CHICAGO BOOTH

The University of Chicago Booth School of Business

Working Paper No. 15-04

Bike-Share Systems: Accessibility and Availability

Ashish Kabra
INSEAD

Elena Belavina
University of Chicago Booth School of Business

Karan Girotra
INSEAD

BIKE-SHARE SYSTEMS: ACCESSIBILITY AND AVAILABILITY

Abstract. The cities of Paris, London, Chicago, and New York (among many others) have set up bike-share systems to facilitate the use of bicycles for urban commuting. This paper estimates the impact of two facets of system performance on bike-share ridership: accessibility (how far the user must walk to reach stations) and bike-availability (the likelihood of finding a bicycle). We obtain these estimates from a structural demand model for ridership estimated using data from the Vélib' system in Paris. We find that every additional meter of walking to a station decreases a user's likelihood of using a bike from that station by 0.194% ($\pm$0.0693%), the reduction is even more significant at higher distances (>300m). These estimates imply that almost 80% of bike-share usage comes from areas within 300m of stations, highlighting the need for dense station networks. We find that a 10% increase in bike-availability would increase ridership by 12.211% ($\pm$1.097%), three-fourths of which comes from fewer abandonments, and the rest from increased user interest. We illustrate the use of our estimates in comparing the effect of adding stations or increasing bike-availabilities in different parts of the city, at different times, and in evaluating other proposed improvements.

1. Introduction

Urban agglomerations across Asia, Europe, and the Americas are faced with unprecedented traffic congestion and poor air quality that threatens their attractiveness to citizens and businesses. An increased use of bicycles for urban commuting can help alleviate both these concerns. The cities of Paris, Barcelona, London, Wuhan, Hangzhou, Shanghai, New York, and Chicago (among many others) have thus set up large-scale *bike-share systems* that facilitate the use of bicycles in cities.[1]

A bike-share system includes a communal stock of sturdy, low-maintenance bicycles and a network of parking stations. Each station provides 10–100 automated parking spots (docks) and a control kiosk. A registered user can "check out" any available bicycle from a station and at the end of her commute, can return the bicycle to *any station* in the network. Bike-share systems eliminate barriers to bike-ownership such as the lack of safe parking spaces for bikes in cities, vandalism and theft of bicycles, and the inconvenience and cost of owning and maintaining a bicycle. They also facilitate one-way trips that make bicycles an effective "last-mile" complement to other public transit systems, such as bus, metro or regional rail.

Although bike-share systems have garnered considerable attention, their promise of urban transformation is far from being fully realized. A key reason is that while providers and operators have focused on bike-design and technology aspects, there is limited rigorous analysis of operational aspects such as

---

[1]As of June 2014, public bike-share systems were operating in 712 cities with approximately 806,200 bicycles and 37,500 stations (Wikipedia entry on "Bicycle-Sharing system").

station density and service-levels, nor are the user responses to such aspects understood [Tangel, 2014]. The aim of this paper is to identify relationships between ridership and operational performance of a bike-share system, and to illustrate the use of these relationships in designing systems that achieve higher ridership.

In particular, we estimate the impact of two facets of operational performance on ridership: station accessibility, or how far a user must walk to reach a station; and availability, or the likelihood of finding bicycles at stations. There are, in turn, two aspects of availability. First, if nearby stations don't have bicycles *at the instance* when a user wants to take a trip, users must substitute to farther stations or abandon using bike-share. The extent of this abandonment, "lost-sales" in traditional operations parlance, is the *short-term effect* of availability. The second *long-term* aspect is that– if users *typically expect* a lower chance of finding a bicycle, they are less likely to even consider bike-share for their commutes and the system will have lower ridership. Put differently, the short-term effect concerns the extent of demand that is converted to sales, whereas the long-term concerns the level of demand itself.

Directly estimating the relationship between ridership and the distances users must walk, requires observing ridership at different stations and the distance users' must walk to access the stations. The latter requires knowledge of the users' starting point or origin locations. Neither we nor any bike-share operator has data on these origin locations. As such, the effect of distance on use can not be estimated from any reduced form regression model (linear, MNL, etc.), as we do not observe the key independent variable— the distances users must walk to access various stations. We follow past work in the marketing, economics, and operations literatures (see Berry et al. [1995], Davis [2006], Pancras et al. [2012], Allon et al. [2011], Guajardo et al. [2015] and the references therein) on estimating the impact of unobserved user characteristics on demand. Specifically, we follow Davis [2006], which itself extends the celebrated work of Berry et al. [1995], and build a structural demand model that allows us to estimate the impact of unobserved walking distances and bike-availability on ridership, using data on ridership at different stations and different times.

Our structural model has two components. The first component, an *origin-density* model, provides the number of potential users originating at a location in a unit time. More specifically, the origin-density model relates the rate of *potential* bike-share users *originating* at any location to the residential population density, the presence of various points of interest (metro stations, cafes, hotels, stores, etc.) at the location, and importantly, to the *typical likelihood* of users originating at this location finding bikes at nearby stations. The typical likelihood captures a user's belief/expectation that she would find a bike.

Users originating at a location can choose between different nearby stations and other options for commuting. The second component, a multinomial-logit *choice model,* captures the probabilities of

these choices. Specifically, this model provides the probabilities with which a user originating *at a given location* chooses between different stations and outside options *as a function of the distances* to different stations from this location.

These two components combine to lead to the predicted ridership at each station. Specifically, the predicted ridership at a station is the sum of users from all possible origin locations that choose this station. The parameters of the origin-density model and those of the choice model are then estimated using observed data on the station ridership. These parameters quantify the effects of walking distance and bike-availability on ridership. Station locations and bike-availabilities are endogenous and we use instruments (as in Berry et al., 1995, Davis, 2006) to correct the resulting bias in estimates. The full description of the model and the estimation strategy is provided in Sections 4 and 5.

Our model follows past work on structural demand models for spatially differentiated products [Davis, 2006, Thomadsen, 2005, Allon et al., 2011]. Yet it departs in two respects: (1) Past work assumes the same number of users at each location in a district or census unit (uniform density). Instead, we use a non-uniform density model which allows for different numbers of users to originate at different locations in the city. (2) Past work assumes that all products/choices are always available, ignoring demand censoring, which is known to introduce a bias [Bruno and Vilcassim, 2008, Conlon and Mortimer, 2013]; we exclude stations that do not have available bikes from user choice sets, which addresses this bias and also allows us to include and measure the short-term impacts of stocked out stations.

Unfortunately, these two differences drastically increase the computational burden of our model. Incorporating information on which stations have available bikes at each instance makes each time-instance unique, while assuming a non-uniform density makes locations unique. We then develop a transformation of our data which significantly reduces the computational burden. In short, rather than consider each time uniquely, we combine times when the set of nearby stations which are stocked with bikes is the same, into one data-point.

We estimate our model using data from the Vélib' bike-share system in Paris, the biggest bike-share system outside of China. Our data is based on observing 946 bike-stations in central Paris for a period of four months, which cover more than 4.35 million trips. We obtain data on residential population density, and locations of various points of interest from Google Places and various French public agencies (statistics: INSEE; transport: RATP; tourism: OTCP).

As one would expect, we find that a users' likelihood of using bike-share decreases with the distances they must walk. For walking distances between 0-300 meters, every additional meter of walking to a station decreases a user's likelihood of using a bike from that station by 0.194% ($\pm$0.069%), i.e. a user that originates 300 meters away from a station is almost 60% less likely to use a station than one that originates close to the station. The effect is even higher after the first 300 meters, every

additional meter of walking decreases this likelihood by 1.307% (±0.426%). These numbers imply that users more than 500m away from a station are highly unlikely to use the system. Overall, these estimates highlight the very local nature of bike-share demand and consequently the need for a very dense network of stations to make bike-share systems effective.

We also find that increasing bike-availability increases ridership. A 10% *increase* in bike-availability at stations would increase system-use by about 12.211% (±1.097%), of which about three-fourths of the effect (9.493%) arises from fewer lost trips (short-term effect) and the rest (2.482%) is due to increased user interest (long-term effect). These estimates imply that only 5.070% (±0.538%) of the demand substitutes to nearby stations when confronted with a stock-out at the station of choice. This highlights the importance of maintaining high bike-availability levels, and accounting for availability information when modeling user choices.

Our estimated structural model provides an out-of-sample fit of over 60%, and significantly outperforms all other reduced form models. The above estimates are robust to multiple alternate model specifications, functional forms of the distance disutility, variable definitions, computational choices, instrument choices, etc.

The estimated model can be used to predict ridership for any given station network and bike-availability levels. This provides a powerful tool for system managers to evaluate and compare changes to the station-network, and/or system management policies that change bike-availabilities. We illustrate some such use cases of our model For example, we can use the model to compare the effect of adding stations or increasing bike-availabilities in different parts of the city and at different times of the day. Among other prescriptions, our model predicts that increasing station density in residential districts will be more useful than increasing it in commercial districts. On the other hand, the benefits of improving bike-availability are higher in commercial districts. Further, adding stations in areas closer to supermarkets provides more benefits than adding them closer to public transit and other points of interest. Stations closest to supermarkets are also the best to target for bike-availability improvements. Mobile stations which could be moved to temporarily increase station density, if available, could be usefully employed in district 12 in the morning, and moved to district 7 in the afternoon and evening. Mobile availability improvement resources (such as transshipment trucks) would be useful in districts 11 and 12 in the morning hours, and then assigned to districts 4 and 7 in the evening hours. Overall, these prescriptions illustrate some ways in which the obtained estimates can be used to improve bike-share systems.

Our study provides the first estimates for the relationship between ridership, and accessibility (walking distance) and bike-availability, in the context of bike-share systems. We illustrate the use of these estimates to provide actionable prescriptions for system improvement. Our analysis adapts methods from

the demand-estimation literature to the context of public-transportation, and extends these methods to account for the effects of changing product availability and non-uniform user origin densities. We hope this can be a template for future research on estimating aspects of user behavior in the context of other disruptive transportation models such as ride-sharing, on-demand public-transport, etc.

2. Literature Review

This paper is related to fledgling research on bike-share systems and to other studies that measure the customer response to accessibility and availability.

Bike-Share Systems: Recent research has employed operations research methods to optimize bike-share system design and operation, considering key decisions such as number of bikes [George and Xia, 2011], station locations [García-Palomares et al., 2012] and transshipment of bikes (Henderson et al., 2016 and the references therein). Another stream is concerned with predicting ridership using demographic and traffic data (see Singhvi et al., 2015 and references therein). Pendem and Deshpande [2016] combine the two streams by using an empirical demand model as an input to optimal bike-allocation. Neither stream has explored the empirical consumer response to operational performance, the focus of our work.

Accessibility and Availability: The notion of accessibility has been studied in the context of motorized public-transportation systems via surveys [El-Geneidy et al., 2014], and in the context of retail networks using archival data, like this paper [Davis, 2006, Pancras et al., 2012, Allon et al., 2011, Thomadsen, 2005]. The impact of product availability has been studied in the context of consumer goods (eg: Musalem et al., 2010) and in a mail-order catalogue context [Anderson et al., 2006].

Transportation Choice: Problems in transportation choice have primarily been modeled using three main approaches– surveys that directly measure user preferences, distances traveled, etc. (see El-Geneidy et al., 2014 and references therein), early work that used gravity models (estimated on city or district level aggregate data, for e.g. Reilly, 1931), and using multinomial logit-like models on individual user-level data (for e.g. McFadden, 1974). Our approach builds on the third class of models by incorporating unobserved consumer heterogeneity to allow estimation from aggregated data and by addressing endogeneity concerns ignored in past work.

Finally, our work is part of a renewed interest in studying facility-location, (sustainable) transportation and spatial competition in the operations management community (see for e.g. Cachon 2014, Lederman et al. 2014, Li et al. 2015, Belavina et al. 2016).

3. Data and Variable Construction

We estimate our model using data from the Vélib' bike-share system in Paris. Vélib', the biggest bike-share system outside of China and the densest system in the world, includes 946 bike stations located in the city of Paris that house roughly 17,000 bicycles.[2] Our data is built by capturing a public data stream that publishes the status of stations in the network, every two minutes.[3] Each two-minute observation that we collect contains the number of bikes and the number of empty docking points at each bike-station. We collect these snapshots for weekdays in a four-month period starting in May 2013.

Next, we describe the use of this data to construct the variables that will enter the model we build in Section 4. We start with the definition of station-use.

3.1. **Station-Use.** S*tation-use is the rate at which trips start at a bike-share station at a given time.*

We consider each decrement in the number of bikes at a station as an instance of a bicycle being checked out and used for a trip. Now, observed station-use at a station f at time t, Λ_{ft} is defined as $\frac{1}{2}(\#\text{Bike}_{ft} - \#\text{Bike}_{f,t+1})^+$ trips/minute, where $\#\text{Bike}_{ft}$ is the number of bikes at station f at the tth snapshot; the leading fraction arises due to the 2-minute gap between snapshots.

Arguably, a decline in the number of bikes at a station could also be the *net result* of simultaneous check-outs and returns of bikes. However, the average rates of both activities within a two-minute interval are low; check-outs and returns at a station also exhibit a *negative* temporal correlation, which implies that the likelihood of such contemporaneous events is extremely small (observed rates of these activities indicate a less than 0.1% chance of these activities).

Vélib' system managers regularly transfer bikes from full stations to empty ones, a procedure that could confound our usage data. We therefore omit the data from any two minute period in which *more* than four bikes are checked out of a station, which we interpret either as transshipment by system managers or as outliers in the usage. This scenario is rare, so even this conservative elimination allows us to retain over 95% of the data. The results of our analysis are unchanged when other nearby thresholds are used for eliminating such instances, or even if the instances are not eliminated.

Overall, we observe ~3.25 million trips. The average station-use for a typical station is 3.78 trips/hour; this rate is doubled during the evening peak hours, and is about one-third as much in the early morning hours.

In Section 4, we will build a model that explains this station-use as a function of the presence of different points of interest such as cafes, etc. in the vicinity of the station, whether this and other

---

[2]"Paris fête les six ans de son Vélib' (en infographie)", Mes Débats, 15 July 2013, http://bit.ly/14Cn6n6
[3]See Oliver O'Brien, "Bike-Share Map", 31 August 2013, http://bikes.oobrien.com/paris/

stations have bikes available for use (are "stocked-in"), and the typical odds of users finding bikes in different parts of the city. We next provide the construction/definitions of variables that operationalize these aspects.

3.2. **Stocked-In Stations.** Following recent analytical work on bike-share systems [Henderson et al., 2016], *we consider a station to be stocked-in if there are more than five bikes at the station.*

The threshold of five bikes accounts for the fact that a handful of bikes at a station are typically in an unusable or undesirable state (e.g., a bike with a broken chain or with bird droppings on its seat); expectedly these are the last remaining bikes at stations. Anecdotally, most users of bike-share systems also think of availability in these terms— a user who sees only a small number of bikes often assumes that those last few are likely unusable or might be checked out by other users by the time she reaches the station. Such behavior also clearly manifests itself in station-use data— even in peak hours, the station-use for most popular stations drops to zero once there are fewer than 4-6 bikes left at the station; indicating a stock-out on account of unusable bikes (or a perception of the same).

We also estimate our models with alternate thresholds for designating a station as stocked-in. We considered specifications where stations with more than four bikes, more than six bikes, stations that are more than 5% full, more than 10% full, or stations that have more bikes than the day's or the week's minimum number are considered stocked-in. Another set of specifications we estimated use different thresholds for small (< 20 docks) and large stations. Small stations are considered stocked-in if they have any bike, more than 1 bike, or more than 2 bikes (large stations are considered stocked-in in these specifications if they have more than 5 bikes). Our findings are unaffected by the choice of this definition, see section 9 for some details.

3.3. **Average Bike-Availability at a *Location*.** The average bike-availability measure captures the *typical likelihood* of a user at a given location finding a usable bike near her. In other words, these are the ex-ante odds, or a user's expectation that at least one nearby station has usable bikes, or is stocked-in.

The typical likelihood of finding a bike nearby can vary considerably by the time of the day. We divide the day into six "time-windows" to capture these differences: early morning (05h30–08h00), morning-rush (08h00–12h00), afternoon (12h00–16h00), evening (16h00–20h00), late-evening (20h00–00h30) and night/metro closed (00h30–05h30). $w(t) : t \rightarrow \{w_1, w_2, ..., w_6\}$ maps each time t to one of the six time windows, $w_1, w_2, ..., w_6$.

We consider bikes to be available at a location l at time t, if any station within 300 meters of location l is stocked-in (as defined above, in §3.2); the indicator variable, bike-available, ba_{lt} is set to 1 in this

case, and is set to 0 otherwise.[4] The *average bike-availability* at a location l in a time-window w_j, $j \in \{1, ..6\}$, aba_{lw_j} is the average of the bike-available indicator variable for all times in the specified time window:

$$aba_{lw_j} = \underset{t|w(t)=w_j}{\text{Average}} \quad ba_{lt}$$

This average provides a measure of the expected odds of finding a bike at each location l in a time-window w_j. The average bike-availability at a typical location in the city is 80.4%, it varies from 75.3% during evening hours to 86.8% in the early morning hours.

3.4. **Control Variables in the Origin-Density Model.** Our model for station-use also includes the residential population density and the presence of various points of interest (metro stations, cafes, hotels, stores, etc.) at a location. For locations with metro stations and tourist locations, we also include the effects of metro ridership and tourist visitor numbers respectively. We elaborate on these variables next:

Residential Population Density: INSEE, the french national statistics bureau, provides data on the population density in each of the 20 arrondissement (districts) of Paris. The residential density at a location l, rd_l is the population density of the district in which location l lies.

Presence of Points of Interest: Using an API provided by Google Places, we collect data on the location and type of ~70,000 points of interest in the city of Paris. The most significant types are metro, bus, and tram transit stations, stores or retail locations, restaurants, bars, cafes, other food-service locations, hotels and lodges, groceries and supermarkets, shopping malls, universities, parks, museums, libraries, and movie theaters. These types cover a majority of the points of interest in the city, the remaining categories and unclassified points of interest are grouped in a catch-all category, others.

For each location l, we construct a vector of dummies, $\vec{pi}_l$, each element of this vector corresponds to a type of point of interest. The dummy takes a value 1 if there is a point of interest of that type at location l and 0 otherwise. For example, consider a location that has a cafe and a store. For this location all elements of $\vec{pi}_l$ take value 0 except the elements corresponding to the two categories– "cafes" and "stores or retail locations". These two elements take the value 1.

Metro Ridership and Tourist Volume: For locations with metro stations and tourist spots, we obtained further data which allows us to take into account the size and importance of these points of interest. We collect annual ridership data for all metro transit locations from the transport operator, RATP's

---

[4]Survey literature, anecdotal accounts and best practices from urban planning suggest that users (on foot) typically consider 300 meters as "nearby" [O'Neill et al., 1992, Zhao et al., 2003, O'Sullivan and Morrall, 1996]. For the small fraction of locations where there are less than 3 stations within 300 meters, we consider the nearest 3 stations to define ba_{lt}.

open data archive. We construct a variable ri_l that takes the value of the relevant annual ridership if location l has a metro station, and the value 0 for locations that don't have metro stations. Similarly, we construct a variable vi_l to capture the number of visitors to key tourist locations. For this we obtain data on the number of visitors to the top 20 tourist locations in Paris from the Office du Tourisme et des Congrès de Paris (OTCP). The variable vi_l takes the value of the number of annual visitors if location l has a tourist location, and the value 0 otherwise.

