INSEAD
The Business School for the World®

Working Paper Series
2015/08/TOM

Bike-Share Systems: Accessibility and Availability

Ashish Kabra
INSEAD, ashish.kabra@insead.edu

Elena Belavina
The University of Chicago Booth School of Business,
elena.belavina@chicagobooth.edu

Karan Girotra
INSEAD, karan.girotra@insead.edu

January 26, 2015

The cities of Paris, London, Chicago, and New York (among others) have recently launched large-scale bike-share systems to facilitate the use of bicycles for urban commuting. This paper estimates the relationship between aspects of bike-share system design and ridership. Specifically, we estimate the effects on ridership of station accessibility (how far the commuter must walk to reach a station) and of bike-availability (the likelihood of finding a bike at the station). Our analysis is based on a structural demand model that considers the random-utility maximizing choices of spatially distributed commuters, and it is estimated using high-frequency system-use data from the bike-share system in Paris. The role of station accessibility is identified using cross-sectional variation in station location and high -frequency changes in commuter choice sets; bike-availability effects are identified using longitudinal variation. Because the scale of our data, (in particular the high-frequency changes in choice sets) render traditional numerical estimation techniques infeasible, we develop a novel transformation of our estimation problem: from the time domain to the "station stockout state" domain. We find that a 10% reduction in distance traveled to access bike-share stations (about 13 meters) can increase system-use by 6.7% and that a 10% increase in bike-availability can increase system-use by nearly 12%. Finally, we use our estimates to develop a calibrated counterfactual simulation demonstrating that the bike-share system in central Paris would have 29.41% more ridership if its station network design had incorporated our estimates of commuter preferences—with no additional spending on bikes or docking points.

BIKE-SHARE SYSTEMS: ACCESSIBILITY AND AVAILABILITY

Abstract. The cities of Paris, London, Chicago, and New York (among others) have recently launched large-scale bike-share systems to facilitate the use of bicycles for urban commuting. This paper estimates the relationship between aspects of bike-share system design and ridership. Specifically, we estimate the effects on ridership of station accessibility (how far the commuter must walk to reach a station) and of bike-availability (the likelihood of finding a bike at the station). Our analysis is based on a structural demand model that considers the random-utility maximizing choices of spatially distributed commuters, and it is estimated using high-frequency system-use data from the bike-share system in Paris. The role of station accessibility is identified using cross-sectional variation in station location and high -frequency changes in commuter choice sets; bike-availability effects are identified using longitudinal variation. Because the scale of our data, (in particular the high-frequency changes in choice sets) render traditional numerical estimation techniques infeasible, we develop a novel transformation of our estimation problem: from the time domain to the "station stockout state" domain. We find that a 10% reduction in distance traveled to access bike-share stations (about 13 meters) can increase system-use by 6.7% and that a 10% increase in bike-availability can increase system-use by nearly 12%. Finally, we use our estimates to develop a calibrated counterfactual simulation demonstrating that the bike-share system in central Paris would have 29.41% more ridership if its station network design had incorporated our estimates of commuter preferences—with no additional spending on bikes or docking points.

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. Only three of the 74 Chinese cities monitored by the central government managed to meet official minimum standards for air quality in 2013 [Wong, 2014]. In March 2014, levels of suspended particulate matter in the air above Paris reached twice the permissible level, which led to driving restrictions [Rubin, 2014]. Likewise, many large US cities are in a state of "non-attainment" with respect to their clean air requirements.[1] Passenger vehicles are the major culprit in each case: 45% of air pollution in European cities can be directly attributed to private passenger vehicles, and that figure reaches 80% for some Asian cities.[2]

Traffic congestion worsens air quality and is a scourge in its own right. An average US commuter loses 34 hours and $750 annually to traffic congestion; commuters in Washington DC, Los Angeles, San Francisco, and Boston lose twice as much time and money.[3] An average resident of Paris loses €2,883 each year because of traffic congestion, costing the French economy some €17 billion

[1] "Green Book", US Environmental Protection Agency, 2 July 2014, http://bit.ly/EPAGreen
[2] "Changing Gears: Green Transport for Cities", World Bank and Asian Development Bank Report, 2012.
[3] "Urban Mobility Report", Texas A&M Transportation Report, 2012.

annually [Negroni, 2014]. Emerging market mega-cities (Bangkok, Manila, Kuala Lumpur, Delhi, Mumbai, and Ho-Chi Minh City, inter alia) routinely break traffic congestion records, and Asian economies lose from 2% to 5% of their annual GDP to traffic congestion.

The use of bicycles helps to alleviate both traffic congestion and poor air quality. Bicycles can substitute for polluting vehicles on short trips, and they facilitate the use of environmentally efficient public transport for long trips by providing effective "last mile" connectivity. The use of bicycles also reduces road congestion: compared with a typical passenger vehicle, which occupies 115 m^3 of road space, a bicycle makes do with only about 6 m^3 [Rosenthal, 2011]. However, the adoption of bicycles by commuters remains low in most major cities. The main barriers are the lack of safe parking spaces for bikes in urban dwellings and at public transit hubs, vandalism and theft of bikes, and the inconvenience and cost of owning and maintaining a bike. Bike-share systems address each of these concerns.[4]

A typical bike-sharing system includes a communal stock of sturdy, low-maintenance bikes distributed over a network of parking stations. Each station provides 10–100 automated parking spots, or *docking points*, and a networked controller interface. A registered commuter can "check out" any available bike from a station and, at the end of her commute, can return the bike to any station in the network. Registration often requires the commuter to pay a security deposit. Usually the first half hour of use is free of charge and subsequent intervals are progressively more expensive.

From an individual commuter's point of view, bike-share systems eliminate the inconvenience of bike ownership, the need to find parking places, and the fear of theft and vandalism. Moreover, being able to take a bike from one station and drop it off at another facilitates one-way trips and the use of different modes on round trips. A crucial system feature is that bicycles can be used as an effective last-mile feeder system to other public transit systems, such as metro rail or bus systems.

While bike-sharing systems have existed since the 1950s, there has been renewed interest since the successful implementations in France in 2006. As of April 2013, these systems had spread across Europe, the United States, and Asia—there were more than 530 bike-sharing systems in operation around the world with a total fleet of about 517,000 bikes. Paris, Barcelona, London, Wuhan, Hangzhou, Shanghai, New York City, and Chicago have all implemented large-scale systems.[5]

Although bike-sharing systems have garnered considerable attention, their promise is far from being fully realized. Despite widespread enthusiasm among citizens, ridership in some systems has fallen short of projections and there is increasing pressure on operator finances. More importantly,

[4]In the words of Chicago transportation commissioner Gabe Klein: "There is no holy grail, but a public bike share is pretty close." http://bit.ly/1pTaagl

[5]Janet Larsen, "Bike-Sharing Programs Hit the Streets in Over 500 Cities Worldwide", *Earth Policy Institute*, 25 April 2013; and Wikipedia entry on the "bicycle sharing system".

current ridership levels are well short of meeting the challenge of transforming urban transportation. A key reason for the lacking ridership is that while providers and operators have focused on bike-design and technology aspects, there is almost no rigorous analysis of operational aspects such as station location, system-reliability, nor are the commuter responses to such aspects understood [Tangel, 2014]. The aim of this paper is to identify relationships between ridership and design aspects 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 on ridership of two factors: station accessibility, or how far a commuter must walk to reach a station; and bike-availability, or the likelihood of finding a bike at the station. There are, in turn, two aspects of bike-availability. The immediate one is that commuters must walk longer (or use other means of transport) if nearby stations don't have bikes. The more subtle, long-term aspect of availability is that—to the extent that the system is less reliable in this regard—commuters are less likely to incorporate bike-sharing into their daily commute or to make long-term commitments (e.g., forgoing their cars, choosing to live in an urban area).

We conceptualize commuter behavior in bike-share systems as a choice between differentiated products. Thus, each commuter is viewed as a consumer, each station is a different product, the distance to a station and its historic bike-availability are product characteristics, and the set of stations with available bikes is the consumer's choice set. Our parameters of interest are commuter preferences for distance and for historic bike-availability. Note here that the first product characteristic (distance to the station) is a characteristic that is both station and commuter specific. A reduced-form, station-level model would need to consider a representative commuter, and in such a model the effect of distance is confounded with the effect of the station's "catchment" area; that is, stations far away from other stations require that commuters walk farther but also have a larger catchment area. An alternate approach would be to build a model at the commuter–station level, but that would require observing the start point of every commuter—data that are available neither to us nor to any bike-share operator. To avoid the pitfalls of either approach, we develop a structural demand model based on a random utility maximization framework (as in [Berry et al., 1995], "BLP" henceforth) that uses only station-level data yet recovers the effect of distance on commuter choice.

Our model considers a population of potential commuters distributed randomly across the operation area with a parametric spatial density. Each individual commuter chooses between different stations and an outside option (i.e., some other means of transport). That choice is based on the commuter's propensity to bike, the commuter's distance from the different stations, and the historic bike-availability at those stations. Next, we aggregate individual commuter decisions to derive the

number of originating trips from different stations, using the spatial density of commuters. This random commuter origin in our model is akin to the random coefficients used in the traditional BLP approach.

Commuter preferences for historic bike-availability are estimated using longitudinal variation in that availability. Preferences for distance are estimated using cross-sectional variation, among stations, in the distance that commuters must travel to access them. As stations run out of bikes and get replenished, the commuters' choice sets change; these changes provide us with another level of variation—in product offering—in addition to variation that arises from the spatial segregation of commuters and stations. In the same way as changing choice sets improve the efficiency of identifying random coefficients in the traditional BLP method, longitudinal variation in the stockout state of stations (and the resulting changes in choice sets) helps us efficiently estimate our parameters in the presence of unobserved commuter heterogeneity in their origin location. This estimation directly reveals the long-term effect of bike-availability (through the historic bike-availability covariate), and the short-term effect of bike-availability can be determined by comparing system-use among different choice sets.

We estimate our model using data from the Vélib' bike-share system in Paris. Our data is based on observing, every two minutes, 349 bike stations in central Paris for a period of four months. There are more than 22 million such observations (or data "snapshots"), which correspond to more than 2.5 million bike trips. As a result, our data is orders of magnitude larger than the data typically used in structural demand models. These high-frequency data can make our estimates of commuter preferences precise; but they also lead to high-frequency changes in the choice sets, which renders the usual numerical approaches to estimating structural demand models computationally infeasible.

To deal with this computational challenge, we develop a novel transformation of the data. We notice that, in the context of our model, the (two-) minute-to-minute variation in use at the station level is affected only by the contemporaneous variation in the choice sets available to commuters and in some fixed effects (i.e., month, time of day). Hence we can transform our data from the time domain to the domain of available choice sets, aggregating different times with the same choice sets (and same fixed effects) into a single data point located in what we call the *stockout state domain*. However, the space of systemwide stockout states is still too large to constitute enough of an improvement over the time domain; after all, in theory that space contains as many as $2^{\#\text{ stations}}$ elements. But since choices by each commuter are made only between nearby stations, we can further improve our setup simply by creating local stockout states for each station and then aggregating data points with common local stockout states (and same fixed effects). This drastically reduces our computational load, though it requires that we carefully account for consistent local

stockout states at neighboring stations. Altogether, this transformation of the data allows us to obtain precise estimates from large spans of data that exploits the high-frequency variation in choice sets for identification while managing the computational burden.

