ECO395M
Final Report
Yefu Chen

Nonlinear relationship between public transit and related short-trip E-scooter usages: A empirical and preliminary study in Austin, Texas

Summary: As an emerging micro-mobility, Dockless E-scooter provides a way for short-distance traveling and can be a solution to the first/last mile issue of public transit. Previous studies investigated the spatial-temporary patterns of E-scooters while lacked the relationship between E-scooter usage and public transit. Also, they studied general E-scooter trips, while the characteristics of E-scooter trips, such as long/short trips, were ignored. Two research questions are asked to fill up the research gaps; what *is the collective importance of built environment attributes? What is the nonlinear relationship between public transit and related short-trip E-scooter usages?* This study answered the research questions by exploring the nonlinear relationship between public transit and related short-trip E-scooter usages through gradient boosting machines in Austin, Texas. Results indicate that the most important predictor is median earnings, while built environment attributes are the second most important predictors. Also, the threshold of important predictors in predicting public transit related short-trip E-scooter usages is significant. Then, this study makes several important research and policy implications suggestions based on these results.

Keywords: E-scooter; public transit; nonlinear relationship

1. Introduction

In the City of Austin, Capital Metropolitan Transportation Authority operates the public transit system. The whole system serves more than 80,000 riders daily in 2018 (Statesman, 2018). Although the public transit system seemed to operate Ill, many areas are still unserved (Jiao, 2017).

Dockless electrical scooter (E-scooter) is a new term defined as the electrical-powered scooters without fixed "stations" or "docks", which has been introduced and has become popular in recent years around the US. This popularity could be convenient, and this new mode may fill the travel gaps of public transit systems (Zariff, et al., 2019). This new travel mode reached Austin in Texas in April 2018 (Lekach, 2019; Roman, 2019). Austin Transportation Department (ATD) has published several regulations for this mode to impose safety and efficiency (City of Austin, 2018). ATD has also mobilized licensed operators to manage the city's downtown project coordination zone and provide additional units in other areas (City of Austin, 2018).

Recently, there were rising numbers of studies involving dockless shared E-scooters (Zou, et al., 2020; Tuncer, et al., 2020; Caspi, et al., 2020; Jiao & Bai, 2020). Zou et al. (2020) found that the E-scooter usages gathered around main streets and mixed land use areas. Tuncer et al. (2020) investigated the riders' behaviors through video recordings and pointed out that built environment, especially road design, significantly affect these usages. Caspi et al. (2020) added sociodemographic as a significant factor. Jiao and Bai (2020) further explored the spatial-temporal patterns of E-scooter usages. They further pointed out that these factors, particularly population density and education level, play an essential role in predicting E-scooter usages.

Studies indicated that E-scooters could offer an economically and environmentally friendly way to travel short distances for riders, dealing with the first/last mile issues of public transit. A report has pointed out that E-scooters expanded the reach of public transit stops, thereby solving the last mile issue in transit (Irfan, 2018). However, another study has warned that E-scooters were substitutive to public transit usages (Lee, et al., 2019). Specifically, riders who were used to take public transit for traveling may transfer to take E-scooters.

Previous research preliminarily explored the relationship between E-scooters and public transit. Some research analyzed patterns from observed E-scooter ridership data and riders' characteristics (Degele, et al., 2018; Smith & Schwieterman, 2018). Other reports (Tillemann & Feasley, 2018; Whitehead, 2019; Barnard, 2019; Bliss, 2019; Lee, et al., 2019) have provided pilot perspectives, exploring the relationship between E-scooters and public transit. Tillemann and Feasley (2018) mentioned that E-scooters could expand access to public transportation and promote public transit usages in low public transit accessibility neighborhoods. Whitehead (2019) concluded the pilot program of E-scooters in Chicago and found that this mode of transportation could promote public transit and decrease the instances of driving alone. Barnard (2019) also indicated that E-scooters generally reduced the ridership in public transit by providing a cheaper but more effective travel mode because they could provide a reliable way for residents in areas where public or affordable transit is less available.

Previous studies have explored the E-scooter usages and tried to build the linkage between Public transit and related short-trip E-scooter usages, while most of them ignored the impacts of trip characteristics, such as long/short trips. The effects of urban environments and public transit service on the spatial distribution of E-scooter trips remain unclear. Both research gaps led policymakers to have a hard time implementing policies to make cooperation between E-scooters and public transit without a clear understanding of the current situation. Further, there is no studies explored the nonlinear relationship between E-scooters and public transit.

