Determining the role of bicycle sharing system infrastructure installation decision on usage: Case study of montreal BIXI system

Ahmadreza Faghih-Imani [a], Naveen Eluru [b],*

[a] Department of Civil Engineering and Applied Mechanics, McGill University, Canada
[b] Department of Civil, Environmental and Construction Engineering, University of Central Florida, United States

A B S T R A C T

The traditional quantitative approach to studying Bicycle Sharing System (BSS) usage involves examining the influence of BSS infrastructure (such as number of BSS stations and capacity), transportation network infrastructure, land use and urban form, meteorological data, and temporal characteristics. These studies, as expected, conclude that BSS infrastructure (number of stations and capacity) have substantial influence on BSS usage. The earlier studies consider usage as a dependent variable and employ BSS infrastructure as an independent variable. Thus, in the models developed, the unobserved factors influencing the measured dependent variable (BSS usage) also strongly influence one of the independent variables (BSS infrastructure). This is a classic violation of the most basic assumption in econometric modeling i.e. the error component in the model is not correlated with any of the exogenous variables. The model estimates obtained with this erroneous assumption are likely to over-estimate the impact of BSS infrastructure. Our research effort proposes an econometric framework that remedies this drawback. We propose a measurement equation to account for the installation process and relate it to the usage equations thus correcting for the bias introduced in earlier research efforts by formulating a multi-level joint econometric framework. The econometric models developed have been estimated using data compiled from April 2012 to August 2012 for the BIXI system in Montreal. The model estimates support our hypothesis and clearly show over-estimation of BSS infrastructure impacts in models that neglect the installation process. An elasticity analysis to highlight the advantages of the proposed econometric model is also conducted.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

There has been growing interest in bicycle sharing systems (BSS) as an alternative and complementary mode of transportation (Faghih-Imani et al., 2014). BSS systems are recognized to offer benefits such as flexible mobility, physical activity benefits, and support for multimodal transport connections (Shaheen et al., 2010). With the growing installation of BSS infrastructure across the world, there is a substantial interest in understanding how these systems impact the urban transportation system. The typical approach for analyzing BSS usage involves examining the influence of BSS infrastructure (e.g. number of BSS stations and stations' capacity), transportation network infrastructure (e.g. length of bicycle facilities, streets

and major roads), land use and built environment (e.g. population density, presence of metro and bus stations, restaurants, businesses and universities), meteorological data (e.g. temperature and humidity), and temporal characteristics (e.g. time of day, day of the week and month).

Several studies demonstrate that increasing BSS infrastructure (number of stations and capacity) increases BSS usage (Faghih-Imani et al., 2014; Wang et al., 2015). Land use and urban form variables such as higher job or population density also contribute to BSS usage (Rixey, 2013; Faghih-Imani et al., 2014). Studies that examined usage at a fine time resolution (within a day) indicated that temporal characteristics affect BSS usage – with peak usage observed during the evening peak hours (Faghih-Imani et al., 2014; Faghih-Imani and Eluru, 2016). These studies have also found that BSS usage is higher on weekdays compared to weekends. While examining the impact of point of interests such as restaurants, retail stores, and universities near BSS stations, studies found evidence that BSS usage was higher for stations with the higher number of point of interests in the vicinity (Rixey, 2013; Faghih-Imani et al., 2014). More studies have explored the effect of temperature and weather on usages (Gebhart and Noland, 2014; Faghih-Imani et al., 2014). These studies, as expected, conclude that BSS usage is lower under adverse weather conditions (presence of rain, lower temperature). Furthermore, the relationship between BSS and other public transportation systems such as subway or bus transit system are also examined by several research efforts (Nair et al., 2013; Faghih-Imani et al., 2014, 2015; Faghih-Imani and Eluru, 2016).

2. Current study

2.1. Earlier research

Earlier research efforts, while providing useful insights on the system level usage patterns, ignore the BSS infrastructure installation decision process i.e. BSS operators installed the infrastructure based on an expectation of system usage. The aforementioned research studies ignore this when they consider usage as a dependent variable and employ BSS infrastructure as an independent variable. Thus, in the models developed, the unobserved factors influencing the measured dependent variable (BSS usage) also strongly influence one of the independent variables (BSS infrastructure). This is a classic violation of the most basic assumption in econometric modeling i.e. the error component in the model is not correlated with any of the exogenous variables (Greene, 2012 pp. 63). The model estimates obtained with this erroneous assumption are very likely to over-estimate the impact of BSS infrastructure - since the estimates related to BSS infrastructure include two impacts: the actual impact of BSS infrastructure on usage and part of the unobserved factors influencing usage. The current research effort proposes a joint econometric framework that remedies this drawback.

To obtain the "true" impact of BSS infrastructure on BSS usage, it is critical to consider infrastructure installation and usage decisions as an interconnected process. To correctly characterize the decision processes at hand, it is necessary to consider the bicycle-sharing infrastructure installation itself as a dependent variable - simultaneously along with usage patterns. To be sure, endogeneity has been widely investigated in transportation fields including travel behavior, transportation safety, and health. A comprehensive review is beyond the scope of the paper. We restrict ourselves to a brief reference to endogeneity related research efforts in various transportation fields. In *travel behavior* literature, the residential self-selection issue tied with travel behavior outcomes is well documented (see Bhat and Guo, 2007; Cao et al., 2009). Earlier research methods addressed potential self-selection between residential location and several travel behavior dimensions such as vehicle ownership (Bhat and Guo, 2007; Eluru et al., 2010), travel frequency and mileage (Greenwald and Boarnet, 2001; Bhat and Eluru, 2009; Ewing et al., 2015), bicycle ownership (Pinjari et al., 2008) and mode choice (Schwanen and Mokhtarian, 2005). Other travel behavior themed endogeneity research efforts include transit influence on automobile ownership and mode choice (Hu, 2016), highway capacity on increased vehicle miles of travel (Noland and Cowart, 2000), and social influence and taste preference variables on mode choice (Walker et al., 2011; Vij and Walker, 2014). In the *transportation safety* field, documented endogeneity research efforts include seat belt use and subsequent injury severity (Eluru and Bhat, 2007; Abay et al., 2013), speed limits in crash count modeling (Cheng et al., 2013), crash type and injury severity (Rana et al., 2010; Yasmin et al., 2014) and Emergency Medical Service response time and the time to fatality (Yasmin et al., 2015). Finally, in recent years, *health* research efforts have considered endogeneity in examining the relation between active transportation and health outcomes (Schauder and Foley, 2015).

2.2. Contributions of the current study

Our research effort is different from earlier studies because the model system required in the context of BSS systems is challenging for multiple reasons. <u>First</u>, in all the earlier studies, self-selection was considered in a choice framework based on a cross-sectional dataset where one measurement equation was coupled with one choice equation to disentangle the influence of endogeneity. However, in the BSS context, the infrastructure installation process has only one occurrence, while the infrastructure usage process has multiple repeated observations. Hence, the econometric model developed should account for self-selection while accounting for the availability of repeated observations of the choice process. To elaborate, the repeated observations of BSS usage from the same spatial aggregation are plausibly influenced by unobserved factors that influence choice processes. The impact of unobserved factors, given the availability of reasonable number of repeated observations, can also be accommodated at multiple levels. Hence, to correctly identify the presence of endogeneity in the choices

process, the econometric framework needs to consider the potential presence of multiple levels of unobserved heterogeneity affecting BSS usage. For example, in our data, repeated observations can be considered to be at three levels – (1) unobserved heterogeneity across the spatial unit, (2) unobserved heterogeneity for the spatial unit in a day and (3) unobserved heterogeneity at a per observation level. The unobserved heterogeneity at multiple levels can be incorporated as random parameters moderating the influence of exogenous variables or error correlation between the BSS usage dimensions. Employing a cross-sectional model or employing a single level of unobserved heterogeneity for repeated BSS usage terms in the analysis might result in confounding the presence of endogeneity with unobserved heterogeneity.

