Transit and Bikeshare:
Evidence on Rider Switching Behaviour from Subway Delays

Vanya Georgieva *

Abstract

In recent years, public bikesharing programs have become a common feature in many cities, offering residents and visitors a flexible transportation service for commuting, errand-running, and leisure. This analysis investigates the relationship between bikeshare and the traditional public transportation system. Using granular, trip-level data, I develop a non-parametric model to describe the use of bikeshare at a specific location. Compared to this baseline ridership, I find that severe subway delays are associated with significantly higher bikeshare use, while moderate delays are not.

Introduction

Bikesharing systems consist of networks of rental bicycles used for individual, point-to-point trips. They have become a staple of many cities and are used for both commuting and recreational purposes (Fuller et al., 2013; El-Assi et al., 2017; Kong et al., 2020). In addition, when combined with transit, bikeshare can modestly decrease the use of private vehicles (Martin and Shaheen, 2014; Fuller et al., 2013). However, its interaction with other modes of transportation is mixed. Studies suggest that bikeshare can simultaneously complement existing transit options by alleviating the 'first and last mile problem' and act as a substitute for public transit (Martin and Shaheen, 2014; Kong et al., 2020; Fuller et al., 2019). This analysis contributes to the debate by (1) documenting the multi-use nature of bikesharing using temporal patterns, and (2) analysing the interplay between transit and bikesharing, using transit delays as natural experiments.

In particular, I combine granular, trip-level data on bikesharing, detailed data on transit system delays, as well as local weather data, to model the impact of subway delays on bikeshare use

*PhD Candidate, University of Toronto. Contact: v.georgieva@mail.utoronto.ca

in Toronto, Canada. I use a non-parametric, generalized additive model to describe the temporal pattern of ridership at the given location, taking into account day of the week, time of day, and weather conditions. First, the results from model confirm the dual-purpose (commute and leisure) of bikeshare. During weekdays ridership cyclicality matches general rush-hour trends. In the weekend, the cyclicality is reversed, with the highest number of trips being made midday. Second, I find a significant increase in bikeshare ridership during severe subway delays. This finding clearly demonstrates overlap in the populations that use public transit and bikesharing. Furthermore, it is consistent with a first/last mile interpretation.

This analysis presents two contributions to the literature. To my knowledge, this is the first analysis using a natural experiment framework to examine the causal effects of unanticipated service disruptions on the ridership at the precise time and location of the delay. In addition, this is the first study to model temporal and spatial patterns of bikesharing using non-parametric techniques.

Literature review

Bikeshare literature has identified certain temporal, spacial, and rider patterns. El-Assi, Mahmoud and Habib analyze bikeshare demand, specifically in the Toronto context (2017). They model the year-round effects of the built environment (bicycle infrastructure, proximity to major intersection, transit or university etc.) and the weather conditions. They find that safety is an important factor in bikeshare use, with higher ridership near bicycle lanes and lower ridership near major intersections (El-Assi et al., 2017). Proximity to transit, university, good weather are factors in higher use (El-Assi et al., 2017). In regards to temporal patterns, the authors find that, on weekdays, the flow of bikes is positively correlated with population and employment density. Whereas, on weekends, it is only correlated with population density (El-Assi et al., 2017). Other studies highlight the proximity to food establishments (Faghih-Iani & Eluru, 2015; Wang et al. 2016). In examining the barriers and attraction factors to bikesharing, Fishman et al, find different spatial patterns between users and non-users, with the latter being significantly more geographically dispersed (2014). To this effect, the main barrier to bikeshare was the lack of docking stations hear home or work, and the comfort of motorized travel (Fishman et al., 2014). Finally, bikeshare riders tend to be young adults with relatively high income (Martin, Shaheen and Cohen, 2013; Fishman et al. 2014).

