertaken across 408 census blockgroups in Minneapolis, Minnesota between July 10, 2018 and November 26, 2019, as well as independent variables obtained from the 2019 US Census ACS, Census LODES WAC, and OpenStreetMap, all at the census blockgroup level.

As is demonstrated in Figure 1, scooter trips are highly concentrated Downtown and near the University of Minnesota and decay with distance from those centers, with a smaller concentration of trip origins in the Uptown Minneapolis entertainment district.

Using this data, we model trip generation at the blockgroup level as a function of the variables listed in Table 1: densities of people, jobs, people aged 18-25, and food/drink establishments, as well as of distance from the CBD.[2] The daily average number of trip generations in a block group is the dependent variable, using NB regression.

Finally, we calculate the marginal effect for each independent variable at three different points:[3] At the median in terms of all variables (somewhat resembling a single family home neighborhood), at densities found in blockgroups in Dinkytown and Uptown Minneapolis (major entertainment districts with a large young population), in Powderhorn Park (a medium density urban neighborhood three miles south of Downtown containing a low-rise commercial corridor), and at densities equal to those in the most job-dense blockgroup of Downtown Minneapolis (the regional CBD).

---

2 The model is purposefully kept simple for demonstration purposes, however, the phenomenon we point out remains with the inclusion of additional dependent variables, or when using ZINB.