Second, we need to recognize the presence of multiple BSS infrastructure variables (number of stations and capacity in a spatial unit) and presence of time varying multiple dependent variables (usage defined as arrivals and departures for many time-varying records). Retaining all BSS infrastructure variables and usage in their original form would result in a large number of choice equations (one per variable). Moreover, introducing joint unobserved effects at multiple levels across the large equation system will result in a computationally intensive probability function that will require higher dimensions of integration to evaluate the simulated log-likelihood function. Hence, to simplify the econometric model development we propose a single measure of BSS infrastructure that simultaneously encapsulates number of stations and total capacity in a traffic analysis zone (TAZ) (more detailed discussion in Section 3). Having a single equation for BSS infrastructure usage allows us to efficiently accommodate for the potential presence of endogeneity in the proposed model system.[1]

2.3. Research approach

The generated BSS infrastructure index is categorized subsequently as an ordinal variable. Other dependent variables in our study are the BSS usage variables defined as arrivals and departures. Rather than employ these variables as continuous values – we adopt an ordinal categorization in our analysis. The rationale behind this characterization is that imposing a strictly linear response structure for usage is restrictive. While one could consider non-linear variable forms (such as square and cube terms) within a linear system – the mapping of the impact of exogenous variables on usage is still strictly linear. Particularly, given the significant variation in the usage patterns spatially and temporally imposing a strictly linear structure (as imposed by linear regression) might result in biased estimation of the "true" influence of exogenous variables. On the other hand, in an ordinal structure, the probability of a category is mapped with the observed response through a much more flexible non-linear response profile. The number of ordered categories can be increased in a simple and intuitive fashion if a fine resolution is deemed necessary.[2] The reader would note that the discretization would also preclude the need for exclusion restrictions (see Section 3.2 of Abay et al., 2013 for more details).

In summary, the proposed joint modeling process thus estimates a more accurate impact of BSS infrastructure on usage. More importantly, the consideration of the installation process and unobserved heterogeneity at multiple levels allows us to generate consistent estimates of the impact of other exogenous variables (such as land use and urban form). We formulate a multi-level joint econometric framework. The framework considers the bicycle-sharing infrastructure installation process (a one-time process) while allowing for a multi-level analysis of arrivals and departures. We consider an ordered representation for all the dependent variables yielding a three dimensional panel ordered formulation. Specifically, we adopt a repeated observation based panel multi-level mixed (or random parameters) ordered logit model. The proposed model is estimated using data compiled from the Montreal bicycle-sharing system, BIXI, from April to August 2012.

The remainder of the paper is organized in the following order. Section 3 provide a discussion on the proposed BSS infrastructure measure. In Section 4, econometric model structure and estimation procedure are described. Section 5 describes the data and the sample formation procedure. Empirical results and policy analysis are presented and discussed in Section 6. Finally, Section 7 summarizes and concludes the paper.

3. BSS infrastructure measure

As discussed earlier, within a TAZ, number of stations and total capacity represent different spatial components of BSS infrastructure. To facilitate the model development process, we propose and compute a single BSS infrastructure measure. This TAZ level measure accounts for the influence of both number of bicycle stations and the total capacity in the zone simultaneously while considering the area of the TAZ. Several forms of the infrastructure index were examined prior to settling on the index presented here.

Specifically, BSS infrastructure index (BSSI) takes the following form:

---

[1] It is not apparent how to define the spatial aggregation of the installation decisions – as it is not possible that each station was decided independently. It is possible to explore a number of spatial aggregation metrics (census tract, block level and Traffic Analysis Zone (TAZ)). However, determining the accurate spatial aggregation employed is quite challenging. Since transportation planners are likely to consider the traffic analysis zone (TAZ), we opt for TAZ system as the spatial aggregation unit. The proposed methodology can be extended appropriately to any spatial aggregation.

[2] The number of threshold parameters required to be estimated increases linearly with the number of ordered alternatives. However, in the BSS context, the sample sizes are large enough to accommodate for such increases in the number of parameters. The reader would note that an ordered response model with a large number of categories is a true non-linear extension of a continuous linear model which restricts the categories through a single slope (see Chakour and Eluru, 2016 for similar ordinal discretization).

$$BSSI = Ln\left(\frac{Number\ of\ Stations\ in\ TAZ}{Average\ Number\ of\ Stations\ in\ TAZs} \times \frac{TAZ\ Capacity}{Average\ Capacity\ of\ TAZs} \times \frac{1}{TAZ\ area}\right) \qquad (1)$$

The proposed index concurrently considers the influence of number of stations and capacity of stations in a TAZ while normalizing for the TAZ area. Considering only number of stations in a TAZ as the measure of BSS infrastructure cannot recognize the difference between two TAZs with the same number of stations but with different capacities. On the other hand, using only total capacity of stations in a TAZ as BSS infrastructure index overlooks the distinction in spatial distribution in the TAZ. The variation of BSS infrastructure for an average sized TAZ in Montreal with respect to the number of stations in the TAZ and TAZ capacity is illustrated in Fig. 1. The average number of stations in TAZs and the average capacity of TAZs for BIXI system and an average TAZ area are assumed for BSSI calculations in Fig. 1. It is evident from the figure that for the same total capacity in a TAZ, higher number of stations has a higher BSSI value – accounting for better spatial distribution.

The distribution of the BSSI variable for the Montreal BIXI system is presented in Fig. 2. As is expected, the highest values of the variable are observed in downtown Montreal and the bicycle-friendlier neighbourhood of Plateau-Mont Royal (highlighted). The BSSI variable defined in our study forms the first dimension of our three-level econometric framework. Specifically, we employ the proposed BSSI measure as the dependent variable to characterize the bicycle-sharing infrastructure installation process.

4. Methodology

This section describes the structure and estimation procedure of the proposed three dimensional panel mixed multi-level ordered logit (3PMMOL) model. The three dimensions correspond to the installation process, arrivals and departures. The framework allows for random parameters, error correlations at multiple levels while accommodating for endogeneity between installation process and BSS usage dimensions.

4.1. Model structure

The mathematical framework is described below (see Fig. 3 for a conceptual representation of the econometric framework):

Let q (q = 1, 2, ..., Q; in our case Q = 235) be an index to represent traffic analysis zones (TAZ), j (j = 1, 2, 3, ..., J) be an index to represent the different levels of bicycle infrastructure, and k (k = 1, 2, 3, ..., K) and l (l = 1, 2, 3, ..., L) be an index to represent the TAZ level bicycle arrival and departure categories, respectively. Further, to accommodate for time period specific bicycle arrivals and departures, let d (d = 1, 2, 3, ..., D; in our case D = 7) represent the different days and t (t = 1, 2, 3, ..., T; in our case T = 5) represent the different time periods for TAZ q. In our sample, each TAZ q, has 35 repeated observations (7 days * 5 time periods).

The existence of endogeneity is a time invariant impact to be examined. On the other hand, for the usage processes (arrivals and departures) the observed impact of exogenous variables are based on whether the variable is time variant or not (such as temperature in the time period). The unobserved heterogeneity (characterized as random parameters or error correlations) on the other hand has three components: (1) TAZ level, (2) TAZ - Day level and (3) TAZ – Day - Time period level. The reasoning behind this decomposition is that there are potential unobserved factors at every resolution (TAZ, day or time period) that influence arrival and departure rates. Ignoring the presence of such multi-level relationships might result in over-estimation of the presence of endogeneity.

The equation system in its most generic form for modeling the bicycle-sharing infrastructure index and the departure and arrivals may be written as follows:

$$u_q^* = (\beta' + \gamma_q')x_q + \eta_q + \varepsilon_q, \quad u_q = j\text{ if }\psi_{j-1} < u_q^* < \psi_j \qquad (2)$$

$$y_{qdt}^* = (\alpha' + \delta_q' + \delta_{qd}' + \delta_{qdt}')f + (\theta' + \mu_q' + \mu_{qd}' + \mu_{qdt}')g \pm \eta_q + \xi_{qdt}, \quad y_{qdt} = k\text{ if }\omega_{k-1} < y_{qdt}^* < \omega_k \qquad (3)$$

$$z_{qdt}^* = (\tau' + \lambda_q' + \lambda_{qd}' + \lambda_{qdt}')h + (\theta' + \mu_q' + \mu_{qd}' + \mu_{qdt}')g \pm \eta_q + \zeta_{qdt}, \quad z_{qdt} = l\text{ if }\omega_{l-1} < z_{qdt}^* < \omega_l \qquad (4)$$

Eq. (2) is associated with the BSSI installation propensity u_q^* for TAZ q, and x_q is an ($M \times 1$)-column vector of attributes associated with TAZ q (for example, job density, bicycle lane density, etc.) and ψ represents threshold vector for BSSI installation. β represents a corresponding ($M \times 1$)-column vector of mean effects of the elements of x_q while γ_q' is another ($M \times 1$)-column vector with its mth element representing unobserved factors specific to TAZ q and its choice environment that moderate the influence of the corresponding mth element of the vector x_q. ε_q is an idiosyncratic random error term assumed to be identically and independently standard Logistic distributed across TAZs.

