Using Propensity Score Matching Technique to Address Self-selection in Transit-Oriented Development (TOD) Areas

Arefeh Nasri
Research Scientist
Department of Civil and Environmental Engineering
University of Maryland
aanasri@umd.edu

Carlos Carrion
Research Scientist
Department of Civil and Environmental Engineering
University of Maryland
carrion@umd.edu

Lei Zhang (Corresponding Author)
Herbert Rabin Distinguished Professor
Department of Civil and Environmental Engineering
University of Maryland
Phone: 301-405-2881
Email: lei@umd.edu

Babak Baghaei
Department of Civil and Environmental Engineering
University of Maryland

Word Count: 1,496 words + 2 Tables

Acknowledgement

This research is funded partially by the National Transportation Center at the University of Maryland. The authors would like to thank the MWCOG for providing household travel survey data for this analysis. Findings presented in this paper do not necessarily represent the official views of the sponsoring agency. The authors are solely responsible for all statements in the paper.

1. INTRODUCTION

Many studies have investigated the effects of transit-oriented development (TOD) on travel behavior, especially on transit ridership. However, most studies do not explicitly and effectively address the issue of residential self-selection in their analyses. The aim of this paper is to use cross-sectional data and propensity score matching (PSM) technique to quantify the contribution of residential self-selection to the analysis of mode choice in TOD areas across the metropolitan areas of Washington, D.C. and Baltimore, MD. The results of PSM indicate that, even though the self-selection effect is considerable in the analysis of mode choice in TOD areas (about 7.65% in Washington, D.C. and 5.05% in Baltimore), living in TOD still has a significant impact on encouraging transit and other active modes of transportation.

Many studies have suggested that, in general, residents of TOD make fewer auto trips, more transit trips, and generate lower VMT (1, *2, 3, 4, 5, 6, 7, 8, 9, among others*). However, among these analyses, only a few address the effect of residential location choice on travel behavior in TOD areas. Failing to address the issue of self-selection would create biased estimations in the analyses of the relationships between urban form and travel behavior.
Due to the higher accessibility and design features associated with TOD, it is generally expected that the share of transit use in TOD areas is higher, compared to non-TOD areas, for various trip purposes. However, it is not clear whether residents of TOD areas tend to drive less (lower VMT) because of their personal preference, or because the TOD setting encourages them to do so. Thus, the effect of TOD on travel behavior should be separated from that of residential self-selection.
In this study, a propensity score matching method is applied to control for self-selection bias and estimate the net effect of living in TOD areas on mode share.

The issue of residential self-selection is addressed in a few recent TOD studies. However, a strong causal link that fully addresses aspects of self-selection has not yet been established, despite various methods and techniques that have been proposed (10, 11, 12, 13).

Chatman (2005) suggested that self-selection plays a considerable role for pro-transit people, but not as much for "auto-oriented" people who move to TOD areas (14). Lund (2006) investigated the effect of self-selection on mode choice by surveying people before and after moving to TOD areas in three California cities. Their analysis found that residents moved to TODs mainly because of its easy access to transit and are more likely to use transit than those who do not live in TODs (*15*).

PSM is based on controlling for the observed characteristics related to a treatment (e.g., a certain travel behavior) and to what extent people in either treatment or control group share similar characteristics (*16*).

2. METHODOLOGY

Propensity score matching is a method for estimating the treatment effect in observational studies. In contrast to controlled experimental studies, the treatment in observational studies is not assigned randomly and there is the possibility for error in estimating the effect, due to issues like self-selection or systematic errors in selecting treated units.

In this study, the treatment is living in TOD, and the outcome of interest is the measure success for TOD policy, which is transit mode share. The average treatment effect would be the difference observed in non-auto (transit, walk, and bike) mode share between TOD and non-TOD residents. PSM would match TOD and non-TOD residents based on their socioeconomic and attitudinal characteristics and compare mode choice between the matched households. The propensity score is the probability of a household choosing to live in the TOD, given their observed characteristics. This probability can be estimated using discrete choice models. Therefore, comparing the matched households—one from TOD (treatment group) and one from non-TOD (control group)—could roughly translate to having an ideal experiment, where the assignment of households to TOD is random. In this setting, the average treatment effect is the average difference in an auto mode share between the matched TOD and non-TOD households.

If the matched households from the treated and controlled groups are systematically different in terms of their travel behavior, it implies that the self-selection effect is minimal and living in TOD truly influences people's travel choices (*13*).