Accordingly, this study aims to fill existing research gaps by providing an empirical and preliminary perspective on the relationship between E-scooters and public transit. Since it is a preliminary study, I only focused on short E-scooter trips starting from areas near bus stops. I will keep working on this topic and plan to develop it as an article. In this study, there are two research questions in this study:

- RQ1: What is the collective importance of built environment attributes?
- RQ2: What is the nonlinear relationship between public transit and related short-trip E-scooter usages?

The rest of the paper was structured as follows. First, I introduced study areas, variables, and analysis methods. Then I described the results of models. At last, I discussed and concluded these results.

2. Material and methods

2.1. Study area

The basic study area in Austin, Texas. It is one of the cities that first launched an E-scooter program on the streets. There is an ordinance to regulate the number of devices and allocation areas for four dominant E-scooter companies (City of Austin, 2018). Figure 1 presents the final study area (E-scooter area), which integrates these four leading operators (JUMP, Lyft, Bird, and Lime).

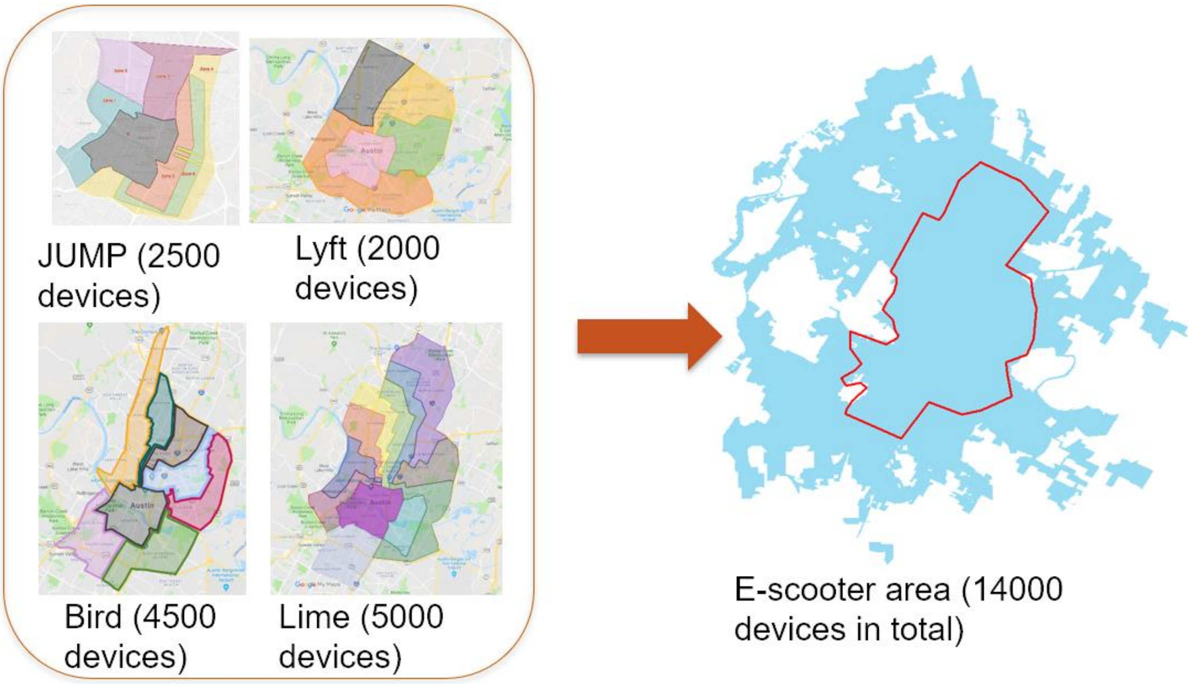

Figure 1. Study area

2.2. Data

Table 1 presents the definitions and statistical description of variables in this study.

2.2.1. E-scooter trips

This study used a dataset that comprised E-scooter trips (n = 2,015,467) recorded for 11 months (April 2018 to February 2019) from the City of Austin data portal; however, this dataset was off due to confidence concern. Each record represented one trip made by E-scooters with essential trip information, including locations of origin and destination, trip durations, and trip distances. The data obtained from Austin include actual origins and destinations of E-scooter trips. This study used the following rules for data cleaning:

- Location rules (either origins or destinations must be within TC; the origin and destination cannot within the same buffer of TC)
- Distance (0.062 miles to 10 miles)
- Duration (60 seconds to 3600 seconds)
- Average speed (less than 25 miles per hour)

The final data contained a total of 1,056,782 E-scooter trips, with 816,324 origins and 818,916 destinations. I only chose the origins of E-scooter trips less than 1 miles as the final dependent variables in this study. It represented the origins of short E-scooter trips.