Our estimates imply that a 10% *decrease* in distance (about 13 m) traveled to access bike-share stations can increase system-use by 6.7% (61,140 additional trips each month). A 10% *increase* in bike-availability can increase system-use by about 12% (109,530 more trips/month). We also find that only 4.4% of the demand substitutes to nearby stations when confronted with a stockout at the station of choice; this low figure is consistent with the significant positive effects of reducing distance to stations and improving bike-availability.

These estimates can be used to improve the performance of extant systems by enabling system managers to estimate and trade off the social and financial benefits of increased ridership with the costs of system improvements that reduce distances (e.g., by adding extra stations) or increase bike-availability (adding bikes to the system, increasing trans-shipment of bikes from one station to another, etc.). At the same time, such estimates can also serve as key inputs for the design of new systems.

We illustrate a use of our estimates by providing a counterfactual study of alternate station network designs. Specifically, we use our estimates to calibrate a simulation that predicts the ridership of different station network designs that incorporate the same number of bikes and docking points but place different priorities on the competing demands of station accessibility versus bike-availability. Whereas a network with more stations but fewer bikes at each one reduces distances to station and so increases accessibility, fewer stations with more bikes at each can achieve higher bike-availability owing to the statistical benefits of holding pooled bike inventory. Hence knowledge of how commuters value these two aspects of system design can lead system designers to make optimal trade-offs between them. We identify the optimal design and find that the bike-share system in central Paris would have 29.41% more ridership (268,440 more trips/month) if it had incorporated our estimates of commuter preferences in the design of its station network. That improvement does not require spending any additional resources on bikes or docking points, which are the main costs to public agencies that institute these systems.

Overall, the large effect of our estimates highlights the considerable improvement opportunities in bike-share systems that could result from greater use of data. More generally, our work illustrates the substantial impact of facility location and inventory availability—two objectives of concern to operations management research since the field's inception—in the context of new models for urban transport.

Although the results reported here are developed in the context of bike-sharing systems, it is likely that our estimated effects of accessibility/distance on consumer utility apply in many

contexts; examples include other public transit systems, retail stores, and consumer services (banks, dry cleaning, hotels, etc.). Our estimates concerning the role of bike-availability could serve as benchmark availability metrics for the design of other on-demand transportation systems, such as as taxi-hailing apps (e.g., Uber, Lyft, EasyTaxi). The accessibility–availability trade-off in station design is hardly unique to bike-sharing systems. Designing networks of retail stores, hotel chains, car-share stations, and other location-based services involves similar trade-offs. In such contexts, "availability" may correspond to product availability, assortment, or quality of service; there are also similar accessibility concerns. Our analysis can inform choices in each of these contexts.

Our study makes three important contributions. We provide the first empirical analysis of commuter response to accessibility (walking distance) and availability (service level) in the context of public transport systems. As we illustrate, this analysis can help design much-improved transport systems. Second, the methodology developed in this paper can be used in a variety of demand estimation contexts where products are spatially differentiated with a high frequency of changing choice-sets. Finally, our research extends the thriving empirical operations management research on estimating the role that service levels play in demand, by adding previously unstudied aspects—availability and accessibility in the context of bike-share systems—to the elements already addressed in the literature (waiting time, product variety, queue length, warranties, etc.). Further, to the best of our knowledge, this is the first study to consider these connected notions jointly.

2. Literature Review

This paper is related to research on bike-sharing systems and on customer response to accessibility and availability. Our work is also related to the operations management tradition of considering spatial issues (facility location, vehicle routing, etc.). Finally, our techniques are inspired by demand estimation models from empirical industrial organization.

2.1. **Bike-Sharing Systems.**

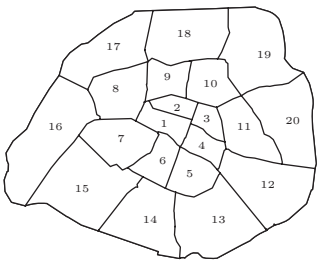
There is an emerging literature on bike-sharing systems in different areas of study. One stream of this literature uses mixed-integer programming and numerical calibration methods to examine policies for managing the trans-shipment of bikes in a bike-share system [Nair and Miller-Hooks, 2011, Nair et al., 2013, Raviv et al., 2013]. George and Xia [2011] use their approximation of a closed queuing network to determine the optimal number of bikes in a system. Daddio [2012] uses data from the Washington DC bike-share system to suggest that demand at a station can be predicted by the demographic characteristics (population, race, number of retail locations) and geographic features (distance to metro/rail) of a station's catchment area. Lathia et al. [2012] examine the nature of increase in bike-share utilization, after the London bike-sharing system was opened to casual users. O'Brien et al. [2013] classify bike-sharing systems based on different intra- and interday usage patterns.

2.2. **Accessibility.** The notion of accessibility has appeared in previous research on transportation systems design. Murray and Wu [2003] formulate an integer linear program for transit-stop location that trades off the distance to transit stops against the inconvenience of transit vehicles that stop too often. They apply this model to real geographic data from Columbus, Ohio; yet absent information about commuter preferences for accessibility or transit time, the model cannot provide recommendations. El-Geneidy et al. [2014] survey commuters on their walking distance to different transit systems (rail, bus) and find that, in Montreal, the 85th percentile walking distance to the bus (resp., rail) transit system is about 524 meters (resp., 1,259 m). Accessibility has also been studied in the context of retail network design; approaches include the use of gravity models (Reilly [1931], Huff [1964]) and discrete choice models of consumer preferences (Craig et al. [1984], Fotheringham [1991], Davis [2006], Pancras et al. [2012]). Unlike our study, these analyses consider only the effect of commuting distance and do not consider related service concerns (e.g., product/bike availability).

2.3. **Availability and Service Level.** Our notion of bike-availability corresponds closely to the well-studied concept of service level in operations management. In the context of retail consumer goods, Musalem et al. [2010] study the effect of product stockouts and Olivares et al. [2011] the effect of waiting times at grocery stores. Anderson et al. [2006] use field experiments to look at the short- and long-term impact of stockouts in the context of a mail-order catalog service. In the fast-food sector, Allon et al. [2011] study the waiting time at drive-through locations and evaluate its effect on demand. In the automobile industry, Guajardo et al. [2014] conceptualize service level as warranty length and quality of after-sales service; these authors examine the effects of service level so defined, and Moreno and Terwiesch [2013] evaluate the impact of product variety. For the retail banking industry, Buell et al. [2014] look at the customer response to service levels and discuss the consequences for competition. Parker et al. [2013] consider the quality of information services and the effect of that quality on marketplaces. Most close to our context, in the transportation sector, is the paper of Arikan et al. [2013], who study the effect of increasing airport capacity on flight delay propagation—reduced service levels to consumers.

To the best of our knowledge, there is no extant research that seeks to identify the drivers of commuter behavior in bike-share systems. Neither are we aware of previous work that simultaneously addresses the notions of availability and accessibility in this (or any other) context.

2.4. **Facility Location and Spatial Models in Operations.** The vast majority of research on facility location provides analytical methods for computing cost-reducing designs given consumer preferences [Melo et al., 2009]; in contrast, our work seeks to estimate those consumer preferences. Some recent work has tackled environmental issues. For example, Cachon [2014] considers network



(a) Paris Arrondissements

Number of Snapshots	Raw Data	22,542,770
	Removing Weekends	15,990,831
	Removing Trans-shipments	15,850,704
Number of Trips	Raw Data	2,765,933
	Removing Weekends	1,954,779
	Removing Trans-shipments	1,906,269

(b) Number of Observations

FIGURE 3.1. Vélib': Data Description

design decisions made by a retail store and analyzes the trade-offs among operating costs, availability, and carbon emissions; Belavina et al. [2014] study the role of city geography and of online grocery shopping on emissions stemming from travel and food waste. More closely related to our research are the papers by Li et al. [2014] and Lederman et al. [2014], which develop data-driven approaches to identifying competitors in spatially differentiated markets.

2.5. **Demand Estimation.** Our estimation technique builds on the seminal work of Berry et al. [1995], where consumer preferences are estimated using only market share data at the product level. Whereas products in BLP are differentiated in terms of their physical attributes, our "products" are stations that are differentiated in terms of their spatial location and their bike-availability. Davis [2006] applies the BLP approach to the spatial differentiation of cinema chains (much as in our setting), but his study does not address issues involving service level or stockouts; recall that the latter alters user choice sets in real time. Pancras et al. [2012], in addition to geographic differentiation as in Davis, also models goodwill dynamics over time. Yet to the best of our knowledge, our paper is the first using empirical methods to capture spatial product differentiation with rapidly changing choices sets. The rapidly changing choice sets necessitates the development of a new estimation procedure.

3. Data Description

We estimate our model using data from the Vélib' bike-share system in Paris. Of all the major systems, Vélib' has the most bikes per capita: about 10 times more than London's bike-share system and 100 times more than the system in lower Manhattan.[6] Vélib' has more than 1,200 stations with some 17,000 bikes on which nearly 173 million trips were made during the system's first six years of existence. The environmental impact is estimated to be a reduction of more than 137,000 tonnes of CO_2 emission equivalents; the effect on commuter health is also significant in that some 19 billion calories were burned by bike riders during this period.

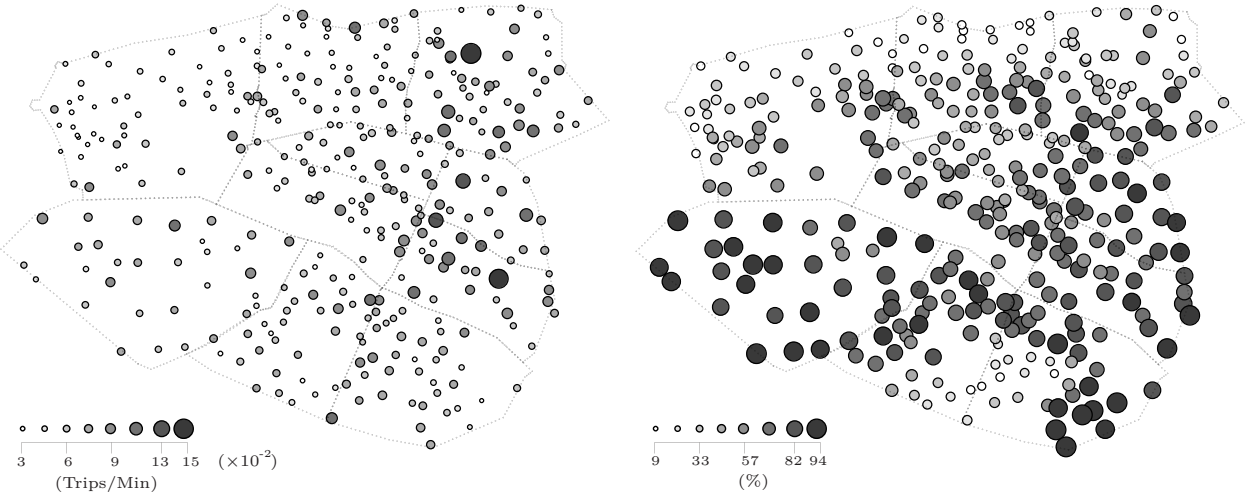
[6]"Paris fête les six ans de son Vélib' (en infographie)", Mes Débats, 15 July 2013, http://bit.ly/14Cn6n6

Our data set is built by capturing the status of each station in the network every two minutes, via the programing interfaces available for Vélib'.[7] Each two-minute observation that we collect contains the number of available bikes and the number of empty docking points at all stations. The city of Paris proper (*Paris intra-muros*) is divided into 20 *arrondissements* or districts that are numbered in a spiral pattern; see Figure 3.1(a). Of these, we restrict our attention to the 10 central districts (1st–10th arrondissements). This area, which covers about 9 square miles (23 km^2), is the most densely populated part of Paris and includes a mix of residential, retail, commercial, public, and historical establishments. And since these districts constitute a contiguous inner core of the city, all adjacent areas are also in the city of Paris and so have equal access to Vélib'. We thus reduce the "edge effects" that might arise from using any of the outer districts (i.e., of the 11th–20th).