Eq. (3) is associated with y_{qdt}^* being the latent (ordered) bicycle propensity for arrivals of TAZ q on day d and tth time occasion. This latent propensity y_{qdt}^* is mapped to the actual grouped arrival category y_{qdt} by the ω thresholds ($\omega_0 = -\infty$ and $\omega_k = \infty$) in the usual ordered-response modeling framework. ξ_{qdt} is an idiosyncratic random error term, assumed identically and independently logistic distributed (across TAZs, days and time periods). Equation (3) similarly posits the departure rates
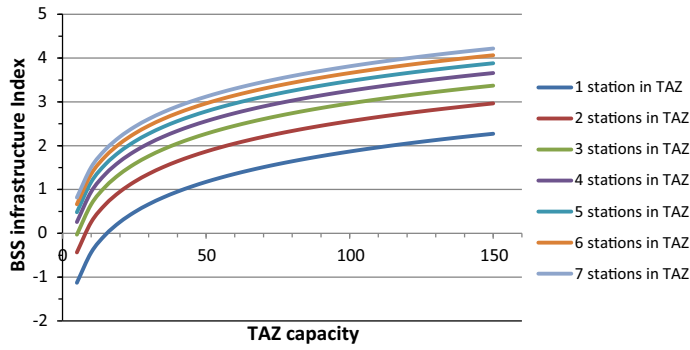
Fig. 1. The variation of BSS infrastructure for an average sized TAZ in Montreal.
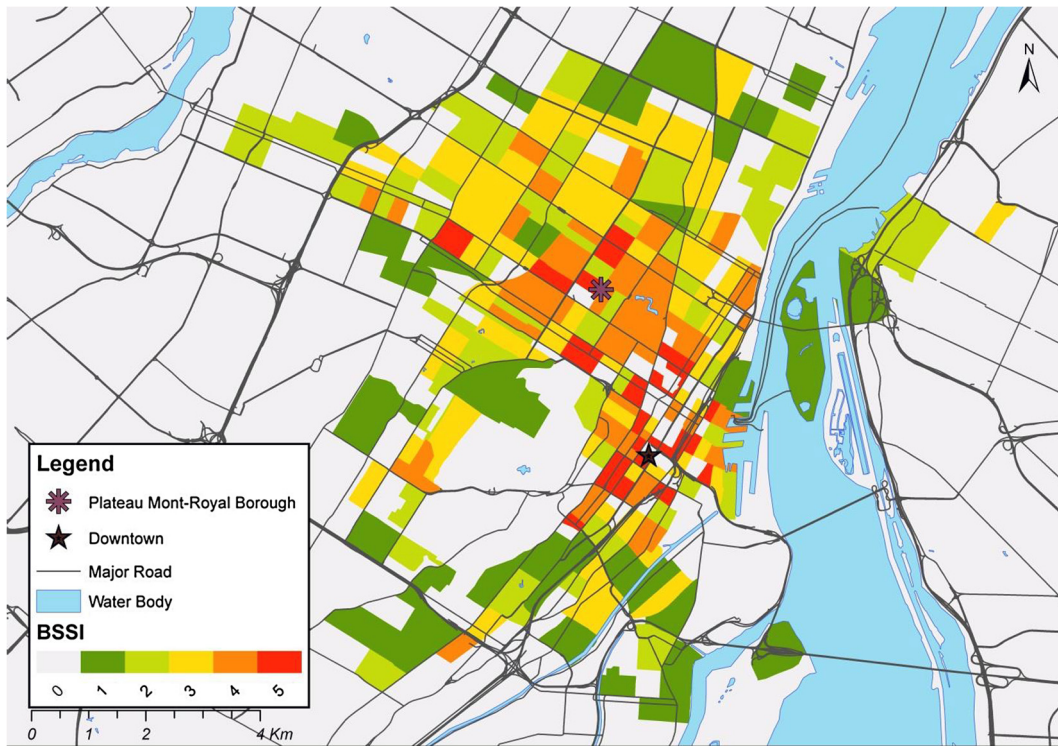

Fig. 2. Bicycle-sharing infrastructure index in montreal.

of TAZ q on day d and tth time occasion with the latent ordered propensity of z_{qdt}^*, ω thresholds ($\omega_0 = -\infty$ and $\omega_l = \infty$), and error term ζ_{qdt}.

f matrix composed of f_q, f_{qd} and f_{qdt} in Eq. (3) is the set of attributes that solely influence the TAZ's bicycle arrival rate while h matrix composed of h_q, h_{qd} and h_{qdt} in Eq. (4) is the set of attributes that solely affect the TAZ's bicycle departure rates. In addition, g matrix composed of g_q, g_{qd} and g_{qdt} in Eq. (3), (4) is the set of common attributes that simultaneously impacts both arrival and departure rates. The reader would note that there is no overlap in the variables in f and g as well as in h and g. α, τ, and θ are corresponding vectors of mean effects for matrices f, g and h respectively. δ, λ and μ are the corresponding vectors for unobserved factors moderating the influence of attributes on f, g and h respectively at multiple levels. The random parameters or error correlations can potentially be separated into three levels of influence as captured by the TAZ level (q), TAZ and Day level (qd) and TAZ – Day and Time Period level (qdt).[3]

