Appendix A. De-Seasonalizing Weather Effects

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

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

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

The regression model is given by,

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

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

Table 3 shows the impact of weather variables. Based on the estimated weather effects, station-use Λ_{ft} is de-seasonalized as follows.

Define the net weather effect at time t as,

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

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

Appendix B. Additional Results

B.1. **Relevance Test for Instruments.** In Table 4 we test for the relevance of instruments. We apply the tests for both– the *station average* bike-availability and the *neighborhood average* bike-availability. Specifically, we compare the change in Adjusted R^2 when the proposed instruments are included as dependent variables to explain bike-availability at the *station* $\times$ *time-window* level (Eq. B.1). We divide our instruments in two sets, the ones based on the neighborhood characteristics of the focal station and the ones based on neighboring station's neighborhood characteristics (or equivalently

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

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

TABLE 3. Weather Variables Effect

whether or not c is 0 in $V_{fwj}(a, b, c, d)$). The population density variable $pd(L_f)$ is included in the focal station's neighborhood characteristics.

$$ba_{fw} = \eta_0 + \eta_1 \cdot pd(L_f) + \vec{\eta}_2 \cdot \vec{V}_{fw}(a, b, cd)) \tag{B.1}$$

Dependent Variable:	Station Bike-Availability			Neighbourhood Bike-Availability		
	(1)	(2)	(3)	(4)	(5)	(6)
TwXDistrict F.E.†	Yes	Yes	Yes	Yes	Yes	Yes
Focal Station's Neighbourhood Instruments		Yes	Yes		Yes	Yes
Neighbouring Station's Neighbourhood Instruments			Yes			Yes
Adjusted R^2	0.174	0.204	0.473	0.217	0.251	0.570
Number of observations	5676	5676	5676	5676	5676	5676

†TwXDistrict F.E. sum upto 0 for each District like in structural model

Table 4. Relevance test of Instruments

where (a, b, c, d) take values of $(0, 25, 0, 25)$, $(25, 50, 0, 50)$, $(50, 100, 0, 100)$, $(0, 100, 0, 100)$, $(100, 300, 0, 300)$, $(300, 500, 0, 500)$, $(0, 100, 100, 300)$, and $(0, 100, 300, 500)$. A similar regression is run with $naba_{fw}$ as dependent variable.

We see that both the neighborhood characteristics of the focal station, and those of the neighboring stations are effective instruments.

B.2. **Estimates from the Density Model.** Recall the spatial density distribution we have used is given by,

$$P_w^D\left(L_i; \alpha\right) = \alpha_0 + \alpha_1 \cdot naba_{L_i,w} + \alpha_2 \cdot pd\left(L_i\right) + \vec{\alpha}_{3,w} \cdot \vec{V}_w\left(L_i\right).$$

We report the estimates (α) of our density model in Table 5. We observe that bike-availability, residential users, metro stations, supermarkets (day time only) and cafes are the major contributors of bike-share use.

B.3. **Robusntess w.r.t. Variable Definitions, Model Specification, Computational Choices and Instruments.** We test the robustness of our effect sizes to alternate variable definitions, model specifications, and to computational choices made in model estimation. Table 6 reports the results of our estimation under many alternate assumptions; row (1) replicates our original estimates (from Table 2) 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), Row (3) considers a station stocked in if it has *more than six* bikes available at the station. The estimates are similar to those obtained under our original regressions.

Row (4) replicates our analysis for data from weekends only. We find that the impact of increasing station density and the short-term effects of increasing availability are marginally higher on weekend

	Value	Std. error
Residential Users	0.004	(0.000)***
Metro Stations	0.499	(0.211)*
Intercept	0.000	0.000
Bike-Availability	0.020	(0.002)***
Non Night Hours		
Store	0.000	(0.000)
Food	0.005	(0.033)
Restaurant	0.000	(0.000)
Bar	0.000	(0.000)
Lodging	0.018	(0.017)
Cafe	0.270	(0.039)***
Supermarket	0.391	(0.028)***
University	0.236	(0.053)***
Park	0.000	(0.000)
Museum	0.278	(0.08)***
Library	0.333	(0.183)
Tourist Locations	4.198	(1.099)***
Movie-theater	0.000	(0.000)
Shopping-mall	0.000	(0.000)
Other points of interest	0.005	(0.005)
Tram Line 3a	0.000	(0.000)
Tram Line 3b	0.000	(0.000)
Night Hours		
Store	0.000	(0.000)
Food	0.000	(0.000)
Restaurant	0.000	(0.000)
Bar	0.694	(0.148)***
Lodging	0.139	(0.099)
Cafe	1.316	(0.223)***
Supermarket	0.000	(0.000)
University	0.036	(0.224)
Park	0.000	(0.000)
Museum	0.000	(0.000)
Library	0.000	(0.000)
Tourist Locations	6.253	(1.369)***
Movie-theater	0.000	(0.000)
Shopping-mall	0.000	(0.000)
Other points of interest	0.084	(0.031)**

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

TABLE 5. Density Variables Effect

days as compared to weekdays (5.295% v/s 5.090%, 9.477% v/s 9.399%) while the total (or long-term effect) is noticeably lower (11.137% v/s 12.293%). There are some interesting differences in density variables. As one would expect, users originating from bars, lodging, museums, residences, tourist locations, cafes and other food locations (at nights only) account for a higher proportion of system use on weekends than on weekdays; while universities, libraries and grocery stores (all in the day time only) are higher demand drivers on weekdays than on weekends.