For example, consider a location in the 7th district of Paris that has a cafe, a store, and a metro station with annual ridership of 500,000 riders. For this location, the three elements of $\vec{pi}_l$ corresponding to the three categories— "transit stations", "stores of retail", and "cafe" would take the value 1; all other elements will be zero. In addition, the variable ri_l will also be non-zero taking the value 500,000 (the ridership at this metro-station). The variable vi_l would be zero as this location is not a tourist location. The residential population density variable rd_l would take the value of the population density of the 7th district.

For the engaged reader, Appendix C.1 further illustrates the construction of all the variables described above and used in the model for a sample of the data.

3.5. **Weather Data.** Finally, the potential number of bike-share users could vary with varying weather conditions. We collect half-hourly weather data for the city of Paris, specifically the temperature, humidity, wind speed and "conditions" (clear, mist, cloudy, etc.) from weatherbase.com. There is little weather variation in our study period, with mild temperatures (between 10^oC and 30^oC) for 92.7% of the period. Nevertheless, we deseasonalize the station-use data using the prevalent weather conditions (see Appendix A).

4. A Structural Model for Station-Use

We follow Davis [2006], which itself extends the celebrated work of Berry et al. [1995] (BLP), and build a structural demand model that explains *station-use.* Our structural model has two components. The first component captures the rate of potential users originating at a location, or the *origin-density.* The second captures the likelihood of these users using bikes from different stations near this location; this is the *user-choice* model. These two components combine to lead to the predicted ridership at each station.

4.1. **Origin-Density Model.** The per minute rate of potential bike-share users originating at a location l in a time-window w_j, $P_{w_j}(l; \alpha_{w_j})$ is given as

$$P_{w_j}\left(l; \alpha_{w_j}\right) = \alpha_0 + \alpha_1 \cdot aba_{lw_j} + \alpha_{2w_j} \cdot rd_l + \vec{\alpha}_{3w_j} \cdot \vec{pi}_l + \alpha_{4w_j} \cdot ri_l + \alpha_{5w_j} \cdot vi_l. \tag{4.1}$$

This rate depends on the average bike-availability aba_{lw_j} at the location in the time-window, the residential population density rd_l at the location, the dummy variables $\vec{pi}_l$ that capture the presence of points of interest at this location, the ridership associated with the metro stop ri_l if there is one at this location, and the number of visitors vi_l if there is a tourist spot at this location (recall sections 3.3 and 3.4 for construction/definition of these variables). α_0, α_1, α_{2w_j}, $\vec{\alpha}_{3w_j}$, α_{4w_j}, α_{5w_j} are the coefficients corresponding to these variables, which we together denote with α_{w_j}.

The origin-density model allows for a baseline rate at all locations in the city, captured by α_0. In addition, the model allows for the rate to vary depending on the average bike-availability at the location. Locations around which bikes are typically more available are likely to generate more interest from users, or have a higher number of potential users originating— we call this the *long-term effect of bike-availability* and it is captured by the coefficient α_1.

The remaining control variables in the model (residential population density, points of interest, metro ridership, and number of visitors at tourist locations), all correspond to different sources of users that may use the bike-share system. Locations with high residential population density, presence of a point of interest, high metro ridership or tourist visitorship, are all likely to be associated with more users originating at these locations.

Note here that the above origin-density model is such that all locations in the city have a non-zero origin-density due to the intercept α_0, the residential density, and the average bike-availability associated with the location– these are non-zero for all locations. The remaining variables corresponding to the presence of points of interest, metro ridership, and tourist visitors and are non-zero only for select locations with relevant points of interest. Also observe that we allow the coefficients of the control variables (α_{2w_j}, $\vec{\alpha}_{3w_j}$, α_{4w_j}, α_{5w_j}) to vary with the time-window (hence the subscript w_j). This allows for locations to be more or less prominent origins at different times of the day. For instance, the model allows for residential population density to matter more in the morning hours, while the presence of cafes/restaurants can be more influential in the evening hours.

4.2. **User-Choice Model.** The user choice model is a multinomial-logit *choice model* that provides the probabilities with which users originating at a *given location* choose between different stations and outside options as a function of the distances to different stations from this location.

Consider a utility-maximizing user i who originates at a location l_i at a time t and chooses between different stations of the bike-share system and other modes of transport. Her utility from accessing a bike from station $f \in \{1, 2, \ldots, F\}$ at time $t \in \{1, 2, \ldots, T\}$ is given by

$$u_{ift} = \beta_0 + h\left(d\left(l_i, l_f\right); \beta_d\right) + \gamma_{w(t) \times di(f)} + \xi_{fw(t)} + \epsilon_{ift}, \tag{4.2}$$

Here, times $1 \ldots T$ denote the discrete time-units at which the station level data is observed (every two minutes in our case). $h()$ is a *piecewise linear* function with a change in slope at 300m, that captures the disutility of walking a given distance. The disutility of walking a distance d kms is given as:

$$h(d; \beta_d) = \mathbb{I}\{d < 0.3\} \cdot \beta_1 d + \mathbb{I}\{d \geq 0.3\} \cdot (\beta_1 \cdot 0.3 + \beta_2(d - 0.3));$$

$\beta_d \equiv \{\beta_1, \beta_2\}$ are the distance coefficients. $d(l_i, l_f)$ gives the walking (manhattan) distance (kms) between user i and station f (located at l_f).[5] Survey literature, anecdotal accounts, and current practices in bike-share network design suggest that users (on foot) are less sensitive to distance until approximately 300 meters, beyond which users are expectedly much more sensitive to distance [Zhao et al., 2003]. We thus assumed that the distance disutility function h is *piecewise linear* with a change in slope or a kink at 300 meters. We examine other kink points and alternate functional forms in Section 9.2.

$\gamma_{w(t) \times di(f)}$ are the *time-window × district* fixed effects, $di(f)$ is the district for station f. Recall $w(t)$ is the time window corresponding to the time t. The term $\xi_{fw(t)}$ denotes the unobservable components of utility that are common to all users for station f in time-window $w(t)$, or the error terms. The ϵ_{ift} are the idiosyncratic *user × station × time*-specific shocks; we assume that these shocks are of type-I extreme value, and are independently and identically distributed (i.i.d.).

The user's utility from using other modes of transport is

$$u_{i0t} = \xi_{0w(t)} + \epsilon_{i0t}; \tag{4.3}$$

here $\xi_{0w(t)}$ is the unobservable component of this utility that is common to all users in the time-window $w(t)$; this captures how the other options become more or less attractive over the course of the day. The ϵ_{i0t} are the idiosyncratic shocks that users derive from other means of transport, which we also assume are type-I extreme value, and independently and identically distributed. $\xi_{0w(t)}$ is normalized to 0 as per standard practice.

In our model, users observe the current stocked-in status of all the stations (widely available via the official Vélib' app/website or many third-party apps such as CityMapper), compare the utility of using a bike from different stocked-in stations and choose the station that earns them the highest utility. Using Eqs. 4.2 and 4.3, the probability of choosing a station follows the Multinomial Logit form. Let S_t be the set of stations that are *stocked-in* at time t; this is the *choice-set* for users at time t. The

---

[5]"Manhattan" distances, are simply the l_1 norm distance or the sum of the absolute differences of the co-ordinates of the station and the user location. These are often used to capture walking distances in built-up environments [Belavina et al., 2016].

probability of a user i using a bike from a stocked-in station $f \in S_t$ at time t is given by

$$p_{ift}\left(\beta, \gamma_{w(t).}, \xi_{\cdot w(t)}\right) = \frac{\exp\left(\beta_0 + h\left(d\left(l_i, l_f\right); \beta_d\right) + \gamma_{w(t)\times di(f)} + \xi_{fw(t)}\right)}{1 + \sum_{g \in S_t} \exp\left(\beta_0 + h\left(d\left(l_i, l_g\right); \beta_d\right) + \gamma_{w(t)\times di(g)} + \xi_{gw(t)}\right)}, \quad (4.4)$$

and is 0 for a stocked-out station, $f \notin S_t$. Here β $(\beta \equiv \{\beta_0, \beta_d\})$ are the coefficients, $\gamma_{w(t).}$ is the vector of all the fixed effects in time-window $w(t)$, and $\xi_{\cdot w(t)}$ is the vector of error terms in time-window $w(t)$.

4.3. **Station-Use.** These origin-density and user choice models combine to determine the predicted station-use at each station in a time-unit, specifically it is the sum of the rates of users originating at all possible origin locations that choose this station. Formally, the predicted use at a station f and time t, i.e. λ_{ft}, is obtained by integrating over all locations in $\mathscr{L}$, which is the set of all locations in the city:

$$\lambda_{ft}\left(\theta, \xi_{\cdot w(t)}\right) = \int_{l_i \in \mathscr{L}} p_{ift}\left(\beta, \gamma_{w(t).}, \xi_{\cdot w(t)}\right) P_{w(t)}\left(l_i; \alpha_{w(t)}\right) dl_i \quad (4.5)$$

Here θ denotes all the coefficients $(\alpha \equiv \{\alpha_{w_j}\}_{j \in \{1,..,6\}}$, β, and all the fixed effects $\gamma_{w(t)\times di(f)})$, and $\xi_{\cdot w(t)}$ is the vector of error terms $\xi_{fw(t)}$.

Note that when a station f is stocked-out, it is not part of any users' choice set at time t, S_t. As such, no user can choose it, i.e. $p_{ift} = 0 \ \forall i$, which implies that predicted station-use $\lambda_{ft}\left(\theta, \xi_{\cdot w(t)}\right) = 0$. The model is thus built for station-use which corresponds to sales in traditional operations parlance (and not demand). That is, our model explicitly recognizes stock-outs and takes into account the demand censoring that they lead to. Put differently, we will estimate the model only for station×time-periods when the station is not stocked-out (when demand is not censored).

4.4. **Discussion: Effects of Distance, and Bike-Availability.** The above model predicts station-use as a function of characteristics of locations in the city and average bike-availabilities. This allows us to use data on observed station-use to estimate the coefficients in the model. These coefficients capture the effects of distance and bike-availability on station-use. The effect of distance is captured in the choice model (Eq 4.2) by the coefficients β_1 and β_2 in the distance disutility function $h()$. Stations that are further away from users would have a higher disutility associated with them and are less likely to be chosen.

The effect of *average* bike-availability on system-use is captured in the origin-density model. Locations with higher average bike-availability are likely to have more user interest which would increase the use at stations close to such locations— this effect is captured by the coefficient α_1 in the origin-density model (Eq 4.1). We call this the *long-term effect* of bike-availability.

Bike-availability has an additional *short-term effect* on station-use— a stocked-out station forces users who would have chosen this station to substitute to other stations or to outside options. When a station is stocked-out, choice probabilities in our model change as this station is excluded from users' choice-set. The probability of users using the stocked-out station drops to zero, while the probability of using other nearby stations and the outside option increases. An increase in bike-availability at a station reduces the extent of stockouts at a station, and consequently the extent of users switching to outside options on observing stockouts. The reduction in the usage of the outside option (abandonment) due to higher bike-availability is called the short-term effect of bike-availability. It depends on the combination of the distance coefficients β_d, and other utility model coefficients β_0 and $\gamma_{w(t)}$.

4.5. **Comparison with Past Work and Alternate Models.**

Reduced Form Models. Directly estimating the relationship between station-use and the distances users must walk using a reduced form model requires observing use at different stations and the distance users' must walk to access the stations. The latter requires knowledge of the users' starting points or origin locations. Neither we nor any bike-share operator has data on these origin locations. As such, the effect of distance on station-use cannot be estimated from any reduced form regression model (linear, MNL, etc.), as we do not observe the key independent variable— the distance users must walk to access various stations.

Nevertheless, one way to use a reduced form model would be to use a "proxy" for the unobserved distances that users must walk. For example, we could use the distance between the station and the next closest station as a proxy for how much users at this station walked. This distance is indeed a measure of how much a typical or representative user must walk to reach the station. Yet, such proxies are poor measures. Continuing with the example: distance to the next nearest station confounds the disutility of walking distances with a catchment area effect— if the next station is very far, the use should be lower as users have to walk more, but it might be higher as this station is the closest station for more users, i.e. it has a bigger catchment area.

An additional reason for not using a linear or a direct MNL model in our setting (and in general for estimating demand of differentiated products) relates to accurately capturing the substitution patterns from station stock-outs. Upon a stock-out at a station, we expect that the users who preferred this station to be more likely to substitute to nearby stations rather than farther away stations. In contrast, a MNL model (due to the independence of irrelevant alternatives (IIA) property) would imply that upon a stock-out, all stocked-in stations experience the same percentage increase in demand irrespective of their proximity to the stocked-out station.

Structural demand models such as Berry et al. [1995], Nevo [2000], and Davis [2006] in the spatially differentiated choices context (on which our model is based), address this concern. Like in Davis [2006], users in our model have heterogenous origin locations. This implies that upon a stock-out, it is the users closest to the stocked-out station that are most influenced, and expectedly they substitute to nearby stations– achieving the desired substitution pattern.

Structural Models. We differ in two ways from prior work on structural demand models for spatially differentiated products; these differences address two known limitations of this work.

First, past work on demand estimation for spatially differentiated products [Davis, 2006, Thomadsen, 2005, Allon et al., 2011] assumes that the same number of users originate at each location within a district or census unit (i.e. a uniform density), effectively only using the residential population density variable in our origin-density model, which takes the same value for all locations within a certain administrative unit. Although residential population density might be sufficient to capture spatial differences in interest for retail stores, restaurants, etc., it is arguably inadequate to capture interest in bike-share systems. Bike-share is often used as a last-mile connection by suburban commuters to the city, and by the large transient population in the city (tourists, office workers, students, etc.), all of whom are not captured well in the residential populations. Including the presence of different points of interests, metro transit stops and their ridership, the visitorship of tourist locations, in addition to the residential population density from past work allows us to consider the unique characteristics of each location, which facilitates more precise estimation.

Second, past work in consumer choice models (including the seminal works of Berry et al. [1995], Nevo [2001], and Davis [2006] in the spatially differentiated retail choice context) assumes that all *offered* products are always stocked-in or available. In the context of our model, this means that the set S_t in Eq. 4.4 above is always assumed to be the set of *all* products, rather than the set of products available at time t, as we do above. This has been shown to substantially bias parameter estimates in the case of consumer goods [Bruno and Vilcassim, 2008, Conlon and Mortimer, 2013]. Bike-availability is typically ~80%, lower than the 90% or so availability in the case of consumer goods, and arguably more important to users. Thus, assuming full availability is likely to bias estimates even more in our context. Further, this would run counter to one of our key goals–measuring the impact of bike-availability. Thus, as discussed above, in our model, we allow the user choice sets to change depending on which stations are stocked-in. This allows us to address this known bias of demand censoring, as well as measure the short-term impacts of station stock-outs.

5. Model Estimation

We follow the GMM procedure, instrumentation, and estimation techniques from Berry et al. [1995], Davis [2006] and Dubé et al. [2012], to estimate the model described in Equations 4.2-4.5. The subsequent sections provide details of the endogeneities present in our model, how we address them, the formal GMM formulation, a computational challenge, and then a transformation to alleviate this challenge.

5.1. **Endogeneity and Instruments.**

Sources of Endogeneity. System managers often choose bike-availabilities and station locations on the basis of the characteristics of the area. While we observe many such area characteristics (ex. presence of points of interest), there are others that might not be observed. Such unobservable characteristics might also influence the use at these stations. This introduces an endogeneity bias. Consider bike-availability. System managers might provide higher bike-availability in areas with politically important stakeholders, or other such unobservable factors. Politically important areas might also have different users that influences the use at stations in the area in ways not captured by our model. There is also reverse-causality: average bike-availability at a location influences use at nearby stations, but a high station-use realization itself leads to more stockouts which could lead to lower average bike-availability in the area. Both of these phenomena could lead to a correlation between the error terms (the component of demand not explained by observed variables) and the likelihood of finding bikes at a given station in a particular time-window. This implies that the error term for a *station×time-window* is correlated with the *average bike-availability* variable at nearby locations for that time-window. Naive estimation could lead to biased estimates.

Similarly, more stations might be located in areas of political importance. Now, the unobserved political importance might influence both the walking distances of a station's users as well as the use at these stations, possibly leading to a correlation between the distances users walk to access a station and the error-terms. This again biases our estimates.

Addressing Endogeneities: Controls and Instruments. First, note that the most prominent factors on which station managers might choose to base their bike-availability or station location decisions (factors such as presence of transit stations, cafes, number of tourists, etc.) are the same as the ones that we have included in our origin-density model. As such, a majority of the important area characteristics are *observed* and incorporated in our analysis leaving relatively few unobserved characteristics that cause the endogeneity problems discussed above. In particular, recall that past work only uses one characteristic of the area, the residential population density [Berry et al., 1995, Davis,

2006, Thomadsen, 2005, Allon et al., 2011], and as such, it has many more unobserved characteristics and more potential for these biases. Overall, our richer origin-density model should lead to much fewer endogeneity problems that in any past work.

Nevertheless, we follow past work that has applied BLP-like instruments in the context of spatially differentiated products [Davis, 2006] and employ the same instruments as in this study. Following BLP, Davis [2006] proposes using exogenous characteristics of the focal and neighboring theaters as instruments for endogenous characteristics of the focal theater to correct for endogenous prices and theater locations. In our context, this translates to using *exogenous characteristics of focal and neighboring stations* (for e.g. number of nearby cafes, stores, bars, etc., ridership of nearby metro stations, etc.) as instruments for *endogenous* characteristics of the focal stations (average bike-availability and station-location). Our arguments for the validity of these instruments mimic those in the past work; an overview is provided after their formalization.

Formally, corresponding to the error term ξ_{fw_j}, the instruments from Davis [2006] adapted to our model context are given as $Z^D_{fw_j} = \hat{Z}_f \otimes \vec{y}_{w_j}$.[6] Here, $\hat{Z}_f = \{rd_f^{(a,b,c,d)}, \vec{pi}_f^{(a,b,c,d)}, ri_f^{(a,b,c,d)}, vi_f^{(a,b,c,d)}\}$, where for $x \in \{rd, \vec{pi}, ri, vi\}$, $x_f^{(a,b,c,d)}$ is the sum of x_l for locations l that are between a and b meters of stations neighboring the focal station f; these neighboring stations are the stations between c and d meters of the focal station f. $\vec{y}_{w_j}$ is a vector that accounts for different time-windows, $\vec{y}_{w_j} = \{y_k\}_{k \in \{1,6\}}$, $y_k = 1$ if $k = j$ and 0 otherwise. For example, corresponding to error term ξ_{fw_2}, there would be six instruments corresponding to $rd_f^{(a,b,c,d)}$ in $Z^D_{fw_2}$. Since $\vec{y}_{w_j}$ in this case of $j = 2$ would be $\{0, 1, 0, 0, 0, 0\}$, the values of these instrument would be $\{0, rd_f(a, b, c, d), 0, 0, 0, 0\}$.

The parameter sets used for (a, b, c, d) are (in meters) $(0, 25, 0, 25)$, $(25, 50, 0, 50)$, $(50, 100, 0, 100)$, $(0, 100, 0, 100)$, $(100, 300, 0, 300)$, $(300, 500, 0, 500)$, $(0, 100, 100, 300)$, and $(0, 100, 300, 500)$. Note that multiple alternate sets of parameters must be used to capture non-linearity of distance [Davis, 2006]. Together, the Davis-instruments $Z^D_{fw_j}$ and the remaining model-covariates (intercept term 1, and dummies $\gamma_{w_j \times di(f)}$) constitute the vector Z_{fw_j} of instruments. See Appendix C.2 for an illustration of the construction of these instruments for a sample station and time-window.