---

[3] For example, a TAZ level variable, population density has one observed variable impact and could potentially have three random parameters – one for each resolution. Thus, we allow for the TAZ population density variable to impact the entire TAZ, TAZ and Day combination or TAZ - Day and Time period combination. Of course, it is possible population density variable does not influence the usage processes at all three possible levels.
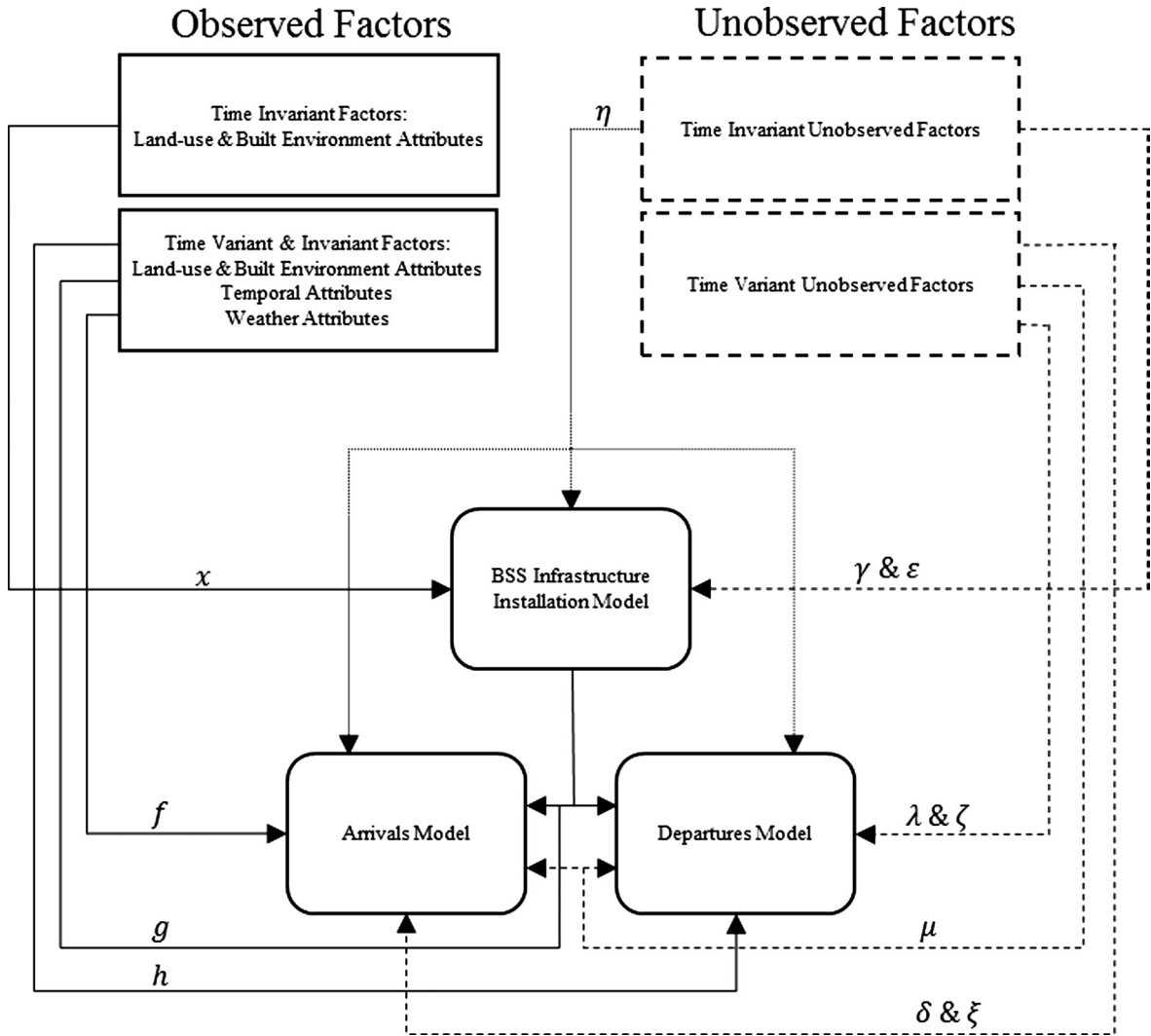
Observed Factors　　　　　　Unobserved Factors

Fig. 3. Three dimensional panel mixed multi-level ordered logit (3PMMOL) framework.

η_q represents the endogeneity effect by capturing the unobserved factors that simultaneously impact BSS infrastructure installation and BSS usage for TAZ q. The $\pm$ sign in front of η_q in the bicycle usage equation indicates that the correlation in unobserved factors between the BSS infrastructure installation and the bicycle arrival and departure rates may be positive or negative. A positive sign implies that unobserved factors that increase the propensity of BSS installation will also increase the bicycle flows, while a negative sign suggests that unobserved factors that increase the propensity of BSS installation for a certain TAZ will decrease the BSS usage. Clearly, one expects, from an intuitive standpoint, that the former case will hold. However, one can empirically test the models with both '+' and '−' signs to determine the best empirical result.

To complete the model structure of the system in Eqs. (2) through (4), it is necessary to specify the structure for the unobserved vectors γ, δ, λ, μ and η. In this analysis, we assume that the elements are independent realizations from normal population distributions and we estimate the standard deviation of the distribution. With these assumptions, the probability expression for the infrastructure installation conditional on γ and η is given as:

$$P_{qj} = \Lambda\Big[\psi_j - ((\beta' + \gamma'_q)x_q + \eta_q)\Big] - \Lambda\Big[\psi_{j-1} - ((\beta' + \gamma'_q)x_q + \eta_q)\Big] \tag{5}$$

where $\Lambda[\cdot]$ is the cumulative distribution of the standard logistic distribution.

Similarly, conditional on δ, λ, μ and η, the probability of bicycle arrivals and departures are respectively given by:

$$P_{qdtk} = \Lambda[\psi_k - ((\alpha' + \delta'_q + \delta'_{qd} + \delta'_{qdt})f + (\theta' + \mu'_q + \mu'_{qd} + \mu'_{qdt})g \pm \eta_q)] - \Lambda[\psi_{k-1} - ((\alpha' + \delta'_q + \delta'_{qd} + \delta'_{qdt})f$$
$$+ (\theta' + \mu'_q + \mu'_{qd} + \mu'_{qdt})g \pm \eta_q)] \tag{6}$$

$$P_{qdtl} = \Lambda[\psi_l - ((\tau' + \lambda'_q + \lambda'_{qd} + \lambda'_{qdt})h + (\theta' + \mu'_q + \mu'_{qd} + \mu'_{qdt})g \pm \eta_q)] - \Lambda[\psi_{l-1} - ((\tau' + \lambda'_q + \lambda'_{qd} + \lambda'_{qdt})h$$
$$+ (\theta' + \mu'_q + \mu'_{qd} + \mu'_{qdt})g \pm \eta_q)] \tag{7}$$

In the joint model of Eqs. 5–7, the parameters to be estimated are the vectors of $\beta, \alpha, \tau, \theta$, the ψ thresholds, and the variance terms of $\gamma^2, \delta^2, \lambda^2, \mu^2, \eta^2$. Thus, the joint likelihood function can be written for TAZ q as follows:

$$L_q|\gamma, \delta, \lambda, \mu, \eta = (P_{qj}|\gamma, \eta)\left(\prod_{d=1}^{D}\prod_{t=1}^{T}(P_{qdtk}|\delta, \mu, \eta) * \prod_{d=1}^{D}\prod_{t=1}^{T}(P_{qdtl}|\lambda, \mu, \eta)\right) \tag{8}$$

The unconditional likelihood function is given by:

$$L_q = \int_{\gamma, \delta, \lambda, \mu, \eta} (P_{qj}|\gamma, \eta)\left(\prod_{d=1}^{D}\prod_{t=1}^{T}(P_{qdtk}|\delta, \mu, \eta) * \prod_{d=1}^{D}\prod_{t=1}^{T}(P_{qdtl}|\lambda, \mu, \eta)\right)dF(\gamma, \delta, \lambda, \mu, \eta) \tag{9}$$

where F is the multidimensional cumulative normal distribution. The log-likelihood function is:

$$Ln\, L = \sum_q L_q \tag{10}$$

The likelihood function in Eq. (9) involves the evaluation of a multi-dimensional integral of size equal to the total number of elements in $\gamma, \delta, \lambda, \mu$ and η.

4.2. Model estimation

We apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals with respect to $\gamma, \delta, \lambda, \mu$ and η (Bhat, 2003). Since, we are decomposing the unobserved heterogeneity influencing arrivals and departures to three levels (i.e. TAZ component, TAZ - Day component, and TAZ - Day - Time period combination), different levels of simulation draws are needed to obtain simulated likelihood function. The code was programmed in Gauss matrix programming language. The maximum simulated likelihood estimator described is consistent under weak regularity conditions (see Hajivassiliou and Ruud, 1994; Lee, 1992; Eluru and Bhat, 2007).

5. Data

5.1. Sample formation

For this study, the arrival and departure rates are obtained from the minute-by-minute BIXI bicycle availability data for all stations in service (410 stations) between May and August 2012. The raw data extracted from the BIXI website was processed to generate the minute-by-minute arrival or departure rate of bicycles for every station (see Faghih-Imani et al., 2014 for more details on basic data preparation methods).

5.2. Dependent variable generation

By observing the bicycle flow pattern for arrivals and departures we categorized a day into five time periods: AM (6:00–10:00), Midday (10:00–16:00), PM (16:00–20:00), Evening (20:00–24:00), and Night (0:00–6:00). By aggregating obtained arrivals and departures for every station, we generated arrival and departure rates for these five time periods. Then, by adding arrivals and departures of all stations in a TAZ for the corresponding time period we obtained the TAZ level flows for every time period. BIXI stations are spatially located in 235 TAZs in Montreal. To retain a manageable data sample, we randomly selected <u>seven consecutive</u> days for every TAZ in our database (distinct set of 7 consecutive days for each TAZ). This sampling process provides reasonable coverage of data from May through August. The final sample consists of 8225 records (5 time periods * 7 days * 235 TAZs) of arrival and departure rates at a TAZ level.