We monitor these stations over a four-month period starting in May 2013; however, the first month's observations are used only to establish historic levels of bike-availability for subsequent months. The Paris bike-share system had been in operation for more than six years at the time of data collection. It is reasonable to assume that, at this stage, the system is likely in a steady state: we expect there to be few perturbations due to changes in the station network, increasing awareness of the system, or change in system management policies. The months of May, June, July, and August have similar average temperatures and precipitation levels, reducing spurious variation caused by changing weather. Furthermore, these months are also the most favorable ones for use of the system and so are the months in which it experiences considerable variation in bike-availability—which facilitates our estimating the effects of that variation. We eliminate data snapshots collected on weekends because commuter preferences on those two days may differ substantially from their weekday preferences. Note also that commuter patterns on weekdays (and in summer months) are the most important because it is at such times that increased bike use has the maximum possible benefit in terms of reducing traffic congestion and improving air quality. For these reasons, the city should design its system so as to maximize its use during that period. Altogether, our data set includes 22 million snapshots of 349 bike stations (Figure 3.1(b)).

Next we convert these snapshots into our variables of interest. We assume that each decrement in a station's number of available bikes (at the two-minute observation interval) is an instance of a bike being checked out and used. One could argue that a declining number of bikes merely signifies 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 are extremely small (observed rates of these activities indicate that such simultaneity occurs at

[7]Oliver O'Brien, "Bike Share Map", 31 August 2013, http://bikes.oobrien.com/paris/



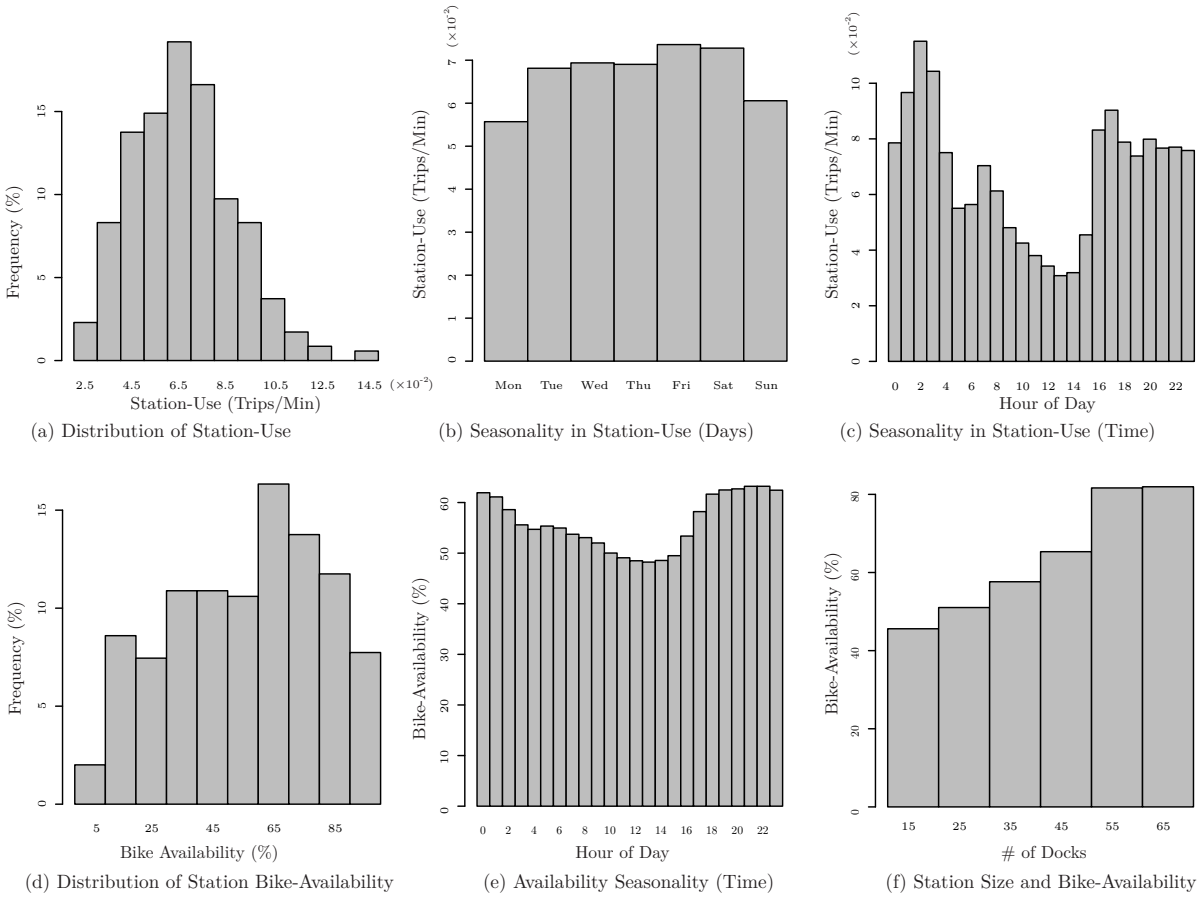
(a) Average Station-Use (b) Average Bike-Availability

FIGURE 3.2. Vélib' Stations: Usage and Bike Availability

a frequency of less than 1%). Therefore, any errors that result from our making this particular assumption will almost certainly be insignificant.

The 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 three bikes are checked out, which we interpret either as trans-shipment by system managers or as outliers in the usage. That scenario rarely occurs, so even this conservative elimination allows us to retain over 95% of the data. Thus we construct our main dependent variable, *station-level use in a given two-minute interval*. Results of our analysis are unchanged when other thresholds are used for eliminating outliers.

When not stocked out, a typical station in central Paris is the starting point for 4.05 rides/hour; this rate can increase by a factor of 15 during the peak late-night hours. Figure 3.2(a) shows the mean use by station; here and henceforth, by "station-use" we mean *the number of trips that originate at a station in a unit time*—conditional on bikes being available at that station. In the figure, bubble sizes correspond to the level of station-use. We observe that stations in different districts have systematically different levels of use, which suggests the need for controls at the district level. Also, stations with fewer neighboring stations generally have higher use. A naive interpretation (as would arise, e.g., from estimating a station-level model) of this result is that the increasing distance to stations increases use; that is, commuters prefer stations that are farther away! Such a conclusion is false because station-level use reflects not only commuter preferences for different station characteristics (where use is decreasing in distance) but also the size of a station's catchment area (where use is increasing in distance). Figure 3.3(a)-(c) shows the distribution of station-use as well as inter- and intraday patterns of station-use.



(a) Distribution of Station-Use (b) Seasonality in Station-Use (Days) (c) Seasonality in Station-Use (Time)

(d) Distribution of Station Bike-Availability (e) Availability Seasonality (Time) (f) Station Size and Bike-Availability

FIGURE 3.3. Station-Use and Bike-Availability Statistics

Our study has three main independent variables, the first of which is the distance that a commuter must walk to reach different stations. We have the GPS coordinates of each station, which allows us to compute the distance ("as the crow flies") to a station from any point in the city. For a dense city like Paris, this approach yields a close approximation of the walking distance. We also obtain the precise boundaries of different districts via Keyhole Markup Language files (available from Google Maps). These boundaries allow us to allocate each point to a particular district.

The other two main independent variables—the set of stations available to a commuter (the choice set) and the historic bike-availability at each station (the service level)—both derive from the state of a station: namely, whether or not there are any usable bikes available at the station. Although we observe the number of bikes available at the station every two minutes, some of these bikes are not actually usable. First, bikes in these systems are regularly removed from service after a certain number of trips for purposes of preventive maintenance; we have data on these bikes and exclude them from our analysis. Second, some bikes are officially in circulation but are in an undesirable state (e.g., bikes with a broken chain or with bird droppings on its seat). Most stations

Variables		Mean	Min	1 Qu.	Median	3 Qu.	Max	Standard Deviation		
								Overall	Between	Within
Station-Use	per min	0.067	0.000	0.018	0.053	0.099	1.000	0.054	0.018	0.051
Distance to nearest station	kms	0.166	0.007	0.118	0.159	0.206	0.479	0.075		
Av. Distance to 2 nearest stations	kms	0.196	0.062	0.146	0.185	0.233	0.493	0.076		
Bike-Availability (#bikes>0)	fraction	0.894	0.000	0.868	1.000	1.000	1.000	0.204	0.107	0.173
Bike-Availability (#bikes>5)	fraction	0.573	0.000	0.158	0.663	1.000	1.000	0.403	0.240	0.324

TABLE 1. Summary Statistics

have a few such bikes, whose condition is such that they are practically unusable and they tend to be the last remaining bikes at stations. We account for this factor by considering a station to have usable bikes in stock only if it has *more than five* available bikes. In addition to accounting for unusable bikes, arguably this specification also better captures how commuters think of a station's bike-availability. A commuter who sees only a small number of bikes may often assume that those last few are likely unusable or might well be checked out (by other commuters) by the time he reaches the station.[8]

Stations that are stocked in at the start of the two-minute period are candidates for the choice set. We operationalize bike-availability as the fraction of two-minute intervals at whose start the station is stocked in. Figure 3.2(b) shows the average bike-availability at each station. Note that stations in district 7 (in the lower left corner of the plot) have much higher availability even though the station-use levels are comparable to neighboring districts (such as the 8th, which is directly above the 7th). The 7th district is home to many ministerial offices, embassies, and other centers of power. Public-sector system managers arguably set higher availability targets depending on such unobserved station characteristics, suggesting that longitudinal variation in bike-availability may be more useful rather than either cross-sectional variation (differences across stations) for identifying the effect of availability. Panels (d), (e), and (f) of Figure 3.3 show (respectively) the distribution of station-level bike-availability, the hourly pattern of average bike-availability, and the dependence of bike-availability on station size. Note that larger stations have higher average availability, which may be an effect of statistical pooling.

Table 1 provides some summary statistics for our data sample. Stations are located 166 meters apart, on average, from the next nearest station, but there is wide variation in this distance. Bike-availability exhibits significant between-station and within-station variance. The substantial variation in nearest-neighbor distances suggests that the cross-sectional variation arising from different station network patterns in different parts of the city can be used to estimate distance effects.

[8]We try many alternate definitions for in-stock stations, such as stations with more than four or with more than six bikes, stations that are more than 5% or more than 10% full, and stations that have more bikes than the day's or the week's minimum number (if less than 5). Similar results are obtained with each of these alternate specifications; some are reported in Section 8 (on robustness).