		Primary variables							
		Walking Distance (0-300mts)	Walking Distance (>300mts)	Bike-Availability (naba)	10% increase in Station Density	10% increase in Bike-Availability		Number of observations	$\chi^2(df)$
						Short-term	Total		
(1)	Original Estimates	-2.700	-15.734	0.005	5.090%	9.399%	12.293%	39,302	0.049 (136)
		(0.495)***	(3.043)***	(0.001)***					
(2)	Stockout: ≤ 4 Bicycles	-2.810	-15.375	0.005	5.079%	9.404%	12.443%	39,320	0.047 (136)
		(0.494)***	(3.039)***	(0.001)***					
(3)	Stockout: ≤ 6 Bicycles	-2.611	-15.398	0.005	5.170%	9.407%	12.103%	39,090	0.052 (136)
		(0.493)***	(2.830)***	(0.000)***					
(4)	Weekends only	-3.963	-14.814	0.003	5.295%	9.477%	11.137%	34,374	0.051 (136)
		(0.495)***	(3.099)***	(0.000)***					
(5)	Metro in outside option	0.737	-14.800	0.005	4.721%	9.364%	12.509%	39,302	0.048 (136)
		(0.689)	(2.574)***	(0.001)***					
		Metro outside option							
			0.770						
			(0.166)***						
(6)	Var. of Bike-Availability	-3.534	-11.816	-0.036	5.577%	9.447%	11.714%	39,302	0.045 (136)
		(0.521)***	(2.308)***	(0.002)***					
(7)	Finer Grid Size (20 meters)	-2.633	-15.732	0.003	5.084%	9.399%	12.265%	39,302	0.048 (136)
		(0.501)***	(3.061)***	(0.000)***					
(8)	16 Top states considered	-2.359	-14.823	0.004	5.077%	9.434%	11.821%	71,994	0.040 (136)
		(0.437)***	(2.700)***	(0.000)***					
(9)	Target density −10%	-2.753	-15.786	0.004	5.038%	9.331%	12.221%	39,302	0.049 (136)
		(0.494)***	(3.071)***	(0.000)***					
(10)	Target density +10%	-2.681	-15.694	0.005	5.137%	9.453%	12.347%	39,302	0.049 (136)
		(0.495)***	(3.030)***	(0.001)***					
(11)	Choice set size = 5	-3.728	-9.959	0.005	5.220%	9.392%	12.675%	38,602	0.046 (136)
		(0.516)***	(1.760)***	(0.001)***					
(12)	Focal station Instruments	-4.33	-16.967	0.005	5.278%	9.404%	12.219%	39,302	0.028 (71)
		(0.518)***	(4.561)***	(0.001)***					
(13)	Demeaned Bikes-in$_{w\text{-}1}$	-3.105	-16.311	0.005	5.055%	9.412%	12.571%	39,302	0.058 (137)
		(0.453)***	(3.236)***	(0.000)***					
(14)	Alt. Instrument Parameters - I	-1.894	-15.974	0.005	4.890%	9.394%	12.442%	39,302	0.039 (70)
		(0.590)**	(3.328)***	(0.001)***					
(15)	Alt. Instrument Parameters - II	-3.325	-15.368	0.005	5.009%	9.403%	12.469%	39,302	0.053 (169)
		(0.461)***	(2.984)***	(0.000)***					

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

Table 6. Robustness Tests

In row (5) we include the distance to the nearest metro station as a covariate in the outside option of each user. Metro stations in addition to acting as feeders to use of bike-share stations, can also act as substitutes to bike-share. While, the feeder effect is already captured in the density parameters, inclusion of metro stations as the outside option can further capture any substitution effects. Although we had hoped to find a substitution effect, we find the opposite effect in the model in Row (5). This positive coefficient suggests that irrespective of where the metro variable is included, the net effect of the presence of a metro station is an increase in bike-share use, i.e. metro stations feed demand to bike-share stations rather than act as substitutes. In row (6) we use the variance of bike-availability instead of its mean value as the variable in the density model. We use the same instruments as those in the original model. We find that variability of bike-availability has a long-term negative effect on bike-share use, consistent with our expectations.