Validity of Instruments. Our arguments for the validity of these instruments mimic those in Berry et al. [1995], Davis [2006]. The instruments need to satisfy two conditions: the exclusion restriction (the instruments should be uncorrelated with the unobserved components of station-use), and the relevance conditions (instruments should influence the endogenous characteristics of the focal station).

We test for the relevance of the instruments by estimating two linear regression models: one with the average bike-availability as the dependent variable and the other with station density. The above

---

[6]$\otimes$ is the Kronecker product.

instruments are covariates in each case. Including instruments in the models increases the R^2 by 35.3% and 37.4% (respectively) in the two models, see Table 2 in Appendix B. This suggests that the proposed instruments are highly relevant in our context.

Our arguments for the exclusion condition follow those in Berry et al. [1995], Davis [2006] and others. In these studies, the instruments are exogenous characteristics (characteristics besides price such as power of the vehicle, etc.) of the focal and other products, which are argued to not affect unobserved component of demand of the focal product. Similarly, our instruments are based on exogenous characteristics (for e.g. number of nearby cafes, stores, bars, etc., ridership of nearby metro stations, etc.) in areas surrounding the focal and neighboring stations. These characteristics such as the presence of metro stations and cafes are static characteristics of a region and are unlikely to change in response to the unexplained realizations of high/low bike-share use at stations. Similarly, the characteristics of neighboring stations are unlikely to affect the unobserved components of focal station-use. Arguably, the plausibility of this exclusion restriction is stronger in our case, than in past work. For example, in Berry et al. [1995]'s context, some of the exogenous product characteristics such as horse power of a car could be changed by model revisions/tune-ups in response to realization of the unexplained component of demand. In our case, exogenous characteristics such as the presence of metro stations are static characteristics of a region and are very unlikely to change in response to the unexplained high/low bike-share use at a station in a time-window.

Robustness: Alternate Instruments. We also estimate our model with alternate instruments: i) Instruments based only on the focal station's nearby characteristics (akin to Davis [2006]'s robustness analysis) ii) A novel instrument based on the average rate of incoming bikes at station f in lagged time-window $w_j - 1$, demeaned at the station level. This rate affects average bike-availability of station f in time-window w_j, but is not correlated with the unobserved characteristics ξ_{fw_j} after the rate has been demeaned of the station characteristics; iii) Alternative parameters formulations for above BLP-like instruments; iv) Including simpler instruments as used by Thomadsen [2005], Allon et al. [2011]: distance to the nearest station, average distance to 5 nearest stations, and number of stations within 500m of a station in addition to our instruments. All instruments provide essentially identical estimates (some reported in Table 5) reinforcing the validity of our analysis.

5.2. **GMM Estimation.** Intuitively, the model parameters are identified using the cross-sectional variation in the control and explanatory variables across locations. Average bike-availability varies across different locations. The difference in use at a station which is close to locations with higher average bike-availability and one which is close to locations with lower average bike-availability identifies

the coefficient for the average bike-availability. Next, consider the control variables in the origin-density (section 3.4). Stations have different presence of nearby points of interest, different residential densities, etc. (control variables in the origin density). For example, there could be a cafe near a station, while there might not be one near other stations. The observed differences in use between these stations identifies the impact of the control-variable, the presence of a nearby cafe. The cross-sectional variation in these control variables also identifies the effect of walking distance, i.e. the distance coefficient. For example, consider a station that has a cafe 10 meters from it while another station which has a cafe which is 50 meters away from it. The observed difference in station-use between these stations is then attributed to the effect of higher walking distance for the users originating at cafes–this identifies the distance coefficient.

Formally, we use a Generalized Method of Moments (GMM, Hansen [1982]) method, which allows us to use instruments for our estimation. Our moment conditions are

$$E_{f,w_j} \left[Z_{fw_j} \sigma_{fw_j} \xi_{fw_j} \left(\theta^* \right) \right] = 0 \tag{5.1}$$

where Z_{fw_j} is the set of instruments defined in section 5.1, ξ_{fw_j} are the error terms in our model, and σ_{fw_j} is the number of stocked-in observations for station f in time-window w_j, which we use as weights. These conditions restrict the unobserved components, ξ_{fw_j} to be uncorrelated with instruments Z_{fw_j} at the true parameter values θ^*. Using the number of observations σ_{fw_j} as weights in the moment conditions allows us to get efficient estimates (refer sec. 6.3.7 Cameron and Trivedi, 2005, $Var(\xi_{fw_j}) \propto 1/\sigma_{fw_j}$).

In addition to the moment conditions above, we also require these coefficients to satisfy some additional constraints. The first set of constraints comes from the balance conditions, as in Berry et al. [1995]. These conditions equate the predicted use and the observed use for each station and time-window, which determines the values of error terms $\xi_{\cdot w_j} (\theta)$ for given parameter values θ:

$$\lambda_{fw_j} \left(\theta, \xi_{\cdot w_j} (\theta) \right) = \Lambda_{fw_j} \quad \forall f, w_j \, ; \tag{5.2}$$

where $\lambda_{fw_j}(\theta, \xi_{\cdot w_j})$ is the predicted station-use at a station f in a time-window w_j, which is obtained by averaging λ_{ft} over all times in w_j when station f is stocked-in. In Eq. 5.2 the LHS is obtained as

$$\lambda_{fw_j}(\theta, \xi_{\cdot w_j}) = \underset{t|w(t)=w_j \,\&\, ba_{ft}=1}{\text{Average}} \lambda_{ft}(\theta, \xi_{\cdot w_j}) \tag{5.3}$$

Here, ba_{ft} is 1, if station f is stocked-in at time t and 0 otherwise. Similarly, the RHS in Eq. 5.2 Λ_{fw_j} is the observed station-use at a station f in a time-window w_j, which is obtained by averaging

the observed station-use at a station f and time-period t, Λ_{ft}, for time-periods when station f is stocked-in, so that $\Lambda_{fw_j} = \text{Average}_{t|w(t)=w_j \& ba_{ft}=1} \Lambda_{ft}$.

Finally, following Berry et al. [1995], for efficiency, we include an additional condition that matches the total *potential* user-interest in a day (by aggregating our origin-density model, eq 4.1) to an estimate of this number obtained from external sources. Formally,

$$\int\limits_{w_j} \int\limits_{l \in \mathscr{L}} P_{w_j}\left(l; \alpha^*_{w_j}\right) \cdot minutes_{w_j} \cdot dl \cdot dw_j = T^D, \tag{5.4}$$

where T^D is the total potential user interest in a day, and $minutes_{w_j}$ is the number of minutes in time-window w_j which is used above to convert the rate of users, $P_{w_j}\left(l; \alpha^*_{w_j}\right)$ to the number of users, while $\alpha^*_{w_j}$ is the true parameter value of α_{w_j}. We base T^D on the working age population of Paris with each person using the system once per day which gives 1,120,320 potential users per day.[7] This follows Davis [2006] and the subsequent literature which set their market size to the population count.

As in past work, the number of moment conditions plus the number of constraints in the above procedure is greater than the number of coefficients. Thus, not all moment conditions and constraints can be exactly satisfied at the estimates. The GMM procedure ensures that the moment conditions (Eq. 5.1) are as close to being satisfied as possible, while the constraints (Eq's. 5.2, 5.4) are exactly satisfied. We follow the MPEC procedure [Dubé et al., 2012] which requires the constraints to be satisfied only at the final estimates, unlike the nested procedure in Berry et al. [1995] and Davis [2006] which requires the constraints to be satisfied at each iteration. This saves on the number of computations and is also known to be robust to numerical inaccuracies. The exact formulation of the MPEC procedure is presented in Appendix D.1, Eq. D.3 .

Finally, to allow for spatial and temporal correlation between nearby stations and adjacent time-windows, we divide the city into 600m-square grids.[8] We allow for the error terms of stations in the same grid to be correlated across adjacent time-windows. Formally, the structure of co-variance for any two *station × time-window* pairs f_1, w_j, and f_2, w_k is such that $cov\left(\xi_{f_1 w_j}, \xi_{f_2 w_k}\right) \neq 0$ if $\text{grid}(f_1) = \text{grid}(f_2)$ and w_j is same as or adjacent to w_k; the co-variance is 0 otherwise. Complete details of the standard error computations are presented in Appendix D.3. The next sections (5.3-5.4) explain a computational challenge and its resolution by a novel transformation. Readers not interested in these issues can safely skip ahead to the results in Section 6.

---

[7]In principle, T^D could be identified by the above procedure, however we find that there is a broad range of T^D where our parameters of interest are essentially the same and using an external estimate for T^D turns out to be more computationally efficient. More interestingly, this implies that our results are not at all sensitive to a broad range of values around T^D, therefore even a rough estimate of T^D suffices. As expected, our estimates are very robust to this choice (Section 9).

[8]Results are qualitatively unchanged with a 300m-square grid specification.

5.3. **A Computational Challenge.** Despite using the techniques proposed by Dubé et al. [2012], numerically estimating the above model is extremely computationally intensive. The driver of the computational burden that is unique to our context is the computation of predicted station-use of a *station* × *time-window*, λ_{fw_j}, used in the balance equations (Eq. 5.2). Computation of λ_{fw_j} in turn requires computing station-use for each station at each time, λ_{ft} (Eq. 5.3). There are over 20 million such functions, on account of $F = 946$ stations, and $T =$22,743 two minute observations for each station. Further, for each λ_{ft} computation, we numerically integrate the choices of users located in the entire city of Paris which is a *105 km²* area that we discretize into nearly 210,000 locations.[9] This implies that about 4 trillion computations are required to compute the predicted station-uses for each iteration of the estimation. For finding optimal estimates, several iterations of the above computations are required which would take the total number of computations to over a quadrillion, a process we estimate would take over a year.[10]

Note that the computational burden can be reduced by ignoring stock-out information or by discretizing the city less finely. Ignoring stock-out information makes all times in a time-window identical, much reducing the computational burden; yet as discussed in Section 4.5, ignoring this information is known to bias estimates. Further a key objective of this paper is to capture the short-term impacts of stock-outs– which requires capturing stock-out information. Less fine discretization of the city is also not desirable as stations can be as close as 50 meters. Past work has typically not incorporated either stock-outs or spatial information, and uses much smaller scales of data; thus typically has 6 orders of magnitude fewer computations than us (see Appendix E.1 for details).

5.4. **A Transformation.** We propose a transformation of our data that reduces the computational burden while still incorporating stock-out information. Note that in our model, station-use λ_{ft} is a function of the set of stocked-in stations, the *time-window* × *district* fixed effects, average bike-availabilities, and the control variables in the origin-density model. Within a *time-window*, only the set of stocked-in stations (user choice sets) change with time. Thus, the values of λ_{ft} for the time periods with the same set of stocked-in stations can be computed together. This implies that we can aggregate these time-periods without losing any information relevant to our estimation. That is, if we aggregate time-periods according to *station* × *time-window* × *"system-stockout-state"*, or equivalently, if times within a time-window for which choice set for all the users (S_t) is the same are considered as one data-point, then we can use all stock-out information while potentially decreasing the computational burden.

---

[9]See Appendix Section D.2 for details.

[10]Based on solving smaller instances of this problem using a highly optimized implementation where all core components are implemented and run with C++ binaries and sparse analytical jacobian implementations fed to an IPOPT optimizer.

The advantage of computing station-use in the stockout-state domain instead of the time-domain arises from there being fewer distinct system-stockout-states during each time-window than there are distinct two-minute time intervals. However, if the stockout-state is defined at the system level, there could be as many as $\approx 2^{946}$ distinct values, and while not all distinct values are realized in the data, the number of realized system-stockout-states is of the same order as distinct two-minute times; hence such a transformation is only slightly superior to the original model.

We notice that the use at station f is not affected equally by *all* the other stations' stocked-out statuses or stockout-states. The stockout-state of *neighboring* stations of station f have a much stronger effect than the stockout-state of far off stations.[11] The implication is that we can construct a *"local-stockout-state"* for each station (which captures the stockout-state of only nearby stations) and aggregate station time-periods within a time-window on such local-stockout-states rather than on the system-stockout-states. The local-stockout-state will have lower dimensionality than the system-stockout-state, so that there would be far fewer distinct local-stockout-states than distinct system-stockout-states. This makes the number of distinct computations much smaller and thus computationally feasible, while retaining almost all the relevant information.

Formally, we construct the local-stockout-state for a station f by working upwards from the choice sets of each user. We limit a user i's choice set to her nearest m_d *stocked-in* stations; the set of her nearest m_d stations is denoted by N_i. Given this, note that for a station f, the only relevant stock-out information is of stations close enough to users who are close enough to station f. For any station f, we can write the set of relevant stations N_f as

$$N_f \equiv \bigcup_{i \mid f \in N_i} N_i.$$

The vector of stockout-states at time t of stations in N_f is given by a binary vector v_{ft}— it is the *local-stockout-state* for station f at time t. Let the set of all such realized local-stockout-states be given by $V_f \equiv \bigcup_t v_{ft}$. Also, let $\sigma_{f w_j v_f}$ denote the number of observations with state v_f at station f, time-window w_j, and S_{v_f} denote the set of stocked-in stations in state v_f; while the set of stocked-in stations in user i's choice-set in state v_f is given by $N_i \cap S_{v_f}$.

---

[11]Note that even though far off stations don't meaningfully affect the use at the focal station, we must estimate the entire system jointly because of the users in the overlapping nearby regions of different stations.

Taken together, the transformed probability of user i choosing station f s.t. $f \in N_i \cap S_{v_f}$, in state v_f, time-window w_j is

$$p_{ifw_jv_f}\left(\beta, \gamma_{w_j\cdot}, \xi_{\cdot w_j}\right) \tag{5.5}$$

$$= \frac{\exp\left(\beta_0 + h\left(d\left(l_i, l_f\right); \beta_d\right) + \gamma_{w_j \times di(f)} + \xi_{fw_j}\right)}{1 + \sum_{g \in N_i \cap S_{v_f}} \exp\left(\beta_0 + h\left(d\left(l_i, l_g\right); \beta_d\right) + \gamma_{w_j \times di(g)} + \xi_{gw_j}\right)},$$

and $p_{ifw_jv_f}\left(\beta, \gamma_{w_j\cdot}, \xi_{\cdot w_j}\right) = 0$ for $f \notin N_i \cap S_{v_f}$. Integrating user's choice probabilities gives the predicted station-use at station f, time-window w_j in state v_f simply as

$$\lambda_{fw_jv_f}\left(\theta, \xi_{\cdot w_j}\right) = \int_{l_i \in \mathscr{L}} p_{ifw_jv_f}\left(\beta, \gamma_{w_j\cdot}, \xi_{\cdot w_j}\right) \cdot P_{w_j}\left(l_i; \alpha_{w_j}\right) dl_i. \tag{5.6}$$

Finally, the transformed predicted station-use at station f in time-window w_j, which is the key predicted value that must be computed for our estimation is given by the weighted average of $\lambda_{fw_jv_f}$'s over state's v_f in which station f is stocked-in. Formally,

$$\lambda_{fw_j}\left(\theta, \xi_{\cdot w_j}\right) = \frac{\sum_{v_f \in V_f \mid f \in S_{v_f}} \sigma_{fw_jv_f} \cdot \lambda_{fw_jv_f}\left(\theta, \xi_{\cdot w_j}\right)}{\sum_{v_f \in V_f \mid f \in S_{v_f}} \sigma_{fw_jv_f}}, \tag{5.7}$$

Our model is now estimated using the procedure described in Section 5.2; except the predicted station-use, $\lambda_{fw_j}\left(\theta, \xi_{\cdot w_j}\right)$ in Eq. 5.2 is computed as per equation 5.7 rather than Eq. 5.3. The full MPEC procedure associated with the estimation is provided in Appendix D.1, Eq. D.3.

The non-linear optimization with constraints is done using the open source package Interior Point Optimizer (IPOPT, Wächter and Biegler [2006]), that is interfaced with R via ipoptr [Ypma, 2010]. A step by step outline of the estimation algorithm is provided in Appendix D; specifically the computational details are in Appendix D.2.

5.5. **Computational Choices.** To integrate over the continuous elements of the spatial density model (the terms α_0, $\alpha_1 \cdot aba_{lw_j}$, and $\alpha_{2w_j} \cdot rd_l$), we discretize the city using a square grid with edge length $\mathscr{D} = 25$ meters. We consider the choice set of each user to be the four nearest stations, $m_d = 4$, this user-level choice drives the size of local-stockout-state of a station. In the robustness analysis (Section 9), we show the estimates obtained by considering the five nearest stations and the consequent local-stockout-states.

Considering local-stockout-states much reduces our computational burden. Yet, for some stations in high station-density areas, there can still be a large number of local-stockout-states that are realized in the data. For 75% of the *station×time-window's*, more than 30 local-stockout-states are realized in data and there are a total of 609,858 different local-stockout-states across all *station × time-windows*. Note

System-Use Model

	Choice Model				Density Model			
Intercept	Walking Distance (0-300mts)	Walking Distance (>300mts)	District x Time-window F.E.		Intercept	Average Bike-Availability	Control Variables (Sec 3.4)	Number of Observations
β_0	β_1	β_2	$\gamma_{w_j}.$		α_0	α_1	$\vec{\alpha}_{\{2,3,4,5\}w_j}$	
-1.801	-2.229	-15.445	Yes		11.666	0.004	Refer Table 3	45,122
(0.096)***	(0.437)***	(2.947)***			(126.214)	(0.001)***		

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)

Table 1. Estimated Coefficients

that the frequency distribution of the realization of each stockout-state is expectedly highly skewed.[12] So, while the computational burden increases proportionally with each extra state considered, the relevant and reliable data-observations come from a small subset of higher frequency states. Thus, for each *station × time-window*, we consider the 8 most frequent states. Typically these cover more than 60% of the data available for a vast majority of the stations. In Section 9, we also reevaluate our model with the top 16 states for each *station × time-window* which typically cover about 75% of the data; we find nearly identical estimates.

6. Estimates and Interpretation

We next provide the estimates from our model. Out-of-sample prediction performance of the model, and comparisons with simpler models are provided in Section 7, illustrations of how model estimates can be used for improving bike-share operations are provided in Section 8, while the robustness of the estimates to computational and modeling choices are in Section 9.

6.1. **Effect of Distance and Bike-Availability.** Table 1 reports the key estimated coefficients while the estimates for the control variables in the origin-density model are in Table 3. Table 2 interprets the estimated coefficients in terms of the marginal effect of changes in user-station distance and bike-availability on use of the bike-share system.

Effect of Distance. We find statistically significant effects of distance. Expectedly, users incur a disutility from walking, $\beta_1, \beta_2 < 0$ (Table 1). Interestingly, the *marginal* disutility of walking is lower for

---

[12]Consider a station-time-window with 5 other stations in the local-stockout-state, say each of these neighboring stations has a 90% average bike-availability in this time-window. The most likely state is 9^5=59049 times more likely to realize than the least likely state.

Marginal Effects	Δ Likelihood of Use at a Station	
	Mean	95% C.I.
+1m Walking Distance (0-300m)[*]	0.194%	(0.125%-0.263%)
+1m Walking Distance (>300m)[*]	1.307%	(0.881%-1.733%)
	Δ Station-use	
+10% Bike-Availability (Short-term)[+]	9.493%	(9.426%-9.533%)
+10% Bike-Availability (Long-term)[#]	2.482%	(1.621%-3.525%)
+10% Bike-Availability (Total)	12.211%	(11.200%-13.394%)

[*]Refer to footnote 12; [+]Refer to footnote 16; [#]Refer to footnote 15, for further details.