A large portion of the bikeshare literature examines the relationship between bikeshare and other modes of transportation. Two common lines of discussion are modal substitution and modal inte-

gration, terms well described by Kong, Jin and Sui (2020). “Modal substitution” refers to replacing traditional public transport with bikeshare trips. “Modal integration” refers to the use of bikeshare in addition to transit. Kong et al. also use the term “modal complementarity” to refer to joint use of transit and bikeshare when the origin and/or destination is not well serviced by public transit (2020). Literature presents evidence for both types of effects. When examining the dominance of each effect some studies highlight location (urban versus periphery) (Martin and Shaheen, 2019), others stress timing and membership type (Kong et al., 2020). Martin and Shaheen use survey data, including respondents’ homes, from Washington DC and Minneapolis to assess modal integration and substitution (2014). They find that in Washington DC, urban residents use less public transit (modal substitution) while those in the periphery used more transit (modal integration/complementarity) (Martin and Shaheen, 2014). By contrast, the distinction is not as clear in Minneapolis, where transit use increased even in the core and substitution was more geographically dispersed (Martin and Shaheen, 2014). The authors indicate that the key distinction is the intensity of the existing public transport system, in particular of rail, and the degree to which transit caters to long trips (2014). In areas with sparse rail service, bikesharing is more likely to be a complement, while, in areas with dense rail, it is more likely to be a substitute (Martin and Shaheen, 2014). Alternatively, Kong et al. find that time of day, weekends/weekdays, and membership type are more important than location when determining the modal substitution, integration, and complementarity in four US cities: Boston, Chicago, Washington DC and New York City (2020).
There is also evidence for more complicated transport substitutions. Fuller et al. (2013) find that roughly half of bikeshare users surveyed in Montreal switched away from transit to bikeshare, and 10% replaced cars. In Toronto, 44% of bikeshare respondants replaced transit and 27% replaced cars (Martin, Shaheen and Cohen, 2013). Notably, they find rail use decreased but bus use was largely unaffected (Martin, Shaheen and Cohen, 2013). Younes et al. exploit variation in subway station closure during transit construction to provide some evidence in favour of modal integration (2019).

Several papers in the literature have examined the effect of transit disruptions on bikeshare ridership. Fuller et al. examine the effect of a week-long Philadelphia public transit strike on bikeshare, using interrupted time series and Bayesian structural time series models (2019). They find that ridership increased by 57% during the strike and returned to the base-level when transit was restored (Fuller et al., 2019). The increase was composed of slightly more non-members than members (Fuller et al., 2019). Similarly, Younes et al, find that in Washington DC bikeshare trips increased during planned transit disruptions from construction (2019). Again, the ridership returned to normal levels after the disruptions (Younes et al., 2019). This finding is confirmed by Kaviti et al. (2020). Younes et al conclude that in the immediate vicinity of a station in areas with high level of rail service, bikeshare can complement transit (2019).

Though significant, the magnitude of these effects remains modest. The decrease in car use in Montreal represents 0.4% of all car trips (Fuller et al, 2013). Similarly, the impact of transit disruptions on bikesharing is comparatively small against the level of disruption. For instance, in Philadelphia the increase in ridership was 86-92 more trips per 100 000 population per day (Fuller et al., 2019). This is modest compared to the scale of a city-wide transit halt.

Data and Methodology

Description of data sets

This analysis combines three separate datasets. Data on bikeshare trips and transit service delays is provided by the City of Toronto. The weather data comes from historical records of Environment and Natural Resources Canada.

Bikeshare trip data

Toronto has had a 3rd generation bikeshare system since 2011. Currently, "Bike Share Toronto", has 625 docking stations and 6850 bikes. Bikes are rented from docking stations using the dock's kiosk or an app. Riders can purchase a (weekly, monthly, or annual) membership with unlimited 30-minute trips or a single 30-minute fare. If a bike is used for more than 30-minutes both members and non-members are subject to additional fees. Hence, brief trips are strongly encouraged. Users can check the live capacity of stations on the Bikeshare app or on Google Maps.

The bike share data contains every instance of bike rental since July 2016. Each entry includes the start and end station, the precise time the bike was rented, returned, trip duration, as well as trip, subscriber identifiers and membership type. The full data set contains several million observations. The observations are restricted by date and location. Observations with start points near Spadina subway station between April 1, 2019 and September 30th, 2019 were selected. In specification 1, six bike share docks in the immediate vicinity of the subway station were used. In the expanded specification, docks as far as 10 minutes away are included. At the trip level, there are 27 067 observations in specification 1, and 70 050 observations in specification 2. These specifications were used to capture the tradeoff between the relevance of a bike station and its capacity. The first specification prioritizes relevance. In the event of a service disruption the closest bike docks would be most useful to affected riders. However, if these docks are empty, those that are further afield may become important. The time frame purposely excludes winter months. The year 2019 is selected to minimize external factors such as bikeshare's initial rise in popularity and, more recently, the effect of COVID-19.