Next, we investigate the role of various computational choices made in estimation. In row (7) we provide estimates obtained by using a finer grid for our numerical integration (viz., one that covers 1.5 times as many simulation points for continuous spatial elements) this produces no qualitative change in the estimated effects. In row (8) we consider 16 top states for each *station*× *time-window*. In row (9) and (10) we test the sensitivity of our estimates with respect to the choice of total market density. In row (11) we increase the definition of local choice set of a user to nearby *five* stations. The estimates are exceptionally robust to these choices.

Finally, we investigate the effect of different instruments on our estimates. In row (12) we use only the focal station's neighborhood characteristics as instruments as used in Davis [2006]. In row (13) we use a novel instrument which is the average realized rate of incoming bikes at station f in lagged time-window $w-1$, demeaned at station level. This instrument affects the starting number of bikes at a *station* × *time-window* and therefore its bike-availability. It however does not affect the unobserved factors influencing station-use at f in time-window w, after controlling for demand sources in the density model and station level factors removed in demeaning process. In row (14) and (15) we use alternative parameters for construction of instruments $V_{wj}\,(a, b, c, d)$. The set used for row (14) is $\{(0, 100, 0, 100),\ (100, 300, 0, 300),\ (300, 500, 0, 500)\}$ and for row (15) is $\{(0, 100, 0, 100),$ $(100, 200, 0, 200),\ (200, 300, 0, 300),\ (300, 500, 0, 500),\ (0, 100, 100, 300),\ (0, 100, 300, 500)\}$. The marginal effects are again very close to the original model.

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

B.4. **Robustness of Distance Disutility Function.** Our main model assumes a piecewise linear form for the disutility of distance. We reran our model with linear and quadratic disutility functions, and a variety of different kink points for the piecewise linear form– 100, 200, 250, 275, 325, 350m instead of 300m in original model.

Irrespective of the specification used for the distance disutility, our estimate for the effect of *bike-availability* is essentially identical. In so far as the effect of *distance* is concerned, depending on the specification we get different estimates for parameters of the distance disutility function, but remarkably irrespective of the specification, the the effects of walking distance on ridership is essentially the same– a 10% increase in station density always leads to between 4.56%-5.19% increase in system use, with all but 3 specifications returning an effect within 2% of our original estimate. Also notably, irrespective of the functional form– all specifications imply the disutility from distance is convex. The extent of substitution or the short-term effect of bike-availability which derives from the accessibility effect is again essentially identical in all specifications ranging from 9.399% to 9.435%. Finally, the total effect ranges from 11.970% to 12.293%, The full results for all these alternate specifications are reported in Table 7 of the Appendix.

B.5. **Robustness of the Distance Disutility Function.** Table 7 reports the estimates with a variety of alternate distance disutility functions– with alternate kink points in the piecewise linear form– 100, 200, 250, 275, 325, 350 meters instead of 300 meters in the original model, a simple linear function, and a quadratic function. Section B.4 provides a discussion of these estimates.

B.6. **Comparison of Estimates.** We compare our estimates with estimates (or in some cases decisions implied by those estimates) from reduced-form analysis and other studies in the bike-share, public-transport and retail-store network design contexts.

Comparison of Distance Estimates. Reduced-Form Analysis: While it is hard to measure the accessibility effect directly from a reduced form analysis,[13] we can use some reduced form analysis for the substitution effect which derives directly from the marginal disutility of distance and the existing station network design. We look at the difference between the use at a station when its neighboring stations have available bikes and when they do not. In a regression model, we explain the log of station-use using the stockout-state of neighboring stations, specifically we include a dummy variable based on whether or not the nearest station to the focal station has any bikes available. We find an average *6.056%* increase in use at a station when the nearest station runs out of available bikes, a number very close to the substitution percentage implied by the distance estimates in our model.

Bike-Share Systems: We are aware of no other econometric study (survey or archival-data based) that has attempted to estimate the disutility of distance in the context of bike-share systems. In

[13]The effect of increasing distances between stations from any reduced form station-level model necessarily confounds the larger catchment area effect with user response effect

		Primary variables: Walking Distance (until kink)	Walking Distance (after kink)	Bike-Availability (naba)	10% increase in Station Density	10% increase in Bike-Availability: Short-term	Total	Number of observations	$\chi^2(df)$
(1)	Original Estimates	-2.700 (0.495)***	-15.734 (3.043)***	0.005 (0.001)***	5.090%	9.399%	12.293%	39,302	0.049 (136)
(2)	Kink at 100mts	-0.148 (2.115)	-7.186 (0.558)***	0.004 (0.000)***	4.830%	9.427%	12.021%	39,302	0.050 (136)
(3)	Kink at 200mts	-2.220 (0.779)**	-9.058 (0.988)***	0.005 (0.000)***	5.010%	9.409%	12.154%	39,302	0.050 (136)
(4)	Kink at 250mts	-2.342 (0.595)***	-11.472 (1.563)***	0.005 (0.000)***	5.103%	9.401%	12.266%	39,302	0.050 (136)
(5)	Kink at 275mts	-2.628 (0.535)***	-13.094 (2.085)***	0.005 (0.001)***	5.103%	9.400%	12.280%	39,302	0.049 (136)
(6)	Kink at 325mts	-2.676 (0.468)***	-19.579 (4.799)***	0.005 (0.001)***	5.009%	9.398%	12.286%	39,302	0.048 (136)
(7)	Kink at 350mts	-2.639 (0.455)***	-25.513 (8.619)**	0.005 (0.001)***	4.817%	9.397%	12.246%	39,302	0.048 (136)
		Walking Distance	Walking Distance Square	Bike-Availability (naba)					
(8)	Linear Model	-6.459 (0.413)***		0.004 (0.000)***	4.560%	9.435%	11.970%	39,302	0.050 (136)
(9)	Quadratic Model	2.158 (1.468)	-16.099 (3.602)***	0.005 (0.001)***	5.193%	9.422%	12.245%	39,302	0.049 (136)