TABLE 2. Interpreting Coefficients: Marginal Effects

the first 300 meters, and is much higher for further distances (-2.229 v/s -15.445), in effect we have increasing marginal disutility of distance (or convex disutility).

Next, we use the estimates of the parameters of the distance disutility function to compute the ratio of the likelihood of using the system by a user that originates at a bike-station versus one that originates further away, in effect capturing how distance to a station reduces ridership (Figure 6.1(a)).[13] For the first 300 meters, every additional meter of walking to a station decreases a user's likelihood of using a bike from that station on average by 0.194% ($\pm$0.0693%).[14] The effect is higher after the first 300 meters, every additional meter decreases the likelihood on average by 1.307% ($\pm$0.426%) (Table 2). A user that originates 300 meters away from a station is about 60% as likely to use the system as one at the station, while a user that originates 500 meters away is much less likely to use the system than one at the station. Overall these estimates imply that even requiring users to walk relatively short distances drastically reduces the likelihood of them using bike-share.

Figure 6.1(b) illustrates the estimated distances traveled by users of the bike-share system. The figure looks at the use predicted at different stations (as per our estimates) and classifies the predicted users based on how far their origin locations are from the station.[15] We find that the median user travels about 220m to reach her preferred station. 13.09% of the users originate within 100m of the station, 30.328% from 100-200m of the station, 34.63% from 200m-300m, and 21.94% from areas more than 300m. Again this calculation highlights that the user-base of the bike-share system is highly local. Overall, the high impact of distance on system-use suggests the need for building dense station

---

[13]This ratio for a user at a distance m and one at the station is the ratio of $p_{ift}(\hat{\beta}, \hat{\gamma}_{w(t)\cdot}, \hat{\xi}_{\cdot w(t)})$ for user-station distances of m, and 0, averaged over stations and times.

[14]95% confidence interval is in brackets. Here and all other places in the paper.

[15]The system-use coming from users who walk a distance m: $\tilde{\lambda}(m, \hat{\theta}, \hat{\xi}) = \sum_t \sum_{f \in S_t} \int_{l_i \in \mathscr{L} | d(l_i, l_f) = m} p_{ift}(\hat{\beta}, \hat{\gamma}_{w(t)\cdot}, \hat{\xi}_{\cdot w(t)}) P_{w(t)}(l_i; \hat{\alpha}_{w(t)}) dl_i$.

(a) Use-Likelihood and Distance
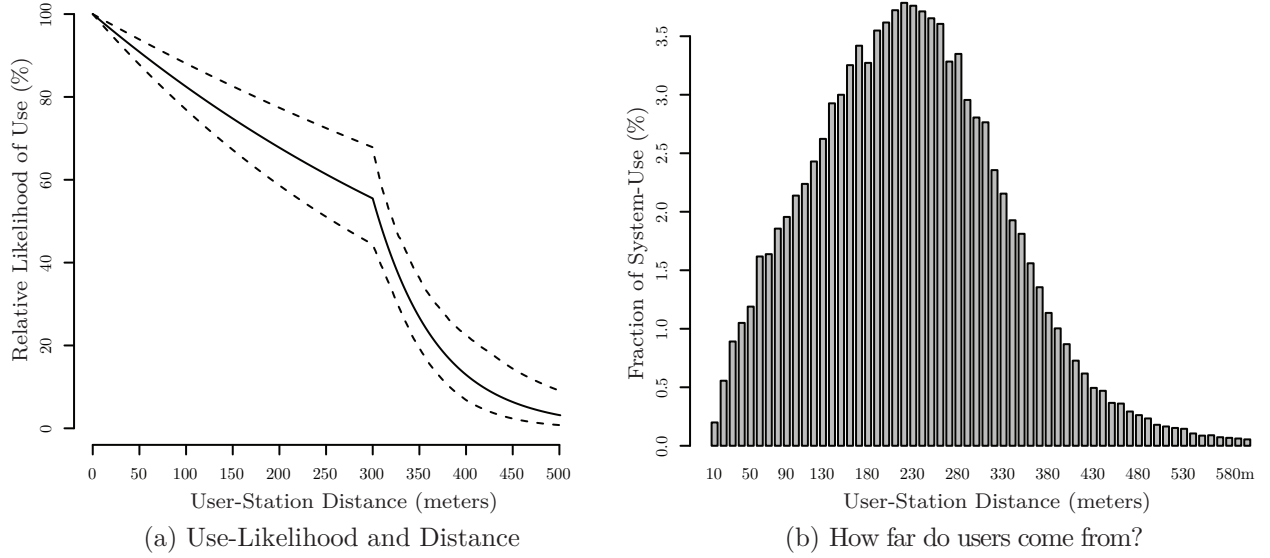
(b) How far do users come from?

FIGURE 6.1. Effects of Distance

networks. In Section 8.2, we provide further analyses that predicts and compares system-use for some proposed (denser) station-networks.

Effect of Bike-Availability. The positive and statistically significant coefficient for average bike-availability $(\hat{\alpha}_1)$ suggests that bike-availability influences the number of potential users of the bike-share system, Table 1. In other words, more users would consider using the bike-share system if they expect a higher chance of finding bikes at stations in their vicinity. The estimates in Table 1 imply that *a 10% increase in bike-availability increases the number of potential users (origin-density) by 2.628%($\pm$0.967%) which translates into a 2.482%($\pm$0.952%) increase in ridership.* This is the *long-term effect* of an increase in bike-availability.[16]

As discussed before, bike-availability also has a short-term effect that arises due to fewer instances of stocked-out stations, and consequently fewer instances that users switch to outside options (abandonment) on observing stockouts. We estimate this effect by using the estimated model to simulate the system-use for a scenario where the odds of finding bikes are increased by 10%.[17] We find that the scenario with 10% higher bike-availability has 9.493% ($\pm$0.0538%) higher use. This is the *short-term*

---

[16]The increased origin-density at each location l due to a 10% increase in average bike-availability is given by $\tilde{P}_{w_j}(l; \hat{\alpha}_{w_j}) = P_{w_j}(l; \hat{\alpha}_{w_j}) + 0.1 \times \hat{\alpha}_1 \cdot aba_{lw_j}$. The modified station-use $\tilde{\lambda}_{fw_j}$ is then given by replacing $P_{w_j}(l; \hat{\alpha}_{w_j})$ by $\tilde{P}_{w_j}(l; \hat{\alpha}_{w_j})$ in Eq. 5.6-5.7. The long-term effect is the percentage difference in total station-use across all stations and time-windows with and without any increase in bike-availabilities. Note that to isolate the long-term effect, this computation considers changes only in the origin-density (# of potential users) but no changes in user choices (specifically it does not include the lower incidence of abandonment). This lower incidence is the short-term effect and computed separately next.

[17]To isolate the short-term effect, this simulation assumes that there are no changes in the origin-density, the only changes in use arise from changes in user-choices (specifically from lower abandonment), see Appendix E.3 for details.

effect of bike-availability. This estimate also implies that only *5.070% (±0.538%)* of a stocked out station's unserved users substitute to other stations.[18]

Overall, a 10% increase in bike-availability would increase system-use by 12.211% (±1.097%); of this, three-quarters of the gains (9.493%) arise immediately on account of a reduced "lost trips", while the rest (2.482%) will be achieved over the long-term on account of increased user interest in the system. In Section 8.1, we provide further scenario analyses that compare predicted increases in system-use due to increases in bike-availability at different locations and times.

Appendix E.2 provides a comparison of our estimates with estimates from other studies in the bike-share, public-transport and retail-store network design contexts (or in some cases decisions implied by those estimates). Our estimates are in line with past work, or differ in ways expected due to the different contexts.

Residential v/s Commercial Districts. We also estimated a version of our model that allows for different average bike-availability and distance coefficients (α_1, β_1, and β_2) for residential and commercial districts. We used three different ways to classify districts as residential or commercial (high/low population density; high/low presence of points of interest; inner/outer districts). The details of these classifications and estimated coefficients are presented in Appendix B.4. We find that the disutility of distance is higher in residential districts than in commercial districts. Perhaps, commercial districts with the presence of retail outlets, etc. are more amenable to walking. On other hand, average bike-availability has a higher impact in commercial districts. Arguably, busy users in commercial districts are more responsive to a higher expected chance of finding bikes.

6.2. **The origin-density model.** The estimated coefficients for the control-variables in the origin-density model (Table 3) quantify the number of potential users that originate at different types of origin locations at different times of the day. During the day, residences, supermarkets, cafés and metro-stops are the most important origin locations for potential users. On the other hand, during the night, bars, cafés, restaurants and hotels are the most important origin locations.

7. MODEL VALIDATION

We validate our model by (i) comparing the out of sample prediction performance of our model with that of simpler regression models, and (ii) by comparing the substitution percentage predicted by our structural model with the level calculated using a reduced form model.

---

[18]The fraction of users that substitute is given as (10-9.493)/10=5.070%. If all users were substituting, the short term effect of a 10% increase would be zero, if none substituted the effect would have been 10%.

		Time-windows				
	00h30–05h30	05h30–08h00	08h00–12h00	12h00–16h00	16h00–20h00	20h00–00h30
Residential Population	0.000	57.714	422.092	204.991	459.845	53.825
Density	115.863	103.307	(164.85)*	(92.097)*	(108.279)***	79.841
Metro Ridership	0.000	0.000	8.101	28.129	4.253	73.375
		42.808	70.446	52.874	49.764	44.819
Tourist Volume	0.000	1.031	7.807	4.277	5.464	7.077
	1.949	0.931	(2.329)***	(0.996)***	(0.992)***	(1.728)***
Presence of Point of Interest Variables						
Groceries and Supermarket	14.979	209.952	481.656	137.100	0.000	0.000
	33.606	(89.07)*	(120.955)***	(25.939)***	22.497	19.634
Cafe	78.727	0.000	12.711	61.640	104.329	136.411
	47.194	16.227	20.322	(13.431)***	(16.617)***	(35.11)***
Metro Location Dummy	0.000	0.000	33.160	18.929	69.960	0.000
		39.714	67.252	53.166	47.688	44.094
Hotels and Lodges	20.497	0.000	0.000	13.500	26.030	48.789
	27.735	22.932	29.660	14.809	15.419	(18.715)**
Library	10.520	0.000	0.000	12.001	32.760	13.161
	9.738	5.437	8.061	(4.767)*	(6.233)***	(5.452)*
Bar	145.687	0.000	0.000	6.973	7.200	43.594
	83.290	15.148	28.285	10.737	12.667	(15.269)**
University	0.000	1.104	1.817	24.193	24.675	0.887
	7.007	4.652	7.135	(6.842)***	(7.768)**	5.215
Park	0.000	1.054	24.807	14.132	0.000	4.735
	9.951	8.554	(11.662)*	(7.088)*	8.195	7.985
Museum	6.166	0.000	0.000	6.240	17.321	18.227
	10.653	5.681	8.232	6.091	(6.897)*	(7.253)*
Other Points of Interest	0.000	1.155	18.599	0.000	0.000	19.005
	42.310	32.573	42.717	21.292	20.448	30.355
Restaurant	36.247	0.000	0.000	11.860	0.000	20.551
	52.444	25.628	38.336	19.846	19.990	25.340
Other Food-service	0.000	0.000	16.078	0.000	0.000	2.756
	28.151	25.842	32.970	15.534	23.032	18.581
Stores or Retail Locations	4.092	0.000	0.000	4.691	11.523	0.000
	27.187	16.576	18.598	10.381	12.010	16.488
Bus Location Dummy	0.000	0.000	6.714	0.000	0.000	0.000
	16.909	7.727	7.973	4.680	6.371	5.767
Shopping Mall	0.000	0.000	0.000	0.000	0.000	4.550
	9.020	4.822	7.024	3.746	4.547	5.890
Tram Line 3a	0.000	0.000	0.000	0.000	0.000	1.121
		1.533	2.792	2.298	2.167	1.206
Tram Line 3b	0.000	0.000	0.000	0.000	0.000	0.000
		1.845	2.002	1.464	2.711	0.737
Movie Theater	0.000	0.000	0.375	0.000	0.000	0.000
	7.199	3.435	3.783	2.450	2.715	3.417

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)

Note: The coefficients are obtained after standardizing the control variables

TABLE 3. Estimates from the Origin-density Model

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	Model Type	Linear	Linear	Linear	MNL	Linear	MNL	Structural Model
(2)	Intercept $(\nu_0; \beta_0)$	0.105 (0.003)***	0.009 (0.003)**	0.000 (0.000)*	-0.037 (0.000)***	0.000 (0.000)*	-0.040 (0.000)***	-1.801 (0.096)***
(3)	Bike-Availability $(\nu_2; \alpha_1)$	0.009 (0.003)**	0.015 (0.003)***	0.013 (0.003)***	3.404 (0.081)***	0.045 (0.005)***	4.114 (0.154)***	0.004 (0.001)***
(4)	Walking Distance $(\nu_1; \beta_1, \beta_2)$	-0.059 (0.006)***	0.060 (0.005)***	0.067 (0.005)***	1.182 (0.137)***	0.068 (0.009)***	1.270 (0.284)***	-2.229$^\frown$; -15.445$^+$ (0.437)***; (2.947)***$^+$
(5)	District x Time-window F.E. $(\eta_{w_j \times di(f)}; \gamma_{w_j \times di(f)})$	No	No	Yes	Yes	Yes	Yes	Yes
(6)	Control Variables (Section 3.4) $(\vec{\nu}_{3w_j};\ \vec{\alpha}_{\{2,3,4,5\}w_j})$	No	Yes	Yes	Yes	Yes	Yes	Yes
(7)	Instruments Z_{fw_j}	No	No	No	No	Yes	Yes	Yes
(8)	In-sample R^2	0.71%	28.23%	52.44%	54.79%	43.92%	54.25%	65.60%
(9)	Out-of-sample R^2	0.52%	25.81%	48.86%	35.31%	38.29%	33.58%	64.40%

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)   Structural distance coefficients: $^\frown$(<300mts) $^+$(>=300mts)

TABLE 4. Model Fit Comparisons

7.1. **Out-of-sample Tests.** We compare the prediction performance of our structural model with the performance of different linear and multinomial-logit models. To perform these analyses we split our sample into two equal parts. We use the first two months of data to estimate our model, and test its prediction performance on the remaining two months of data. The prediction performance of all models is assessed using the R^2 measure,[19] which captures the goodness of fit.

Alternate Models. Table 4 provides the main coefficient estimates and the fit statistics for each model. The in-sample and out-of-sample R^2 metrics are provided in the last two rows of the table.

Model 1 is a linear model that regresses observed station-use, at station f in time window w_j, Λ_{fw_j} on (i) the distance to the nearest station dis_f and (ii) the average bike-availability at station f in time window w_j, $\underline{aba}_{fw_j}$. Model 2 is this model plus the control variables from the origin-density model (Section 3.4: residential population density, presence of points of interest, ridership at metro stops, and the number of visitors at tourist locations), $\underline{rd}_f, \vec{\underline{pi}}_f, \underline{ri}_f, \underline{vi}_f$, where for $x \in \{rd, \vec{pi}, ri, vi\}$, $\underline{x}_f$ is the sum of x_l for locations l that are within 300 meters of station f. To simplify notation below, we denote the vector of these control variables by $\vec{\underline{c}}_f = \left(\underline{rd}_f, \vec{\underline{pi}}_f, \underline{ri}_f, \underline{vi}_f\right)$. In model 3, we add the

---

[19]The formulation for this fit-statistic is a weighted R^2, formally given by: $1 - (\sum_{fw_j} \sigma_{fw_j}(\ln(\Lambda_{fw_j}) - \ln(\hat{\Lambda}_{fw_j}))^2)(\sum_{fw_j} \sigma_{fw_j}(\ln(\Lambda_{fw_j}) - \overline{\ln(\Lambda)})^2)^{-1}$, where Λ_{fw_j} and $\hat{\Lambda}_{fw_j}$ are the actual and predicted station-use for station f, time-window w_j, and $\overline{\ln(\Lambda)}$ is the weighted mean of $\ln(\Lambda_{fw_j})$.

time-window × *district* fixed effects. The overall model 3 is:

$$\Lambda_{fw_j} = \nu_0 + \nu_1 \cdot dis_f + \nu_2 \cdot \underline{aba}_{fw_j} + \vec{\nu}_{3w_j} \cdot \vec{\underline{c}}_f + \eta_{w_j \times di(f)} + \psi_{fw_j}.$$

Model 4 is a multinomial logit model (MNL), which is often used when choice data is given. The variables we use are the same as those in linear models, except they are now assumed to affect the utility of using a station, instead of directly affecting the station-use. The station-use in this model is given by,

$$\Lambda_{fw_j} = \frac{\exp\left(\nu_0 + \nu_1 \cdot dis_f + \nu_2 \cdot \underline{aba}_{fw_j} + \vec{\nu}_{3w_j} \cdot \vec{\underline{c}}_f + \eta_{w_j \times di(f)} + \psi_{fw_j}\right)}{1 + \sum_{g \in F} \exp\left(\nu_0 + \nu_1 \cdot dis_g + \nu_2 \cdot \underline{aba}_{gw_j} + \vec{\nu}_{3w_j} \cdot \vec{\underline{c}}_g + \eta_{w_j \times di(g)} + \psi_{gw_j}\right)}.$$

The linear and multinomial models face the same endogeneity issues as our structural model, those that arise from endogenous station locations and bike-availabilities. We correct for these endogeneities in the linear model and the MNL model using the same set of instruments as those in our structural model (described in Section 5.1). Models 5 and 6 are the models with instruments.[20] Model 7 is the structural model of the paper.

Comparisons with Alternate Models. The structural model provides an *out-of-sample fit of 64.4%* which is significantly higher than that of other models, whose fit ranges from 0.5% to 48.8%. Comparing the fit with models that include instruments— the linear model provides an out-of-sample fit of 38.3%, while the MNL model has a fit of 33.6%. Thus, *among the models that provide unbiased estimates (Models 5, 6, 7), the structural model is almost twice as good as other models.*

The coefficient for bike-availability is positive and highly significant in all of the models. However, the coefficient for proxies of walking distance have different signs in different models. In fact, most reduced form models suggest that more walking distance is associated with more use! As discussed before, we can not correctly measure the effect of distance in reduced form models as these proxies confound many different aspects beyond walking distance.

7.2. **Comparing predicted substitution levels with those from a reduced form model.** We build a reduced form model that can assess the extent of substitution. This model estimates the increase in station-use when any nearby station stocks-out giving us an assessment of the extent of substitution, without imposing the choice structure of the structural model.

Formally, let $q_{ft} = 1$ if at time t, any station near station f is stocked out, and 0 otherwise.[21] The dependent variable Λ_{fqw_j} is the observed average station-use at station f, in time window w_j, when

---

[20]We use the IV-Logit method as in Petrin [2002] to correct for the endogeneity in the MNL model.

[21]The set of nearby stations N_f is the same set of nearby stations we had used in our main analysis to construct a local-stockout state for each station. See section 5.4 for details.