The three dependent variables employed in our study - BSSI, arrivals and departures - are categorized as three ordered variables. The BSSI categories considered in our analysis (and sample shares) are $\leqslant$0 (16.2%), 0–1 (25.5%), 1–2 (26.8%), 2–3 (20.4%), >3 (11.1%). The arrivals and departures are categorized in four groups: very low-rate (<1 bicycle per hour), low-rate (1–5 bicycles per hour), medium-rate (6–10 bicycles per hour), and high-rate (10+ bicycles per hour). The proportion of arrivals and departures categories are as follows, respectively: very low-rate (31.4%, 32.1%), low-rate (47.3%, 45.6%), medium-rate (11.7%, 12.4%), and high-rate (9.6%, 9.9%).

5.3. Independent variables considered

Various independent variables are used to examine the determinants of bicycle usage for each TAZ. We controlled for <u>weather and temporal variables</u> to better capture determinants of BSS usage of every TAZ in our analysis. Weather variables

include average temperature and average relative humidity for every time period. Also, weather condition is represented as a categorical variable - whether it rained or not for every time period. Temporal variables considered include time of the day and day of the week effects. The influence of weekend and weekday were also accounted for. <u>Land-use and built environment characteristics</u> considered include the distance from central business district (CBD), an indicator for TAZs in the downtown area, the presence of transit in the TAZ, various points of interest, and job and population density. The size of TAZ is considered by normalizing the spatial variables with respect to the TAZ's area. The length of transportation network including major and minor roads and bicycle facilities (such as bicycle lanes, bicycle paths) is considered to identify cyclist preference of routes. The bicycle-sharing infrastructure is taken into account by using our proposed BSSI measure. To provide an illustration of the data compiled, a descriptive analysis of the sample is presented in Table 1.

6. Empirical analysis

As part of the model estimation exercise, four model structures are estimated: (1) three simple ordered logit models (3OL) model, (2) two dependent variable panel mixed ordered logit (2PMOL) model which only has TAZ level unobserved components (3) two dependent variable panel mixed multi-level ordered logit (2PMMOL) model which has all the three level unobserved components and (4) three dimensional panel mixed multi-level ordered logit (3PMMOL) model which has the three level unobserved components and the endogeneity parameter. The final specification of each of these models is obtained based on a systematic process of removing statistically insignificant variables guided by prior research, intuitiveness and parsimony considerations.

6.1. Model fit measures

To compare these four model frameworks, we employ three statistical measures: (1) Log-likelihood ratio test, (2) Akaike Information Criterion (AIC) and (3) Bayesian Information Criterion (BIC). The AIC and BIC penalize the modeling frameworks for additional parameters. For a given empirical model, $AIC = 2K - 2ln(L)$ where K is the number of parameters and $ln(L)$ is the log–likelihood value at convergence. The model with the lowest value of AIC is preferred. For a given empirical model, $BIC = -2ln(L) + Kln(Q)$ where Q is the number of observations. The BIC imposes higher penalty than AIC for over-fitting. The model with the lowest BIC values is the best model in terms of goodness of fit. Model fit measures are summarized in Table 2.

All the statistical measures employed - Log-likelihood ratio test, AIC, and BIC - clearly illustrate the superiority of the 3PMMOL model in terms of data fit. This comparison provides strong evidence in support of our hypothesis that model incorporating the common unobserved factors and endogeneity effect in the modeling of BSS usage offers an enhanced estimation framework.[4] For the sake of brevity, only the 3PMMOL model estimation results are discussed in this section and presented in Table 3.

6.2. Bicycle-sharing infrastructure installation model

As expected, bicycle facility density has a positive impact on the propensity of bicycle-sharing infrastructure installation in a TAZ. On the contrary, the propensity of installing BIXI infrastructure decreases as highway density and rail length increase. Highway and railway reflect locations that hinder cyclist's movements. Thus, it is expected that bicycle-sharing operators allocate stations away from such infrastructure. Moreover, distance to CBD negatively influences the propensity of BSS infrastructure installation while TAZs in the downtown area are more likely to have higher BSSI. Population density has a positive impact on bicycle-sharing infrastructure installation which is expected as the bicycle-sharing operators provide more stations and capacity where more people reside. The positive impact of job density and the number of restaurants in a TAZ on the propensity of BSS infrastructure installation illustrate that bicycle-sharing operators are likely to consider not only where people reside but also on activity opportunities in their decision process.

The results evidently indicate that bicycle-sharing infrastructure is not randomly allocated in the urban region. Considering Montreal's BIXI system as a mature system with reasonable success, the estimated BSS installation model results provide a guiding template for transportation planners and engineers across the world to model their installation decisions for a TAZ (or neighbourhood) as a function of existing land-use, built environment and bicycle infrastructure attributes.

6.3. Arrivals and departures model

Prior to discussing the arrival and departure models, it is important to recognize that several possible specifications were tested in model development. Specifically, the variables were considered to have distinct effects on arrivals (f matrix from Eq. (3)) and departures (h matrix from Eq. (4)). In cases where the mean parameter impacts across the two equations were not different a model that constrains the impact to be same across the two models was estimated (based on g matrix in Eq. (3)

---

[4] It is also important to recognize that the 2PMMOL and 2PMOL models perform substantially better than the 3OL model indicating that considering the impact of unobserved factors influencing arrivals and departures offers substantial improvement in fit. At the same time, we can clearly see that the 2PMMOL model outperforms the 2PMOL model in terms of goodness of fit, demonstrating the benefits of decomposing the unobserved effects into the three levels.

Table 1
Descriptive summary of sample characteristics.

Continuous variables	Min	Max	Mean	Std. deviation
Temperature (°C)	5.2	33	20.26	5.32
Relative humidity (%)	24	99	63.03	16.96
Elevation (m)	14.8	139.3	47.54	24.60
TAZ distance to CBD (km)	0.17	9.00	3.53	1.96
Length of bicycle facility in TAZ (km)	0	9.08	1.26	1.87
Length of streets in TAZ (km)	0.62	39.48	8.01	5.95
Length of major roads in TAZ (km)	0	10.30	1.03	1.23
Length of highways in TAZ (km)	0	8.36	0.39	0.93
Length of bus lines in TAZ (km)	0	12.81	2.42	2.22
Length of railways in TAZ (km)	0	6.00	0.42	0.84
Area of parks in TAZ (km^2)	0	1.45	0.03	0.14
Number of restaurants in TAZ	0	110	14.26	16.22
Number of other commercial enterprises in TAZ	0	1882	97.93	140.03
Number of BIXI stations in TAZ	1	6	1.74	0.94
Capacity of BIXI stations in TAZ	11	141	34.07	22.27
Station capacity	7	65	19.53	7.95
TAZ pop density (people per m^2 × 1000)	1.01	187.79	59.38	31.62
TAZ job density (jobs per m^2 × 1000)	0.07	4078.13	141.19	528.96

Categorical variables	Percentage
Rainy weather	8.3
Weekends	28.9
Metro station in TAZ	15.3
TAZ in downtown area	19.1
University in TAZ	12.8

Table 2
Model fit measures.

	3OL	2PMOL	2PMMOL	3PMMOL
Model description	3 Independent Ordered Logit	2 dimensional panel mixed ordered logit model	2 dimensional panel mixed multi-level ordered logit	3 dimensional panel mixed multi-level ordered logit
Number of parameters	38	42	47	48
LL at convergence	−14725.2	−11528.2	−11474.9	−11452.7
AIC	29526.4	23140.4	23043.7	23001.3
BIC	29793.0	23435.0	23373.4	23338.0
LL ratio test				
3PMMOL vs 3OL	6545 (23.2[a])			
2PMMOL vs 2PMOL	106.6 (16.8[a])			
3PMMOL vs 2PMMOL	44.4 (6.64[a])			

[a] The corresponding chi-square value at 99% level of significance.

and (4)). The final estimation results clearly show variable impacts from *f*, *g* and *h* matrices. The same approach was employed for estimation of random parameters.