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

Table 7. Robustness of Distance Effect

practice, European bike-share system designers follow a common handbook (Büttner and Petersen, 2011) which suggests that very few users walk further than 300 meters and provides station location guidelines based on this assumption. This guideline is squarely in line with our estimates, our distance estimate also implies that only 11.019% of the use of a station comes from users who walk further than 300 meters.

Other Public Transport Systems: Several studies survey the walking distance of users of bus, light rail and metro systems. [O'Neill et al., 1992, Zhao et al., 2003] report that 75-80% users of public-transit systems walk less than 400 meters. El-Geneidy et al. [2014] finds that, in Montreal, the 85th percentile walking distance to the bus (resp., rail) transit system is about 524m (resp., 1,259 m). O'Sullivan and Morrall [1996] reports that transit planners in several Canadian and American cities consider the catchment area to be no further than 300-900m, with the median light rail user in Calgary, Canada walking 320m. Alshalalfah and Shalaby [2007] report that median access distance of bus users in Toronto is about 200m and that of subway users is 350m. In comparison, our estimates are marginally lower, our median user walks about 186m, and almost 90% of the demand comes from the first 300m. Since bike-share systems are used for much shorter trips than those taken by other public transport systems, bike-share systems exist in more densely populated areas, and have a much denser station network; it is expected that our users walk less.

Zhao et al. [2003], based on survey data of about one thousand users' transit use (bus or rail) in southeast Florida, determines that usage decreased exponentially with a coefficient of *-4.265/km.* Gutiérrez et al. [2011] uses survey data from the Madrid metro network to estimate the effect of walking distance and finds an average distance disutility coefficient of about *-1.689/km.* Our comparable estimate, *-2.7/km,* is squarely in line with these observed coefficients.

Retail Store Networks: While the context of retail store networks provides us with multiple past studies to compare our estimates, these estimates typically consider the disutility of distance for users who *drive* to the retail locations. One way to compare with these estimates is to convert distances to commuting time using average walking and driving speeds. We compare our estimates with those in Davis [2006] (driving to movie theaters), Pancras et al. [2012] (driving to grocery stores), and [Thomadsen, 2005, Allon et al., 2011] (driving to drive-through fast food outlets). Our estimate is higher than that of Davis [2006] and Pancras et al. [2012], comparable to those in Allon et al. [2011] and lower than that in Thomadsen [2005]. Together, our estimate is again squarely in line with the average of these past studies. The average disutility of commuting time from these studies is *14.497/hr* while, for majority of the users in our study, it is *16.875/hr.*[14]

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

Anderson et al. [2006] in their study of a home-bedding catalogue retailer find that a 10% decrease in stockouts leads to a 7.2% short-term increase in product sales. The lost-demand in their case is much lower than ours (28% in their case compared to 94% in our case) probably because users are more

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

willing to wait for bedding ordered via a catalogue, than for bikes to get somewhere. The long-term impact of all items ordered by a customer in their setting being out of stock compared to none is 22% lower future demand, i.e. a 10% decrease in stockouts leads to a 2.2% long-term increase in product demand. This estimate is comparable to our estimate of a 2.645% increase in long-term demand due to a 10% higher bike-availability.

Musalem et al. [2010] in their study on estimating the effect of stockouts in the shampoo category find that almost no sales are lost when a few brands stock out, but as much as 20.02% of sales might be lost when multiple brands stockout, suggesting there is more than 80% substitution to adjacent shampoo brands. Not surprisingly, there is much less substitution to adjacent stations in our context (only about 6%), perhaps because users must walk to other stations rather than just simply switch to comparable, adjacent brands.

Appendix C. Estimation Details

C.1. **Estimation Procedure.**

Estimation. The estimation procedure introduced in section 5.3 is as follows. The set of moment conditions are given by,

$$\begin{aligned} E\left[Z_{fw}\sigma_{fwv_f}\xi_{fwv_f}\left(\theta^*\right)\right] &= 0 \quad \text{, and} \\ E_w\left[\gamma^*_{w\times di}\right] &= 0 \text{ for } \forall di \end{aligned}$$

The constraints used to determine values of all $\xi_{\cdot w}$ $\left(\xi'_{fwv_f}s\right)$ are,

$$\lambda_{fwv_f}\left(\theta, \xi_{\cdot w}\right) = \Lambda_{fwv_f} \quad \forall f, w, v_f \,.$$

We rewrite above moment conditions and constraints in a GMM formulation in Eq's. C.1 below.