$q_{ft} = q$, $q \in \{0, 1\}$:

$$log\left(\Lambda_{fqw_j}\right) \sim \rho_0 + \rho_1 \cdot \mathbb{I}\{q = 1\} + \rho_f + \rho_{w_j} + \epsilon_{fqw_j}$$

Here the variable ρ_1 captures the effect of stock-outs at nearby stations on the use at station f. ρ_f and ρ_{w_j} are the station and time-window fixed effects, ϵ_{fqw_j} are the error terms. The estimated value of ρ_1 is 0.094*** (s.e. =0.018).[22] This implies that a typical station's demand increases by about 9.862% when any neighboring station stocks out. This suggests that the fraction of users that substitute on a station being stocked is limited, likely less than 9.862%– as the increase arises from users substituting from *all* the nearby stations that might be stocked-out when $q_{ft} = 1$. Alternate specifications of q_{ft}[23] lead to substitution percentage estimates ranging from 6.562%-8.389%. Overall this small increase in use on nearby stockouts is very much in line with our estimated substitution percentage of 5.068% in the structural model.

8. Managerial Use Cases

Our estimated model can be used to predict station-level system use for any given station network and any bike-availabilities at the stations in that network. This provides us with a powerful tool to compare alternate station-networks and/or system management policies and identify improvement opportunities. We illustrate such use next.

The analyses provided are intended to be illustrative of the potential different uses of our estimates, nevertheless in the interest of simplicity these analyses necessarily exclude a number of other factors not considered in our study such as political and geographical constraints on station locations and sizes, management challenges in increasing availability, etc. A complete analysis of the below issues remains an open subject for future study; what follows is simply indicative of how the estimated model can be used by system operators.

8.1. **Improvements to Bike-Availability.** System managers can improve bike-availability at specific stations by giving higher priority to these stations in trans-shipments, scheduling of preventive maintenance, etc. Figure 8.1(a) provides the increase in system-use on a 0.1 increase in bike-availability at stations in different districts (total effect, as in Section 6.1). The same investments in improving availability have more than twice the benefit in the young, densely populated and touristy districts 3 or 4 than in the district 16, which is relatively older and sparsely populated.

Figure 8.1(b) compares the effect of increasing bike-availability by 0.1 in different districts and time-windows. Improving bike-availabilities in the evening time-windows (1600-2000) is the most useful.

---

[22]N= 7320, Adjusted R^2= 67.7%

[23]$q_{ft} = 1$ iff any station within 300, or 500 meters of station f is stocked out.

‡(a),(b),(c) show effect of 0.1 increase in station's bike-availabilities †(d),(e),(f) show effect of 10% uniform increase in station-density. Refer footnote #23.
*Increase in use is relative to average station-use in the system. These effects are normalized by number of stations affected in each case.
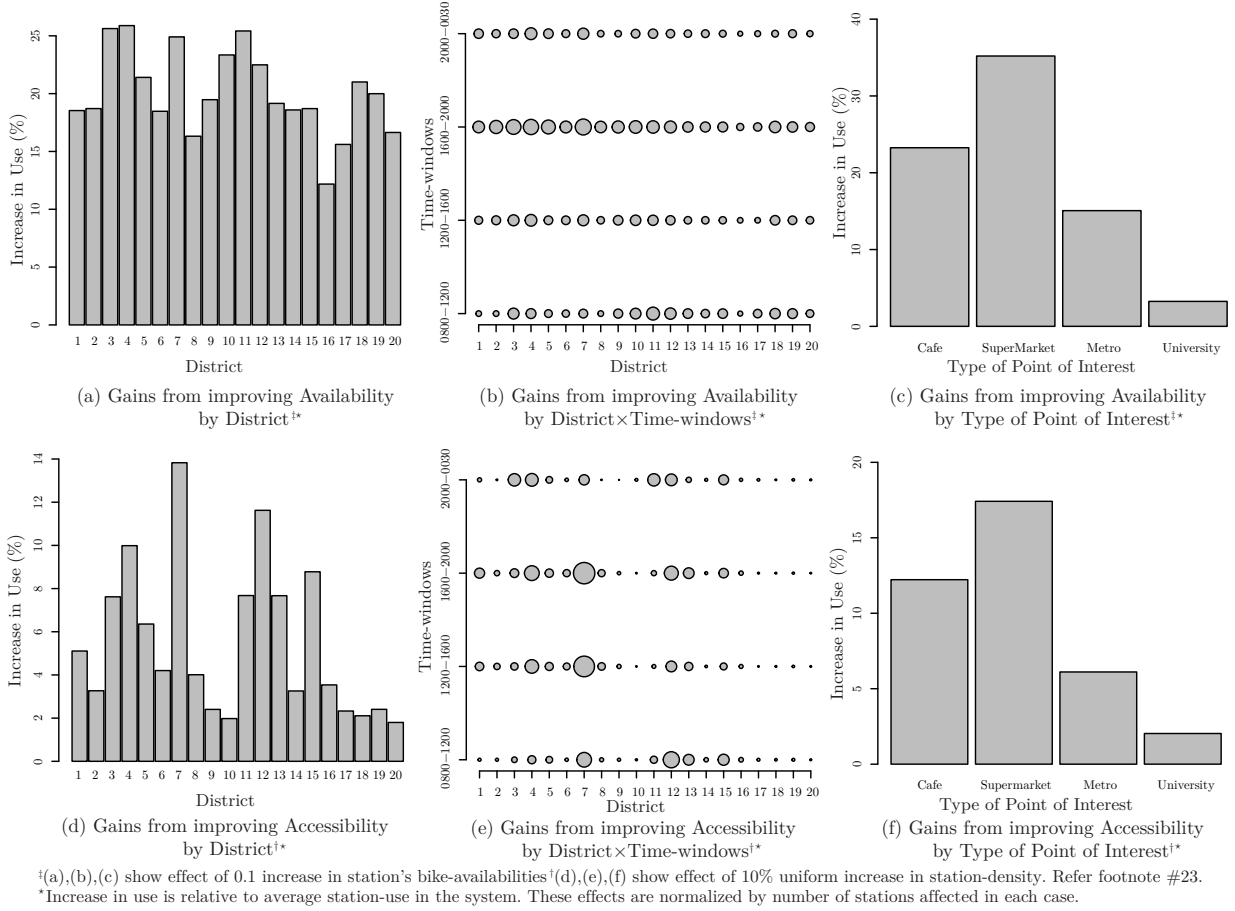
FIGURE 8.1. Different Impacts of Improving Station Density and Bike-Availability

System managers should allocate availability-improving resources (transshipment trucks, etc.) to districts 11 and 12 in the morning hours, and move them to districts 4 and 7 in the evening hours.

Finally, we compare the effect of improving availability at stations near users originating from a particular kind of location. Figure 8.1(c) compares increase in system-use that comes from increasing availability of the four stations closest to all locations of a specific type (eg: all cafes) by 0.1. We find that the effect of improving availability is most prominent for users who originate at *supermarkets*.

8.2. **Denser Station Networks.** We consider alternate station networks that are denser than the existing network. While our model can help us predict the use from any alternate station network, we illustrate the use by considering some specific proposals for increasing station density.

First, as a benchmark consider the hypothetical case when the station density is increased all over the system by 10% (done by adding about 95 new stations "uniformly" to the city).[24] Our estimates

---

[24]We effectuate this by reducing all user–station distances by 4.653%, which increases density by 10% while shrinking the city's area by 9.090% $(1 - (1 - 4.653\%)^2)$, and then scaling-up system-use from this shrunken city to estimate the effect in our actual city. This process increases the number of stations such that any given area has 10% more stations than before, thus illustrating the effect of distance while respecting the distribution of stations in the current network.
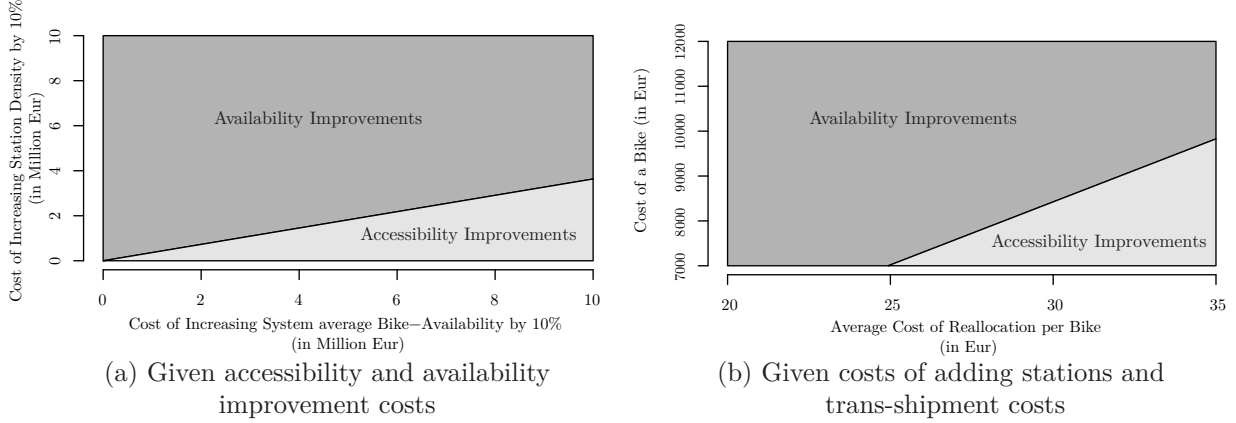
(a) Given accessibility and availability
improvement costs

(b) Given costs of adding stations and
trans-shipment costs

FIGURE 8.2. Preferred Improvements

suggest that such a 10% increase in station density results in a 4.436% ($\pm$0.529%) increase in system-use. Figure 8.1(d) shows the effect of increasing station density in the same way as the benchmark, but now only in *specific districts* by 10%. The effect of increasing density is generally higher in districts where the existing system has a lower station-density. Interestingly, districts 1 and 2, although quite busy and densely populated, reap lower benefits of increasing density; likely these districts are already saturated with stations.

System operators also have the ability to temporarily increase station density in some time-windows by the use of so-called mobile stations or "valet/manned" stations. The effect of a 10% uniform increase in station density in different districts *in different time-windows* (Figure 8.1(e)) helps us identify opportunities for the use of such mobile resources– for e.g. mobile stations could be employed in district 12 in the morning, and moved to district 7 in the afternoon and evening.

Finally, we consider if there are specific locations where increasing access to stations would be most useful. Figure 8.1(f) provides increases in system-use that come from increasing the accessibility for users originating from a particular kind of location. We consider the increases from bringing the four stations closest to all points of interest of a specific type, 10% closer to the point of interest; for e.g. bringing the four closest stations of all cafes, 10% closer to the cafes. We find that bringing stations closer to users originating at *supermarkets* has the highest impact followed by bringing stations closer to metros, cafes, or universities.

8.3. **Comparison of Accessibility and Availability Improvements.** System operators with fixed budgets are concerned with identifying whether accessibility or availability improvement give higher returns in terms of increases in system-use per dollar spent. Figure 8.2 identifies the preferred improvements, by combining the predicted benefits of improvements with the potential costs of achieving these improvements. Panel (a) performs this comparison in terms of the costs of system-wide accessibility

and availability improvements, while Panel (b) considers costs under the assumption that accessibility improvements are achieved by adding more bikes at new stations while availability is increased by extra transshipments. While the preferred strategy would depend on the precise cost estimates, anecdotally, for the city of Paris, the costs are in the ranges where this analysis predicts availability improvements are much preferred.[25]

9. Robustness

9.1. **Variable Definitions, Model Specification, Computational Choices and Instruments.**

We test the robustness of our effect sizes to alternate variable definitions, model specifications, and to computational choices made in model estimation. Table 5 reports the results of our estimation under many alternate assumptions; row (1) replicates our original estimates (from Table 1) for easy comparison. Rows (2) and (3) of the table report the estimates obtained under alternate definitions of station stocked-in or bike-available. Row (2) gives estimates from a model where a station is said to be stocked-in if there are *more than four* bikes available at the station (versus five bikes in the original estimation), Row (3) considers a station stocked-in if it has *more than six* bikes available at the station. The estimates are similar to those obtained under our original regressions.

Row (4) replicates our analysis for data from weekends for the same four month period as our main analysis. We find that the impact of increasing station density are higher on weekend days as compared to weekdays (5.049% v/s 4.436%), For the total bike-availability effect (or long-term effect) the weekend effect is somewhat lower (11.117% v/s 12.211%), while short term effect of increasing availability are comparable (9.506% v/s 9.493%). These slight differences probably arise due to differences in the user-base; weekdays are perhaps dominated by work-commuters from outer-boroughs, while weekends are probably Parisians.

In row (5) we include the distance to the nearest metro station as a covariate in the outside option of each user. Metro stations in addition to acting as feeders to use of bike-share stations, can also act as substitutes to bike-share. While the feeder effect is already captured in the density parameters, inclusion of metro stations as the outside option can further capture any substitution effects. Although we had hoped to find a substitution effect, we find the opposite effect in the model in Row (5). This positive coefficient suggests that irrespective of where the metro variable is included, the net effect of the presence of a metro station is an increase in bike-share use, i.e. metro stations feed demand to bike-share stations rather than act as substitutes.

---

[25]Anecdotal cost estimates provided by the operator are: for adding new stations, 10,000€/system-bike/yr, and for transshipments cost roughly 25€/bike-transshipment.

		Primary variables			10% increase in Station Density	10% increase in Average Bike-Availability			Number of observations
		Walking Distance (0-300mts)	Walking Distance (>300mts)	Average Bike-Availability					
		β_1	β_2	α_1		Short-term	Long-term	Total	
(1)	Original Estimates	-2.229 (0.437)***	-15.445 (2.947)***	0.004 (0.001)***	4.436%	9.493%	2.482%	12.211%	45,122
(2)	Stocked-in: > 4 Bikes	-2.378 (0.388)***	-15.432 (3.369)***	0.003 (0.001)***	4.380%	9.552%	2.409%	12.191%	45,240
(3)	Stocked-in: > 6 Bikes	-1.863 (0.448)***	-16.276 (3.554)***	0.004 (0.001)***	4.533%	9.470%	2.674%	12.397%	44,936
(4)	Weekends only	-3.360 (0.42)***	-18.187 (4.051)***	0.002 (0.001)**	5.049%	9.506%	1.471%	11.117%	43,324
(5)	Metro in outside option	-1.761 (0.467)***	-14.798 (3.057)***	0.004 (0.001)***	4.335%	9.486%	2.621%	12.355%	45,122
			Metro outside option 0.460 (0.198)*						
(6)	Finer Grid Size (20 meters)	-2.201 (0.429)***	-15.397 (3.269)***	0.002 (0)***	4.427%	9.496%	2.468%	12.199%	45,122
(7)	16 Top states considered	-2.347 (0.405)***	-15.671 (3.334)***	0.003 (0.001)***	4.510%	9.298%	2.341%	11.857%	89,333
(8)	Total user interest -10%	-2.336 (0.432)***	-16.157 (3.482)***	0.003 (0.001)***	4.707%	9.572%	2.630%	12.454%	45,122
(9)	Total user interest +10%	-2.293 (0.43)***	-15.860 (3.403)***	0.004 (0.001)***	4.461%	9.536%	2.474%	12.245%	45,122
(10)	Choice set size = 5	-2.275 (0.41)***	-8.448 (1.425)***	0.004 (0.001)***	4.150%	9.566%	2.587%	12.400%	45,201
(11)	Focal Station Instruments	-3.690 (0.523)***	-15.356 (6.167)*	0.003 (0.001)*	4.386%	9.482%	2.193%	11.883%	45,122
(12)	Demeaned rate Incoming-Bikes$_{wj-1}$	-2.137 (0.376)***	-15.656 (3.208)***	0.004 (0.001)***	4.424%	9.488%	2.710%	12.455%	45,122
(13)	Alt. Instrument Parameters - I	-1.876 (0.638)**	-15.139 (5.649)**	0.003 (0.001)**	4.193%	9.487%	2.585%	12.317%	45,122
(14)	Alt. Instrument Parameters - II	-1.811 (0.334)***	-16.518 (2.44)***	0.004 (0.001)***	4.477%	9.492%	2.512%	12.243%	45,122
(15)	Lower granularity data - 4 mins	-2.088 (0.431)***	-15.326 (3.22)***	0.004 (0.001)***	4.605%	9.546%	2.661%	12.460%	45,122

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)

TABLE 5. Robustness Tests

Next, we investigate the role of various computational choices made in estimation. In row (6) we provide estimates obtained by using a finer grid for our numerical integration (viz., one that covers 1.5 times as many simulation points for continuous spatial elements) this produces no qualitative change in the estimated effects. In row (7) we consider 16 top local stock-out states for each *station × time-window*. In row (8) and (9) we test the sensitivity of our estimates with respect to the choice of total potential user interest in a day. In row (10) we increase the definition of local choice set of a user to nearby *five* stations. The estimates are exceptionally robust to these choices.

Finally, we investigate the effect of different instruments on our estimates. In row (11) we use only the focal station's exogenous characteristics as instruments, as used in Davis [2006]. In row (12) we use a novel instrument which is the average realized rate of incoming bikes at station f in lagged time-window $w_j - 1$, demeaned at station level. This instrument affects the starting number of bikes at station f, time-window w_j, and therefore its average bike-availability. It however does not affect the unobserved factors influencing station-use at f in time-window w_j, after controlling for demand sources in the origin-density model and station level factors removed in demeaning process. In rows (13) and (14) we use alternative parameters for construction of instruments $\hat{Z}_{fw_j} = \{rd_f(a,b,c,d), \vec{pi}_f(a,b,c,d), ri_f(a,b,c,d), vi_f(a,b,c,d)\}$. The set used for row (13) is $\{(0,100,0,100), (100,300,0,300), (300,500,0,500)\}$ and for row (14) is $\{(0,100,0,100), (100,200,0,200), (200,300,0,300), (300,500,0,500), (0,100,100,300), (0,100,300,500)\}$. The marginal effects are again very close to the original model.

In row (15), we report estimates from an artificially limited 4-minute snapshot dataset (we drop every alternate observation). Arguably, the 2-minute snapshots in our original analysis might not be frequent enough; system-use is substantially underestimated due to many 2-minute intervals where both checkouts and returns happen. If this were the case, 4-minute snapshots would then be even worse and lead to substantially different estimates. This is not the case with our data suggesting that, on average, the flows are not high enough and 2 (and even 4) minute snapshots are sufficient for our estimation.

In summary, we find that our estimates are robust to various model specifications, variable definitions, computational, and instrument choices.

9.2. **Robustness of Distance Disutility Function.** Our main model assumes a piecewise linear form for the disutility of distance. We reran our model with linear and quadratic disutility functions, and a variety of different kink points for the piecewise linear form— 100, 200, 250, 275, 325, 350m instead of 300m in original model.

The estimates and full results for all these alternate specifications are reported in Table 3 in Appendix B.2. Irrespective of the specification used for the distance disutility, our estimate for the effect of *average bike-availability* is essentially identical. In so far as the effect of *distance* is concerned, depending on the specification we get different estimates for parameters of the distance disutility function, but remarkably irrespective of the specification, the effects of walking distance on ridership is essentially the same– a 10% uniform increase in station density always leads to between 3.926%-4.494% increase in system use. Also notably, irrespective of the functional form– all specifications imply that the disutility from distance is convex. The extent of substitution or the short-term effect of bike-availability which derives from the distance effect is again essentially identical in all specifications ranging from 9.490% to 9.525%. Finally, the total bike-availability effect ranges from 11.771% to 12.284%.