There are small differences between the docks used in the analysis and those that are currently on the network due to service changes since 2019. Within the restricted set of immediate stations, all of the current stations were present in 2019. In the expanded set, there are two docks that do not appear in the data. The map in figure 1 shows the dock stations used in each specification. Entrances to Spadina Station are marked with a red 'X'. The two docks not found in the data set are crossed out. The number inside each pin refers to the current number of bikes at the given dock. This information is available in the bikeshare website, app, and, most notably, on Google Maps.

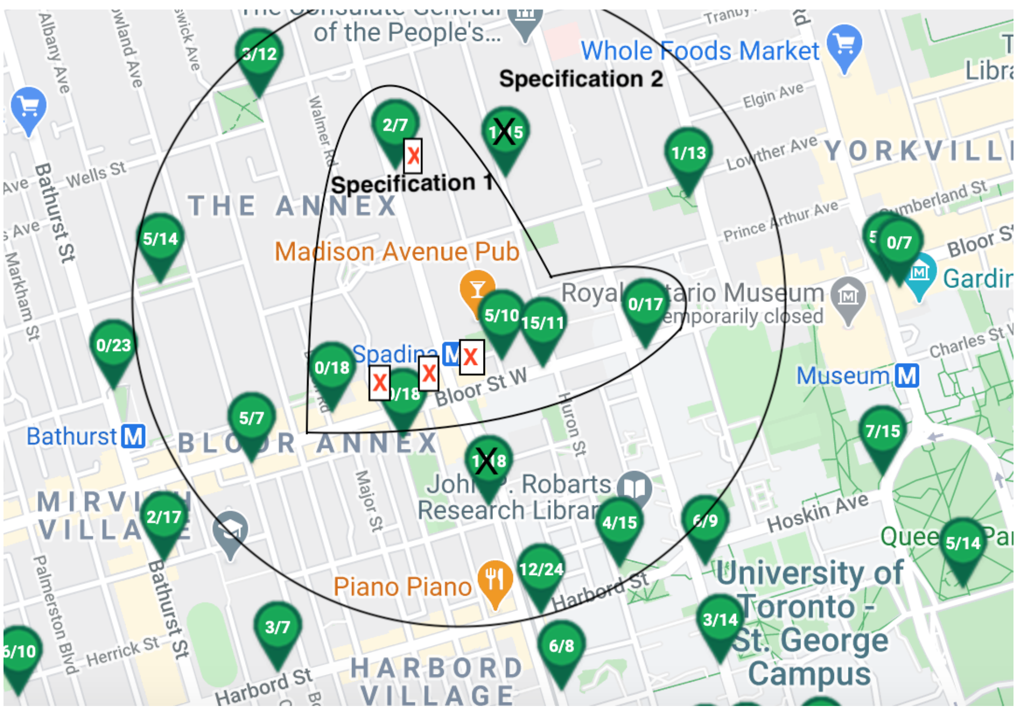

Figure 1: Bikeshare docks included in each specification

Transit delay data

The second set of data used are the transit delay reports. Each incident on a Toronto Transit Commission (TTC) bus, streetcar, or subway since 2014 is logged in these files. Each entry lists the time, location, direction bound, cause, and length of delay. In this analysis, I use subway delays at Spadina Station, a main interchange station in the city centre. All incidents causing a delay of 10 minutes or more are selected. Further, the delays are divided into "average" and "severe". Average delays last 10-20 minutes. There are two severe delays: one of 185 minutes and another of 108 minutes. To be sure that these events are not correlated with bike use at that time, I verify the cause of the delay. In the first case, it is due to a fire at the station. The second case is a tragic accident on the tracks. Consequently, the subway delays are plausibly exogenous of bikeshare use at the time.