3 Marginal effects are calculated using the "margins" command in Stata.

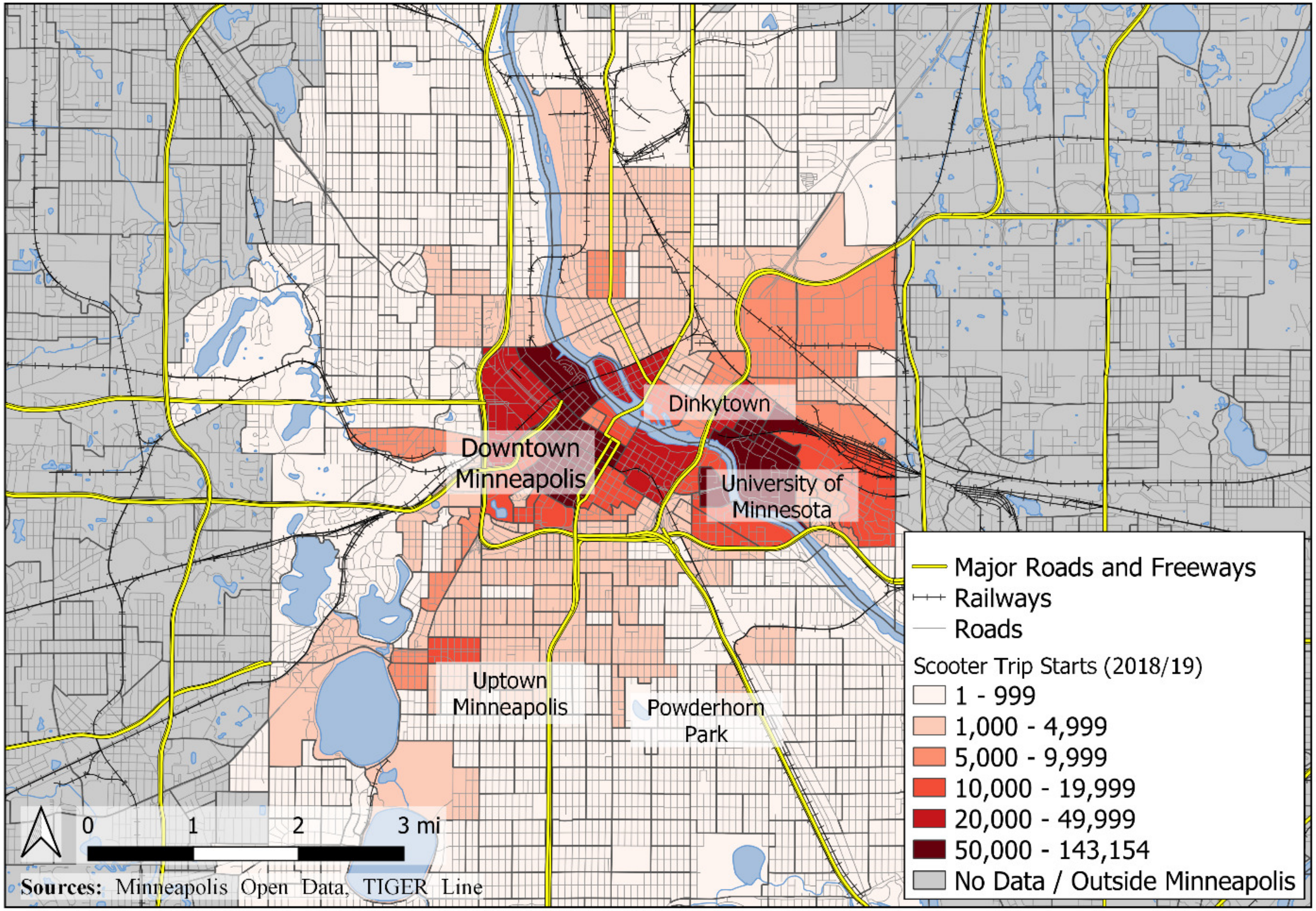

Figure 1. Total Scooter Trip Origins by Blockgroup in Minneapolis, MN, 7/10/2018 – 11/26/2019

Table 1. List of Variables

Variable	Source	Description
Trip Count	City of Minneapolis Open Data Portal	The average number of trips starting in a blockgroup per day between July 10, 2018 and November 26, 2019 (The duration during which the scooter program was active).
Job Density	Census Bureau LEHD LODES WAC	Number of jobs located inside a blockgroup, divided by land area.
Population Density	2019 American Community Survey	Number of people residing inside a blockgroup, divided by land area.
Density of Young People	2019 American Community Survey	Number of people aged 18-25 residing inside a blockgroup, divided by land area.
Density of Food Establishments	OpenStreetMap	Number of food and drink establishments located inside a blockgroup, divided by land area.
Distance to Downtown	Own Calculations	Distance from a blockgroup's centroid to the center of Minneapolis' most job-dense census tract per LODES WAC.

Table 2. Coefficients for NB Regression

Variable	Coefficient	Standard Error
Job Density	0.0568***	(0.0174)
Population Density	-0.0845***	(0.0166)
Density of People Aged 18-25	0.000384***	(6.65e-05)
Density of Food Establishments	0.0236*	(0.0136)
Distance to Downtown	-0.649***	(0.0438)
Constant Term	3.621***	(0.238)
ln Alpha	-0.0226	(0.0903)
Observations	408	

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

3. Findings

Table 2 presents the regression model as is commonly displayed in the literature, while Table 3 shows marginal effects when using the "starting point" value shown for each independent variable. Table 3 also shows where that starting point is in the distribution of all values in the data. Figure 2 presents marginal effects and confidence intervals for a single variable – distance to downtown – across its entire range, holding all other variables constant at their medians.[4]

4 Differences between the extreme marginal effects in Table 3 for Downtown and Dinkytown and the values shown in Figure 2 as distance approaches zero exist because Figure 2 presents marginal effects for Distance to Downtown with all other variables held at their sample medians, while Table 3 uses values for the specific neighborhoods.