10. Discussion

Each use of a bike-share system involves two transactions: the user must choose a station with available bikes; and she must also be able to return the bicycle to a station with empty docking points. Thus each station features two streams of use—outgoing and incoming—and so there are two kinds of availabilities, bike-availability and docking-point availability. System-use presumably depends on both kinds of availability, but our analysis has focused on *outgoing* use and bike-availability.

Observe that at the system level, incoming and outgoing use must be equal and each corresponds to the number of trips; therefore, either use type can be analyzed to develop important prescriptions for system-use. Yet bike-availability and dock-availability can have different and significant effects on system-use. There are two important differences between these effects that make the analysis of bike-availability far more relevant. First, when bikes are not available, the user has the option of either seeking out another station or forgoing the bike-share system entirely. However, the same cannot be said when docking points are not available: the affected user does not have the option of abandoning the bicycle and she can complete her trip only by finding another station (users using Vélib' get an extra 15 free minutes when the preferred station has no available docking points). Note that in this case the user can ride the bicycle to an alternate station, which is presumably easier than walking there. So in the short term, use is affected more by bike-availability than by the availability of docking points.

Second, bike-share systems are designed with many more docking points than bikes (to accommodate demand asymmetries at different times of the day, etc.); there are usually almost twice as many docking points as bikes. Hence not finding an available dock is much rarer (in our data) than not finding an available bicycle. So even though an under-supply of docking points will degrade the user experience and, in the long run, have a negative effect on system-use, from a practical standpoint we expect that

docking point availability has a much weaker impact. Together these trends indicate that, in the short run and over the long run, system-use is much more likely to be affected by bike-availability than dock-availability; hence our analysis focuses on the former. It is theoretically possible to extend our model so that it includes docking point availability, but by doing so, we expect to find no significant differences than our current model despite much higher computational complexity.

This paper provides the first empirical estimates of user response to accessibility and availability in the context of bike-share systems. We illustrate the use of our estimates in identifying a number of different system improvement efforts. Furthermore, the methodology developed here can be used in a variety of demand estimation contexts where products are spatially differentiated and with availabilities that change frequently. It is important to highlight that sufficing of nearby choices is a unique feature of spatially differentiated markets and might not be applicable to traditional demand estimation problems like choice of products. But given the proliferation of spatially differentiated markets in form of car-sharing, cab-hailing, food delivery platforms, these ideas might be generalizable and used in these other contexts.

While our 2-minute snapshots are adequate for the purposes of our study, they might not be sufficient in a system with higher flow rates. In future work, we hope to address the limitations of this study. First, a more detailed data set on user starting locations would improve the precision of estimates of the effects we study. Second, a larger study comparing many cities could provide insight not only into how user preferences vary by city but also into how those preferences might be driven by different demographic and/or geographic factors. Such analyses could help bike-share systems fully deliver on their promise of transforming urban lifestyles.

References

G. Allon, A. Federgruen, and M. Pierson. How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management*, 13(4):489–507, 2011.

E. T. Anderson, G. J. Fitzsimons, and D. Simester. Measuring and mitigating the costs of stockouts. *Management Science*, 52(11):1751–1763, 2006.

E. Belavina, K. Girotra, and A. Kabra. Online grocery retail: Revenue models and environmental impact. *Management Science*, 2016.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, pages 841–890, 1995.

H. A. Bruno and N. J. Vilcassim. Research note-structural demand estimation with varying product availability. *Marketing Science*, 27(6):1126–1131, 2008.

G. Cachon. Retail store density and the cost of greenhouse gas emissions. *Management Science*, 2014.

A. C. Cameron and P. K. Trivedi. Microeconometrics: Methods and applications. *Cambridge university press*, 2005.

C. T. Conlon and J. H. Mortimer. Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30, 2013.

P. Davis. Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics*, 37(4):964–982, 2006.

J.-P. Dubé, J. T. Fox, and C.-L. Su. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267, 2012.

A. El-Geneidy, M. Grimsrud, R. Wasfi, P. Tétreault, and J. Surprenant-Legault. New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, 41(1):193–210, 2014.

J. C. García-Palomares, J. Gutiérrez, and M. Latorre. Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography*, 35(1), 2012.

D. K. George and C. H. Xia. Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. *European Journal of Operational Research*, 211(1):198–207, 2011.

J. A. Guajardo, M. A. Cohen, and S. Netessine. Service competition and product quality in the US automobile industry. *Management Science*, 2015.

L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

S. Henderson, E. O'Mahony, and D. B. Shmoys. (Citi)Bike sharing. *Submitted*, 2016.

R. Lederman, M. Olivares, and G. Van Ryzin. Identifying competitors in markets with fixed product offerings. *Working Paper, SSRN 2374497*, 2014.

J. Li, S. Netessine, and S. Koulayev. Price to compete... with many: How to identify price competition in high dimensional space. 2015.

D. McFadden. The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328, 1974.

A. Musalem, M. Olivares, E. T. Bradlow, C. Terwiesch, and D. Corsten. Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197, 2010.

A. Nevo. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548, 2000.

A. Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342, 2001.

W. A. O'Neill, R. D. Ramsey, and J. Chou. Analysis of transit service areas using geographic information systems. *Transportation Research Record*, 1364, 1992.

S. O'Sullivan and J. Morrall. Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*, 1538:19–26, 1996.

J. Pancras, S. Sriram, and V. Kumar. Empirical investigation of retail expansion and cannibalization in a dynamic environment. *Management Science*, 2012.

P. Pendem and V. Deshpande. Maximizing ridership in bike-sharing systems using empirical data and stochastic models. *UNC Working Paper*, 2016.

A. Petrin. Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, 110(4):705–729, 2002.

W. J. Reilly. *The law of retail gravitation*. WJ Reilly, 1931.

D. Singhvi, S. Singhvi, P. I. Frazier, S. G. Henderson, E. O'Mahony, D. B. Shmoys, and D. B. Woodard. Predicting bike usage for new york citys bike sharing system. *In AAAI 2015 Workshop on Computational Sustainability.*, 2015.

A. Tangel. City bike-sharing programs hit speed bumps. *The Wall Street Journal*, July 9, 2014. URL `http://on.wsj.com/1oJWjuv`.

R. Thomadsen. The effect of ownership structure on prices in geographically differentiated industries. *RAND Journal of Economics*, pages 908–929, 2005.

A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.

J. Ypma. Introduction to ipoptr: an R interface to Ipopt. 2010.

F. Zhao, L.-F. Chow, M.-T. Li, I. Ubaka, and A. Gan. Forecasting transit walk accessibility: Regression model alternative to buffer method. *Transportation Research Record*, 1835:34–41, 2003.

ONLINE APPENDIX

Appendix A. De-Seasonalizing Weather Effects

We report here the effect of de-seasonalizing the system-use data with prevalent weather conditions.

The weather data is collected at a half-hourly frequency for the city of Paris, specifically the Temperature, Humidity, Wind Speed and "Conditions" (clear, mist, cloudy, etc.) from weatherbase.com. We incorporate each weather condition as dummy variables of ranges of their different expected impacts.

The Temperature variable is divided in three ranges of: ≤ 10, $(10, 30]$, and > 30; Humidity in the ranges of: ≤ 40, $(40, 80]$, and > 80; Wind Speed in the ranges of ≤ 20, $(20, 30]$, and > 30, while weather conditions are classified into clear, fog, heavy fog, heavy rain showers, light drizzle, light rain, light rain showers, light thunderstorms and rain, mist, mostly cloudy, overcast, partial fog, partly cloudy, rain, scattered clouds, shallow fog, heavy thunderstorms and rain, light thunderstorm, thunderstorm, thunderstorms and rain, light fog, and patches of fog.

	Value	Std. error
Temperature		
Range [,10)	0.000	
Range [10, 30]	0.301	(0.036)***
Range (30,]	0.211	(0.113).
Humidity		
Range (,40]	0.000	
Range (40, 80]	-0.114	(0.046)*
Range (80,]	-0.225	(0.054)***
Wind Speed		
Range (, 20]	0.000	
Range (20, 30]	0.060	(0.03)*
Range (30,)	-0.074	(0.257)
Conditions		
Clear	0.000	
Fog	-1.497	(0.161)
Heavy Fog	-1.064	(0.671)
Heavy Rain Showers	-0.134	(0.487)
Light Drizzle	0.097	(0.126)
Light Rain	-0.651	(0.046)
Light Rain Showers	-0.238	(0.084)**
Light Thunderstorms and Rain	-0.441	(0.306)
Mist	-1.029	(0.346)**
Mostly Cloudy	-0.142	(0.024)***
Overcast	-0.391	(0.085)***
Partial Fog	-1.161	(1.362)
Partly Cloudy	0.024	(0.043)
Rain	-1.804	(0.22)***
Scattered Clouds	-0.058	(0.039)
Shallow Fog	0.163	(0.141)
Heavy Thunderstorms and Rain	-1.073	(0.507)*
Light Thunderstorm	0.108	(0.367)
Thunderstorm	-0.263	(0.361)
Thunderstorms and Rain	0.125	(0.579)
Light Fog	-0.166	(0.52)
Patches of Fog	-0.594	(0.429)

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)

TABLE 1. Weather Variables Effect

	Average Bike-Availability		Station Density	
	(1)	(2)	(3)	(4)
District x Time-window F.E.	Yes	Yes	Yes	Yes
Instrumental Variables		Yes		Yes
Adjusted R^2	0.240	0.593	0.368	0.742
Number of observations	5676	5676	5676	5676

TABLE 2. Relevance test of Instruments

The regression model is given by,

$$\ln(\Lambda_t) = \rho_0 + \vec{\rho}_1 \cdot Temp_t + \vec{\rho}_2 \cdot Humidity_t + \vec{\rho}_3 \cdot Wind\_Speed_t + \vec{\rho}_4 \cdot Condition_t + \rho_{h(t)} + \epsilon_t \qquad (A.1)$$

where t denotes a two minute interval, Λ_t system-use in time t, and $h(t)$ denotes half-hourly index within a day (48 in total) corresponding to t. Each of the $\vec{\rho}_1, \vec{\rho}_2, \vec{\rho}_3$, and $\vec{\rho}_4$ is a vector of effects of different levels (dummies) for our weather variables. Finally, we also include half-hourly fixed effects, $\rho_{h(t)}$.

Table 1 shows the impact of weather variables. Based on the estimated weather effects, station-use Λ_{ft} is de-seasonalized as follows.

Define the net weather effect at time t as,

$$\rho_t^w = \hat{\vec{\rho}}_1 \cdot Temp_t + \hat{\vec{\rho}}_2 \cdot Humidity_t + \hat{\vec{\rho}}_3 \cdot Wind\_Speed_t + \hat{\vec{\rho}}_4 \cdot Condition_t$$

The de-seasonalized station-use, $\hat{\Lambda}_{ft}$ is given by, $\hat{\Lambda}_{ft} = \Lambda_{ft}/\exp(\rho_t^w)$. For all our analysis in paper, we have used this de-seasonalized station-use $\hat{\Lambda}_{ft}$ in place of Λ_{ft}.

APPENDIX B. ADDITIONAL RESULTS

B.1. **Relevance Test for Instruments.** In Table 2, we report the test for relevance of the instruments by estimating a linear regression model with the average bike-availability for each $station \times time-window$ as a dependent variable, and the instrumental variables described in section 5.1 as covariates. We control for other model covariates– intercept and $time\text{-}window \times district$ fixed effects. Inclusion of instrumental variables (Z_{fw_j}) in Model 2 is able to predict an additional 35.3% of the variation.

$$\underline{aba}_{fw_j} = \eta_0 + \eta_{w_j \times di(f)} + \vec{\eta}_1 \cdot Z_{fw_j} + \epsilon_{fw_j}^{\eta} \qquad (B.1)$$

Similarly, we estimate a linear regression with station density as a dependent variable and using the same set of covariates. We capture station density as the number of stations within 500m of a station. Inclusion of instrumental variables (Z_{fw_j}) in Model 4 is able to predict an additional 37.4% of the variation. Very similar results are obtained by using other measures for station density (number of stations within 300m of a station, distance to nearest station from a station, average distance to five nearest station from a station).

$$station\_density_{fw_j} = \zeta_0 + \zeta_{w_j \times di(f)} + \vec{\zeta}_1 \cdot Z_{fw_j} + \epsilon_{fw_j}^{\zeta} \qquad (B.2)$$

B.2. **Robustness of the Distance Disutility Function.** Table 3 reports the estimates with a variety of alternate distance disutility functions– with alternate kink points in the piecewise linear form– 100, 200, 250, 275, 325, 350 meters

		Primary variables			10% increase in Station Density	10% increase in Average Bike-Availability			Number of observations
		Walking Distance (until kink) β_1	Walking Distance (after kink) β_2	Average Bike-Availability α_1		Short-term	Long-term	Total	
(1)	Original Estimates	-2.229 (0.437)***	-15.445 (2.947)***	0.004 (0.001)***	4.436%	9.493%	2.482%	12.211%	45,122
(2)	Kink at 100mts	3.072 (2.56)	-6.654 (0.551)***	0.003 (0.001)***	4.082%	9.523%	2.235%	11.970%	45,122
(3)	Kink at 200mts	-0.261 (0.841)	-9.429 (1.255)***	0.003 (0.001)***	4.419%	9.502%	2.395%	12.125%	45,122
(4)	Kink at 250mts	-1.258 (0.562)*	-12.175 (1.951)***	0.004 (0.001)***	4.533%	9.496%	2.508%	12.243%	45,122
(5)	Kink at 275mts	-1.869 (0.479)***	-13.477 (2.433)***	0.004 (0.001)***	4.494%	9.497%	2.545%	12.284%	45,122
(6)	Kink at 325mts	-2.436 (0.405)***	-18.186 (4.798)***	0.004 (0.001)***	4.326%	9.490%	2.442%	12.164%	45,122
(7)	Kink at 350mts	-2.531 (0.397)***	-22.679 (7.714)**	0.003 (0.001)***	4.179%	9.489%	2.447%	12.168%	45,122
		Walking Distance	Walking Distance Square	Average Bike-Availability					
(8)	Linear Model	-5.829 (0.338)***		0.003 (0.001)***	3.926%	9.525%	2.051%	11.771%	45,122
(9)	Quadratic Model	4.217 (1.714)*	-19.395 (4.469)***	0.003 (0.001)***	4.235%	9.500%	2.373%	12.099%	45,122

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)

TABLE 3. Robustness of Distance Effect

instead of 300 meters in the original model, a simple linear function, and a quadratic function. Section 9.2 provides a discussion of these estimates.

B.3. **Effect of District Demographic variables on Accessibility and Availability marginal effects.** We quantitively relate the marginal effects of increasing station density (from Paper Section 8.2) and that of increasing bike-availability (from Paper Section 8.1) with some of the characteristics of nearby area. In particular, we regress the marginal effects of increasing station density/increasing bike-availability (dependent variables) on the following nearby characteristics–– percentage of young population (between ages of 18 and 34 years from INSEE), existing density of stations, residential population density (from INSEE), total number of annual tourists (from Office du Tourisme et des Congrès de Paris (OTCP)), and whether the district is one of the outer districts (district # 11 to 20) (independent variables). The results of these regressions are presented in Table 4.

We find that districts which have a lower density of stations experience higher gains from increasing station density. Districts that are *younger*, have *more tourists*, or a *higher population density* obtain higher gains from increasing bike-availability. On the other hand, districts which are *outer*, or already have *high station-density* accrue smaller gains from increasing bike-availability.

District Characteristics	Increase in System-Use from +10% Station Density (By Districts)	Increase in System-Use from +10% Bike-Availability (By Districts)
Intercept	5.498 (0.708)	20.124 (0.516)
Percentage Young Population (18-34)	1.788 (1.145)	1.869 (0.835)*
Station Density	-3.231 (1.49)*	-2.691 (1.086)*
Population Density	-0.536 (1.072)	2.319 (0.782)*
Number of Tourists	0.645 (0.889)	1.541 (0.648)*
Outer District	-1.492 (1.473)	-2.943 (1.074)*
Number of Observations	20	20
Adj-R^2	0.197	0.604

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)

The explanatory variables are standardized by substracting their mean and dividing by their standard deviation

TABLE 4. District Demographic variables on marginal effects

			Primary variables			10% increase in Station Density	10% increase in Average Bike-Availability			Number of Observations
	Method of Classification		Walking Distance (0-300mts) β_1	Walking Distance (>300mts) β_2	Average Bike-Availability α_1		Short-term	Long-term	Total	
(1)	Inner vs Outer Districts	Residential Districts	-0.962 (0.517)	-25.089 (8.803)**	0.002 (0.001)***	5.447%	9.431%	1.956%	11.571%	45,122
		Commerical Districts	-5.666 (0.486)***	-3.135 (1.451)*	0.009 (0.001)***	3.217%	9.664%	3.555%	13.562%	
(2)	Low vs High Population Density	Residential Districts	0.559 (0.437)	-15.797 (3.94)***	0.002 (0.001)***	4.203%	9.324%	1.457%	10.917%	45,122
		Commerical Districts	-4.305 (0.522)***	-6.755 (1.632)***	0.008 (0.001)***	4.088%	9.667%	4.484%	14.585%	
(3)	Density of Points of Interests	Residential Districts	-1.764 (0.553)**	-23.806 (9.599)*	0.004 (0.001)***	5.766%	9.558%	3.017%	12.864%	45,122
		Commerical Districts	-3.314 (0.51)***	-3.868 (1.799)*	0.006 (0.001)***	2.317%	9.476%	3.084%	12.852%	

*(p-value<0.05)   **(p-value<0.01)   ***(p-value<0.001)

TABLE 5. Estimates by Different Geographic Subsamples

B.4. **Insights on geographic sub-samples of data.** We estimate a version of our model that allows for different availability and distance coefficients for residential and commercial districts. We try a number of different ways to classify districts as residential and commercial.

Below we present the details of our analysis using three different ways of classifying districts as residential and commercial. In table 5, we report the main estimates and the marginal effects.

High and low population density: We classify the 10 districts with the highest resident population density as the residential group, the remaining 10 districts are classified as commercial.

High and low number of points of interest: We classify the 10 districts with the highest number of points of interest as the commercial districts, the remaining 10 districts are classified as residential.

Inner and outer districts: The 10 inner districts (districts 1 to 10) are classified as commercial while the 10 outer districts are classified as residential districts.

We find that the effect of disutility of distance is higher in residential districts than in commercial districts. Perhaps, commercial districts with the presence of retail outlets, etc. are more amenable to walking. On other hand, availability has a higher impact in commercial districts. Arguably, busy users in commercial districts are more responsive to a higher expected chance of finding bikes.

Appendix C. Variable Construction

	(1) Station Id	(2) Time	(3) Bikes	(4) Spaces Available
(1)	846	30/12/2012 18:52	18	16
(2)	846	30/12/2012 18:54	16	18
(3)	846	30/12/2012 18:56	16	18
(4)	846	30/12/2012 18:58	6	28
(5)	846	30/12/2012 19:00	5	29
(6)	846	30/12/2012 19:02	5	29
(7)	846	30/12/2012 19:04	7	27
(8)	846	30/12/2012 19:06	6	28

TABLE 6. Sample Data

C.1. **Origin-Density and Choice Model Variables.**

C.1.1. *Sample Data.* Our data is snapshots every two-minutes of the number of bikes and the number of empty docking points at each bike-station. We collect these snapshots for a four-month period starting in May 2013.

Table 6 shows a sample of our data, which contains the number of available bikes and the number of empty docking points at a bike-station.