6.3.1. Weather and temporal variables

There is a positive correlation between temperature and BSS usage. On the other hand, rainy weather conditions and humidity have negative impacts on arrival and departure rates, expectedly. People tend to bicycle more on weekdays than weekends as highlighted by the negative coefficient of the weekend variable. The result indicates that bicycle-sharing system is used more for daily activities than weekend leisure activities. The interpretation of time of day variables needs to be judiciously undertaken due to the presence of interaction effects with population density, job density, and university variables. Nevertheless, we clearly observe that BIXI system is predominantly used during the PM and less during night period relative to other times of the day. One plausible explanation for higher usage of the system in PM is that employed people might also consider riding the BIXI as a useful exercise after work or might make short trips within the system area—for instance, going from work to a restaurant. Furthermore, it is also possible that during the evening peak hour the population using BIXI includes tourists, non-member users and other individuals without the typical work schedule (e.g., students).

6.3.2. Land-use and built environment characteristics

It is expected that the arrival and departure rates decrease when the station location is farther from CBD as highlighted by the negative coefficient of distance to CBD variable. The positive coefficient of the presence of metro station variable indicates that BIXI stations near metro system are more likely to have higher usage; a strong evidence in favour of the bicycle-sharing systems support of multimodal transport. The number of restaurants in a TAZ is associated with a positive

Table 3
3PMMOL model estimation results.

Parameter		BSS infrastructure		Arrival rate		Departure Rate	
		Estimate	t-stat.	Estimate	t-stat.	Estimate	t-stat.
Threshold 1		−2.0132	−2.866	−3.1451	−7.621	−3.1451	−7.621
Threshold 2		0.0756	0.113	1.7097	4.145	1.7097	4.145
Threshold 3		1.9573	3.086	4.5364	10.945	4.5364	10.945
Threshold 4		4.1434	5.489	N/A	N/A	N/A	N/A
Weather & time variables							
Temperature				0.5297	7.794	0.5297	7.794
Relative humidity				−3.6787	−19.086	−3.6787	−19.086
Rainy				−0.5267	−8.464	−0.5267	−8.464
Weekend				−0.7867	−19.091	−0.7867	−19.091
PM				1.6961	20.314	1.6961	20.314
Night				−3.2944	−26.013	−3.2944	−26.013
TAZ variables							
Bicycle facility density		0.0714	2.097	0.0865	5.203	0.0865	5.203
Highway density		−0.1058	−1.757	−0.1642	−5.865	−0.1642	−5.865
Rail length		−0.5133	−2.623				
Downtown		1.0132	1.722				
Distance to CBD		−0.2931	−2.506	−0.2759	−4.737	−0.2759	−4.737
Metro Station in TAZ				0.9273	4.525	0.9273	4.525
Number of restaurants in TAZ		3.1938	2.703	0.9764	2.278	0.9764	2.278
University in TAZ ∗ AM				0.6593	3.13	−0.8526	−4.221
University in TAZ ∗ PM				−0.5204	−2.448	0.6458	2.776
TAZ job density		0.7556	2.054				
TAZ job density ∗ AM				1.0594	9.955	−0.289	−3.698
TAZ population density		14.6527	2.618				
TAZ population density ∗ AM				−9.6899	−10.245	10.2661	9.933
TAZ population density ∗ PM						−6.662	−4.941
BSSI (BSSI Low is base)							
BSSI medium				0.5911	1.776	0.5911	1.776
BSSI High				3.1816	9.861	3.1816	9.861
Standard deviation estimates							
TAZ	BSSI low			1.3771	10.896	1.3771	10.896
	BSSI medium			1.4776	11.938	1.4776	11.938
	BSSI high			2.1339	17.452	2.1339	17.452
	Night time Period			1.1828	12.073	1.1828	12.073
	AM and PM Time Period			0.4267	5.941	0.4267	5.941
TAZ-Day	Constant			0.2549	7.702	0.2549	7.702
	Bicycle facility density			0.0246	4.701	0.0246	4.701
TAZ-Day-Time Period	Constant			0.1692	3.273	0.1692	3.273
	TAZ Job Density			4.295	8.277	4.295	8.277
	TAZ Population Density			0.3023	7.940	0.3023	7.940
Standard deviation of endogeneity component (η_q)		1.1585	10.294	1.1585	10.294	1.1585	10.294

impact on usage of BIXI system. The coefficients for the presence of university in a TAZ show interesting results. In AM period, the presence of university in a TAZ has a positive impact on arrivals while a negative impact on departures propensity. On the other hand, it has a negative impact on arrivals and a positive impact on departure propensity in PM period. These findings evidently demonstrate the use of BIXI system for commuting to/from universities. A similar pattern also can be observed from the interaction of population and job density variables with time periods. TAZs with higher population density tend to have higher departure and lower arrival rates in the AM and higher arrival rates in the PM. People use BIXI service as commute mode to their work in the morning peak hour which is recognized by the coefficient of TAZ job density in AM period.

6.3.3. Bicycle infrastructure variables

Bicycle-sharing infrastructure variables are introduced as the following dummy variables: Low BSS (BSSI < 1), Medium BSS (1 ⩽ BSSI < 2) and High BSS (2 ⩽ BSSI). As expected, the medium and high BSS variables have a significant positive effect on arrivals and departures propensity. As expected, the propensity of arrivals and departures increase in zones with higher bicycle facilities (bicycle lanes, bicycle paths, etc.) while the density of highways in TAZ has a negative impact on arrivals and departures.

6.4. Endogeneity and Unobserved Heterogeneity

The estimated standard deviation results for unobserved heterogeneity as well as endogeneity are presented at the bottom part of Table 3. In the model estimation exercise, several alternative specifications have been examined for the unob-

served effect variables (γ, δ, λ, and μ) at the three level of variations. The specification that offered superior fit was retained; i.e. 10 parameters capturing common unobserved effects between usage dimensions, and the endogeneity parameter recognizing the BSS installation impact on usage. The final specification has 5 common unobserved effect parameters estimated influencing arrivals and departures at TAZ level (μ_q), 2 at the TAZ - Day level (μ_{qd}), and 3 at TAZ – Day - Time period level (μ_{qdt}).

6.4.1. TAZ level estimates

The estimates for the standard deviation of unobserved factors exhibit significant variability across the various TAZs in impacting usage for several exogenous variables. The BSSI levels (Low, Medium and High) exhibit reasonably varied impact across the region. For example, while the mean impact of BSSI Low category is 0 the parameter estimate in the unobserved heterogeneity component indicates the existence of a substantial variance of the BSSI low variable across the Montreal region. Similarly, BSSI Medium and High categories also exhibit variation across the region. It is important to note while there exists substantial variability across the three BSSI levels, the distribution of parameters (mean and standard deviation) are such that for a large number of cases the relationship between three levels is along expected lines (BSSI High > BSSI Medium > BSSI Low). The results also indicate significant variation temporally across the TAZs. Specifically, the Night time period and Peak periods (AM and PM) exhibit large variations across the TAZs. These parameter estimates confirm the presence of unobserved factors that jointly affect arrivals and departures at a TAZ level.

6.4.2. TAZ – Day level estimates

We observe significant variability in the usage for each day across the TAZs highlighted by the unobserved effect for a constant term at the day level. The results also demonstrate significant variation in the impact of bicycle facility density variable. While the bicycle facility density variable is a TAZ specific variable, the random parameter is significant for the TAZ – Day level highlighting how the unobserved heterogeneity of exogenous variable can be at a different resolution.

6.4.3. TAZ – Day – Time period level estimates

The results for model estimation at the finest resolution indicate the presence of significant variations across time periods. There is an overall unobserved effect captured by the constant term. Further, the population and job density variables also reveal substantial variability at this resolution. The results indicate that although population and job density variables are time invariant variables, their effects vary for every time period; thus highlighting how the influence of population and job density can be affected by time period. The existence of such multi-level variations supports our hypothesis that decomposing the unobserved effects is beneficial. The population, job, or bicycle facility density variables did not have significant variation when we considered them at the TAZ level variation. However, by employing different levels in the specification of unobserved effects, we can capture the impact of these variables better.