Then, I used the 0.25-mile buffers surrounding the bus stop to capture the number of short E-scooter trips' origins for each bus stop since 0.25 miles is used as the standard walking distance (Yang & Diez-Roux, 2012; Jia, et al., 2019; Ferrer & Ruiz, 2018).

2.2.2. Public transits

In this study, I measured public transit service's influence on E-scooter trips close to public transit infrastructures using Transit deserts index (Jiao & Dillivan, 2013; Jiao, 2017). There are two basic concepts in this measurement, transit demand and transit supply. The transit demand refers to the transit-dependent population which is the sum of the transit-dependent household population, population aging from 12 to 15, and non-institutional population. The transit-dependent household population is the population who have to take public transit for moving, and the noninstitutionalized population was defined by the United States Census Bureau (the Census Bureau, 2016). The z-scored transit-dependent population density represent the transit demand.

Transit supply is measured using the number of public transit facilities, the average frequency of transit service, number of transit routes, length of sidewalks, length of bike routes, length of low-speed limit roads, and intersections. Each factor is measured as related densities. All densities are z-scored and summed as the index. Then, the z-scored index represents the transit supply.

The differences between the transit supply and demand are the indexes of transit gaps, which represents the level of public transit services in this study.

2.2.3. Land use factors

Land use factors is a part of built environment attributes. I applied 0.6 miles buffers to capture land use factors referencing previous studies (Frank, et al., 2006; Saelens & Handy, 2008; Vojnovic, et al., 2013). Frank et al. (2007) indicated that land-use factors, including single-family residential, multifamily residential, commercial, office, institutional, industrial, and park or recreation, can affect individual travel behaviors. Following their study, I chose six types of land uses: single-family, multifamily, commerce, office, government and education, and green spaces. I excluded industrial because the study area had only a few industrial lands.

Then, I measure the land use mix index to present the land diversity situations, capturing the distribution of development area on six types of land uses in each buffer as follows:

$$Land\ use\ mix\ index\ =\ -A\ /\ (ln(N));$$

$$A = (b1/a) * ln(b1/a)\ +\ (b2/a) * ln(b2/a)\ +\ (b3/a) * ln(b3/a)\ +\ (b4/a) * ln(b4/a)\ +\ (b5/a) * ln(b5/a)\ +\ (b6/a) * ln(b6/a),$$

where a equals the total square mile of land for all six types of land uses present in our buffers. bi measures areas of land use. N means the number of six types of land uses with areas larger than zero.

2.2.4. Road design

Road design is the other part of built environment attributes. Similar to land use factors, the 0.6-mile buffers were used to capture rod design factors. There are two factors in road design. The first one is road. It refers to the total length of road within the 0.6 miles buffer of bus stops. The lengths of the primary road, secondary road, tertiary road, residential road, and motorway were summed as the total road length. Second, I captured the number of intersections of road networks within 0.6-mile buffers of bus stops as the junction variable.

2.2.5. Sociodemographic factors

Previous studies indicated that sociodemographic and economic factors affect travel behaviors (Lu & Pas, 1999; Yang, et al., 2016). Lu and Pas (1999) mentioned that household-related factors could affect travel behaviors given that transportation expenses cannot depend on individual income. Yang et al. (2016) found that sociodemographic characteristics and the chosen context could determine travel choice. Although bike shares are different from E-scooters, both transit modes play significant micro-mobility roles, offering an alternative travel mode for short-distance trips.

I integrated the census block group data from the 2017 American Community Survey, because there is a ready to go dataset. I captured population, median personal earnings, number of household, and median income in households.

Table 1. Descriptive statistics

	Variables	**N**	**Description (units)**	**Mean**	**St. Dev.**
Dependent variable	Origins of short trips	1802	Number of origins with distance shorter than 1 mile (1000 count) within 0.25-mile buffer surrounding bus stops	2.95	10.71
Independent variables	Public transits		Differences between z-scored transit supply and z-scored demand	0.65	4.72
	Land-use mix index		Referencing Frank et al. (Frank, et al., 2007)	0.69	0.11
	Road		Length of road within the census block group where each bus stop locates	18.39	4.72
	Junction		Number of junctions within the census block group where each bus stop locates	155.13	50.51
Controlling variables	Population		Number of people of census block group where each bus stop locates (per 1000)	1.86	1.23
	Median earnings		Median personal earning annually of census block group where each bus stop locates (per 1000 dollars)	38.75	23.91
	Household number		Number of households of census block group where each bus stop locates (per 1000)	0.78	0.48
	Household median income		Median household income annually of census block group where each bus stop locates (per 1000 dollars)	556.07	28.94