	(1) Station Id	(2) Time	(3) Bikes	(4) Spaces Available	(5) Station-use	(6) Stocked-in
(1)	846	30/12/2012 18:52	18	16	1	1 (Yes)
(2)	846	30/12/2012 18:54	16	18	0	1
(3)	846	30/12/2012 18:56	16	18	0	1
(4)	846	30/12/2012 18:58	6	28	0.5	1
(5)	846	30/12/2012 19:00	5	29	0	0 (No)
(6)	846	30/12/2012 19:02	5	29	0	0
(7)	846	30/12/2012 19:04	7	27	0.5	1
(8)	846	30/12/2012 19:06	6	28	-	1

TABLE 7. Variables Constructed from Sample Data

C.1.2. *Variables constructed using Snapshot Data.* From this snapshot data, we construct two of our variables: station-use and stocked-in. Columns 5 and 6 in Table 7 shows their values.

Station-use. Station-use is the rate at which trips start at a bike-share station at a given time. For station f at time t, station-use Λ_{ft} is defined as $\frac{1}{2}(\#\text{Bike}_{ft} - \#\text{Bike}_{f,t+1})^+$ trips/minute, where $\#\text{Bike}_{ft}$ is the number of bikes at station f at the tth snapshot; the leading fraction arises due to the 2-minute gap between snapshots. Further, we omit the data from any two minute period in which *more* than four bikes are checked out or brought in to a station, which we interpret either as transshipment by system managers or as outliers in the usage.

Station-use is in Column 5 of Table 7 for our sample data. At time-period 18:52 as shown in row 1 of Table 7, the number of bikes drops from 18 to 16. The formula above implies a station-use of 1. At time-period 18:56 the number of bikes changes from 16 to 6, thus ten bikes are checked-out, which is interpreted as transshipment by system managers and not counted as station-use.

Stocked-in. We consider a station to be stocked-in if there are more than five bikes. Column 6 in Table 7 shows the stocked-in status of stations. The time-periods of 19:00 and 19:02 are marked as not stocked-in (or stocked-out), while rest of the time-periods are stocked-in.

	(1)	(2)	(3)	(4)	(5)
	Time	Stocked-in (Station 846)	Stocked-in (Station 847)	Stocked-in (Station 848)	Bike-available at l_l ($ba_{l,t}$)
(1)	30/12/2012 18:52	1	0	1	1
(2)	30/12/2012 18:54	1	1	1	1
(3)	30/12/2012 18:56	1	0	1	1
(4)	30/12/2012 18:58	1	0	1	1
(5)	30/12/2012 19:00	0	0	0	0
(6)	30/12/2012 19:02	0	0	0	0
(7)	30/12/2012 19:04	1	0	1	1
(8)	30/12/2012 19:06	1	1	0	1

TABLE 8. Bike-Available at a Location

C.1.3. *Bike-available at a location.* We consider bikes to be available at a location l at time t, if any station within 300 meters of location l is *stocked-in*; the indicator variable, bike-available, ba_{lt} is set to 1 in this case, and is set to 0 otherwise. For the small fraction of locations where there are less than 3 stations within the 300 meters, we consider if any of the the nearest 3 stations is stocked-in to define ba_{lt}.

Consider a location l_1 which is at 50m from station-id 846, while the other stations within 300m of l_1 are 847, and 848. Columns 2-4 in Table 8 shows the stock-in status of these stations while Column 5 shows the value of *bike-available* at l_1 for some sample time-periods. In row 1, which is time-period '18:52', two of its nearby stations are stocked-in and therefore its bike-available indicator is equal to 1. On the other hand, at '19:00', none of it's nearby stations are stocked-in and thus the bike-available indicator then is equal to 0.

Average bike-availability at a location. The *average bike-availability* at a location l in a time-window w_j, $j \in \{1,..6\}$, aba_{lw_j} is the average of the bike-available indicator variable at this location l for all times in the time window w_j.

Thus at the location l_1 above and for time-window w_4, $aba_{l_1 w_4}$ is constructed by taking average over $ba_{l_1 t}$, so that $w(t) = w_4$ including the time-periods shown in table 8. Formally, $aba_{l_1 w_4} = \underset{t|w(t)=w_4}{\text{Average}} \ ba_{l_1 t}$.

District 10 | District 11

l_3, Store

Station 847

Station 846

l_1, Cafe

100m   200m   300m   400m

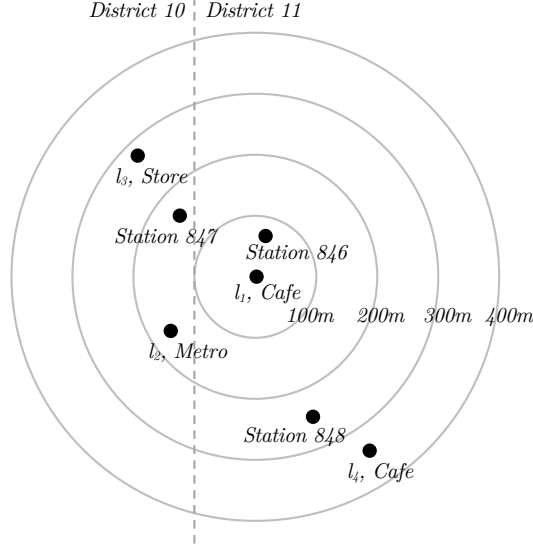l_2, Metro

Station 848

l_4, Cafe

FIGURE C.1. Example for Control Variables Construction

C.1.4. *Control Variables.* Figure C.1 shows the locations of three stations with ids' 846 to 848. Location l_1 is at the center from which concentric circles of radius 100m, 200m, 300m and 400m are drawn to show relative spacing. There is a cafe at l_1 and l_4, a metro stop with annual ridership of 500,000 at l_2, and a store at l_3. The dotted vertical line divides the regions of district 10 and 11; the region to the left of the line is district 10 while to the right is district 11. We next construct control variables in Section 3.4 of the paper for the location l_1-l_4.

Residential Population Density: The residential density at a location l, rd_l is the population density of the district in which location l lies. Location l_1 belongs to 11th district of Paris; this district has a population density of 42,138 per km^2. This is the value of the variable rd_{l_1}.

Presence of Points of Interest: For each location l, we construct a vector of dummies, $\vec{pi}_l$, each element of this vector corresponds to a type of point of interest. The different points of interests are metro, bus, and tram transit stations, stores or retail locations, restaurants, bars, cafes, other food-service locations, hotels and lodges, groceries and supermarkets, shopping malls, universities, parks, museums, libraries, and movie theaters. The dummy for a type of point of interest takes a value 1 if there is a point of interest of that type at location l and 0 otherwise. There is a cafe at l_1. Thus the element of $\vec{pi}_{l_1}$ corresponding to cafe takes a value of 1, while it is 0 for all other elements.

Metro Ridership: We construct a variable ri_l that takes the value of the annual ridership of the metro station at the location if a location l has a metro station, and the value 0 for all other locations. Location l_2 in our example has a metro station with ridership of 500,000, while location l_1 has none. Thus, $ri_{l_1} = 0$ and $ri_{l_2} =$500,000.

Tourist Volume: We construct a variable vi_l to capture the number of visitors to key tourist locations. Since none of the location in our example is a tourist location, the value of vi_l is 0 for all of them.

Specifically, for each of the locations in our example, the values of our control variables are:

l_1: rd_{l_1} is population density of 11th district, $rd_{l_1} =$42,138; $pi_{l_1}^{cafe} = 1$, while rest of the vector $\vec{pi}_{l_1}$ is 0; metro ridership ri_l is 0 and tourist volume vi_l is 0.

l_2: rd_{l_2} is population density of 10th district, $rd_{l_2} =$32,535; $pi_{l_2}^{metro} = 1$, while rest of the vector $\vec{pi}_{l_2}$ is 0; metro ridership ri_{l_2} is 500,000 and tourist volume vi_{l_2} is 0.

l_3: rd_{l_3} is population density of 10th district, $rd_{l_3} = 32{,}535$; $pi_{l_3}^{store} = 1$, while rest of the vector $\vec{pi}_{l_3}$ is 0; metro ridership ri_l is 0 and tourist volume vi_l is 0.

l_4: rd_{l_4} is population density of 11th district, $rd_{l_4} = 42{,}138$; $pi_{l_4}^{cafe} = 1$, while rest of the vector $\vec{pi}_{l_4}$ is 0; metro ridership ri_l is 0 and tourist volume vi_l is 0.

C.2. **Instrumental Variables.** We give an example of constructing our instrumental variables from Section 5.1 of the paper. We use the setup from Figure C.1 for this purpose. For brevity, we restrict ourselves to constructing the Davis instrumental variables $Z_{fw_j}^D$ corresponding to location type cafe and the values of $(a, b, c, d) = (0, 100, 0, 100)$. The construction for other location types and (a, b, c, d) values is very similar.

Our focal station is station-id 846 which we denote by f, and our focal time-window is w_4. We first measure the value of $pi_f^{cafe,\,(0,100,0,100)}$, which is the element corresponding to cafe in $\vec{pi}_f^{\,(0,100,0,100)}$. The value of $pi_f^{cafe,\,(0,100,0,100)}$ is equal to number of cafes that are between 0 and 100 meters of stations neighboring the focal station f; these neighboring stations are the stations between 0 and 100 meters of f. In Figure C.1, the neighboring stations within 0 to 100m of station f are station-ids 846 and 847. There is only one cafe which is within 0 to 100m of station-ids 846, and 847, that at the location l_1. Thus, $pi_f^{cafe,\,(0,100,0,100)} = 1$. To illustrate another example– if instead $(a, b, c, d) = (0, 100, 100, 300)$, then the stations within 100 to 300m of station 846 is just the station-id 848. The only cafe within 0 to 100m of station 848 is that at location l_4. Thus, the value of $pi_f^{cafe,\,(0,100,0,300)} = 1$. For our time-window w_4, the instruments corresponding to $pi_f^{cafe,\,(0,100,0,100)}$ in $Z_{fw_j}^D$ are obtained by Kronecker multiplication of $pi_f^{cafe,\,(0,100,0,100)}$ with $\vec{y}_{w_4}$. Here, $\vec{y}_{w_4} = \{0, 0, 0, 1, 0, 0\}$. Since $pi_f^{cafe,\,(0,100,0,100)}$ is equal to 1, these instruments which are given by $\{0, 0, 0, 1, 0, 0\}$.

The vector $Z_{fw_j}^D$ is constructed by collating together all the instrumental variables corresponding to all our control variables types and different values of (a, b, c, d) we use for our main analysis, for station f and time-window w_j. In addition, instruments vector Z_{fw_j} has an intercept term and *time-window$\times$district* dummies. The vector corresponding to *time-window$\times$district* dummies is a vector of size $6 \times 20 - 1$ (as there are 6 time-windows and 20 districts, and -1 is for time-window w_1 and district 1 element which is not identified), with 1 at the element corresponding to time-window w_4 and district 11 and 0 everywhere else, in our example where station f is 846 and time-window is w_4 of Figure C.1.
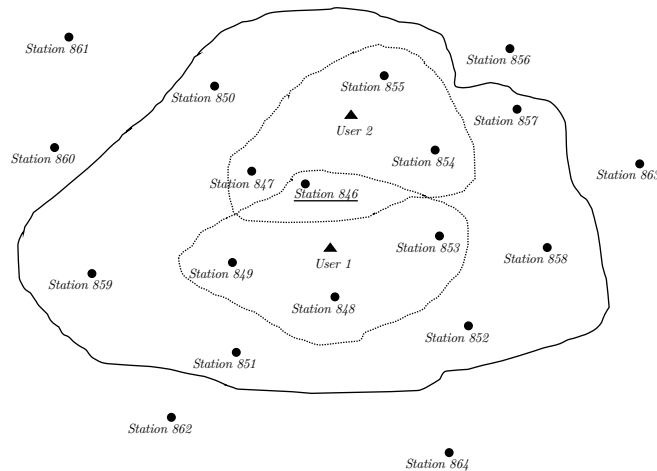

FIGURE C.2. Example for Local-stockout state Construction

C.3. **Local-stockout state.** Recall that in section 5.4, we construct the local-stockout-state for a station f by working upwards from the choice sets of each user. We limit a user i's choice set to her nearest *four* stocked-in stations; the set

of her nearest *four* stations is denoted by N_i. For a station f, the only relevant station are stations close enough to users who are close enough to station f. For any station f, we can write the set of relevant stations N_f as

$$N_f \equiv \bigcup_{i\mid f \in N_i} N_i.$$

The vector of stockout-states at time t of stations in N_f is given by a binary vector v_{ft}— it is the *local-stockout-state* for station f at time t.

Consider the example in Fig. C.2, which shows the locations of stations near our focal station 846, and two representative users to illustrate. Denote our focal station 846 by f. The four nearest stations, (N_i) to user 1 in this figure are stations $\{846, 848, 849, 853\}$. Similarly for user 2, N_i is $\{846, 847, 854, 855\}$. By taking union over these nearby stations of all users who have station f as their nearby station, we get the set N_f given by the region covered in thick line in the figure. These are stations $N_f = \{846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 857, 858, 859\}$. The local-stockout-state v_{ft} for station f is given by a binary vector of stocked-out status of stations in N_f at time t.

Appendix D. Estimation Algorithm

D.1. **GMM Estimation.** Here we provide more details of the estimation process which was described in Paper Section 5.2 and 5.4.

The set of our moment conditions the true coefficients values θ^* is supposed to satisfy are (Eq. 5.1 in paper):

$$E_{f,w_j}\left[Z_{fw_j}\sigma_{fw_j}\xi_{fw_j}(\theta^*)\right] = 0$$

We denote the expectation on the LHS above in our data set for given values of θ and corresponding error terms ξ by $G(\theta,\xi)$ (where ξ is the vector of error terms $\xi_{fw(t)}$) and is given by

$$G(\theta,\xi) = \frac{1}{N}\sum_f\sum_{w_j}Z_{fw_j}\sigma_{fw_j}\xi_{fw_j}$$
$$= \frac{1}{N}Z^T\Sigma\xi$$

Using matrix notations shall be useful later. Z is a matrix with rows Z_{fw_j}, Σ is a $N \times N$ square matrix with diagonal elements equal to σ_{fw_j} and N is $F \times 6$, which is the number of stations times number of time-windows.

The balance conditions constraint, following Berry et al. [1995], which determines the error terms $\xi(\theta)$ for given parameters values θ is given by:

$$\lambda_{fw_j}\left(\theta,\xi_{.w_j}(\theta)\right) = \Lambda_{fw_j} \quad \forall f, w_j\,; \tag{D.1}$$

Please refer Nevo [2000], Berry et al. [1995], Davis [2006] for identifiability and uniqueness of $\xi(\theta)$. Details on computation of $\lambda_{fw_j}\left(\theta,\xi_{.w_j}\right)$ by discretizing the city area is given in Appendix Section D.2.

The constraint on the total *potential* user-interest in a day is given by,

$$\int\limits_{w_j}\int\limits_{l \in \mathscr{L}} P_{w_j}\left(l;\alpha^*_{w_j}\right)\cdot minutes_{w_j}\cdot dl\cdot dw_j = T^D. \tag{D.2}$$

As discussed in the paper, the number of moment conditions plus the number of constraints in the above procedure is greater than the number of coefficients. Thus not all the elements in $G(\theta,\xi)$ can be set to 0 at our coefficient estimates $\hat\theta, \xi\left(\hat\theta\right)$. To ensure that the moment conditions are as close to being satisfied as possible, we minimize their weighted

sum of squares given by:

$$G\left(\theta,\xi\right)^{T}AG\left(\theta,\xi\right)$$

while ensuring the two set of constraints (Eq. D.1 and D.2) are satisfied. Here A is the GMM weighing matrix. We write the estimation as a Mathematical Program with Equilibrium Constraints [Dubé et al., 2012]. We also introduce a change of variable, so that the moment vector $G\left(\theta,\xi\right)$ are treated as additional parameters (η), as suggested by Dubé et al. [2012], which makes the hessian matrix of the objective function sparse.

MPEC problem. The Math Program with Equilibrium Constraints (MPEC) which is solved to obtain our estimates $\hat{\theta}$ is given by

$$\hat{\theta},\xi\left(\hat{\theta}\right),\hat{\eta} = \arg\min_{\theta,\xi,\eta}\eta^{'}A\eta \tag{D.3}$$

s.t.

$$G\left(\theta,\xi\right) = \eta$$

$$\lambda_{fw_{j}}\left(\theta,\xi._{w_{j}}\right) = \Lambda_{fw_{j}}$$

$$\int\limits_{w_{j}}\int\limits_{l\in\mathscr{L}}P_{w_{j}}\left(l;\alpha_{w_{j}}\right)\cdot minutes_{w_{j}}\cdot dl\cdot dw_{j} = T^{D}$$

where, $\xi\left(\hat{\theta}\right)$ is the value of the error terms ξ at coefficient estimate $\hat{\theta}$.

We further simplify this MPEC procedure in Section D.2 to reduce the number of parameters to be iteratively estimated.

D.2. **Solving for Estimates.**

Discretizing the city area. Recall that computation of $\lambda_{fw_{j}}\left(\theta,\xi._{w_{j}}\right)$ which is the LHS of the balance equation constraint above is the weighted average of station-use in different states $\lambda_{fw_{j}v_{f}}\left(\theta,\xi._{w_{j}}\right)$. Formally, given by (also Eq. 5.7 in paper),

$$\lambda_{fw_{j}}\left(\theta,\xi._{w_{j}}\right) = \frac{\sum_{v_{f}\in V_{f}\mid f\in S_{v_{f}}}\sigma_{fw_{j}v_{f}}\cdot\lambda_{fw_{j}v_{f}}\left(\theta,\xi._{w_{j}}\right)}{\sum_{v_{f}\in V_{f}\mid f\in S_{v_{f}}}\sigma_{fw_{j}v_{f}}}.$$

$\lambda_{fw_{j}v_{f}}\left(\theta,\xi._{w_{j}}\right)$ is in turn obtained by integrating all the user's choice probabilities (Eq. 5.6 in paper)

$$\lambda_{fw_{j}v_{f}}\left(\theta,\xi._{w_{j}}\right) = \int\limits_{l_{i}\in\mathscr{L}}p_{ifw_{j}v_{f}}\left(\beta,\gamma_{w_{j}.},\xi._{w_{j}}\right)\cdot P_{w_{j}}\left(l_{i};\alpha_{w_{j}}\right)dl_{i}.$$

To perform this integration, we divide the elements in our origin-density model ($P_{w_{j}}$) into two components, the continuous elements (intercept, $aba_{lw_{j}}$, and rd_{l}), and the discrete elements ($\vec{pi}_{l}$, ri_{l}, and vi_{l}). The integration over former continuous elements is performed numerically. We discretize the area of the city of Paris into a grid composed of squares with length $\mathscr{D}$ meters; we consider the center of each such square to be a point mass of users. Denote $\mathscr{L}^{discrete}$ as the set of locations l in the city where one of the discrete elements (rd_{l}, $\vec{pi}_{l}$, ri_{l}, and vi_{l}) is non-zero, and $\mathscr{L}^{Grid(\mathscr{D})}$ as the set of center locations of above square grids. Predicted use is then

$$\lambda_{f w_j v_f}\left(\theta, \xi_{\cdot w_j}\right) = \sum_{l_i \in \mathscr{L}^{discrete}} p_{i f w_j v_f}\left(\beta, \gamma_{w_j \cdot}, \xi_{\cdot w_j}\right) \cdot \left(\vec{\alpha}_{3 w_j} \cdot \vec{pi}_{l_i} + \alpha_{4 w_j} \cdot ri_{l_i} + \alpha_{5 w_j} \cdot vi_{l_i}\right) +$$

$$\sum_{l_k \in \mathscr{L}^{Grid(\mathscr{D})}} p_{k f w_j v_f}\left(\beta, \gamma_{w_j \cdot}, \xi_{\cdot w_j}\right) \cdot \left(\alpha_0 + \alpha_1 \cdot aba_{l_k w_j} + \alpha_{2 w_j} \cdot rd_{l_k}\right) \cdot \mathscr{D}^2,$$

where $\mathscr{D}^2$ is the area of each grid square. The value of $\mathscr{D}$ used in our main analysis is 25 meters and a robustness test in Section 9.1 considers finer discretization as well.