6.4.4. Endogeneity variable

The final row of Table 3 presents the standard deviation estimate related to the endogeneity component of the 3PMMOL model with a positive sign in usage equations. The magnitude and significance of the endogeneity parameter indicate the presence of the impact of BSS installation process on usage. More significantly, the introduction of the endogeneity component substantially reduces the magnitude of the BSSI variables in the model (relative to the estimates in 3OL, 2PMOL and 2PMMOL models). This reduction in magnitude (which is further confirmed and presented through elasticity analysis in next section) supports our hypothesis that ignoring the BSS installation process results in over-estimating the impact of BSS infrastructure while under-estimating the impact of other exogenous variables.

6.5. Elasticity effects

The proposed approach focuses on accounting for the influence of the bicycle-sharing infrastructure installation process on the usage models while recognizing the effect of bicycle-sharing infrastructure itself on the arrival and departure rates. To further examine the model results obtained, we conduct an elasticity analysis for a select set of exogenous variables. The elasticity measures are computed by calculating the percentage change in the TAZ bicycle arrivals and departures due to the change in the exogenous variables. The elasticity computation is conducted by computing the difference in the probability distribution for the ordered alternatives and multiplying it with the mean value of the category. Rather than employ a point estimation of elasticity approach, we consider estimating the distribution of the elasticity effects (see Eluru et al., 2008 for a similar exercise in a different context). These measures are computed for the entire sample as well as for the specific time periods of AM and PM. The scenarios considered include: (a) increasing *only* the number of stations in a TAZ without increasing the capacity in a TAZ, i.e., reallocating current capacity to add new stations, (b) increasing *only* the capacity in a TAZ without increasing the number of stations in a TAZ, (c) increasing number of stations and capacity in a TAZ, (d) increasing population density by 25%, (e) increasing job density by 25%, and (f) increasing bicycle facility density by 10% and 25%.

To conduct a comparison across all model frameworks the elasticity measures and their standard deviations are generated for three models - 3OL, 2PMMOL and 3PMMOL – and presented in Table 4. Several significant observations can be made based on the results. First, the 3OL and 2PMMOL models over-estimate the expected impact of BSS infrastructure. For

Table 4
Elasticity effects (and its standard deviation) for TAZ arrival and departure rates.

	3OL			2PMMOL			3PMMOL		
	Overall	AM	PM	Overall	AM	PM	Overall	AM	PM
Arrivals									
No. of stations + 3, same capacity	3.86 (0.04)	3.97 (0.04)	3.72 (0.03)	3.88 (0.32)	4.21 (0.34)	3.76 (0.34)	2.33 (0.33)	2.62 (0.39)	2.17 (0.32)
Capacity + 25, same No. of Stations	4.4 (0.04)	4.42 (0.05)	4.46 (0.04)	4.25 (0.45)	4.39 (0.43)	4.44 (0.53)	2.07 (0.45)	2.3 (0.44)	2.09 (0.47)
No. of stations + 3, capacity + 15	6.93 (0.06)	7.03 (0.07)	6.85 (0.06)	6.79 (0.65)	7.29 (0.67)	6.72 (0.74)	3.6 (0.65)	4.07 (0.68)	3.39 (0.68)
Population density + 25%	−0.71 (0.06)	−3.99 (0.37)	0 (0)	−0.67 (0.05)	−3.75 (0.38)	0 (0)	−0.69 (0.07)	−3.82 (0.48)	0 (0)
Job density + 25%	0.16 (0.03)	0.88 (0.14)	0 (0)	0.33 (0.05)	1.84 (0.28)	0 (0)	0.37 (0.07)	2.06 (0.39)	0 (0)
Bicycle facility density + 10%	1.19 (0.02)	1.3 (0.02)	1.00 (0.02)	1.69 (0.36)	1.81 (0.4)	1.45 (0.29)	1.46 (0.26)	1.49 (0.28)	1.36 (0.23)
Bicycle facility density + 25%	2.99 (0.05)	3.28 (0.06)	2.52 (0.06)	4.27 (0.9)	4.6 (1.03)	3.64 (0.71)	3.71 (0.66)	3.78 (0.73)	3.41 (0.59)
Departures									
No. of Stations + 3, same capacity	3.86 (0.04)	3.94 (0.06)	3.71 (0.04)	3.83 (0.3)	3.91 (0.36)	3.77 (0.28)	2.29 (0.39)	2.39 (0.42)	2.2 (0.39)
Capacity + 25, same no. of stations	4.36 (0.05)	4.44 (0.07)	4.35 (0.05)	4.2 (0.46)	4.14 (0.52)	4.35 (0.47)	2.00 (0.58)	1.93 (0.61)	2.09 (0.58)
No. of stations + 3, capacity + 15	6.9 (0.08)	7.11 (0.1)	6.72 (0.08)	6.73 (0.64)	6.85 (0.73)	6.65 (0.66)	3.52 (0.8)	3.65 (0.84)	3.39 (0.8)
Population density + 25%	0.09 (0.16)	3.97 (0.31)	−2.49 (0.46)	0.2 (0.19)	4.75 (0.56)	−2.55 (0.49)	0.19 (0.17)	4.68 (0.63)	−2.51 (0.49)
Job density + 25%	−0.18 (0.01)	−0.83 (0.05)	0 (0)	−0.11 (0.02)	−0.53 (0.12)	0 (0)	−0.08 (0.02)	−0.4 (0.12)	0 (0)
Bicycle facility density + 10%	1.19 (0.02)	1.25 (0.03)	1.03 (0.02)	1.78 (0.41)	1.89 (0.46)	1.55 (0.33)	1.48 (0.32)	1.52 (0.34)	1.37 (0.27)
Bicycle facility density + 25%	2.99 (0.06)	3.16 (0.07)	2.58 (0.06)	4.5 (1.06)	4.79 (1.15)	3.88 (0.83)	3.74 (0.81)	3.86 (0.88)	3.46 (0.69)

instance, for arrivals' model, the over-estimation in the average effect is 66% for the number of stations scenario while the corresponding value is 103.5% for station capacity scenario. Similar trends are observed for departures' model as well. Second, the impact of BSS infrastructure is very similar in 3OL and 2PMMOL models highlighting that it is not adequate to consider common unobserved factors affecting arrivals and departures in the modeling exercise. Third, although the mean expected impact of population, job and bicycle facility density are marginally different for the three models, the differences between 2PMMOL and 3PMMOL are not statistically significant. Fourth, we observe that AM period has the highest sensitivity to changes in exogenous variables across the board. Finally, we observe that the models offer very slight differences across arrivals and departures indicating a symmetric impact on the two usage dimensions.

Overall, the elasticity analysis clearly demonstrates the need to incorporate the BSS installation process in order to estimate the "true" impact of BSS infrastructure on the BSS usage. The analysis shows that ignoring the BSS installation process leads to a statistically significant over-estimation of the BSS infrastructure variable impact. However, comparisons for the land-use and urban form variables impacts on BSS usage do not yield statistically significant differences to characterize them either as an under-estimation or an over-estimation.

7. Conclusions

The typical approach for analyzing BSS usage involves examining the influence of BSS infrastructure (e.g. number of BSS stations and stations' capacity), transportation network infrastructure, land use and urban form, meteorological data, and temporal characteristics. Several studies demonstrate that increasing BSS infrastructure (number of stations and capacity) increases BSS usage. Earlier research efforts ignore the BSS infrastructure installation decision process i.e. BSS operators installed the infrastructure based on an expectation of system usage. The model estimates obtained with this erroneous assumption are very likely to over-estimate the impact of BSS infrastructure. The current research effort proposes a joint econometric framework that remedies this drawback.

We opt for TAZ system as the spatial aggregation unit for our study. Considering all BSS infrastructure variables and usage variables in their original form would result in a large number of choice equations (one per variable). Hence, to simplify the econometric model development we propose a single measure of BSS infrastructure that simultaneously encapsulates the number of stations and total capacity in a TAZ. Having a single equation for BSS infrastructure allows us to efficiently accommodate for the potential presence of endogeneity in the model system. Of course, alternate BSSI index structures could be explored in future research.