2.3. Methods

I applied Gradient Boosting Machines (GBM) to explore the nonlinear relationship between E-scooters and public transit. It combines decision tree learning and gradient boosting, and widely accepted in current transportation research (Tao, et al., 2020). Compared to traditional linear regression, there are multiple advantages of GBM models. First, it can affect performance better in multicollinearity situations than traditional linear regression models. Typically, multicollinearity is a problem for inference in statistics and analysis. Nevertheless, the GBM is more like a black-box model ignoring the process of inference but focusing on predictions and evaluations. It applies stepwise strategies for the selection of features to robustly go around multicollinearity. Second, GBM uses the importance of features instead of statistical numbers, such as p-value, in measuring the significance. It produces the importance of predictors in predicting outcome variables, which can be a better way to measure the actual effects of predictors on the outcome variables. In terms of this, I used the GBM to explore the nonlinear relationships. I used R 4.0.4 and “gbm” package to perform the GBM models.

To define the input hyperparameters, I introduced the grid searching (Joseph, 2018). It allows researchers to get the best combination of hyperparameters based on the least root means square errors (RMSE). In this study, the test pool of hyperparameters was based on my experience. In the grid setting, I considered five hyperparameters: number of trees, shrinkage, interaction depth, number of minimum observations, and bag fraction. The number of trees refers to the number of trees. Typically, a higher number of trees can improve the accuracy of models but may cause overfitting issues. I tested two numbers of trees, 1000 and 5000, in this study. The shrinkage is a hyperparameter of gradient boosting regularization. It can help to promote the generalization ability of models. I tested two shrinkages, 0.01 and 0.1, in this study. Interaction depth refers to the maximum nodes of each tree. I tested three interaction depths, 2, 5, and 8, in this study. The number of minimum observations refers to the minimum number of observations in trees’ terminal nodes. I introduced two numbers (2 and 5) to test the best performance of models. Finally, I tested three bag fractions (0.1, 0.2, and 0.3) which are part of the training set observations randomly selected to propose the next tree in the expansion. Hence, the overall grid search pool includes 72 combinations. Due to the limited performance of the laptop, I cannot introduce an extra number of combinations in this study.

The analysis method was processed as following. First, I divided the raw data into train pool and test pool, using 0.8 as the rate. Then, I introduced the above settings intro the grid search pool and applied the grid searching process. After grid searching, I get the best combination of hyperparameters (number of trees = 1000, interaction depth = 8, shrinkage = 0.1, number of minimum observations = 5, and bag fraction = 0.3). Also, it returned the optimal trees as 268. I built the GBM based on grid searching results and optimal trees. Then, I estimated the RMSE separately based on train pool and test pool. The scores were closed (3.16 vs. 4.51).

3. Results

Before the GBM results, I applied a correlation analysis. The correlation graphic is shown as Figure 2. It demonstrates that the dependent variable, origins of short trips, is highly correlated with sociodemographic factors (population and median earnings) and road design (road and junction). Both play negative role in the correlations.

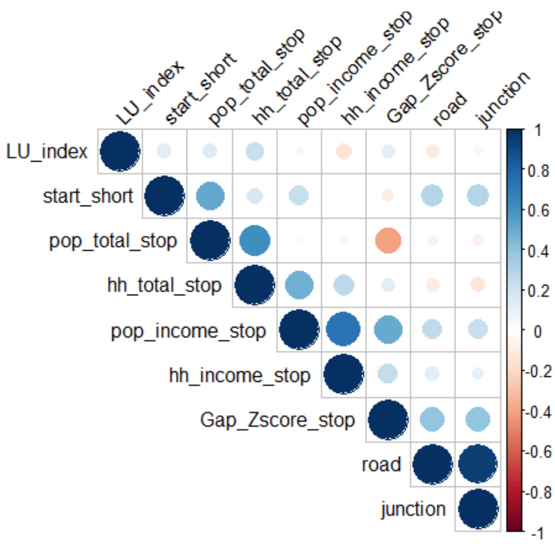

Figure 2. Correlation results

Figure 3 shows the relative importance of independent variables and controlling variables in predicting the origins of short E-scooter trips. The top three variables are median earnings (47.45%), public transit (14.85%), and road (9.87%). The collective importance of built environment attributes is 23.84% (sum of road, junction, and land use mix index).