Reducing number of parameters. The simplest way of estimating our model would be to search over the parameters θ for values that provide the best fit. This would require a search over a space with as many dimensions as parameters (including numerous fixed-effects parameters), resulting in several search iterations each of which is computationally expensive. We instead estimate our model using a process that relies on some parameters (all except density model parameters α_{w_j} and distance coefficients β_d) entering our model in a "user-location–agnostic" way (Berry et al. [1995]). We thus group our parameters in two classes, first as $\theta_1 = (\alpha_{w_j}, \beta_d)$, and the parameters that are "linear" (in $\xi_{f w_j}$) as $\theta_2 = (\beta_0, \gamma_{w_j})$, $\forall w_j$. $\delta_{f w_j}$ denotes the composite of these "linear" parameters and error terms and is given by,

$$\delta_{f w_j} = \beta_0 + \gamma_{w_j \times di(f)} + \xi_{f w_j}.$$

Thus $\lambda_{f w_j}\left(\theta, \xi_{\cdot w_j}\right)$ could be easily re-written in as a function of θ_1 and $\delta_{\cdot w_j}$. For given values of δ, where δ is vector of $\delta_{f w_j}$, the values of coefficients θ_2 and ξ are determined so that the objective function of the MPEC in Eq. D.3 is minimized. This is given as a closed-form expression below:

$$\theta_2\left(\delta\right) = \left(\left(X_2^T \Sigma Z\right) A \left(Z^T \Sigma X_2\right)\right)^{-1} \left(\left(X_2^T \Sigma Z\right) A \left(Z^T \Sigma\right)\right) \delta. \tag{D.4}$$

where X_2 is the co-variate matrix corresponding to the equation, $\delta_{f w_j} = \beta_0 + \gamma_{w_j \times di(f)} + \xi_{f w_j}$, consisting of an intercept column and *time-window$\times$district* dummies.

Thus, we can write

$$\xi\left(\delta\right) = \delta - X_2 \theta_2\left(\delta\right) \tag{D.5}$$

Thus, in each iteration of our estimation process below, values of θ_2 and ξ are determined for given values of θ_1 and δ, and GMM objective function is computed. The re-formulation of our estimation is given by,

$$\hat{\theta}_1, \delta\left(\hat{\theta}_1\right), \hat{\eta} = \arg \min_{\theta_1, \delta, \eta} \eta' A \eta$$

s.t.

$$G\left(\left(\theta_1, \theta_2\left(\delta\right)\right), \xi\left(\delta\right)\right) = \eta$$

$$\lambda_{f w_j}\left(\left(\theta_1, \theta_2\left(\delta\right)\right), \xi_{\cdot w_j}\left(\delta\right)\right) = \Lambda_{f w_j}$$

$$\int_{w_j} \int_{l \in \mathscr{L}} P_{w_j}\left(l; \alpha_{w_j}\right) \cdot minutes_{w_j} \cdot dl \cdot dw_j = T^D$$

where $\theta_2\left(\delta\right)$ and $\xi\left(\delta\right)$ are given by Eq. D.4 and Eq. D.5. The final estimates are given by $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2(\delta(\hat{\theta}_1)))$.

GMM weighing matrix. We use a two-step optimal GMM approach to solve this. In the first step of the GMM, we use $\left(Z^T \Sigma^2 Z\right)^{-1}$ as the weighing matrix A. We find the condition number of the matrix inversion step in Eq. D.4 to be low when using this weighing matrix, in comparison to say an identity matrix which renders the conditions number quite high. This is analogous to the weighing matrix used in the 2SLS procedure. In the second step, we use $A = \left(\tilde{S}\right)^{-1}$, where $\left(\frac{1}{N} \cdot \tilde{S}\right)$ is the variance-covariance matrix of moment conditions vector $G\left(\theta, \xi\right)$ calculated at the estimated values of θ in the first step of this GMM procedure. Please refer to Appendix Section D.3 for an exact formulation of $\tilde{S}$.

Computational Details. The procedure was implemented in R. The open-source package IPOPT (Interior Point Optimizer) (interfaced with R via "ipoptr" [Ypma, 2010]) was used for nonlinear optimization with constraints. The "ffdf" class in R was employed to accommodate the large scale of our data set. Even though we transformed our problem from the time domain to the local stockout state domain, computing the choice probabilities for each user, and then summing over them, was computationally expensive; the initial runtime was of the order of tens of days on a contemporary computer of the workstation class. Implementing the station-use computation function (Eq. 5.7) in C++ and then interfacing with R reduced the computation almost 100 times, to about 50 hours for the Paris data set.

D.3. **Standard errors.** *Spatial and Temporally Clustering of Standard Errors*

To allow for spatial and temporal correlation between error terms of nearby stations and adjacent time-windows, we divide the city into 600m-square grids.[1] We allow for the error terms for stations in the same grid to be correlated across adjacent time-windows. Formally, the co-variance for any two stations f_1 and f_2 and time windows w_j and w_k,

$$cov\left(\xi_{f_1 w_j}, \xi_{f_2 w_k}\right) \neq 0 \quad \text{if grid}\left(f_1\right) = \text{grid}\left(f_2\right), \text{ and } w_j \text{ is same as or adjacent to } w_k$$

$$= 0 \quad \text{otherwise}$$

Computation of Standard Errors

Standard errors are computed using the optimal-GMM framework. The variance-covariance estimate of all the coefficient estimates $\hat{\theta}$ is given by,

$$V\left[\hat{\theta}\right] = \frac{1}{N}\left(\hat{G}^T \tilde{S}^{-1} \hat{G}\right)^{-1}$$

where, G is the vector of moment conditions, $\hat{G} = \frac{\partial G(\theta, \xi(\theta))}{\partial \theta}\big|_{\theta = \hat{\theta}}$ is the first derivative of moment conditions w.r.t. θ, and $\left(\frac{1}{N} \cdot \tilde{S}\right)$ is the variance-covariance matrix of moment conditions. $\tilde{S}$ is where the temporal and spatial correlation of error terms is incorporated. It is given by,

$$\tilde{S} = \frac{1}{N} \sum_{c \in C} \sum_{f_1 \in c} \sum_{f_2 \in c} \sum_{w_j} \sum_{w_k \mid w_j \text{ is same as or adjacent to } w_k} \left(Z_{f_1 w_j} \cdot \sigma_{f_1 w_j} \cdot cov\left(\xi_{f_1 w_j}, \xi_{f_2 w_k}\right) \cdot \sigma_{f_2 w_k} \cdot Z_{f_2 w_k}^T\right)$$

where c is one of the clusters of stations which belong to the same grid, C is the set of all clusters, f_1 and f_2 are stations within cluster c, and w_j and w_k are time-windows which are either same or adjacent to each other. For a station f, time-window w_j, $Z_{f w_j}$ is the set of instruments, $\sigma_{f w_j}$ are the number of stocked-in observations, and $\xi_{f w_j}$ are the error terms. We equate the $cov\left(\xi_{f_1 w_j}, \xi_{f_2 w_k}\right)$ to be $\xi_{f_1 w_j} \cdot \xi_{f_2 w_k}$, which is the value of their empirically observed co-variance. Thus, we allow these residuals to be correlated flexibly and don't restrict with any of the typical correlation structures (such as one lag auto-regression). Please refer to Chapter 6.3 Cameron and Trivedi, 2005 for more details.

---

[1]Results are qualitatively unchanged with a 300m-square grid specification.

	Number of Products	Number of Time periods	(Full/Limited) Availability Information	Spatial Density
Bruno and Vilcassim [2008]	24	113	Yes	No
Conlon and Mortimer [2013]	44	44,458	Yes	No
Musalem et al. [2010]	24	15	Yes	No
Davis [2006]	607	7	No	Population Density
Thomadsen [2005]	103	1	No	Population Density
Allon et al. [2011]	388	1	No	Population Density
Our Model	946	22,743	Yes	Several Demand Sources

TABLE 9. Comparison of data size

Appendix E. Estimation and Estimates Interpretation: Additional Details

E.1. **Comparison of the Computation Challenge.** As noted in Paper Section 5.3 that the computational burden can be reduced by ignoring stock-out information or by discretizing the city less finely. Bruno and Vilcassim [2008], Conlon and Mortimer [2013] have shown that the estimates could be substantially biased because of not taking into account availability information (as is the case in BLP and most applications of it). A fine grained user density model is also necessary in our context because the usage of bike-share systems tends to be quite local in nature, stations can be as close as 50 meters, and the number of potential users can substantially change within a matter of 50 meters.

Since the past works have typically not incorporated either stock-outs or spatial information, and use much smaller scales of data; they typically have 6 orders of magnitude fewer computations than us. We next compare the number of computations in some of the representative past studies with that of our setup.

Comparison with work that has accounted for availability information (None in spatial context). Bruno and Vilcassim [2008] have access to only average product availability information. Assuming independent availabilities, Bruno and Vilcassim [2008] extend the BLP model to account for them. In presence of exact availability information, their model resembles our model in time-domain. Bruno and Vilcassim [2008] consider 24 products (as compared to 946 in our case) for 113 four week periods (as compared to over 22,000 in our case). Conlon and Mortimer [2013] consider 44 products in a vending machine application in 44,458 four-hour time periods. Musalem et al. [2010] consider 24 products for 15 days of data. These papers have users which were not spatially differentiated. There is heterogeneity in user tastes due to normally distributed random coefficients, however the number of draws required to aggregate over these heterogeneous users is much lower. For example, the supplementary code in Nevo [2000] uses 20 draws, and Dubé et al. [2012] use 1000 draws as compared to the more than 210,000 spatially heterogenous users in our case.

Comparison with work in spatial context (None account for availability). On the other hand, models that have accounted for spatially different users (Davis [2006], Thomadsen [2005], Allon et al. [2011]) have not accounted for product availabilities. Davis [2006] considers daily data for 607 theaters for a period of 7 days; Thomadsen [2005] considers 103 fast food locations in Santa Clara county with a single observation per location; Allon et al. [2011] considers 388 fast food outlets in Cook County with one observation per location. Note that in absence of sales data, Thomadsen [2005] and Allon et al. [2011] estimate parameters based on observed prices and other outlet characteristics.

The comparison with past work is summarized in Table 9. The comparison illustrates how the combination of rapidly changing choice sets and a fine grained user density model, in a relatively large scale data set, leads to an explosion in computational requirements.

E.2. **Comparison of Estimates.** We compare our estimates with estimates (or in some cases decisions implied by those estimates) from other studies in the bike-share, public-transport, and retail-store network design contexts.

Comparison of Distance Estimates. Bike-Share Systems: We are aware of no other econometric study (survey or archival-data based) that has attempted to estimate the disutility of distance in the context of bike-share systems. In practice, European bike-share system designers follow a common handbook (Büttner and Petersen, 2011) which suggests that very few users walk further than 300 meters and provides station location guidelines based on this assumption. This guideline is squarely in line with our estimates, our distance estimate also implies that only 21.94% of the use of a station comes from users who walk further than 300 meters.

Other Public Transport Systems: Several studies survey the walking distance of users of bus, light rail and metro systems. [O'Neill et al., 1992, Zhao et al., 2003] report that 75-80% users of public-transit systems walk less than 400 meters. El-Geneidy et al. [2014] finds that, in Montreal, the 85th percentile walking distance to the bus (resp., rail) transit system is about 524m (resp., 1,259 m). O'Sullivan and Morrall [1996] reports that transit planners in several Canadian and American cities consider the catchment area to be no further than 300-900m, with the median light rail user in Calgary, Canada walking 320m. Alshalalfah and Shalaby [2007] report that median access distance of bus users in Toronto is about 200m and that of subway users is 350m. In comparison, our estimates are marginally lower, our median user walks about 220m, and almost 80% of the demand comes from the first 300m. Since bike-share systems are used for much shorter trips than those taken by other public transport systems, bike-share systems exist in more densely populated areas, and have a much denser station network; it is expected that our users walk less.

Zhao et al. [2003], based on survey data of about one thousand users' transit use (bus or rail) in southeast Florida, determines that usage decreased exponentially with a coefficient of *-4.265/km.* Gutiérrez et al. [2011] uses survey data from the Madrid metro network to estimate the effect of walking distance and finds an average distance disutility coefficient of about *-1.689/km.* Our comparable estimate, *-2.2/km,* is squarely in line with these observed coefficients.

Retail Store Networks: While the context of retail store networks provides us with multiple past studies to compare our estimates, these estimates typically consider the disutility of distance for users who *drive* to the retail locations. One way to compare with these estimates is to convert distances to commuting time using average walking and driving speeds. We compare our estimates with those in Davis [2006] (driving to movie theaters), Pancras et al. [2012] (driving to grocery stores), and [Thomadsen, 2005, Allon et al., 2011] (driving to drive-through fast food outlets). Our estimate is higher than that of Davis [2006] and Pancras et al. [2012], comparable to those in Allon et al. [2011] and lower than that in Thomadsen [2005]. Together, our estimate is again squarely in line with these past studies. The average disutility of commuting time from these studies is *14.497/hr* while, for majority of the users in our study, it is *8.9116/hr.*[2]

Comparison of Bike-Availability Estimates. We are not aware of any study that has looked at the impact of availability in the context of bike-share systems. Availability in bike-share systems is not directly comparable with that of other public-transportation systems where it concerns the frequency or reliability of a service. The only somewhat comparable estimates are from the long-term and short-term effects of product availability in the context of consumer goods. Note however that demand for customer goods is much less time-sensitive than that for transportation, products are not modeled as spatially differentiated, and as such availability is expected to play a much smaller role.

Anderson et al. [2006] in their study of a home-bedding catalogue retailer find that a 10% decrease in stockouts leads to a 7.2% short-term increase in product sales. The lost-demand in their case is much lower than ours (28% in their case compared to 95% in our case) probably because users are more willing to wait for bedding ordered via a catalogue, than

---

[2]We assume dense city-driving speeds of *25 km/hr* and walking speeds of *4 km/hr.*

for bikes to get somewhere. The long-term impact of all items ordered by a customer in their setting being out of stock compared to none is 22% lower future demand, i.e. a 10% decrease in stockouts leads to a 2.2% long-term increase in product demand. This estimate is comparable to our estimate of a 2.482% increase in long-term demand due to a 10% higher bike-availability.

Musalem et al. [2010] in their study on estimating the effect of stockouts in the shampoo category find that almost no sales are lost when a few brands stock out, but as much as 20.02% of sales might be lost when multiple brands stockout, suggesting there is more than 80% substitution to adjacent shampoo brands. Not surprisingly, there is much less substitution to adjacent stations in our context (only about 5%), perhaps because users must walk to other stations rather than just simply switch to comparable, adjacent brands.

E.3. **Computation of short-term effect of Bike-Availability.** To compute the short-term effect of bike-availability we simulate the system-use for a scenario where the odds of finding bikes are increased by 10%. Only the stock-out status of stations is changed while the bike-availability in the origin-density model is kept the same. That is, this simulation assumes that there are no changes in origin-density, the only changes in use arise from changes in user-choices (specifically from lower abandonment). The effect of change in system-use due to change in bike-availability in the origin-density is captured in the long-term effect of bike-availability instead.

The simulation proceeds as follows. $\underline{a\tilde{b}a}_{fw_j}$ denotes the average bike-availabilities of *station* × *time-window*'s which are increased by 10%.[3] For a simulation time-period $t|w(t) = w_j$, $\underline{ba}_{ft}$, denotes the stocked-in status at station f at time t. It's value is drawn from a Bernoulli distribution $Bern(\underline{a\tilde{b}a}_{fw_j})$. This process determines which of the station are stocked-in. $\tilde{\lambda}_{ft}\left(\hat{\theta}, \hat{\xi}_{\cdot w(t)}\right)$ is computed for each station using the optimal estimates $\hat{\theta}$, $\hat{\xi}_{\cdot w(t)}$, while taking into account the simulated station stocked-in status ($\underline{ba}_{ft}$ draws) to form user choice-sets.

We perform 1000 simulations in each of the time-windows. The percentage difference in total station-use across all stations and time-windows with and without any increase in bike-availabilities determines the short-term effect of bike-availability. The confidence interval around the short-term effect is calculated by performing 1000 different set of above simulations for alternative values of estimates $\hat{\theta}$, $\hat{\xi}_{\cdot w(t)}$ drawn from a normal distribution of their point estimates and estimated variances.

References

G. Allon, A. Federgruen, and M. Pierson. How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Operations Management*, 13(4):489–507, 2011.

B. W. Alshalalfah and A. S. Shalaby. Case study: Relationship of walk access distance to transit with service, travel, and personal characteristics. *Journal of urban planning and development*, 133, no. 2(114-118), 2007.

E. T. Anderson, G. J. Fitzsimons, and D. Simester. Measuring and mitigating the costs of stockouts. *Management Science*, 52(11):1751–1763, 2006.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, pages 841–890, 1995.

H. A. Bruno and N. J. Vilcassim. Research note-structural demand estimation with varying product availability. *Marketing Science*, 27(6):1126–1131, 2008.

J. Büttner and T. Petersen. Optimising bike sharing in european cities-a handbook. 2011.

A. C. Cameron and P. K. Trivedi. Microeconometrics: Methods and applications. *Cambridge university press*, 2005.

---

[3]They are set to 1 if 10% increase leads them to be higher than 1.

C. T. Conlon and J. H. Mortimer. Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30, 2013.

P. Davis. Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics*, 37(4):964–982, 2006.

J.-P. Dubé, J. T. Fox, and C.-L. Su. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267, 2012.

A. El-Geneidy, M. Grimsrud, R. Wasfi, P. Tétreault, and J. Surprenant-Legault. New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, 41(1):193–210, 2014.

J. Gutiérrez, O. D. Cardozo, and J. C. García-Palomares. Transit ridership forecasting at station level: An approach based on distance-decay weighted regression. *Journal of Transport Geography*, 19(6)(1081-1092), 2011.

A. Musalem, M. Olivares, E. T. Bradlow, C. Terwiesch, and D. Corsten. Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197, 2010.

A. Nevo. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548, 2000.

W. A. O'Neill, R. D. Ramsey, and J. Chou. Analysis of transit service areas using geographic information systems. *Transportation Research Record*, 1364, 1992.

S. O'Sullivan and J. Morrall. Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*, 1538:19–26, 1996.

J. Pancras, S. Sriram, and V. Kumar. Empirical investigation of retail expansion and cannibalization in a dynamic environment. *Management Science*, 2012.

R. Thomadsen. The effect of ownership structure on prices in geographically differentiated industries. *RAND Journal of Economics*, pages 908–929, 2005.

J. Ypma. Introduction to ipoptr: an R interface to Ipopt. 2010.

F. Zhao, L.-F. Chow, M.-T. Li, I. Ubaka, and A. Gan. Forecasting transit walk accessibility: Regression model alternative to buffer method. *Transportation Research Record*, 1835:34–41, 2003.