The moment conditions vector given by $G\left(\theta, \xi\right)$ is,

$$\begin{aligned} G\left(\theta, \xi\right) &= \frac{1}{N}\sum_{f,w,v_f} Z_{fw}\cdot\sigma_{fwv_f}\cdot\xi_{fwv_f} \\ &= \frac{1}{N}Z^T\Sigma\xi \end{aligned}$$

where Z, Σ, and ξ are matrix and vector notations for Z_{fw}, σ_{f,w,v_f} and ξ_{fwv_f} respectively over all observations. Note that Σ is a diagonal matrix and N is number of observations f, w, v_f.

We also introduce a change of variable, so that the moment conditions are treated as additional parameters as suggested by Dubé et al. [2012], which makes the hessian matrix sparse.

The GMM estimator is given by,

$$\hat{\theta}^{*} = \arg\min_{\theta} \eta^{'} A \eta \tag{C.1}$$

$$\text{s.t.}$$

$$\sum_{w} \gamma_{w \times di} = 0 \quad \forall di$$

$$\lambda_{fwv_f}\left(\theta, \xi_{.w.}\right) = \Lambda_{fwv_f}$$

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

$$\int_{w} \int_{L_i} P_w^D \left(L_i; \theta\right) dL_i = T^D$$

where A is the GMM weighing matrix.

Note that each computation of λ_{fwv_f} as per Eq. 5.5 involves integrating over the spatial density of users. We divide the density elements into two components, the discrete density elements, $\vec{V}_w$, such as metro stations, movie theaters, etc. and the continuous density elements which are the population density, bike-availability and intercept term. The integration over latter density elements is performed numerically. We discretize the physical area of the city of Paris into a grid composed of squares with length $\mathscr{D}$ meters; we consider the center of each such square to be a point mass of users. Predicted use is then

$$\lambda_{fwv_f}\left(\theta, \xi_{.w.}\right) = \sum_{i \in Points\_of\_Interests} p_{ifwv_f}\left(\theta, \xi_{.w.}\right) \cdot \left(\vec{\alpha}_{3,w} \cdot \vec{V}_w\left(L_i\right)\right) + \sum_{j \in Grid(\mathscr{D})} p_{jfwv_f}\left(\theta, \xi_{.w.}\right) \cdot \left(\alpha_0 + \alpha_1 \cdot naba_{L_j,w} + \alpha_2 \cdot pd\left(L_j\right)\right) \cdot \mathscr{D}^2,$$

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

The simplest way of estimating our model would be to search over the parameters θ for values that provide the best fit. This would require a search over a space with as many dimensions as parameters (including numerous fixed-effects parameters), resulting in several search iterations each of which is computationally expensive. We instead estimate our model using a process that relies on some parameters (all except density model parameters $\vec{\alpha}$ and distance coefficient β_d) entering our model in a "user-location–agnostic" way (Berry et al. [1995]). We thus group our parameters in two classes, first as $\theta_1 = (\alpha,\, \beta_d)$, and the parameters that are "linear" (in ξ_{fwv_f}) as $\theta_2 = (\beta_0, \vec{\gamma})$.

We rewrite the p_{ifwv_f} and λ_{fwv_f} in terms of composite terms δ_{fwv_f} :

$$\delta_{fwv_f} = \beta_0 + \gamma_{w\times di(f)} + \xi_{fwv_f}$$

The user choice probabilities and station-use are now written as a function of θ_1 and δ. The user choice probabilities p_{ifwv_f} is now given as,

$$p_{ifwv_f}(\theta_1, \delta_{\cdot w \cdot}) = \frac{\exp\left(h(\beta_d; d(L_i, L_f)) + \delta_{fwv_f}\right)}{1 + \sum_{g\in N_i \cap S_{v_f}} \exp\left(h(\beta_d; d(L_i, L_g)) + \delta_{gwv_f}\right)},$$

where S_{v_f} denotes the set of stations with available bikes in state v_f. δ_{gwv_f} are estimated as

$$\hat{\delta}_{gwv_f} = \frac{\sum_{v_g} \sigma_{gwv_g}\delta_{gwv_g}}{\sum_{v_g}\sigma_{gwv_g}}.$$

Then station-use λ_{fwv_f} is given by,

$$\lambda_{fwv_f}(\theta_1, \delta_{\cdot w \cdot}) = \int_{L_i} p_{ifwv_f}(\theta, \delta_{\cdot w \cdot}) \cdot P_w^D(L_i; \alpha)\, dL_i\,.$$