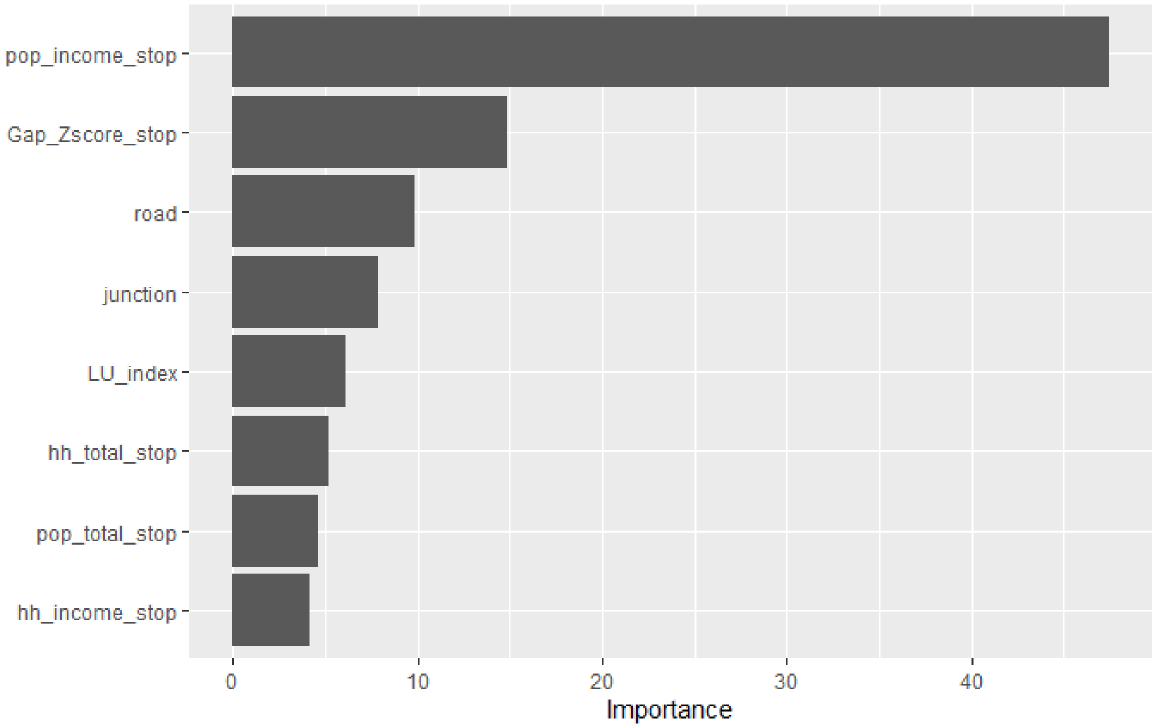

Figure 3. Relative importance of variables

Then, I focused on the partial dependence plots of median earnings, public transit, and road (shown as Figures 4-6). The results of these plots indicate that the overfitting issues may be significant in this study since the fluctuations are clear. When interpreting the plots, I spend more time on the general trends of the relationships between variables than the fluctuations.

Figure 4 indicates that the relationship between median earnings and short E-scooter trips is negative at first and positive at last. This variable shows a clear threshold effect. When the median earnings grow from 0 to 14, the origins of short E-scooter trips decrease from around 50 to nearly 0. When the median earning grows from 14 to around 80, the origins of short E-scooter trips increase from around 0 to 2. When the median earning grows from around 80 to around 90, the origins of short E-scooter trips increase from around 2 to 11. After this, the increase of median earning is positively correlated with the increase of origins.

According to Figure 5, public transit has a positive relationship with short E-scooter usages. When the level of public transit service is bad, its effects on short E-scooter usages are insignificant. Nevertheless, when it grows from around -3 to -2, the short E-scooter usages increased from no usage to around 5. Furthermore, there is a phase that short E-scooter usages do not change according to the increase of public transit service level when it grows from -2 to around 3. When the public transit service is larger than 3, the relationship turns positive again.