Historical weather data

The final data set is the historical weather data from Environment and Natural Resources Canada. It provides a variety of hourly weather indicators from every weather station. There are several weather stations in Toronto. Coincidentally, the "Toronto City" weather station is located within walking distance of Spadina Station. Hence, the precipitation record is very accurate.

Model specification and Methodology

The model specification for the rate of bike rentals is linearly additive, containing parametric and non-parametric terms:

$$RentalRate = \alpha + f_1(Time) + f_2(Time,\, by\, weekend) + \beta_1 Weekend + \beta_2 Rain + \beta_3 Avg.\, Delay + \beta_4 Severe\, Delay$$

Since ridership is very cyclical throughout the day, the non-parametric smoothing function is applied to the time of day. In addition, the temporal patterns are likely to vary between weekdays and weekends. So, the smoothing function is also applied to time of day by the weekend/weekday variable. For identifiability, "weekend" is also included as a parametric binary variable. The final control variable is the binary "rain" variable. The coefficients of interest are β_3 and β_4, demonstrating the effect of average and severe delays on the rental rate, respectively.

The trip data provides a detailed picture of each rental event. However, unlike conventional transport, each bike rental event is separate and unscheduled. To gain a perspective on ridership, a flow measure of bicycles is needed. A straight-forward way of achieving this is to count the number of bikes rented from the set of stations over a given period of time. I use 15-minute intervals. This transforms the dataset from rental events to flows: each "observation" becomes a 15-minute interval.

The key variable becomes the number of bikes rented in that period. In essence, this is a simple poisson process. But, rather than imposing assumptions on the rental rate, the objective is to model it. Hence, the trip data is transformed into flow/interval data. This requires only a minor level of aggregation. The resulting data set contains 17 586 "observations": $4 * 24\, hours * 183\, days = 17586 = N$. Based on previous literature, factors that can be influential when modeling the rate include the time of day, the weather, the built environment, proximity to prominent locations, availability of other transport etc. (El-Assi et al., 2017). Since the analysis is restricted to a single station, there will not be any variation in location-based factors. These effects will be captured by the constant. Of the remaining variables, the most significant are likely to be time of day, weekend/weekday, and weather. Since the horizon is restricted to warm months (April to September), inclement weather will be in the form of rain. Finally, the variable of interest is binary, taking value of 1 when there is subway delay.

To combine the data sets, first I merge the hourly precipitation. Each 15-minute period in the hour is assigned the same precipitation value. For simplicity, I format the precipitation (given in millimeters) into a binary variable. Using the transit data, I create a binary variable for the delay. It takes value 1 from the starting time of the delay until 15 minutes after the delay ends. The weekend binary variable is obtained from the interval (date and time) using a built-in STATA function. Finally, I reformat the time-of-day into a number indicating hour and quarter hour, based on the 24 hour clock. For example, 5:30pm becomes the number 17.5. Once the data is cleaned, the analysis is conducted in R to make use of the Generalized Additive Model, 'gam', function. I use the Restricted Maximum Likelihood (REML) method for the smoothing parameters. Due to the large sample size, there is no meaningful difference between using REML and the default smoothing method in 'gam' (a modified Newton optimizer). The results from estimating specification 1 with and without REML are compared in the Appendix.

Results

Table 1 presents the key results from modeling the rate of bike rentals. Specification 1 and the linear model use six bike docks immediately beside the subway station. Specification 2 expands the number of bike docks to 13, all within a 10-minute walking radius.

Each specification has the same number of observations, 17568, because this is the number of 15-

minute intervals between April 1st and September 30th. However, since specification 1 (and the linear model) include fewer bike docks, fewer trips are used to inform the same number of intervals. In particular, specification 1 uses 27 067 trips to form the 17 568 intervals. In specification 2, the number of trips is 70 050. This drives the difference in explanatory power. Specification 1 captures 32.3% of the variation, while specification 2 captures 43.9%. Both non-parametric models far outperform the linear model, with an R^2 of only 13.9%. By allowing for cyclical patterns in the time effect, the model is significantly improved. The strong cyclicality in figure 2 is evidence against the use of a linear model and demonstrates its low explanatory power. If a parametric specification is required, a sinusoidal model would be more appropriate.