The estimation process searches over values of θ_1 and δ. Given the values of θ_1 and δ, the values of coefficients θ_2 are determined non-iteratively from the closed-form expression below which follows from our moment conditions in Eq. C.1:

$$\hat{\theta}_2\left(\delta_{fwv_f}\right) = \left(\left(X_2^T\Sigma Z\right)A\left(Z^T\Sigma X_2\right)\right)^{-1}\left(\left(X_2^T\Sigma Z\right)A\left(Z^T\Sigma\right)\right)\delta_{fwv_f}. \tag{C.2}$$

where X_2 is the co-variate matrix corresponding to the equation, $\delta_{fwv_f} = \beta_0 + \gamma_{w,di(f)} + \xi_{fwv_f}$, consisting of an intercept column and *time-window*$\times$*district* dummies. Thus, in each iteration, values of θ and ξ are obtained for given values of θ_1 and δ, and GMM objective function is computed.

In the first step of the GMM, we use $\left(Z^T\Sigma^2 Z\right)^{-1}$ as the weighing matrix A. We find the condition number of the matrix inversion step in Eq. C.2 to be low when using this weighing matrix, in comparison to say an identity matrix which renders the conditions number quite high. This is analogous to the weighing matrix used in the 2SLS procedure.

Standard error. The variance estimate of $\hat{\theta}^*$ is given by,

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

where, $\hat{G} = \frac{\partial G}{\partial \theta}|_{\theta=\hat{\theta}^*}$ is the first derivative of moment conditions G and $\tilde{S}^{-1}$ is the optimal GMM weighing matrix (sec 6.3.5. Cameron and Trivedi [2005]).

$\tilde{S}$ is given by,

$$\tilde{S} = \frac{1}{N} \sum_{f,w,v_f} \left(Z_{fw} \cdot \sigma_{fwv_f} \cdot \xi_{fwv_f} \right) \cdot \left(Z_{fw} \cdot \sigma_{fwv_f} \cdot \xi_{fwv_f} \right)^T .$$

Implementation Details. The procedure was implemented in R. The open-source package IPOPT (Interior Point Optimizer) (interfaced with R via "ipoptr" [Ypma, 2010]) was used for nonlinear optimization with constraints. The "ffdf" class in R was employed to accommodate the large scale of our data set. Even though we transformed our problem from the time domain to the local stockout state domain, computing the choice probabilities for each user, and then summing over them, was computationally expensive; the initial runtime was of the order of tens of days on a contemporary computer of the workstation class. Implementing the station-use computation function (Eq. 5.5) in C++ and then interfacing with R reduced the computation almost 100 times, to about 70 hours for the Paris data set.

C.2. **Comparison with Past work.** As discussed in the introduction, our model follows Davis [2006]. As in Davis [2006], in our model different service locations (stations in our context, movie theaters in Davis [2006]) are the differentiated products, the differentiating characteristic is the distance a user must walk to access the locations, and the unobserved heterogeneity is the user's origin location. Incorporating this heterogeneity in our model ensures that when a station stocks out, its users are more likely to substitute to nearby rather than distant stations. Station locations and bike-availabilities are the endogenously determined attributes, akin to theater locations and prices in Davis [2006].

Past work in consumer choice models (including the seminal works of [Berry et al., 1995, Nevo, 2001], and Davis [2006] in the spatially differentiated retail choice context) assumes that all *offered* products are always available. This has been shown to substantially bias parameter estimates in the case of consumer goods [Bruno and Vilcassim, 2008, Conlon and Mortimer, 2013]. Bike-availability is typically ~60-70%, much lower than the 90% or so availability in the case of consumer goods, and arguably more important to users, thus assuming full availability is likely to bias estimates even more in our context. Further, this would run counter to one of our key goals– measuring the distinct short and long-term impacts of bike-availability. Anupindi et al. [1998], Musalem et al. [2010], Bruno and Vilcassim [2008] and Conlon and Mortimer [2013] address product availability issues by developing methods to estimate model parameters in presence of limited or no product availability information to the econometrician. Yet, we have information on the actual realizations of product (bike)-availability and we include it directly in our model.

The real-time bike-availability enters our model indirectly via the relevant choice set that is realized at each time, S_t, in Eq. 4.2. When stations are stocked-out they do not enter any user's choice set and serve no users. The effect of this on system-use depends on the extent of customer substitution to nearby stations– if all customers substituted then there would be no effect on system-use; if none substituted, system-use would change by the same amount as a fraction of time the station is stocked out. We call this the short-term impact of bike-availability and it is estimated through the substitution pattern that is embedded in our choice model.

Demand for bike-share is hyper-local, the typical catchment area that a bike-share station serves is much smaller than that for retail stores considered in past work, as users must walk rather than drive and the networks themselves are much more dense. We thus build and estimate a much finer hyper-local parametric spatial density model for potential-user origins as compared to Davis [2006] (Eq. 4.4), aided by the much freer availability of mapping data today from a variety of Google products. More importantly, this hyper-local density model allows us to include a measure of the typical or average bike-availability in a neighborhood, which likely drives user interest in bike-share, or the above discussed long-term effect of bike-availability.