Data for this analysis comes mainly from two sources: first, the 2007-2008 household travel survey data for Washington, D.C. and Baltimore, MD, and the 2005 land use data, both obtained from Metropolitan Washington Council of Government (MWCOG) and the Baltimore Metropolitan Council (BMC); and second, the geocoded rail transit station data for the two metro areas from the national TOD database. TODs are defined as the traffic analysis zones (TAZ)s containing at least one rail transit station surrounded by high-density, mixed use and walkable development (within its half-mile buffer), as seen in previous studies (1, 2*, 13*).

Treatment variable D is a binary variable that determines whether the household lives in TOD: D=1 for treated observations and D=0 for controlled observation. The outcome of interest is the non-auto mode share for each household based on the travel survey.

The propensity score is then estimated for each household using the probit regression model. Socioeconomic variables used for estimation are household size, number of workers in the household, auto ownership, and household income. Probit model D is the dependent variable (whether the household lives in TOD or not) and x is the vector of independent variables. The propensity score here is defined as the conditional probability of a household receiving treatment (i.e., living in TOD) given the background variables (i.e., socio-demographic variables):

$$p(x) = Pr\ (D = 1|x) = E\ (D|x) \tag{1}$$

P(x): Propensity score
$Pr\ (D = 1|x)$*: probability of a household living in TOD*
x= Vector of the households' sociodemographic characteristics
D= treatment effect (i.e., living in TOD)

After estimating the propensity score for each household, observations are matched between the treatment and controlled group, using the nearest neighbor matching technique. After matching on propensity score, the average treatment effect (ATE) is calculated as the difference between the outcomes if treated, and the outcome if they had not been treated:

$$ATE = E\ (\Delta|p(x), D = 1) = E\ (y_1|p(x), D = 1) - E(y_0|p(x), D = 0) \tag{2}$$

P(x): propensity score
D: binary treatment variable
Y: outcome variable

3. RESULTS AND DISCUSSIONS

In the first step, a linear regression model was built to estimate the effect of living in TOD on the non-auto mode share without controlling for self-selection. Non-auto mode share was modeled as a function of whether a household lives in a TOD zone, combined with the sociodemographic characteristics of the household and the results are shown in Table 1. As the table indicates, households living in TOD have a 24.3% higher non-auto mode share in Washington, D.C., and an 11.1% higher non-auto mode share in Baltimore, MD. This implies that people who live in TOD with better and more efficient access to transit tend to drive less.

Table 1. Non-Auto Mode Share Models' Results

Linear Regression Model				
	Washington, D.C.		**Baltimore, MD**	
	Coefficient	P>t	Coefficient	P>t
Dependent Variable: % Non-Auto Mode Share				
TOD	24.28	0.00	11.07	0.00
HH size	1.79	0.00	1.72	0.00
HH #worker	5.67	0.00	6.06	0.00
HH car ownership	-13.28	0.00	-16.77	0.00
HH income	-0.27	0.02	-0.15	0.26
Constant	34.43	0.00	41.44	0.00
Number of observations	10,719		9,029	
Probit Regression Model				
HH size	-0.064	0.001	-0.12	0.000
HH #worker	0.18	0.000	0.21	0.000
HH car ownership	-0.56	0.000	-0.16	0.000
HH income	0.055	0.000	-0.07	0.000
Constant	-0.98	0.000	-1.28	0.000

The next step is to build the probit regression model to estimate the probability of each household living in TOD, representing the propensity score. The range of propensity score for the Baltimore area is from 0.0032 to 0.12. The mean propensity score is not different between the treatment group and controlled group.

Table 2 shows the percentiles of the propensity score distributions of the households, including the smallest and largest propensity scores. Each household is classified s into five blocks based on the propensity scores, and the households from the treatment and controlled groups are matched using the nearest neighbor method within each block.

Table 2. Propensity Score Percentiles

	Washington, D.C.		Baltimore, MD	
	Percentiles	Smallest	Percentiles	Smallest
1%	0.0010500	0.0000104	0.00105	0.0000104
5%	0.0070982	0.0000106	0.007098	0.0000106
10%	0.0167195	0.0000168	0.01672	0.0000168
25%	0.0376509	0.0000168	0.037651	0.0000168
50%	0.0755127		0.075513	
		Largest		Largest
75%	0.1361158	0.421146	0.136116	0.421146
90%	0.1767684	0.442826	0.176768	0.442826
95%	0.2217830	0.442826	0.221783	0.442826
99%	0.3359757	0.442826	0.335976	0.442826

The average treatment effect (ATE) estimate after controlling for self-selection in Washington, D.C. is 16.65% using the propensity score matching, and 24.3% using linear regression. In Baltimore, it is 6.02% using the propensity score matching and 11.07% using linear regression (much smaller effect than in Washington, D.C.). There are differences in the ATE estimates using both PSM and linear regression, and an important reason is that PSM is a nonparametric method while linear regression assumes a linearity relationship with the variables.