		District									
Variables		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Number of Stations		29	22	16	25	38	34	30	55	48	52
Station-Use	per min	0.065	0.065	0.082	0.085	0.071	0.063	0.069	0.050	0.062	0.079
Distance to nearest station	kms	0.109	0.136	0.199	0.172	0.174	0.174	0.272	0.156	0.142	0.154
Av. Distance to nearest 2 stations	kms	0.139	0.163	0.235	0.200	0.207	0.198	0.319	0.191	0.166	0.180
Bike-Availability (#bikes>0)	fraction	0.914	0.883	0.982	0.965	0.905	0.942	0.986	0.838	0.825	0.852
Bike-Availability (#bikes>5)	fraction	0.620	0.554	0.812	0.761	0.617	0.644	0.850	0.427	0.412	0.454

TABLE 2. Summary Statistics by District

Similarly, the substantial within-station variance in bike-availability suggests that we can use longitudinal variation to derive robust estimates of the bike-availability effect. Table 2 reports the mean statistics grouped by district. Because most districts have more than 30 stations, we can use within-district variation as a source of identification; that allows us to use district fixed effects for cleaner identification.

4. A Discrete Choice Model

Our goal is to estimate the effect on station-use of the distance that commuters must walk to gain access and of bike-availability. As mentioned before, the walking distance is a characteristic that is both station and commuter specific. Therefore, estimating its effect directly would require data on use at the *station* $\times$ *commuter-origin-location* level—that is, on how many commuters originating at each location in the city use the system at each station. Yet we have data only on use by station, or the same kind of data that most system operators have access to—station level and not commuter-origin-location level.

One method of using data of this type is to estimate the effects from a station-level reduced form model. In such a model, distance can be included by considering a representative commuter whose trips originate somewhere between this and the next nearest station—say, 25% of the distance between this and the next nearest station. The "representative distance" of stations whose neighbors are farther away would be greater and vice-verse. Recall, however, that any estimate of the role played by this representative distance would also include catchment area effects. In the case of a station with relatively distant neighbors, its longer representative distance should decrease use while its larger catchment area should increase use. This is the case with all such proxies for distance: the effect of commuter preference for distance is confounded with the effect of station catchment areas. Even if we constructed proxies that distinguished between these two effects, the utility of such proxies for estimating precise commuter preference parameters would be limited because they are but indirect measures of the distance that the commuter experiences directly. Also note that such proxies would fail to account for the two-dimensional spatial configuration of stations.

Finally, even for characteristics such as bike-availability that are station specific, a simple station-level model poses a number of challenges. It is likely that use at any particular station is influenced by the level of bike-availability (and/or other characteristics) not only of the focal station but also of other "neighborhood" stations. On the one hand, if these characteristics are observable then a model can include the covariate for the station and its neighbors but the effect of different neighbors must be appropriately weighted to account for different proximity essentially involving the above distance effect even for station-specific characteristics. On the other hand, unobservable station-specific characteristics make it difficult to estimate a spatially dependent error structure.

To avoid these pitfalls of a reduced-form model, we take a structural model approach that starts with a choice model for individual commuters at different origin locations and then aggregates commuter choices to obtain station-level use. This method allows us to recover the structural parameters of commuter choice while using only station-level data.

Commuter Choice Model. We conceptualize commuter behavior in bike-share systems as a choice between differentiated products; thus, each commuter is a consumer, each station is a different product, these stations (products) have certain characteristics (distance from the commuter, historic bike-availability, and unobserved station characteristics at a given time), and the set of stations with available bikes is the consumer's choice set. Our parameters of interest are commuter preferences for distance and historic bike-availability. Our model adapts the approach of BLP to consider spatially differentiated products when choice sets change frequently. The spatial components of our model build on the work of Davis [2006].

Consider a population of utility-maximizing commuters distributed spatially over a given area. Commuters choose between using a bike-share system accessible through a network of stations and other modes of transport. The indirect utility of commuter i from accessing the bike-share system at station $f \in \{1, ..., F\}$ at time $t \in \{1, ...T\}$ is given by

$$u_{ift} = \beta_0 + \beta_1 d\left(L_i, L_f\right) + \beta_2 ba_{fm'w} + \gamma_f + \gamma_m + \gamma_{w \times di(f)} + \xi_{ft} + \epsilon_{ift}, \tag{4.1}$$

where L_i is commuter i's origin location and $d\left(L_i, L_f\right)$ gives the distance between commuter i and station f, which is located at L_f. The operator $m(t)\colon \{1, \ldots, T\} \to \{1, \ldots, M\}$ gives the month corresponding to the time t; the operator $w(t)\colon \{1, \ldots, T\} \to \{1, \ldots, 6\}$ maps the time to one of six four-hour "time-windows" in a day (00h00–04h00, 04h00–08h00, etc.); and $ba_{fm'w}$ denotes the historic bike-availability at station f in time-window $w\left(t\right)$ and month $m\left(t\right) - 1$. (We simplify $m\left(t\right)$ and $w\left(t\right)$ to m and w wherever possible, and we use m' as shorthand for $m\left(t\right) - 1$.) The γ_f are *station* fixed effects, γ_m are the *month* fixed effects, and $\gamma_{w \times di(f)}$ are the *time-window*$\times$ *district* fixed effects; here $di\left(f\right)$ is the district for station f. The term ξ_{ft} denotes the unobservable components

of utility that are common to all commuters for station f at time t, which is the *station* $\times$ *time*-specific shock. The ϵ_{ift} are the idiosyncratic *commuter* $\times$ *station* $\times$ *time*-specific error terms; we assume that these errors are of type I extreme value, and are independent and identically distributed (i.i.d.). The covariate $ba_{f,m',w}$ captures the long-term effect of bike-availability. Our results are given for one-month lagged availability, but similar results are obtained with longer and shorter lag durations.

The commuter could also use other means of transport, in which case her utility is

$$u_{i0t} = \xi_{0t} + \epsilon_{i0t};$$

here ξ_{0t} is the unobservable component of this utility that is common to all commuters at time t, normalized to zero. The ϵ_{i0t} are the idiosyncratic utilities that commuters derive from other means of transport, which we also assume are type I extreme value, and i.i.d.

We limit the commuter's potential choice to his nearest m_d stations within distance dis^{max}; the set of such nearby stations is denoted by N_i. Let S_t be the set of stations that are in stock at time t. The choice-set for commuter i at time t is then $N_i \cap S_t$. Hence the probability of utility-maximizing commuter i using a bike from station $f \in N_i \cap S_t$ at time t is given by

$$p_{ift}\left(L_i, x_{\cdot mwt}; \theta, \xi_{\cdot t}\right) = \frac{\exp\left(\beta_0 + \beta_1 d\left(L_i, L_f\right) + \beta_2 ba_{fm'w} + \gamma_f + \gamma_m + \gamma_{w\times di(f)} + \xi_{ft}\right)}{1 + \sum_{g\in N_i\cap S_t} \exp\left(\beta_0 + \beta_1 d\left(L_i, L_g\right) + \beta_2 ba_{gm'w} + \gamma_g + \gamma_m + \gamma_{w\times di(g)} + \xi_{gt}\right)}.$$

Here $x_{\cdot mwt}$ is F-row dimensional matrix of observed station-level covariates (historic bike-availability $ba_{fm'w}$ and the *station*, *month*, and *time* $-$ *window* $\times$ *district* dummies); $\xi_{\cdot t}$ is the vector of unobservable characteristics at time t; and θ represents the parameter values (β_0, β_1, β_2, and all fixed effects γ).

The net use at station f at time t, or λ_{ft}, is obtained by aggregating choice probabilities of all commuters in the population:

$$\lambda_{ft}\left(x_{\cdot mwt}; \theta, \xi_{\cdot t}\right) = \int_{L_i} p_{ift}\left(L_i, x_{\cdot mwt}; \theta, \xi_{\cdot t}\right) \cdot P_D\left(L_i\right)\, dL_i,$$

where $P_D\left(L_i\right)$ is the spatial density distribution of potential commuters' origin location. We assume that this density is uniform and calibrated such that the market share of the bike-share system is 10% of the potential commuters in each district (as suggested by survey data on market share). The heterogeneous origin location L_i of commuters in the model implies that, when a station stocks out, its commuters are then more likely to substitute to nearby rather than distant stations. Hence our model does not exhibit the independence of irrelevant alternatives property, which is a major drawback of simpler (e.g., multinomial logit) models. The heterogeneous commuter location in our model essentially plays the role of the random coefficient in traditional demand models based on

random utility maximization. Finally, note that use at the stocked-out stations will be zero because they are not part of any commuter's choice set.

Note that the model just described captures the impact of bike-availability in two ways. The first is the long-term effect. From a commuter's standpoint, consistently being able to find a bike at a station allows her to reliably plan an itinerary around this trip and to integrate the bike-share system into her lifestyle choices (where to live and work, whether or not to own a car, etc.). Higher ex ante probabilities of finding a bike can therefore increase the utility from bike-share use, and this effect is captured directly by including the historical bike-availability in a commuters' use utility.

The second impact of bike-availability is the immediate, short-term effect of a stockout, which has the effect of removing that station from the commuter's choice set. A stocked-out station cannot serve potential commuters, but this need not imply that this trip from the system is lost. Some commuters might decide to change their itinerary and take a bike from the next most desirable station. The stockout however lowers the utility from bike-share use, leading to more commuters choosing the outside option; hence some fraction of demand at the stocked-out station would indeed be lost. This immediate effect of stockouts is indirectly captured in our model through the choice sets changed thereby and other components of the utility model.

There are some potential endogeneity concerns with our model as regards identifying the effect of bike-availability. There could be unobserved station- or time-level factors that affect both station-use and bike-availability. For instance, a fleeting usage shock—due, say, to a $station \times time$ specific unobservable event such as a concert—will lead to both higher use and reduced bike-availability. Furthermore, there are more persistent unobservable station- or time-level factors that induce system managers to employ policies to ensure different bike-availability at the station at given times; leading to a correlation between these unobservable factors and bike-availability. Our model addresses concerns around fleeting unobserved factors by using *lagged* bike-availability. Even though contemporaneous bike-availability and usage shocks may be correlated owing to special events, such as a concert, it is unlikely that these usage shocks are correlated with lagged or historic bike-availability. The more persistent unobserved factors are dealt with by our use of both *station* and *time-window* $\times$ *district* fixed effects. Together these ensure our estimates on the effect of bike-availability are unbiased.

Similarly, consider endogeneity in our estimates for the preferences for distance. One could argue for the possibility of unobserved station characteristics that lead to a specific design of station locations and hence to higher or lower use. Our data's panel structure does not help here; unlike the case of bike-availability, there is no change over time in the pattern of station locations. Our model includes *time-window* $\times$ *district*-level fixed effects, which should allay concerns arising from any unobservable district-level characteristics in the pattern of station location. We obtain

similar results if we include fixed effects at the level of *quartier* (a finer classification of districts that is roughly equivalent to "neighborhoods"; each Paris district has four *quartiers*). Together these analyses ensure that our estimates on the effects of distance are unbiased on account of any unobserved neighborhood characteristics that systematically drive station location and station-use.

5. Model Estimation

Our estimation procedure builds on the demand estimation algorithm proposed by Berry et al. [1995] but departs in two important ways. First, we use different identifying assumptions. Whereas BLP rely on the functional form of the *supply* side for *demand*-side identification, we identify our model using the data's cross-sectional and longitudinal variation. We identify the bike-availability coefficient β_2 using the longitudinal variation in bike-availability; the distance effect β_1 is identified using the cross-sectional variation in distances across different pairs of stations and commuter origin. As stations stock out and replenish, the choice sets of commuters change longitudinally—further facilitating identification of the distance effect. The longitudinal variation in choice sets is akin to having different markets in the BLP approach.