We formulate a multi-level joint econometric framework to study the BSS installation process and usage. In all the earlier studies, self-selection was considered in a choice framework based on a cross-sectional dataset where one measurement equation was coupled with one choice equation to disentangle the influence of endogeneity. However, in the BSS context, the infrastructure installation process has only one occurrence, while the infrastructure usage process (arrivals and departures) has multiple repeated observations. Hence, the econometric model developed should account for self-selection while accounting for repeated measures of the choice process. Thus, to correctly identify the confounding factors the econometric framework needs to accommodate for endogeneity while considering the influence of unobserved factors on repeated observations at multiple levels (all records for the spatial unit, all records for the spatial unit in a day and unobserved heterogeneity for each observation). Hence, a common spatial unit specific component across the BSS infrastructure measurement and repeated usage equations needs to be augmented with multi-level common unobserved effects.

The proposed joint modeling process thus estimates a more accurate impact of BSS infrastructure on usage. We consider an ordered representation for all the dependent variables yielding a three dimensional panel ordered formulation. The proposed model is estimated using data compiled from the Montreal bicycle-sharing system, BIXI, from April to August 2012. The proposed joint econometric model – a 3 dimensional panel multi-level mixed ordered logit model (3PMMOL) was compared with the three dimensional independent ordered logit model (3OL) that does not accommodate for common unobserved heterogeneity, a 2 dimensional panel multi-level mixed ordered logit model 2PMMOL model that ignores the presence of endogeneity of BSS infrastructure on the usage and 2PMOL model that restricts the common unobserved effects between arrivals and departures to the TAZ level. The model fit measures clearly provide evidence to support our hypothesis that ignoring the installation process and decomposing the unobserved effects into finer resolutions affect model estimates. To further examine the advantages of the proposed model framework, elasticity impacts for a host of policy variables were computed. The elasticity effects predicted for changes in BSS infrastructure variables indicate that ignoring the potential relationship between BSS infrastructure installation process and usage could result in significant over-estimation of the impact of BSS infrastructure (varies between 66% and 118%). While the elasticity analysis results for population, job and bicycle facilities density variables do not yield statistically significant differences between different models, the mean errors can range from −57% to 125%.

Besides, the empirical analysis has several findings. The results evidently indicate that bicycle-sharing infrastructure is not randomly allocated in the urban region. Considering Montreal's BIXI system as a mature system with reasonable success, the estimated BSS installation model results can be used as a guiding template to model BSS installation decisions for a TAZ (or neighbourhood) as a function of existing land-use, built environment attributes. Moreover, our estimated results on factors influencing arrivals and departures provide interesting findings. The results demonstrate the significant impact of weather characteristics as well as the time of the day and weekend variables on BSS usage. The use of BIXI system for the daily commute to and from work and universities is also highlighted by the results. The findings provide useful information for designing or modifying bicycle-sharing systems with the goal of maximizing usage.

Acknowledgements

References

Abay, K.A., Paleti, R., Bhat, C.R., 2013. The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. Transport. Res. Part B Method. 50, 74–89.

Bhat, C.R., Guo, J.Y., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transp. Res. Part B 41 (5), 506–552.

Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transp. Res. Part B 37 (9), 837–855.

Bhat, C.R., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transp. Res. Part B 43, 749–765.

Cao, X., Mokhtarian, P.L., Handy, S.L., 2009. Examining the impacts of residential self-selection on travel behavior: a focus on empirical findings. Transp. Rev. 29, 359–395.

Chakour, V., Eluru, N., 2016. Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal. J. Transp. Geogr. 51, 205–217.

Cheng, W., Wang, J.H., Bryden, G., Ye, X., Jia, X., 2013. An examination of the endogeneity of speed limits and accident counts in crash models. J. Transport. Saf. Secur. 5, 314–326.

Eluru, N., Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accid. Anal. Prev. 39 (5), 1037–1049.

Eluru, N., Bhat, C.R., Pendyala, R.M., Konduri, K.C., 2010. A Joint Flexible Econometric Model System of Household Residential Location and Vehicle Fleet Composition/Usage Choices, Transportation, 2010 TRB Special Issue, vol. 37, no. 4, pp. 603–626.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accid. Anal. Prev. 40, 1033–1054.

Ewing, R., Tian, G., Goates, J.P., Zhang, M., Greenwald, M.J., Joyce, A., Kircher, J., Greene, W., 2015. Varying influences of the built environment on household travel in 15 diverse regions of the United States. Urban Stud. 52, 2330–2348.

Faghih-Imani, A., Eluru, N., 2015. Analyzing bicycle sharing system user destination choice preferences: an investigation of Chicago's Divvy system. J. Transp. Geogr. 44, 53–64.

Faghih-Imani, A., Eluru, N., 2016. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: a case study of New York CitiBike system. J. Transp. Geogr. 54, 218–227.

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. J. Transp. Geogr. 41, 306–314.

Gebhart, K., Noland, R., 2014. The impact of weather conditions on bikeshare trips in Washington, DC. Transportation 41, 1205–1225.

Greene, W.H., 2012. Econometric Analysis. Prentice-Hall, New Jersey.

Greenwald, M.J., Boarnet, M.G., 2001. Built environment as determinant of walking behavior-analyzing nonwork pedestrian travel in Portland, Oregon. Transp. Res. Rec. 1780, 33–42.

Hajivassiliou, V.A., Ruud, P.A., 1994. Classical estimation methods for LDV Models using simulations. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. IV. Elsevier, New York, pp. 2383–2441.

Hu, L., 2016. Transit, automobile, and commuting: do the relationships differ across income groups?. Presented in Transportation Research Board 95th Annual Meeting (No. 16-4347).

Lee, L.-F., 1992. On the efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Econ. Theory 8, 518–552.

Nair, R., Miller-Hooks, E., Hampshire, R., Busic, A., 2013. Large-scale vehicle sharing systems: analysis of velib. Int. J. Sustain. Transport. 7, 85–106.

Noland, R.B., Cowart, W.A., 2000. Analysis of metropolitan highway capacity and the growth in vehicle miles of travel. Transportation 27, 363–390.

Pinjari, A.R., Eluru, N., Bhat, C.R., Pendyala, R.M., Spissu, E., 2008. A joint model of residential location and bicycle ownership: accounting for self-selection and unobserved heterogeneity. Transp. Res. Rec. 2082, 17–26.

Rana, T., Sikder, S., Pinjari, A., 2010. Copula-based method for addressing endogeneity in models of severity of traffic crash injuries: application to two-vehicle crashes. Transport. Res. Rec. J. Transport. Res. Board, 75–87 (No. 2147).

Rixey, R., 2013. Station-level forecasting of bikesharing ridership: station network effects in three U.S., Systems. Transport. Res. Rec. 2387, 46–55.

Schauder, S.A., Foley, M.C., 2015. The relationship between active transportation and health. J. Transp. Heal. 2, 343–349.

Schwanen, T., Mokhtarian, P.L., 2005. What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? J. Transp. Geogr. 13, 83–99.

Shaheen, S., Guzman, S., Zhang, H., 2010. Bikesharing in Europe, the Americas, and Asia Past, present, and future. Transp. Res. Rec. 2143, 159–167.

Vij, A., Walker, J.L., 2014. Preference endogeneity in discrete choice models. Transport. Res. Part B Method. 64, 90–105.

Walker, J.L., Ehlers, E., Banerjee, I., Dugundji, E.R., 2011. Correcting for endogeneity in behavioral choice models with social influence variables. Transport. Res. Part A Policy Pract. 45, 362–374.

Wang, X., Lindsey, G., Schoner, J., Harrison, A., 2015. Modeling bike share station activity: effects of nearby businesses and jobs on trips to and from stations. Forthcoming J. Urban Plan. Develop.

Yasmin, S., Eluru, N., Pinjari, A.R., 2015. Analyzing the continuum of fatal crashes: a generalized ordered approach. Anal. Meth. Accident Res. 7, 1–15.

Yasmin, S., Eluru, N., Pinjari, A.R., Tay, R., 2014. Examining driver injury severity in two vehicle crashes–a copula based approach. Accid. Anal. Prev. 66, 120–135.