4. CONCLUSIONS

Propensity score matching method tries to simulate perfect experimental conditions for evaluating the effect of TOD by matching residents of TOD and non-TOD areas based on their socio-economic characteristics. In summary, the findings indicated that in both Washington, D.C. and Baltimore, self-selection accounts for about 40% of the effect of the TOD in reducing auto-mode share. Although the effect of self-selection is significant, it is still probable that TOD plays an important role in influencing the mode choice of its residents. Therefore, self-selection is a significant factor in the association between TOD and low auto-mode share and should be considered in any prediction.

However, this approach is limited, in that it cannot account for selection-on-unobservables. In other words, it is possible that some other unobservable variables are leading households to choose to live in TOD, increasing their likelihood of using other travel modes besides auto. This issue can be investigated in more detail in a future analysis to capture the unobserved factors using detailed data on housing, attitudes, safety, etc.

REFERENCES:

1. Nasri, A., & Zhang, L. (2014). The analysis of transit-oriented development (TOD) in Washington, D.C. and Baltimore metropolitan areas. *Transport Policy*, *32*, 172-179.

2. Zamir, K., Nasri, A., Baghaei, B., Mahapatra, S., & Zhang, L. (2014). Effects of Transit-Oriented Development on Trip Generation, Distribution, and Mode Share in Washington, D.C., and Baltimore, Maryland. *Transportation Research Record: Journal of the Transportation Research Board*, (2413), 45-53.
3. Arrington, G. B., & Cervero, R. (2008). TCRP Report 128: Effects of TOD on Housing, Parking, and Travel. *Transportation Research Board of the National Academies, Washington, D.C.*, *3*.
4. Shatu, F. M., & Kamruzzaman, M. (2014). Investigating the link between transit-oriented development and sustainable travel behavior in Brisbane: A case-control study. *Journal of Sustainable Development*, *7*(4), 61-70.
5. Faghri, A., & Venigalla, M. (2013). Measuring travel behavior and transit trip generation characteristics of transit-oriented developments. *Transportation Research Record: Journal of the Transportation Research Board*, (2397), 72-79.
6. Chatman, D. G. (2008). Deconstructing development density: Quality, quantity and price effects on household non-work travel. *Transportation Research Part A: Policy and Practice*, *42*(7), 1008-1030.
7. Nixon, H., Boarnet, M., Houston, D., Spears, S., & Lee, J. (2015). *Changes in Transit Use and Service and Associated Changes in Driving Near a New Light Rail Transit Line* (No. CA-MTI-1108).
8. Shen, Q., Chen, P., & Pan, H. (2016). Factors affecting car ownership and mode choice in rail transit-supported suburbs of a large Chinese city. *Transportation Research Part A: Policy and Practice*, *94*, 31-44.
9. Huang, X., Cao, X., & Cao, J. (2016). The association between transit access and auto ownership: evidence from Guangzhou, China. *Transportation Planning and Technology*, *39*(3), 269-283.
10. Lund, H., Willson, R. W., & Cervero, R. (2006). A re-evaluation of travel behavior in California TODs. *Journal of Architectural and Planning Research*, 247-263.
11. Cervero, R. (2007). Transit-oriented development's ridership bonus: a product of self-selection and public policies. *Environment and planning A*, *39*(9), 2068-2085.
12. Cervero, R. (2008). *Effects of TOD on housing, parking, and travel*. Transit Cooperative Research Program (TCRP) report 128.
13. Park, K., Ewing, R., Scheer, B. C., & Khan, S. S. A. (2017). Travel Behavior in TODs vs. non-TODs: Using Cluster Analysis and Propensity Score Matching. *Presented at the Transportation Research Board 96th Annual meeting* (No. 17-00899).
14. Chatman, D. G. (2005). *How the built environment influences non-work travel: theoretical and empirical essays* (Doctoral dissertation: University of California, Los Angeles).
15. Lund, H. (2006). Reasons for living in a transit-oriented development, and associated transit use. *Journal of the American Planning Association*, *72*(3), 357-366.

16. Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport Reviews*, *29*(3), 359-395.