Our second main departure from the BLP approach arises from the high frequency of our data or, more precisely, the high frequency with which choice sets change. Traditional numerical estimation procedures are thus rendered computationally infeasible. We therefore introduce the notion of a (local) station stockout state, and our estimation is feasible when we transform from the time domain to this stockout state domain.

Estimation Procedure. 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), each iteration of which involving multiple numerical integrations over the spatial density. We instead estimate our model using a *nested* iteration process that relies on all parameters (except β_1) entering our model in a "user-location–agnostic" way. We thus group our parameters in two classes, β_1 and the parameters that are "linear" (in ξ_{ft}). The outer loop of our nested procedure searches over β_1 and the inner loop estimates the linear parameters (Berry et al. [1995]).

The linear terms in the utility model are grouped together into the composite terms δ_{ft}:

$$\delta_{ft} = \beta_0 + \beta_2 ba_{fm'w} + \gamma_f + \gamma_m + \gamma_{w\times di(f)} + \xi_{ft}. \tag{5.1}$$

Our choice model now becomes: $\forall f \in N_i \cap S_t$ (the choice set of commuter i at time t), commuter choice probabilities are given as

$$p_{ift}\left(L_i;\beta_1,\delta_{\cdot t}\right) = \frac{\exp\left(\delta_{ft} + \beta_1 d\left(L_i, L_f\right)\right)}{1 + \sum_{g\in N_i\cap S_t}\exp\left(\delta_{gt} + \beta_1 d\left(L_i, L_g\right)\right)}.$$

Use at station $f \in S_t$ at time t is written as

$$\lambda_{ft}\left(\beta_1, \delta_{\cdot t}\right) = \int_{L_i} p_{ift}\left(L_i; \beta_1, \delta_{\cdot t}\right) \cdot P_D\left(L_i\right)\, dL_i, \tag{5.2}$$

where $\delta_{\cdot t}$ is the vector of composite terms δ_{ft}. We start our search with a guess for the value of β_1. Given this β_1, we obtain $\delta_{\cdot t}$ by equating the actual and predicted use rates for each station–time pair; that is, we solve the following $F \times T$ equations to compute the $\delta_{\cdot t}$ for each time:

$$\lambda_{ft}\left(\beta_1, \delta_{\cdot t}\right) = \Lambda_{ft} \quad \forall f\,; \tag{5.3}$$

here Λ_{ft} is the observed rate of use for station $f \in S_t$ at time t.

Equation 5.4 is then an iterative search process (Berry et al. [1995], Davis [2006]) that converges to the actual value of $\delta_{\cdot t}$:

$$\delta_{\cdot t}^{new} = \delta_{\cdot t}^{old} + \left(\log\left(\Lambda_{\cdot t}\right) - \log\left(\lambda_{\cdot t}\left(\beta_1, \delta_{\cdot t}^{old}\right)\right)\right). \tag{5.4}$$

Note that each search iteration (Eq. 5.4) will require computation of $\lambda_{\cdot t}\left(\beta_1, \delta_{\cdot t}\right)$—which, per Equation 5.2, involves integrating over the spatial density of commuters. We perform this integration numerically. We discretize the physical area of the ten central districts 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 commuters. Predicted use is then

$$\lambda_{ft}\left(x_{\cdot mwt}, \delta_{\cdot t}; \beta_1\right) = \sum_i p_{ift}\left(L_i, x_{\cdot mwt}, \delta_{\cdot t}; \beta_1\right) \cdot P_D\left(L_i\right) \cdot \mathscr{D}^2,$$

where $\mathscr{D}^2$ is the area of each grid square.

Using this predicted use in Equation 5.4 allows us to search and obtain the $\delta_{\cdot t}$ for our guessed value of β_1. Next we estimate the constituent components of $\delta_{\cdot t}$ (i.e.,β_0, β_2, γ_f, γ_m, $\gamma_{w \times di(f)}$, and ξ_{ft}) by using Equation 5.1; this is essentially a standard fixed-effects regression in which $\delta_{\cdot t}$ is the dependent variable and the ξ_{ft} are error terms. This completes the inner loop by estimating all other parameters for a given value of β_1.

The outer loop then iterates over different values of β_1 to identify the β_1 that minimizes squared errors:

$$\hat{\beta}_1 = \arg\min_{\beta_1} \sum_{f,t} \left(\gamma_f + \xi_{ft}\right)^2. \tag{5.5}$$

The estimated value for the bike-availability coefficient β_2 (from the inner loop, Eq. 5.1) that corresponds to the least-squares estimator of the distance coefficient $\hat{\beta}_1$ completes our estimation procedure. Observe that, when estimating β_1, we define the squared errors to include the station fixed effects γ_f (Eq. 5.5). The reason is that—although we want to identify the availability effect β_2 using the longitudinal variation in bike-availability alone (i.e., in order to avoid bias stemming from

endogenously set bike-availability)—for the distance coefficient β_1 we want to use the cross-sectional variation also.

Note that even though the nested procedure effectively reduces the dimensionality of the parameter search space, the procedure just described remains computationally challenging. Each inner loop involves an iterative search–based computation of $\delta_{\cdot t}$ followed by the estimation of its constituent components. Thus, each iteration of β_1 requires an iterative search of $F \times T$ parameters, and each step of this search requires—besides the fixed-effects regression—numerical integrations over a grid with nearly 9,000 points. To make matters worse, recall that our data are collected at an extremely high frequency; hence that data include millions of two-minute intervals T. In effect we must run roughly a billion regressions, iterative searches, and numerical integrations. Such an approach is computationally infeasible.

The computational burden can be reduced by aggregating data over time or by considering shorter time periods of data. The latter will reduce the precision of our estimates, especially since the variability in two-minute use is very high (which entails that large spans of data are needed to infer robust estimates). The former approach is not helpful, either. Aggregating data over time would needlessly use aggregate metrics to proxy for information we have on the actual choice sets available to commuters, and that would introduce unnecessary noise. Moreover, the minute-to-minute variation in station stockouts creates the changing choice sets that dictate the changing distances commuters must walk to access stations. This is important for efficiently estimating the distance effect; aggregating data would eliminate these variations, in which case we would have to rely only on the cross-sectional variation in distances between stations.

We propose an alternate aggregation in our data to reduce the computational burden while still being able to exploit the information and variation in choice sets. Note that in our model of station-use, λ_{ft} is a function only of the choice set of stations (along with the time-window and month fixed effects). We can therefore aggregate data points that are in same time-window and month, and have same choice set of stations, without losing much precision in our estimates—that is, we aggregate data according to *system-stockout-state* $\times$ *month* $\times$ *time-window*. Formally, we define the stockout state v_t as an F-dimensional binary vector of the stockout status of each station at time t. We combine the data points at times in the same *month* and *time-window* that share the same system-level stockout state: Λ_{fmwv} is the average Λ_{ft}, for all t, where t is such that $w(t) = w$, $m(t) = m$, and $v_t = v$. On the model side, the probability of commuter i using a bike from station

$f \in N_i \cap S_v$ at a system state v in month m and time-window w is given by

$$p_{ifmwv}\left(L_i, x_{\cdot mw}; \theta, \xi_{\cdot mwv}\right) = \frac{\exp\left(\beta_0 + \beta_1 d\left(L_i, L_f\right) + \beta_2 ba_{fm'w} + \gamma_f + \gamma_m + \gamma_{w\times di(f)} + \xi_{fmwv}\right)}{1 + \sum_{g\in N_i\cap S_v} \exp\left(\beta_0 + \beta_1 d\left(L_i, L_g\right) + \beta_2 ba_{gm'w} + \gamma_g + \gamma_m + \gamma_{w\times di(g)} + \xi_{gmwv}\right)}.$$

Here N_i is the set of stations that are near commuter i and S_v is the set of stations with positive bike stock in state v, so that $N_i \cap S_v$ is the set of nearby stocked stations for commuter i when the system is in state v. The station-level use can similarly be written as

$$\lambda_{fmwv}\left(\beta_1, \delta_{\cdot mwv}\right) = \int_{L_i} p_{ifmwV}\left(L_i; \beta_1, \delta_{\cdot mwV}\right) \cdot P_D\left(L_i\right)\, dL_i.$$

The rest of the estimation proceeds as before.

The anticipated advantage of estimating our model in the stockout state domain (instead of the time domain) is that there would be fewer distinct stockout states during each month and time window than there are distinct two-minute time intervals, which scales back the computational burden significantly. However, since v is defined over the set of all stations in the data ($F \approx 350$), there could be as many as 2^F distinct values of v. Many distinct values are realized in the data, and their number in data is of the same order as T; hence the transformed estimation is only slightly superior to the original model.

We notice that the use at station f is not affected by *all* the other stations' stockout states. Recall that commuters' choice sets are limited to the nearest m_d stations within dis^{max}. The implication is that we can construct a local stockout state for each station and aggregate our data on such local stockout states rather than on the systemwide stockout states v. The local stockout state will have lower dimensionality than the systemwide stockout state, so there will be far fewer distinct local than distinct systemwide stockout states. This approach enables us to reduce the computational burden drastically even as we continue to utilize the important variation in choice sets. We next explain the procedure formally.

Note first that, for a station f, the only relevant bike-availability information is the availability *at stations close enough to commuters who are close enough to station* f. For any station f, we can write the set of relevant stations N_f as

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

The stockout state at time t of stations in N_f is given by v_{ft}—it is the "local" stockout state at station f. Let the set of all such realized local stockout states be given by $V_f \equiv \bigcup_t v_{ft}$.

Next we aggregate the use at station f for all times where the local stockout state was v_f, a typical element in V_f . We use Λ_{fmwv_f} to denote the average observed use at station f in month m and time-window w over all times when the local stockout state is v_f. Accounting for the salience of state v_f shall prove useful, so let ω_{fmwv_f} denote the number of observations that were averaged to obtain Λ_{fmwv_f}; these numbers will serve as weights in subsequent analysis.

There is one final complication. The predicted use λ_{fmwv_f} arises from the commuter choice probabilities p_{ifmwv_f}, which depend not only on the utility of using station f but also on the utility of using other stations in commuter $i's$ choice set (i.e., stations g such that $g \in N_i \cap S_{v_f}$). The set of stations local to station g is not the same as the set of stations local to station f, which means that the local stockout state of station g is not fully determined by v_f, the stockout state of our focal station f. In other words, multiple (different) local stockout states of station g could correspond to a given stockout state v_f of station f. In computing the predicted choice probabilities, we aggregate unobservables of all states of V_g that are consistent with v_f, where a state v_g is *consistent* with focal stockout state v_f for commuter i if and only if the stockout state of all stations in N_i is the same in state v_g as in v_f. Our aggregation presumes that the likelihood of these multiple consistent elements of V_g being realized when state v_f is realized is the same as the unconditional likelihood of any of these elements arising. We can therefore compute p_{ifmwv_f} by using the weighted average value of station utilities of all the states of gmw that are consistent with v_f in the sense just described.