The non-parametric temporal effect is highly significant on its own and when interacted with weekend. Both effects have a p-value approaching zero. The weekday cyclical pattern of bike rentals (fig. 2, left panel) coincides with general traffic trends. Peak use occurs in the before and after work periods, with lower daytime use. Night time use is very low, except for around 1am. Though outside the scope of analysis, this timing coincides with the nightly closure of subways.

Figure 2, left panel should be interpreted as the ceteris-paribus effect of time of day. For illustration, the peak evening use is 2 additional bikes per 15-minutes, all else held equal. So, on a non-rainy, weekday evening, this would result in roughly 3.5 bikes rented per 15 minutes (1.5 from the intercept plus 2 from the weekday hour effect).

The right panel of figure 2 shows the *difference* between the time effect on weekdays and weekends. The time of day effect is inverted. The highest use occurs mid-day, while morning and evening use is lower. Following from the previous example, a dry *weekend* evening would have 2.5 bike rentals per 15 minutes: 1.5 from the intercept, plus 2 from the weekday hour effect, less 1 from the weekend effect. Note that figure 2 is based on specification 1.

When looking at the parametric control variables, specifications 1 and 2 show that weekend and rain are highly significant, both with p-values approaching zero. Bike rentals are slightly, though significantly, lower on weekends. This is likely a location-specific characteristic. As expected, rain has a significant, large, negative effect on bike rentals.

Lastly, table 1 shows that delays can have an impact on the bikeshare rental rate. Average delays of 10-20 minutes do not have a measurable effect on bikeshare use. However, severe delays of 2-3 hours, have a positive, significant (at 5% level) effect in the immediate vicinity. Severe delays result in 0.7 additional bikes rented per 15 minutes. If the scope is expanded to include docks further away, the coefficient remains positive but loses significance. It's worth noting that this effect is fairly noisy since only two severe delays took place over the period of analysis.

	Specification 1	Specification 2	Linear model
adj R^2	0.323	0.439	0.139
N	17568	17568	17568
# trips	27 067	70 050	27 067
Parametric coefficients			
(Intercept)	$1.5219^{\dagger\dagger\dagger}$	$2.9015^{\dagger\dagger\dagger}$	$0.49675^{\dagger\dagger\dagger}$
	(0.0131)	(0.0202)	(0.02884)
Weekend	$-0.2544^{\dagger\dagger\dagger}$	$-0.4922^{\dagger\dagger\dagger}$	-0.06752^{*}
	(0.0181)	(0.0278)	(0.0408)
Rain	$-0.9201^{\dagger\dagger\dagger}$	$-1.7359^{\dagger\dagger\dagger}$	$-0.845098^{\dagger\dagger\dagger}$
	(0.0469)	(0.0723)	(0.05283)
Average Delay	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Severe Delay	0.6913^{**}	0.3210	0.70422^{*}
	(0.3276)	(0.5045)	(0.36774)
Time of day	–	–	$0.08592^{\dagger\dagger\dagger}$
	–	–	(0.00208)
Time of day, by weekend	–	–	$-0.01571^{\dagger\dagger\dagger}$
	–	–	(0.00294)
Non-parametric effects			
	F (p-value)		
Time of day	$695^{\dagger\dagger\dagger}$	$1163^{\dagger\dagger\dagger}$	–
	0.0000	0.0000	–
Time of day, by weekend	$106^{\dagger\dagger\dagger}$	$201^{\dagger\dagger\dagger}$	–
	0.0000	0.0000	–

Table 1: Non-parametric models of bike renting rate, compared to basic linear model. (††† significant at $<$ 0.00001% level, ** significant at 5% level, * significant at 10% level)