Table 3. Marginal Effects for NB Regression

Using as a starting point:	Downtown	Dinkytown	Uptown	Powderhorn Park	Citywide Medians
Job Density (in 1000s per Square Kilometer)					
Starting Point	89.03	7.08	4.93	0.76	0.54
Percentile	100%	96%	93%	59%	50%
Marginal Effect	417.9	336.6	0.990***	0.145***	0.0894***
Standard Error of Marginal Effect	(754.7)	(364.9)	(0.339)	(0.0409)	(0.0266)
Population Density (in 1000s per Square Kilometer)					
Starting Point	1.36	4.70	3.46	2.78	1.42
Percentile	48%	91%	85%	78%	50%
Marginal Effect	-578.9	-466.3	-1.371***	-0.201***	-0.124***
Standard Error of Marginal Effect	(870.7)	(535.9)	(0.414)	(0.0462)	(0.0253)
Density of Young People (People aged 18-25 per Square Kilometer)					
Starting Point	310.76	15,511.80	2,111.45	355.41	195.03
Percentile	62%	100%	96%	66%	50%
Marginal Effect	2.634	2.121	0.00624***	0.000915***	0.000563***
Standard Error of Marginal Effect	(3.963)	(2.573)	(0.00180)	(0.000163)	(9.98e-05)
Density of Food Establishments (per Square Kilometer)					
Starting Point	7.55	42.61	15.59	9.11	0.00
Percentile	83%	100%	95%	87%	0%
Marginal Effect	162.0	130.5	0.384	0.0563	0.0347*
Standard Error of Marginal Effect	(250.7)	(182.0)	(0.273)	(0.0366)	(0.0187)
Distance to Central Business District (in 1000 Meters)					
Starting Point	0.62	3.10	3.12	4.49	4.97
Percentile	1%	22%	22%	43%	50%
Marginal Effect	-4,446	-3,581	-10.53***	-1.544***	-0.951***
Standard Error of Marginal Effect	(6,612)	(3,750)	(1.850)	(0.177)	(0.102)