Formally, $v_g \overset{i}{=} v_f$ iff the stockout state of all stations N_i is the same in v_g and v_f. Let

$$\xi_{gmwv_f^i} = \frac{\sum_{v_g \overset{i}{=} v_f} \omega_{gmwv_g} \xi_{gmwv_g}}{\sum_{v_g \overset{i}{=} v_f} \omega_{gmwv_g}}.$$

For $f \in N_i \cap S_{v_f}$, this equality yields

$$p_{ifmwv_f}\left(L_i; \theta, \xi_{\cdot mw\cdot}\right) = \frac{\exp\left(\beta_0 + \beta_1 d\left(L_i, L_f\right) + \beta_2 ba_{fm'w} + \gamma_f + \gamma_m + \gamma_{w\times di(f)} + \xi_{fmwv_f}\right)}{1 + \sum_{g \in N_i \cap S_{v_f}} \exp\left(\beta_0 + \beta_1 d\left(L_i, L_g\right) + \beta_2 ba_{gm'w} + \gamma_g + \gamma_m + \gamma_{w\times di(g)} + \xi_{gmwv_f^i}\right)},$$

where S_{v_f} denotes the set of set of stations with available bikes in state v_f. Then station-use is given by

$$\lambda_{fmwv_f}\left(\beta_1, \delta_{\cdot mw\cdot}\right) = \int_{L_i} p_{ifmwv_f}\left(L_i; \beta_1, \delta_{\cdot mw\cdot}\right) \cdot P_D\left(L_i\right)\, dL_i. \tag{5.6}$$

We now apply the nested procedure described before but with one change. Previously, we used a fixed-effects estimator for the decomposition of δ_{ft} (Eq. 5.1) and the estimation of β_1 (Eq. 5.5); now we use the weighted fixed-effects estimator for

$$\delta_{fmwv_f} = \beta_0 + \beta_2 ba_{fmw} + \gamma_f + \gamma_m + \gamma_{w\times di(f)} + \xi_{fmwv_f},$$

with weights $\sqrt{\omega_{fmwv_f}}$, to obtain β_2, β_0, and the fixed effects. Similarly, β_1 is now given by the weighted estimator

$$\hat{\beta}_1 = \arg\min_{\beta_1} \sum_{fmwv_f} \omega_{fmwv_f} \left(\gamma_f + \xi_{fmwv_f}\right)^2.$$

Implementation Details. The procedure was implemented in R. The open-source package IPOPT (interfaced with R via ipoptr [Ypma, 2010]) was used for nonlinear optimization—in particular, the weighted least-squares estimator of β_1. 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 commuter, 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. 5.6 as a function of β_1 and δ) in C++ and then integrating with R reduced computation time by a factor of nearly 100, or to about 11 hours for the central Paris data set.

6. Results

To facilitate comparison with our main model, we also provide the results from estimation of the following simple station-level model:

$$\lambda_{fmw} \sim \eta_0 + \eta_1\, d_f + \eta_2\, ba_{fm'w} + fixed\ effects;$$

here λ_{fmw} is the average of all λ_{ft} such that $m(t) = m$, $w(t) = w$, and $f \in S_t$—that is, the average use at station f in month m in time-window w conditional on the station being stocked in. The d_f term is the distance to the station nearest to station f, a proxy for the distance that commuters must travel to reach the station. As already discussed, stations with far-off neighbors would see increased use from the focal station's larger catchment area but also decreased use because commuters must walk farther. The coefficient η_1 captures both of these effects, yet the estimate of η_1 is not in itself sufficient to infer commuter disutility for distance. Note, however, that if commuters were totally unconcerned about distance then the effects stemming from the catchment area and from the disutility of walking would both disappear; in that case, we would expect η_1 to be zero. As in the structural model, historic bike-availability is included through the covariate $ba_{fm'w}$, and the estimate for coefficient η_2 can be directly compared with the estimate from our main model.

We consider two variants of this station-level model. The first is a so-called pooled specification that introduces fixed effects at the *district* × *time-window* level. Variation across stations identifies the distance effect, while variation across stations, time-windows, and months identifies the historical bike-availability effect. Yet because station bike-availability could depend on unobserved station and time-window characteristics, the estimated coefficients might exhibit an endogeneity bias. The second specification addresses this problem by introducing finer fixed effects at the

	(1) Pooled		(2) Fixed Effects		(3) Structural Model	
	Value	Std. error[$]	Value	Std. error[$]	Value	Std. error[#]
Bike-Availability	0.037	(0.005)***	0.018	(0.003)***	0.304	(0.061)***
Walking Distance	0.059	(0.012)***	0.082	(0.023)***	-4.813	(0.024)***
TimeWindow × District F. E.	Yes					
Station × TimeWindow F. E.			Yes			
Station F. E. for Service Level					Yes	
Adjusted R^2	0.69		0.14		0.65	
F-stat (p-value)	0.00		0.00			
Wald Test (p-value)					0.00	

Marginal Effects	%Increase in Demand	
	Short Term	Long Term
10% increase in Bike-Availability	9.56%	11.73%
10% decrease in Walking Distance	6.65%	

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

[$]Robust Standard Errors

[#]Bootstrapped Standard Errors

TABLE 3. Estimation Results

station × *time-window* level. This specification is closer to our structural model. In this variant and also in the structural model, only the longitudinal variation is used to estimate the impact of bike-availability; the distance coefficient is then estimated in a second step by decomposing the estimated station fixed effects.

Table 3 reports the estimation results. Columns (1) and (2) give the estimated coefficients from the two variants of the station-level model; column (3) presents the results from estimating our main structural model. Both the distance effects and the bike-availability effects turn out to be statistically and economically significant in the two station-level models, as seen in columns (1) and (2). The coefficients change significantly between these two models, which underscores the importance of adding station-level fixed effects to address concerns about the endogeneity of bike-availability. The estimates derived from our model with station-level fixed effects imply that a 0.1 increase in bike-availability leads to a 1.33% increase in the use of a typical station. Increasing the distance to the next nearest station by 100 meters increases use at the focal station by 6.04%. This latter finding captures the catchment area effect as well as commuter sensitivity to distances, and

the increase is likely driven by substitution from other stations. Estimates from the simple model are not that informative about commuter sensitivity to station distances, which motivates our use of the structural model.

Column (3) of Table 3 reports the estimation results for the structural model. These estimates are for $m_d = 3$ and $dis^{max} = 600\ m$; that is, each commuter considers the three nearest stations within 600 meters in his choice set. The numerical integration is carried out using a square grid with $\mathscr{D} = 50$ meters. Further, for computational efficiency we include only the eight most frequently realized local stockout states when a station is stocked in; in other words, for each *station* $\times$ *month* $\times$ *time-window* we use only the eight most frequent local stockout states (these states account for 75.82% of our data). In Section 8 we check the robustness of our estimates to each of these computational choices. Standard errors are computed via bootstrapping with replacement; in the table, they are given in parentheses.

Our structural model finds that the bike-availability effect is positive and significant and that the effect of distance is negative and significant. In other words, commuters incur a significant disutility from more distant stations and derive a higher utility when stations are more reliably in stock. While the direct estimates from this model describe the effect of changing distances and of bike-availability on utility, the effects of these factors on system-use—which we discuss next—may be of more practical value.

First consider the effect of changing distance on system-use. We take a hypothetical city with the same station network structure but in which all commuter–station distances are reduced by 10%. In essence, this shrinks the city's area by 19% $(1 - 0.9^2)$ and increases the density of stations by 23.4%. We then use our commuter choice demand model (now with parameter estimates) to predict system-use while accounting for commuters who choose the outside option. Finally, we scale the system-use from this shrunken city to account for the smaller area. This analysis is akin to reducing distances—by adding more stations in the network—while preserving all spatial relationships between stations, which are critical in determining the demand pattern. From a computational perspective, this approach is equivalent to using the demand model with $\hat{\beta}_1 = 0.9\beta_1$. The rate of system-use per two-minute period is then given by

$$\sum_{f,m,w} \left(\frac{\sum_{v_f \in V_f} \omega_{fmwv_f} \lambda_{fmwv_f} \left(x_{\cdot mw}, \delta_{\cdot mw\cdot}; \hat{\beta}_1 \right)}{\sum_{v_f \in V_f} \omega_{fmwv_f}} ba_{fmw} \right),$$

where the first part gives the rate of demand at each *station* $\times$ *month* $\times$ *time-window*and then multiplying by bike-availability yields the rate of system-use. We find that *a 10% reduction in distances results in a 6.65% increase in system-use.*

Second, the estimates from our structural model also allow us to estimate how commuters behave when a station stocks out. We find that, on average, *95.6% of a stocked-out station's demand is*

lost (so only 4.4% of its unserved commuters substitute other stations). This figure is calculated by removing one station at a time from the network and then re-computing total demand from our demand model for remaining stations. We follow this procedure for all stations, one by one; the reported estimate is the average effect, or the effect of removing a typical station. The implication is that a 10% increase in bike-availability would lead to an immediate 9.56% increase in system-use—yet this is only the short-term (direct) effect.

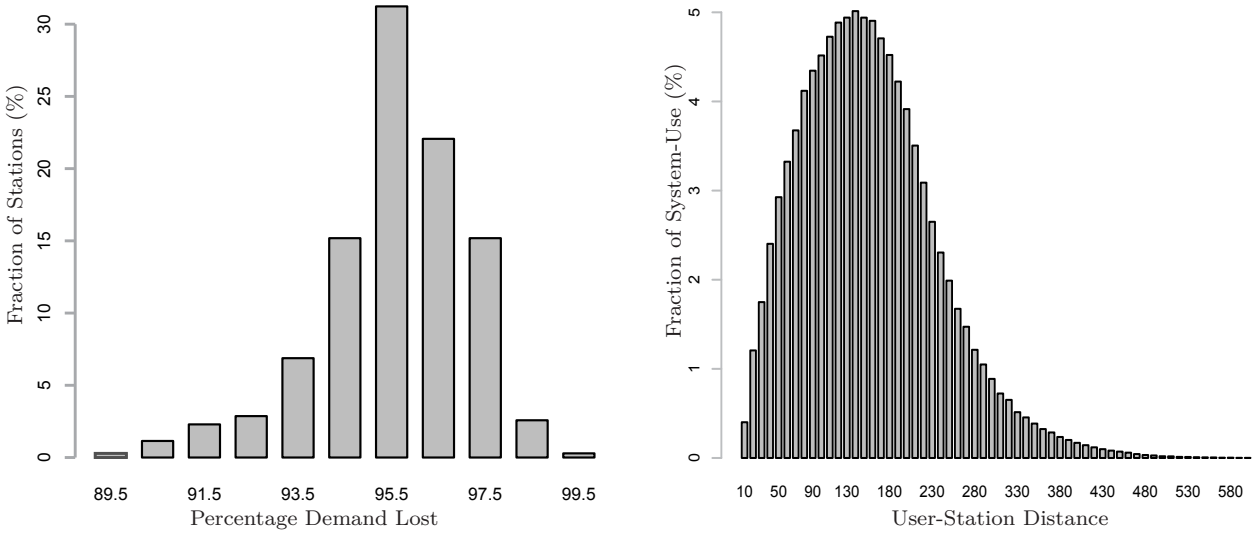
Our estimated model reveals, in addition, a long-term effect that may be due to the changing long-term behavior of commuters (e.g., more commuters incorporating bike-share systems into their lifestyles). As with the distance effect, we compute this effect by considering a network of stations in which each station has 10% higher bike-availability than the status quo. We then use our commuter-level choice model (again with estimated parameters) to compute the new level of system-use as follows:

$$\sum_{f,m,w} \frac{\sum_{v_f \in V_f} \omega_{fmwv_f} \lambda_{fmwv_f} \left(x_{\cdot mw}, \hat{\delta}_{\cdot mw\cdot}; \beta_1 \right)}{\sum_{v_f \in V_f} \omega_{fmwv_f}} \hat{ba}_{fmw};$$

here both $\hat{\delta}_{\cdot mw\cdot}$ and $\hat{ba}_{fmw}$ incorporate the higher bike-availability. In sum: *increasing the bike-availability of all stations by 10% would increase system-use by 11.73%*; of this, 9.56% is the immediate short-term effect. The same estimates can also be interpreted in terms of the absolute changes to station network design. We find that *reducing the distance between stations by 100 meters increases system-use by 37.57%*; also, *an increase in the bike-availability by 0.1 increases system-use by 12.63%.*

We have reported that about 95% of a station's demand is lost if it is stocked out and thus effectively removed from the system. This number indicates that station demand is extremely local: commuters rarely substitute. Next we explore whether this finding is a truly broad effect or is instead driven by a few critical stations that are located far from other stations. Whereas the former explanation suggests that systemwide changes are needed, the latter implies that local changes could improve system-use. To estimate the influence of any one station on this lost demand, we iteratively remove individual stations from the network and then use our fitted commuter choice model to estimate resulting system-use; this procedure is followed for each station in the network. Figure 6.1(a) shows the distribution of demand lost in response to removing different stations from the network. We find low rates of substitution for the vast majority of stations, so any increase in system-use would require systemwide changes; that is, the low substitution rate is not driven by few isolated stations.