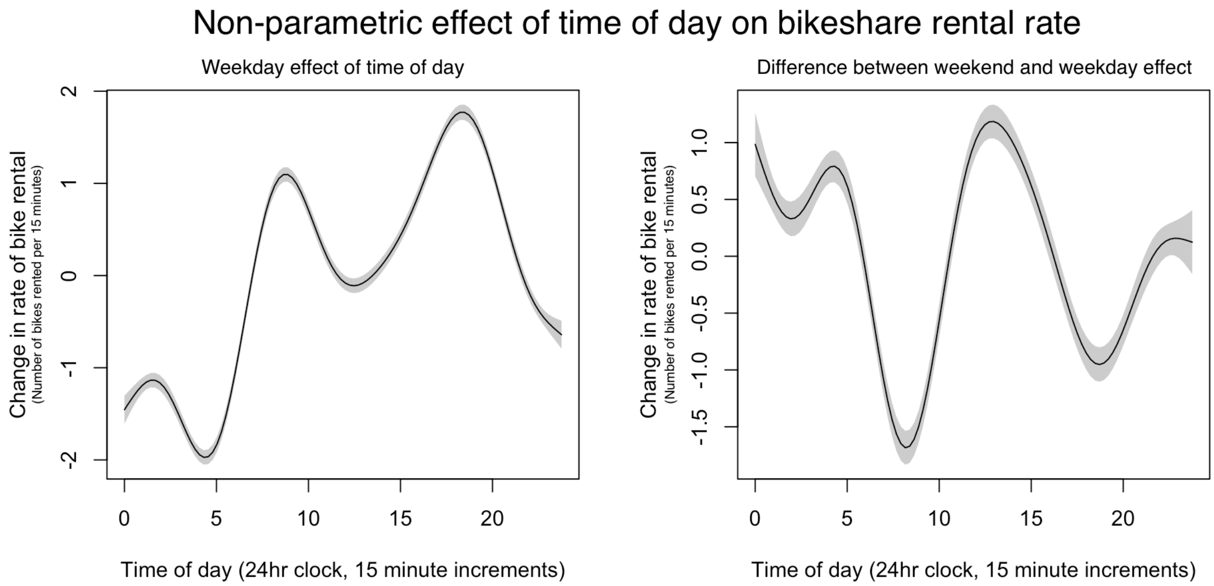

Figure 2: Non-parametric effect of "time of day". Left: total effect during weekday (base). Right: difference in the effect of "time of day" between weekday and weekend

Discussion

The analysis presented here contributes to two discussions: the patterns of bikeshare use and the observed interaction with the public transport system.

On the first point, this analysis benefits from availability of rich data allowing for non-parametric methods. This method reveals strong cyclicality of use, which varies between weekday and weekend. It is indicative of the dual-function of bikeshare: as a commuting and leisure device during the week and weekend, respectively. While this pattern is cited in literature (El-Assi et al., 2017, Kong et al., 2020), to my knowledge, this is the first analysis to use non-parametric methods to explicitly model temporal patterns in bikeshare use.

This analysis demonstrates a statistically significant interaction between transit and bikeshare. In contrast to previous studies examining modal shifts in transportation, I am able to distinguish the real-time interaction between subway and bikeshare systems. When there is a severe transit delay, bikeshare use in the immediate vicinity increases. This effect is in agreement with previous literature capturing the effect of transit disruptions: bikeshare use increases for the duration of the disturbance (Fuller et al., 2019; Younes et al, 2019; Kaviti et al., 2020). The main distinction is that the disruption in my analysis is localized, unforeseen and plausibly exogenous. It demonstrates that at least some riders are actively optimizing their use of public infrastructure, as opposed to adhering to a fixed, planned method of transport.

In addition, certain inferences can be drawn from the magnitude of the effect. At roughly 3 additional bikes per hour, it is modest compared to the impact of subway delays. However, this is in accordance with previous literature (Fuller et al., 2019). Further, the magnitude could indicate that bikeshare is a suitable alternative for few subway users. To this effect, Martin and Shaheen note that the substitutability of bikeshare and transit is related to the degree to which transit specializes in long or short commutes (2014).

Finally, the spatial pattern of the effect, captured by the difference between specifications one and two indicates a highly localized effect. First, this reinforces the relationship between the subway delay and increased bikeshare use. It also demonstrates the importance of convenient dock locations. This is congruent with Fishman et al. who cite lack of docking stations as a high barrier to using bikeshare services (2014).

Limitations and future research

A potential limiting factor in this analysis is the capacity of the dock stations. The effect of the subway delay could be underestimated if bike docks near the subway station are empty at the time. Future analysis could benefit from tracking dock capacity in addition to the outflow of bikes. Furthermore, this analysis was limited to a single subway station. As a result, only two sufficiently large delays occurred during the period of interest. If this analysis were expanded to include multiple stations or modes of transportation, a more robust conclusion about rider switching behaviour could be established.