C.3. **Comparison of the Computation Challenge.** We noted two modeling choices that result in extremely large computational burden necessitating us to devise our local-stockout state based transformation procedure. These were

1) the rapidly changing choice sets of stations available in the high frequency data-set, and
2) the spatial nature of the product which requires a fine grained spatial user density model.

Bruno and Vilcassim [2008], Conlon and Mortimer [2013] have shown that the estimates could be substantially biased because of not taking into account real-time availability information (as is implicitly the case in BLP and most applications of it). A fine grained user density model is also necessary in our context because the usage of bike-share systems tends to be quite local in nature, i.e. the the size of potential users could substantially change within a matter of 100-200 meters.

Comparison with work that has accounted for availability information (None in spatial context). Bruno and Vilcassim [2008] have access to only average product availability information. Assuming independent availabilities, Bruno and Vilcassim [2008] extend the BLP model to account for them. In presence of exact availability information, their model resembles our model in time-domain. Bruno and Vilcassim [2008] consider 24 products (as compared to 946 in our case) for 113 four week periods (as compared to over 22,000 in our case). Conlon and Mortimer [2013] consider 44 products in a vending machine application in 44,458 four-hour time periods. Musalem et al. [2010] consider

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

Table 8. Comparison of data size

24 products for 15 days of data. These papers have users which were not spatially differentiated. There is heterogeneity in user tastes due to normally distributed random coefficients, however the number of draws required to aggregate over these heterogeneous users is much lower. For example, the supplementary code in Nevo [2000] uses 20 draws, and Dubé et al. [2012] use 1000 draws as compared to the more than 210,000 spatially heterogenous users in our case.

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

The comparison with past work is summarized in Table 8. The comparison illustrates how the combination of rapidly changing choice sets and a fine grained user density model, in a relatively large scale data set, leads to an explosion in computational requirements.

C.4. **Validation in Simulated Datasets.** While the full validation of our approach remains the subject of a dedicated study that considers many alternate contexts, we provide a limited validation in our context. Specifically, we validate the use of the local-stockout state transformation and our computational choices–the limits on the choice set and the use of top states– on smaller simulated datasets, where both our approach and the full approach (time-domain, no limits on choice sets, all states) are computationally feasible. We created a number of small simulated datasets for demand at 30 stations around the city-center (Hôtel de Ville) for 50 two-minute time-intervals in the evening-rush time window. The detailed data generating process, analysis and the results are provided in Section C.5 of the Online Appendix.

We estimate our model on the simulated datasets in three ways: (1) The benchmark estimation procedure that uses the untransformed time-domain based moment conditions (Eq. 5.1) and places no limits on the choice set of customers. (2) Using the transformed local stockout state based conditions (Eq. 5.6) and imposing a consistent limit on the choice set of the customer and (3) Approach (2) plus focusing on just enough local stockout states to cover 75% of the data for the typical station (the approach of this paper).

We find that all three approaches recover seed estimates from the demand model. Specifically, the recovery from the first procedure provides support for the moment conditions used in our estimation and validates our approach, while the recovery from the last two procedures validates the use of the *top* states among the *local* stockout states. Interestingly, while all three procedures recover the seed estimates, the computational burden of the third approach is an order of magnitude less than that of the untransformed approach, even in these small datasets. We expect the difference to be much larger in a dataset comparable in size to the one in our study. Taken together, while this analysis provides some validation of our approach– a full validation of such transformations in other contexts remains the subject of a future study focused on further developing the methodological ideas here.

The estimation procedure and model described in the preceding sections provide us consistent estimates of the accessibility (distance) effect using the variation across stations and between different local-stockout states for stations, while the long-term bike-availability effect is identified using the variation across stations and time-windows.[15] The short-term effect of bike-availability derives from the estimated consumer utility (specifically the marginal disutility of distance) and the structure of the station network.

Together, our approach allows us to impute preferences of *heterogenous customers'* (in origin location) from *aggregate* choices (we only observe aggregate station-use), to efficiently include instruments for potentially *endogenous attributes* of bike-stations (for e.g. bike-availability determined by system managers), and to include a large number of fixed effects in our model (the numerous *station* $\times$ *time-window* $\times$ *local-stockout-state* (ξ_{fwv_f}'s) and *time-window* $\times$ *district* ($\gamma_{w\times di}$'s)). These are all key advantages of BLP-like models. Further, our enhancements allow us to bring these features

[15]To ensure cross-section variation in bike-availabilities is meaningful and persistent over time, we look at the correlation in bike-availabilities defined at monthly level for a *station* $\times$ *time-window* and find it to be *0.89*, high enough to justify using it as such.

to the study of the hitherto unexplored operational issue of availability (stockouts) in a hyper-local smart-transportation setting.