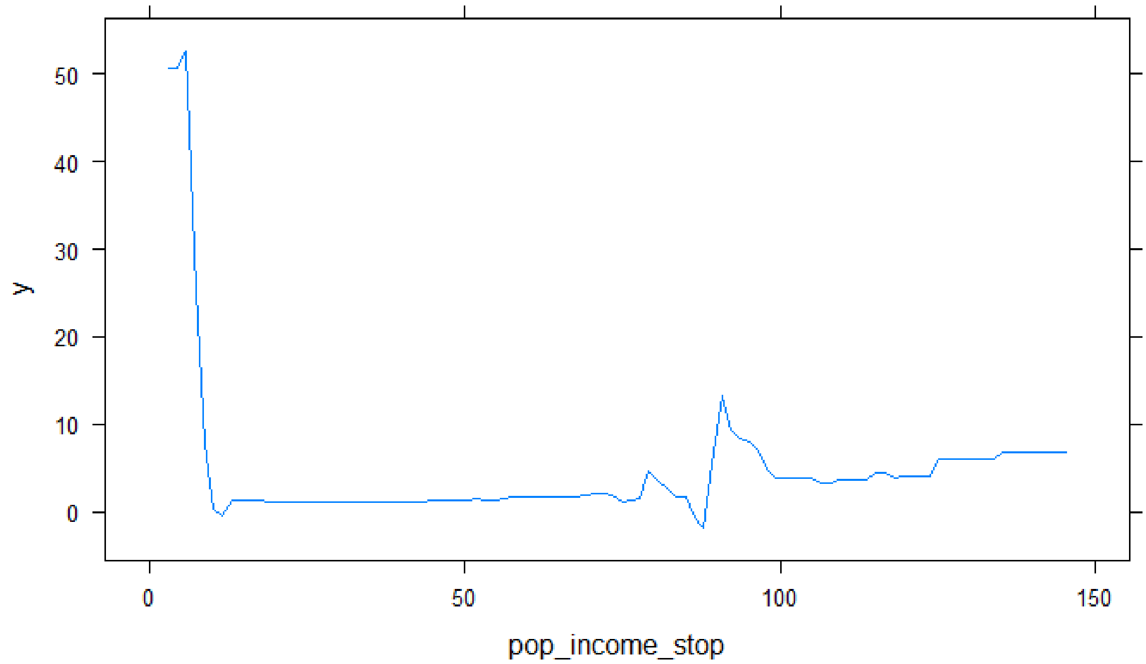

Figure 4. The relationships between median earnings and origins of short E-scooters trips

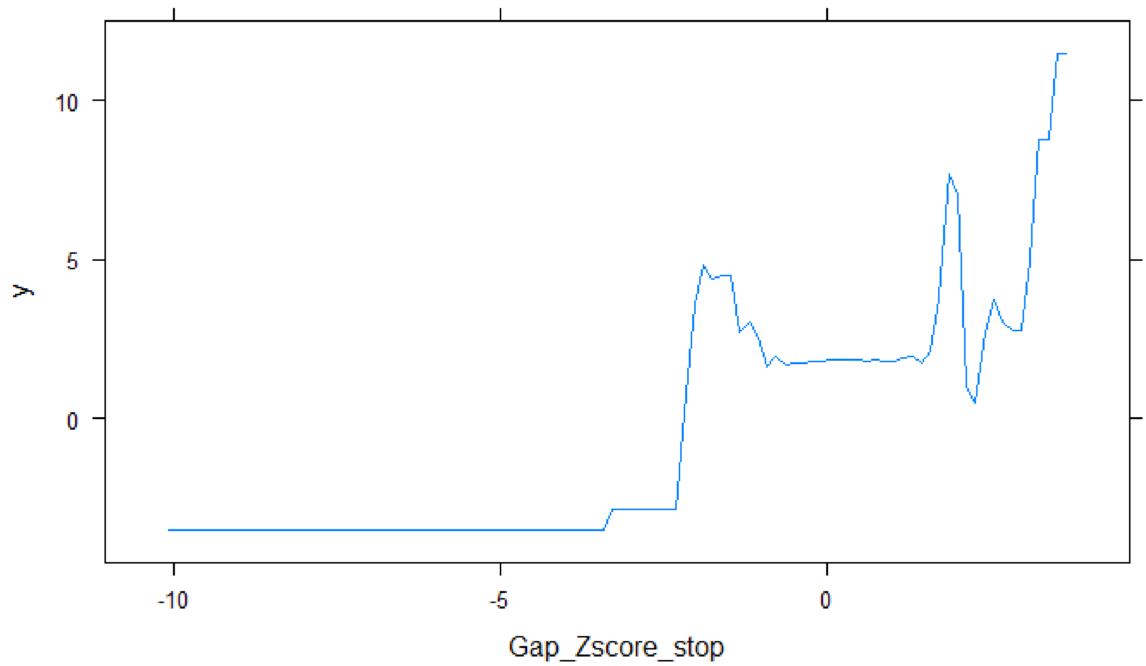

Figure 5. The relationships between public transit and origins of short E-scooters trips