We also investigate the distance effect more thoroughly by examining the distribution of distances that commuters walk to reach a station. Because these distances are not observed directly, their



(a) Distribution of Demand Lost on Stockout

(b) System Use and Distance

FIGURE 6.1. Interpretation of Results

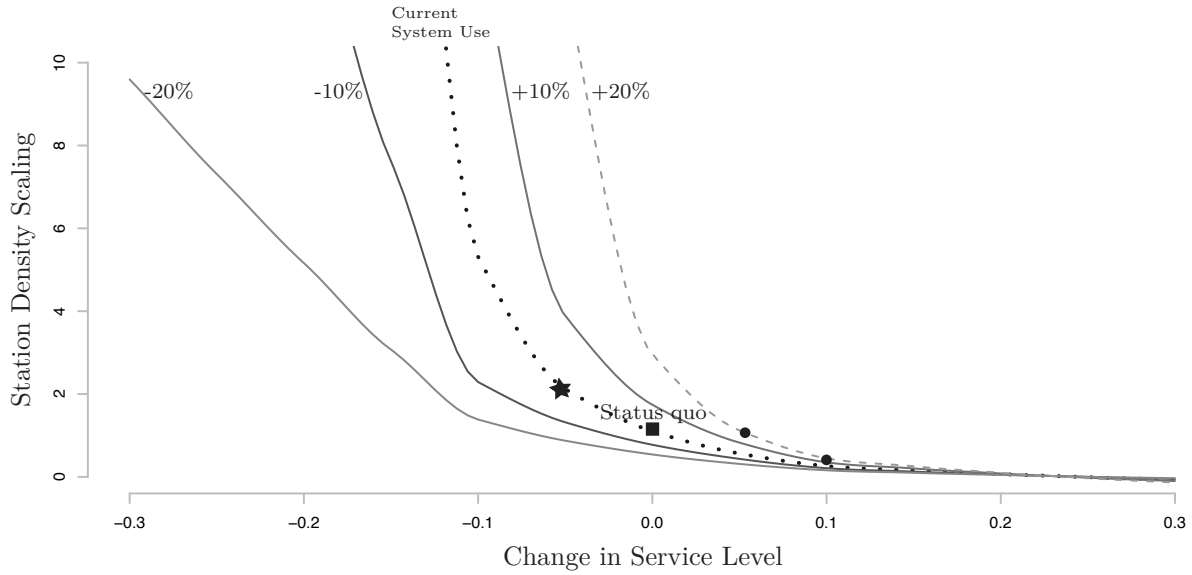


FIGURE 6.2. Iso-Ridership Curves

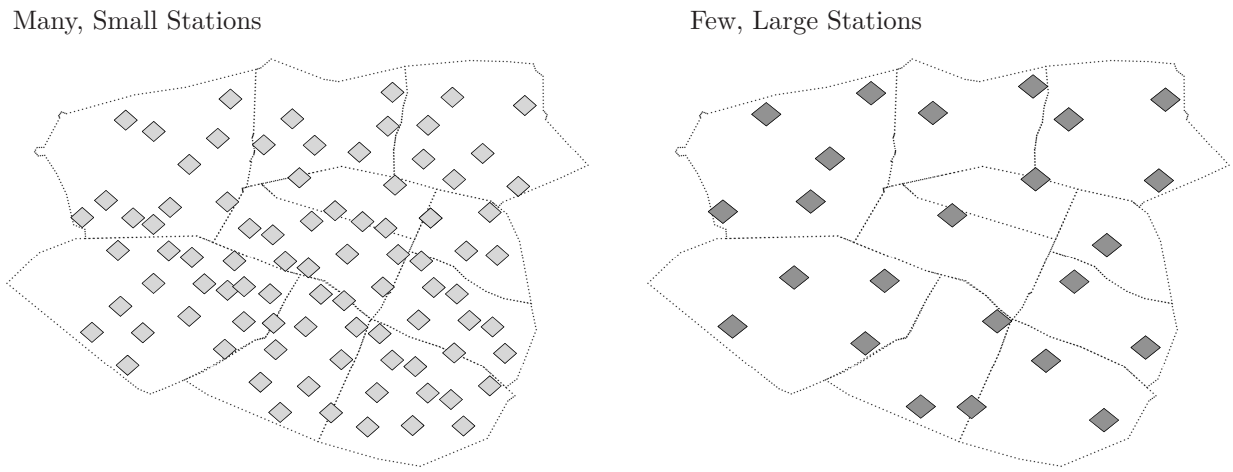
distribution is derived from the estimated demand model. In particular, the probabilistic decisions of each latent commuter are combined to yield the distribution curve plotted in Figure 6.1(b). We find that the median commuter travels about 150 m to reach her preferred station; but some commuters walk longer distances.

The preceding estimates predict the effects of changing station density or bike-availability. Achieving changes of this type requires costly investments, such as adding stations and/or bikes and changing system management policies (increasing bike trans-shipment, offering demand-balancing incentives, etc.). A system manager with limited resources will likely be unable to make all these investments and so must identify improvements that lead to the most improvement. Towards that end, comparing the effects of increased bike-availability and increased station density is key to identifying investments that yield the highest return. A complete analysis along such lines requires data that can be used for relating investment amounts to the changes achieved; Figure 6.2 plots iso-ridership curves, which can help with this decision. Each curve represents a combination of bike-availability and distance changes that lead to the same ridership. The dotted curve represents the status quo system-use, and the point denoted by a star indicates that a 105% increase in station density, when combined with an 0.05 *decrease* in bike-availability, would lead to the same system-use as the status quo. Alternately, 20% increases in use can be achieved by all changes along the dashed curve—for example, by increasing station density by 17% and bike-availability by 0.05, or by decreasing station density by 26% and increasing bike-availability by 0.1, and so forth. This curve shows that, on average, the effect on ridership from an 0.05 increase in bike-availability is equivalent to the effect of increasing station density by 52%.

7. Counterfactual Analysis: Alternate System Designs

The foregoing analysis discussed improvements in accessibility or availability, all of which come at the cost of adding bikes or making other changes. In this section, we consider another analysis that illustrates how our estimates can be used. In particular, we consider alternate station network designs that each include the same number of bikes and docking points but place different emphasis on satisfying the competing demands of station accessibility and bike-availability. The number of bikes and the corresponding number of docks primarily determine the costs of bike-share systems; bikes and docks are often the main costs, or all other other costs such as operation, maintenance and land are proportional to the number of bikes. Thus, all system designs compared in this section have the same costs, yet some might have higher ridership.

Given a fixed budget, the system designer is constrained to build station networks with a fixed number of bikes and docks, where the number of docks is typically set at a fixed multiple of the number of bikes. That being said, a given number of bikes can be used to build alternate station networks that prioritize either accessibility or availability. On the one hand, a network with many distributed stations but relatively fewer bikes at each station (high density; Figure 7.1(a)) reduces commuter distances to stations, which increases accessibility. On the other hand, a network with fewer stations but with more bikes at each station (low density; Figure 7.1(b)) can achieve higher



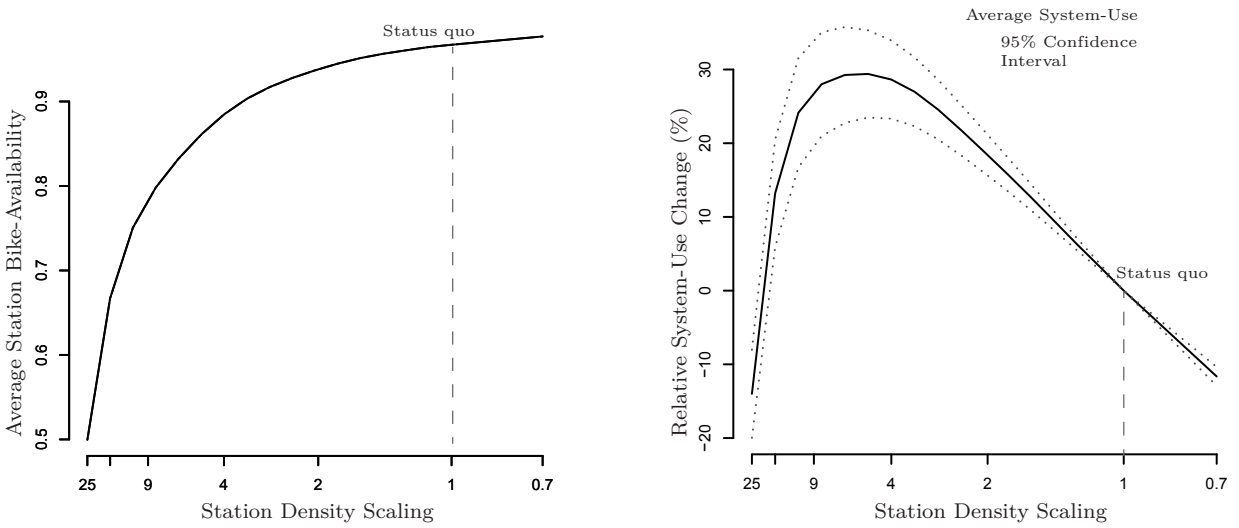
(a) High Accessibility (b) High Availability

FIGURE 7.1. Accessibility–Availability Trade-Off in Station Design

bike-availability owing to the well-known statistical benefits of holding pooled inventory in systems with demand variability (Cachon and Terwiesch [2009]). Thus, there is a trade-off in network-design between the demands of accessibility and availability. Hence information about how commuters value these two aspects can help system designers make optimal trade-offs between them.

In order to compare station networks with the same number of bikes but different density or size of stations, we must first calculate two effects on ridership. The first is directly computable using previously developed estimates and analysis; changing the station density alters the distances that commuters must travel and thereby changes system-use. The second effect is more involved. Changing station density while keeping the number of bikes and docks fixed requires that each station increase or decrease its number of bikes. The result is a change in the statistical pooling effect, which in turn alters bike-availability. So far we have used our estimates to translate availability into ridership; now, however, we must estimate how station size or density affects bike-availability. For instance, how would bike-availabilities change if we split a station in two and divided both bikes and anticipated demand equally between them? We obtain this relationship via a simulation that assumes demand to be Poisson distributed and calibrates the system's use rate to the average use rate at stations. Figure 7.2(a) plots the results of this simulation. As expected, when density is increased and each station has fewer bikes, average bike-availability levels across the system falls; or bike-availability increases in the number of bikes that stations have due to statistical scale economies.