C.5. **Validation in Simulated Datasets.** While the full validation of our approach remains the subject of a dedicated study that considers many alternate contexts, we provide a limited validation in our context. Specifically, we validate the use of the local-stockout state transformation and our other computational choices- the use of top states and limits on the choice set– on smaller simulated datasets, where both our approach and the full approach (time-domain, all states, no limits on choice sets) are computationally feasible.

Data Generation. We created a number of small simulated datasets for demand at 30 stations around the city-center (Hôtel de Ville) for 50 two-minute time-intervals in the evening-rush time window. In particular, using the average bike-availabilities for each of these stations, we simulate the real time bike-availabilities for each station and 2 minute-interval, resulting in the relevant choice set information. The unobserved station-time characteristics ξ_{ft} are drawn from a Normal distribution with mean 0 and variance of 0.1. We combine these with the distance coefficients $(-2.700, -15.734)$ from our estimated structural model to obtain the choice probabilities for infinitesimal users located at each point in the city. Finally, for each of the 6400 density-relevant grid-points (points of interest, transit, etc.), we generate the rate of a potential trip originating in a 2 minute interval using a lognormal distribution with mean 1 and variance 0.1.[16] The choice probabilities combined with this density model that incorporates point of interest location data give us the simulated station-level demand for our station network for each two-minute interval.

Alternate Estimation Methods. We estimate our model on the simulated datasets in three ways illustrated in Table 9:

(1) The benchmark estimation procedure that uses the untransformed time-domain based moment conditions (Eq. 5.1) and places no limits on the choice set of customers. (2) Using the transformed local stockout state based conditions (Eq. 5.6) and imposing a consistent limit on the choice set of the customer and (3) Approach (2) plus focusing on just enough local stockout states to cover 75% of the data for the typical station (the approach of this paper). The results for 10 sample datasets are reported in Table 10. We find that all three approaches recover estimates that are reasonably close to our seed estimates. Specifically, note that the time-domain based estimation procedure is able to recover the seed estimates from the demand model (see mean estimate from columns "Time-Domain" in Table 10) thus providing support for the moment conditions used in our estimation

[16]We also test alternate (more and less dispersed) distributions for the unobserved station-time characteristics and the rate of potential trips originating.

Estimation Method	Moment Conditions	Limit on Choice Set	Number of States Considered	City Discretization
Data Generation	N.A.	None	N.A.	25 m
No Transformation or Computational Choices (Benchmark)	Time-Domain	None	N.A.	25 m
Local Stockout State	Local Stockout-State domain	Closest 4, $m_d = 4$	All	25 m
Top Local Stockout States	Local Stockout-State domain	Closest 4, $m_d = 4$	75% data	25 m

Table 9. Alternate Computational Models

	No Transformation (Benchmark)		Local Stockout State		Top Local Stockout States	
Simulated Dataset #	Walking Distance (0-300mts)	Walking Distance (>300mts)	Walking Distance (0-300mts)	Walking Distance (>300mts)	Walking Distance (0-300mts)	Walking Distance (>300mts)
(1)	-3.487	-15.122	-3.341	-15.106	-3.486	-14.858
(2)	-3.026	-15.680	-2.987	-15.317	-2.673	-16.068
(3)	-2.598	-16.868	-2.594	-16.529	-2.376	-17.428
(4)	-3.116	-14.731	-3.145	-14.215	-2.781	-14.972
(5)	-2.723	-14.415	-2.725	-13.869	-2.417	-14.999
(6)	-2.392	-18.289	-2.263	-18.567	-2.204	-18.352
(7)	-3.275	-14.585	-3.251	-14.228	-3.270	-14.153
(8)	-2.729	-15.163	-2.728	-14.698	-2.547	-15.377
(9)	-2.151	-15.904	-2.077	-15.644	-2.139	-15.244
(10)	-3.712	-14.962	-3.756	-14.261	-3.701	-14.114
Mean	-2.921	-15.572	-2.887	-15.244	-2.760	-15.556
Std. Dev.	0.491	1.199	0.511	1.419	0.546	1.370
Computation Time	41534 sec		4869 sec		2709 sec	

Table 10. Simulation analysis results

and generally validating our approach. The recovery of estimates by the local stockout state based estimation (columns "Local-stockout state" in Table 10) validates the use of our local stockout state transformation and the computational choices of limiting the choice set to 4 closest stations embedded therein. Finally, the estimates obtained by looking at just the top local stockout states (column "Top Local Stockout States" i.e. the approach in our paper) demonstrates that using the top states is sufficient.

Interestingly, while all three procedures recover the seed estimates, the computation burden of the third approach is an order of magnitude less than that of the untransformed approach, even in this small dataset. We expect the difference to be much larger in a dataset of the scale of our study. Taken together, while this analysis provides some validation of our approach– a full validation of such transformations in other contexts remains the subject of a future study focused on further developing the methodological ideas here.