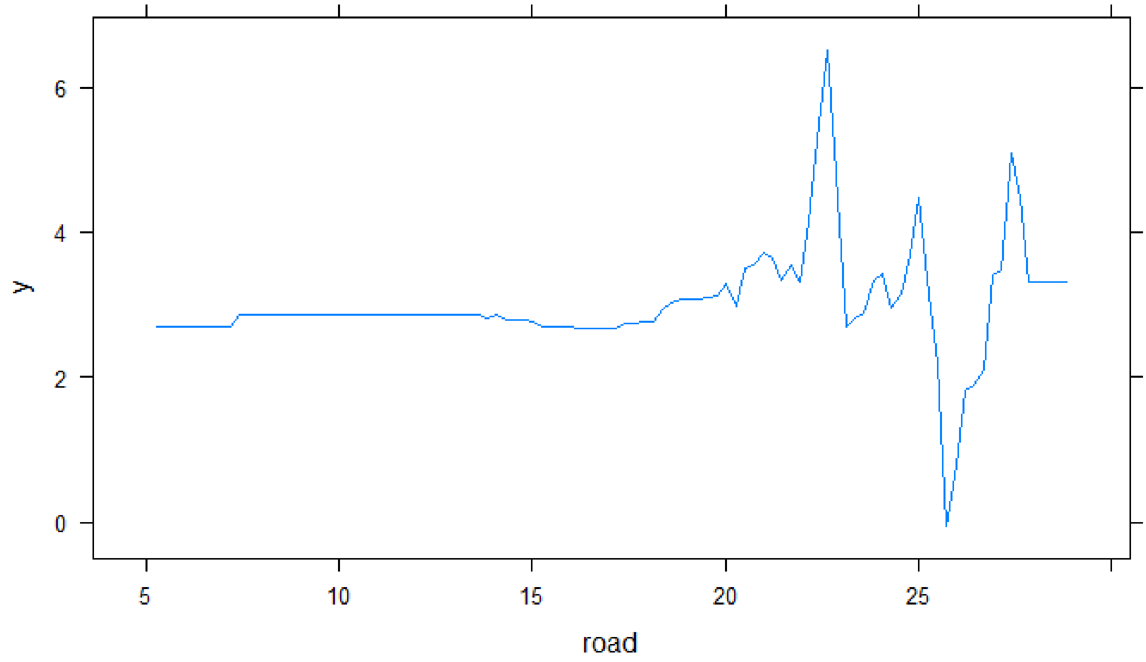

Figure 6. The relationships between road and origins of short E-scooters trips

Figure 6 shows that the relationship between road length and short E-scooter trips is insignificant at first and complicated at last. When road length grows from 5 to 17, the short E-scooter trips do not change a lot. When the road length grows from 17 to 23, the short E-scooter trips increase from around 3 to more than 6. However, when the road length is larger than 6, the relationship is complicated with significant fluctuations.

4. Conclusion remarks

Previous studies explored the relationship between the Public transit and related short-trip E-scooter usages, while few investigated the nonlinear relationships. This study aims to fill up the research gaps by asking: *What is the collective importance of built environment attributes? What is the nonlinear relationship between public transit and related short-trip E-scooter usages?* Based on the American Community Survey 2017 data and Austin data portal, I investigated the nonlinear relationship between Public transit and related short-trip E-scooter usages through GBM. Results indicate that the median earnings can be regarded as the most important factor in predicting short E-scooter usage, and the collective important of built environment attribute is the second most important predictor. Also, the level of public transit services plays a significant role in predicting public transit related short-trip E-scooter usages. Furthermore, all top three important predictors have clear threshold effects in predicting the usages.

Limitations were worth further investigation. First, this study gathered trip information and local sociodemographic characteristics, but it was not confident that all these trips were about connecting to/from public transit facilities. Further qualitative research, such as surveys and interviews, is needed. Second, this study did not consider the effects of temporal patterns, especially the lag or log effects since I used the data from April 2018 to February 2019 to denote an early stage for E-scooter launching. Further studies should consider these effects and maybe introduce a longitudinal analysis to represent these effects. Moreover, the overfitting issues are significant since the fluctuations in partial dependence plots are obvious. Further cross-validation is needed in the next step.

Albeit the limitations, results were worth noting. First, this study pointed out that the importance of built environment attribute is less than sociodemographic factors, which is inconsistent with a recent study (Tao, et al., 2020). Further studies need to consider the reason for this disagreement and try to find a way to unify the results.

Second, all important variables have significant threshold effects in predicting public transit related short-trip E-scooter usages. To maximize the benefits and minimize the cost, transportation agencies should consider these threshold effects and avoid inefficient interventions.

Third, the level of public transit is important in predicting public transit related short-trip E-scooter usages. When the service is terrible, there are no E-scooter usages, which means that riders may need to depend on their foot for the first/last mile issues. Hence, transportation agencies and E-scooter operators need to think about these thresholds and make out strategies to promote the cooperation between public transit and E-scooters.

For me, it is a good chance to formally execute a machine learning project from data collection to analysis, as a good practice. I have learned a lot from a rookie to a “senior” rookie. I plan to adopt these technologies in my future research since they are powerful in solving complicated questions.