Armed with this simulation (which relates density to availability) and our estimates (which relate density and availability to ridership), we have all the elements needed to compare alternate station designs and identify the optimal one. Formally, consider a counterfactual scenario in which station



(a) Effect of Station Density on Bike Avaialability

(b) Counterfactual System-Use for Alternate Station Network Designs

FIGURE 7.2. Density-Availability Simulation and Counterfactual System-Use

density is scaled by a factor of σ. As before, this is done by replicating the current configuration so that distances between stations are scaled by a factor of $\sigma^{-1/2}$. For example, if $\sigma = 4$ then station density increases by 4 times and all distances between stations become half of their current distances. This approach allows the counterfactual station design to retain the spatial properties of station placement, including unobserved practical constraints in station placement. Counterfactual bike-availability at each *station*$\times$ *time-window*$\times$ *month* is the original level scaled by the ratio of the bike-availability at density scaling σ and at density scaling 1 (from the simulation, Figure 7.2(a)), and is capped between 0 and 1.

We now use our estimated model to compute system-use for the counterfactual network at station density scaling σ:

$$\sum_{f,m,w} \frac{\sum_{v_f \in V_f} \omega_{fmwv_f} \lambda_{fmwv_f} \left(x_{\cdot mw}, \tilde{\delta}_{\cdot mw\cdot}; \tilde{\beta}_1 \right)}{\sum_{v_f \in V_f} \omega_{fmwv_f}} \tilde{ba}_{fmw},$$

where $\tilde{\beta}_1 = \sigma^{-1/2}\beta$ and where both $\tilde{\delta}_{\cdot mw}$ and $\tilde{ba}_{fmw}$ include the changed bike-availability as in Figure 7.2(a).

Figure 7.2(b) shows the system-use predicted for station networks with different station density. The horizontal axis represents density scaling as a multiple of existing density; for instance, $x = 1$ corresponds to the status quo network and $x = 2$ corresponds to a network with twice the density of the status quo. The vertical axis shows the percentage change in use as compared with the status quo network. Dotted lines indicate the 95% confidence level around our estimates.

Comparing multiple alternate station designs reveals that the bike-share system in central Paris could *increase system-use by as much as 29.41% (268,440 more trips/month) via deploying designs that are denser than the status quo* (i.e., if the system split existing stations into smaller stations, increasing station accessibility while decreasing bike-availability). This finding states, in essence, that the existing system does not optimally trade off commuter preferences for accessibility with those for availability. In particular, at the status quo the gains from increasing accessibility dominate those from increasing availability. Not surprisingly, the benefits from increasing station density and accessibility are diminishing: beyond a certain point, any increase in density actually leads to declining system-use; thus the trade-off is reversed, and the availability effect then dominates the accessibility effect. Recall that each of these alternate systems has the same number of bikes and of docks, which constitute the main capital investment in these systems. That is to say: If the city of Paris had been given access to analysis and estimates of commuter preferences for accessibility and availability, such as those provided in this paper, then it could have designed a system that achieved from 20% to 30% more use without adding a single new bike or dock to the system.

8. Robustness

We test the robustness of our effect sizes to alternate model specifications and to computational choices made in model estimation. Table 4 reports the results of our estimation under many alternate assumptions; row (1) replicates our original estimates (from Table 3) for easy comparison. Rows (2) and (3) of the table report the estimates obtained under alternate definitions of bike-availability. Row (2) gives estimates from a model where a station is said to be in-stock or have bikes available if there are *more than four* bikes available at the station (versus five bikes in the original estimation); the results are nearly identical to the base estimates. Row (3) uses a station-specific threshold for stockouts whereby a station is stocked in if it has more bikes than the minimum achieved over the day (if that minimum is less than five bikes). Again, the estimates are similar to those obtained under our original regressions *except* for the long-term bike-availability effect, which is higher here.

Next we consider a quadratic effect of distance. So in addition to all the original components of the a commuter's utility from using a bike at station f (Eq. 4.1), the utility now includes a quadratic effect of distance:

$$\hat{u}_{ifmwv_f} = \beta_0 + \beta_{11} d\left(L_i, L_f\right) + \beta_{12} d\left(L_i, L_f\right)^2 + \beta_2 ba_{f,m',w} + \gamma_f + \gamma_m + \gamma_{w \times di(f)} + \xi_{fmwv_f} + \epsilon_{ift}.$$

Row (4) of Table 4 reports the results from this model. We find that the disutility from distance is locally concave; that is, the marginal disutility of walking decreases with distance. The marginal

	Bike-Availability	Walking Distance	10% Increase in Bike-Availability		10% Decrease in Walking Distance
			Short-Term	Long-Term	
(1) Original Estimates	0.304 (0.061)***	-4.813 (0.024)***	9.56%	11.73%	6.65%
(2) Stockout: ≤ 4 Bikes	0.332 (0.059)***	-4.795 (0.273)***	9.57%	12.01%	6.63%
(3) Stockout: Min. Bikes	1.093 (0.184)***	-4.893 (0.266)***	9.57%	18.05%	6.75%
(4) Quadractic Effect of Distance	0.27 (0.056)***	-16.15[1] (1.979)*** 25.86[2] (1.707)***	9.64%	11.53%	7.50%
(5) Choice Set: m_d=4	0.38 (0.072)***	-3.23 (0.283)***	9.53%	12.28%	5.35%
(6) Number of states: top 16	0.41 (0.080)***	-3.91 (0.540)***	9.60%	12.29%	5.53%
(7) Finer Grid Size	0.29 (0.070)***	-4.59 (0.299)***	9.57%	11.65%	6.40%

[1]Linear distance effect

[2]Quadratic distance effect

TABLE 4. Robustness Tests

effects of bike-availability and station distance are again close to those derived from our original model.

Finally, we investigate the role of various computational choices made in estimation. Row (5) gives the results when we set m_d equal to 4 instead of to 3; in this case, then, commuters consider the *four* nearest stations in their choice set within distance $\text{dis}^{\max}$. Using this alternate value, we find that the bike-availability effect remains almost the same while the distance effect decreases by a single percentage point. Row (6) of the table reports results from using the 16 most frequently occurring local stockout states in each *station* $\times$ *month* $\times$ *time-window* instead of the top 8 stock states. That change increases the coverage of our sample from 75.82% to 82.14% of relevant observations, yet once again the result is virtually the same in our estimated effects. In row (7) we

provide estimates obtained by using a finer grid for our numerical integration (viz., one that covers 4 times as many points); this produces no qualitative change in the estimated effects.

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

9. Discussion

Each commuter use of a bike-share system involves two transactions: the commuter must choose a station with available bikes; and he must also be able to return the bike 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 availability, 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 commuter 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 commuter does not have the option of abandoning the bike and she can complete her trip only by finding another station (commuters using Vélib' get an extra 15 free minutes when the preferred station has no available docking points). Note that in this case the commuter can ride the bike 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 bike. So even though an under-supply of docking points will degrade the commuter 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 commuter response to accessibility (walking distance) and availability (service level) in the context of bike-share systems. Our analysis shows that incorporating these estimates into system design can result in much-improved systems. Our estimates are also applicable in contexts beyond bike-sharing—in particular, for the design of retail distribution networks. Furthermore, the methodology developed here can be used in a variety of demand estimation contexts where products are spatially differentiated and with choice sets that change frequently.

In future work, we hope to address the limitations of this study. First, a more detailed data set on commuter starting locations would improve the precision of estimates such as those obtained here. Second, a larger study comparing many cities could provide insight not only into how commuter 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.

M. Arikan, V. Deshpande, and M. Sohoni. Building reliable air-travel infrastructure using empirical data and stochastic models of airline networks. *Operations Research*, 61(1):45–64, 2013.

E. Belavina, K. Girotra, and A. Kabra. Online fresh grocery retail: A la carte or buffet? *Working paper*, 2014.

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

R. W. Buell, D. Campbell, and F. Frei. How do customers respond to increased service quality competition. *Harvard Business School Accounting & Management Unit Working Paper*, (11-084): 11–084, 2014.

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

G. Cachon and C. Terwiesch. *Matching supply with demand*, volume 2. McGraw-Hill Singapore, 2009.

C. S. Craig, A. Ghosh, and S. McLafferty. Models of the retail location process: A review. *Journal of Retailing*, 60(1):5–36, 1984.

D. W. Daddio. Maximizing bicycle sharing: An empirical analysis of capital bikeshare usage. Master's thesis, University of North Carolina at Chapel Hill, 2012.

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

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.

A. S. Fotheringham. Statistical modeling of spatial choice: An overview. *Spatial Analysis in Marketing: Theory, Methods, and*, pages 95–118, 1991.

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. 2014.

D. L. Huff. Defining and estimating a trading area. *The Journal of Marketing*, pages 34–38, 1964.

Lathia, Ahmed, and Capra. Measuring the impact of opening the london shared bicycle scheme to casual users. *Transportation Research Part C*, 2012.

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

J. Li, N. Granados, and S. Netessine. Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 2014.

M. T. Melo, S. Nickel, and F. Saldanha-da Gama. Facility location and supply chain management–a review. *European Journal of Operational Research*, 196(2):401–412, 2009.

A. Moreno and C. Terwiesch. The effects of product line breadth: Evidence from the automotive industry. *Available at SSRN 2241384*, 2013.

A. T. Murray and X. Wu. Accessibility tradeoffs in public transit planning. *Journal of Geographical Systems*, 5:93–107, 2003.

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

R. Nair and E. Miller-Hooks. Fleet management for vehicle sharing operations. *Transportation Science*, 45(4):524–540, 2011.

R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Bušić. Large-scale vehicle sharing systems: Analysis of vélib'. *International Journal of Sustainable Transportation*, 7(1):85–106, 2013.

A. Negroni. Les embouteillages ont coute 17 milliards d'euros en 2013. *Le Figaro*, October 15, 2014. URL `http://bit.ly/TraffLoss`.

O. O'Brien, J. Cheshire, and M. Batty. Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, 2013.

M. Olivares, Y. Lu, A. Musalem, and A. Schilkrut. Measuring the effect of queues on customer purchases. *Management Science*, (2013), 2011.

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

C. Parker, K. Ramdas, and N. Savva. Is it enough? evidence from a natural experiment in indiaÕs agriculture markets. *Evidence from a Natural Experiment in India's Agriculture Markets (August 1, 2013)*, 2013.

T. Raviv, M. Tzur, and I. A. Forma. Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics*, 2(3):187–229, 2013.

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

E. Rosenthal. Across europe, irking drivers is urban policy. *The New York Times*, June 26, 2011. URL `http://bit.ly/EuroIrk`.

A. Rubin. High levels of pollution spur paris to action. *The New York Times*, March 14, 2014. URL `http://bit.ly/NyTParis`.

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

E. Wong. Most chinese cities fail minimum air quality standards. *The New York Times*, March 27, 2014. URL `http://bit.ly/CNCities`.

J. Ypma. Introduction to ipoptr: an r interface to ipopt. 2010.