Finally, future investigation could expand this analysis to profit from the network nature of the data. Currently, the literature uses regression analysis to draw inference about general temporal and spatial patterns. The granularity of the dataset would allow for modeling of a transportation network. This could contribute significantly to the modal integration/substitution debate by investigating the spatial overlap (or complementarity) of trips made using bikeshare against the existing transit routes. Other potential avenues could include investigating the effects of traffic, road construction, or changes in parking availability on the localized use of bikeshare.

References

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. Transportation, 44(3), 589–613.

Environment and Natural Resources, Government of Canada. (2020). Historical Climate Data. [Data file] Retrieved from https://climate.weather.gc.ca/historical_data/search_historic_data_e.html

Faghih-Imani, A., Eluru, N., El-Geneidy, A., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography, 41, 306–314

Fishman, E., Washington, S., Haworth, N., & Mazzei, A. (2014). Barriers to bikesharing: an analysis from Melbourne and Brisbane. Journal of Transport Geography, 41, 325–337.

Fuller, D., Gauvin, L., Kestens, Y., Morency, P., & Drouin, L. (2013). The potential modal shift and health benefits of implementing a public bicycle share program in Montreal, Canada. The International Journal of Behavioral Nutrition and Physical Activity, 10(1), 66.

Fuller, D., Luan, H., Buote, R., & Auchincloss, A. H. (2019). Impact of a public transit strike on public bicycle share use: An interrupted time series natural experiment study. Journal of Transport & Health, 13, 137–142.

Kaviti, S., Venigalla, M. M., Zhu, S., Lucas, K., & Brodie, S. (2020). Impact of pricing and transit disruptions on bikeshare ridership and revenue. Transportation, 47(2), 641–662.

Kong, H., Jin, S. T., & Sui, D. Z. (2020). Deciphering the relationship between bikesharing and public transit: Modal substitution, integration, and complementation. Transportation Research. Part D, Transport and Environment, 85(102392), 102392.

Martin, E. W., & Shaheen, S. A. (2014). Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. Journal of Transport Geography, 41, 315–324.

Shaheen, S., Martin, E., & Cohen, A. (2013). Public bikesharing and modal shift behavior: A comparative study of early bikesharing systems in north America. International Journal of Transportation, 1(1), 35–54.

Toronto Parking Authority. (2020). Bike Share Toronto Ridership Data. [Data file] Retrieved from: https://open.toronto.ca/dataset/bike-share-toronto-ridership-data/

Toronto Transit Commission. (2020). TTC Subway Delay Data. [Data file] Retrieved from: https://open.toronto.ca/dataset/ttc-subway-delay-data/

Wang, X., Lindsey, G. Schoner, J., & Harrison, A. (2016). Modeling bike share station activity:

Effects of nearby businesses and jobs on trips to and from stations. Journal of Urban Planning and Development, 142(1), 1–9.

Younes, H., Nasri, A., Baiocchi, G., & Zhang, L. (2019). How transit service closures influence bikesharing demand; lessons learned from SafeTrack project in Washington, D.C. metropolitan area. Journal of Transport Geography, 76, 83–92.

Appendix

The following table compares the regression results for specification 1 using REML and the default smoothing method.

	Spec. 1 with REML	**Spec. 1 Default**
adj R^2	0.323	0.323
N	17568	17568
# trips	27 067	70 050
Parametric coefficients		
(Intercept)	1.5219	1.5219
	(0.0131)	(0.0131)
Weekend	-0.2544	-0.2544
	(0.0181)	(0.0181)
Rain	-0.9201	-0.9201
	(0.0469)	(0.0469)
Average Delay	0.0000	0.0000
	(0.0000)	(0.0000)
Severe Delay	0.6913	0.6918
	(0.3276)	(0.3277)
Time of day	–	–
	–	–
Time of day, by weekend	–	–
	–	–
Non-parametric effects		
	F (p-value)	
Time of day	695	697
	0.0000	0.0000
Time of day, by weekend	106	107
	0.0000	0.0000

Table 2: Non-parametric models of bike renting rate, compared to basic linear model.