References

Barnard, E., 2019. *Love 'em or hate 'em, electric scooters are changing urban transit.* [Online]
Available at: https://www.planetforward.org/idea/electric-scooters-are-transforming-urban-transit

Bliss, L., 2019. *Electric Scooters Aren’t a Transportation Revolution Yet.* [Online]
Available at: https://www.citylab.com/transportation/2019/04/electric-scooter-micromobility-transportation-revolution/587440/

Caspi, O., Smart, M. J. & Noland, R. B., 2020. Spatial associations of dockless shared e-scooter usage. *Transportation Research Part D: Transport and Environment,* Volume 86.

City of Austin, 2018. *Micromobility.* [Online]
Available at: https://austintexas.gov/micromobility

City of Austin, 2018. *Micromobility Laws and Safe.* [Online]
Available at: http://austintexas.gov/micromobilitylaws

Degele, J., Gorr, A., Haas, K. & Kormann, D., 2018. *Identifying E-scooter sharing customer segments using clustering.* s.l., s.n., pp. 1-8.

Ferrer, S. & Ruiz, T., 2018. The impact of the built environment on the decision to walk for short trips: Evidence from two Spanish cities. *Transport Policy,* pp. 111-120.

Frank, L., Saelens, B., Powell, K. & Chapman, J., 2007. Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity?. *Social Science & Medicine,* pp. 1898-1914.

Irfan, U., 2018. *Electric scooters’ sudden invasion of American cities, explained.* [Online]
Available at: https://www.vox.com/2018/8/27/17676670/electric-scooter-rental-bird-lime-

skip-spin-cities
[Accessed 26 Jan 2020].

Jiao, J., 2017. Identifying transit deserts in major Texas cities where the supplies missed the demands. *the Journal of Transportation and Land Use,* pp. 529-540.

Jiao, J. & Bai, S., 2020. Understanding the Shared E-scooter Travels in Austin, TX. *Int. J. Geo-Inf,* 9(135).

Jiao, J. & Dillivan, M., 2013. Transit Deserts: The Gap between Demand and Supply. *Journal of Public Transportation,* pp. 23-39.

Jia, Y. et al., 2019. Effects of new dock-less bicycle-sharing programs on cycling: a retrospective study in Shanghai. *Sports and exercise medicine.*

Lee, M., Chow, J., Yoon, G. & He, B., 2019. *Forecasting e-scooter competition with direct and access trips by mode and distance in New York City,* Ithaca: arXiv, Cornell University.

Lu, X. & Pas, E., 1999. Socio-demographics, activity participation and travel behavior. *Transportation Research Part A: Policy and Practice,* pp. 1-18.

Smith, S. & Schwieterman, J., 2018. *E-Scooter Scenarios: Evaluating the Potential Mobility Benefits of Shared Dockless Scooters in Chicago.,* Chicago: s.n.

Statesman, 2018. *Wear: Bus overhaul, while hurting some, needed to blunt ridership plunge.* [Online].

Tao, T., Wang, J. & Cao, J., 2020. Exploring the non-linear associations between spatial attributes and walking distance to transit. *Journal of Transport Geography.*

the Census Bureau, 2016. *Civilian noninstitutionalized population.* [Online]
Available at: https://www.census.gov/glossary/#term_Civiliannoninstitutionalizedpopulation
[Accessed 23 July 2020].

Tillemann, L. & Feasley, L., 2018. *Let's Count the Ways E-Scooters Could Save the City.* [Online]
Available at: https://www.wired.com/story/e-scooter-micromobility-infographics-cost-emissions/

Tuncer, S., Laurier, E., Brown, B. & Licoppe, C., 2020. Notes on the practices and appearances of e-scooter users in public space. *Journal of Transport Geography,* Volume 85.

Whitehead, K., 2019. *E-scooter pilot can help win street space back from cars.* [Online]
Available at: https://activetrans.org/blog/e-scooter-pilot-can-help-win-street-space-back-from-cars

Yang, L., Shen, Q. & Li, Z., 2016. Comparing travel mode and trip chain choices between holidays and weekdays. *Transportation Research Part A,* pp. 273-285.

Yang, Y. & Diez-Roux, A., 2012. Walking Distance by Trip Purpose and Population Subgroups. *American Journal of Preventive Medicine,* pp. 11-19.

Zariff, R., Pankratz, D. & Kelman, B., 2019. *Making micromobility work for citizens, cities, and service providers,* s.l.: Deloitte.

Zou, Z., Younes, H., Erdogan, S. & Wu, J., 2020. Exploratory Analysis of Real-Time E-Scooter Trip Data in Washington, D.C.. *Transportation Research Record.*