Note: *** p<0.01, ** p<0.05, * p<0.1

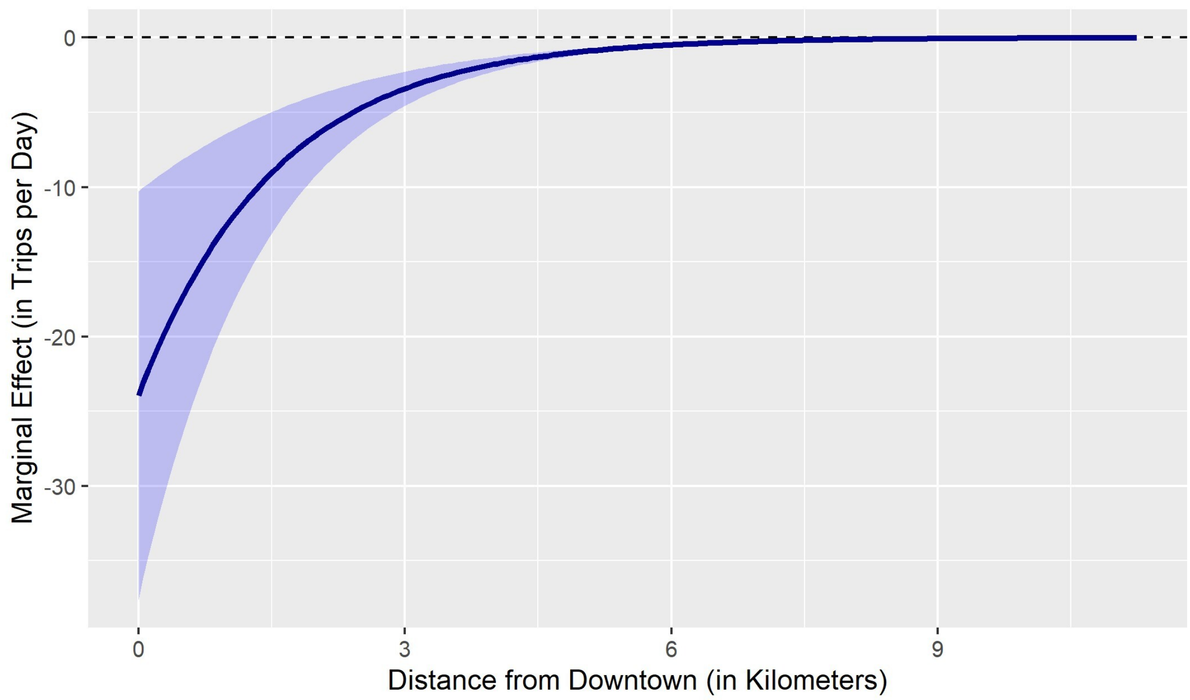

Figure 2. Marginal Effects & Confidence Intervals for Distance to Downtown

All other variables held at citywide medians.

As shown in Table 3 and Figure 2, magnitudes of relationships – when expressed as the number of trips per day associated with a one unit change in one of the land use or urban form variables - can vary dramatically across their own range and across different values for other independent variables as a consequence of the mathematical properties of the NB model. This is not immediately obvious from Table 2: Associations are of an economically almost irrelevant magnitude when variables are held at their median, whereas they are very large but statistically insignificant at a starting point resembling a downtown location. Comparing the implied marginal effects between the medians scenario and a scenario resembling the Uptown Minneapolis entertainment district, almost every association differs by approximately one order of magnitude.

The findings from our simple model have several implications for the literature that should be explored, even in models that have a larger set of independent variables: First, relationships between density and trip generation are nonlinear. Secondly and resulting from this first finding, studies using NB or similar models to study trip generation should report associations for several different starting points resembling real-world locations. Doing so would convey information relevant to practitioners – such as the number of incremental scooter trips associated with a particular change in land use – in a more accessible manner, without the risk of misleading by referring to one particular location. Lastly, a property of non-linear models, such as NB, appears to be that the confidence interval around marginal effects grows large near the upper limit of the data. Unfortunately, these are precisely the locations where scooter trips tend to be most common.

Acknowledgements

This research flowed from an earlier research project funded by the METRANS Transportation Consortium. We thank METRANS for funding to support collection of the scooter and some of the land use data used here. Zakhary Mallett was a participant in that earlier research, and we thank him for background research and econometric analysis related to that earlier work.

Submitted: March 03, 2022 AEST, Accepted: May 01, 2022 AEST

REFERENCES

Bai, Shunhua, and Junfeng Jiao. 2020. "Dockless E-Scooter Usage Patterns and Urban Built Environments: A Comparison Study of Austin, TX, and Minneapolis, MN." *Travel Behaviour and Society* 20 (July): 264–72. https://doi.org/10.1016/j.tbs.2020.04.005.

Gehrke, Steven R., Brendan J. Russo, Bita Sadeghinasr, Katherine R. Riffle, Edward J. Smaglik, and Timothy G. Reardon. 2021. "Spatial Interactions of Shared E-Scooter Trip Generation and Vulnerable Road User Crash Frequency." *Journal of Transportation Safety & Security* 0 (0): 1–17. https://doi.org/10.1080/19439962.2021.1971813.

Huo, Jinghai, Hongtai Yang, Chaojing Li, Rong Zheng, Linchuan Yang, and Yi Wen. 2021. "Influence of the Built Environment on E-Scooter Sharing Ridership: A Tale of Five Cities." *Journal of Transport Geography* 93 (May): 103084. https://doi.org/10.1016/j.jtrangeo.2021.103084.

Jiao, Junfeng, and Shunhua Bai. 2020. "Understanding the Shared E-Scooter Travels in Austin, TX." *ISPRS International Journal of Geo-Information* 9 (2): 135. https://doi.org/10.3390/ijgi9020135.

Merlin, Louis A., Xiang Yan, Yiming Xu, and Xilei Zhao. 2021. "A Segment-Level Model of Shared, Electric Scooter Origins and Destinations." *Transportation Research Part D: Transport and Environment* 92 (March): 102709. https://doi.org/10.1016/j.trd.2021.102709.