

An Empirical Analysis of Bike Sharing Usage and Rebalancing: Evidence from Barcelona and Seville

Ahmadreza Faghih-Imani[†], Robert Hampshire^{*}, Lavanya Marla^{**}, and Naveen Eluru^{††}

[†] Department of Civil Engineering and Applied Mechanics, McGill University,
seyed.faghihimani@mail.mcgill.ca

^{*} Transportation Research Institute, University of Michigan, hamp@umich.edu

^{**} Department of Industrial and Systems Engineering, University of Illinois at Urbana-Champaign, lavanyam@illinois.edu

^{††} Department of Civil, Environmental and Construction Engineering, University of Central Florida, naveen.eluru@ucf.edu

Abstract

Over 400 cities around the world have deployed or have plans to deploy a bike sharing system. However, the factors that drive their usage and the amount of rebalancing they require are not known precisely. A knowledge of these factors would allow cities to design or modify their systems to increase usage while lowering rebalancing costs. We collect station-level occupancy data from two cities and transform station occupancy snapshot data into station level customer arrivals and departures to perform our analysis. Specifically, we postulate that arrivals and departures from stations can be separated into: (i) arrivals (and departures) due to consumers, and (ii) arrivals (and departures) due to the system operators for rebalancing the system. We then develop a mixed linear model to estimate the influence of bicycle infrastructure, socio-demographic characteristics and land-use characteristics on customer arrivals and departures. Further, we develop a binary logit model to identify rebalancing time periods and a regression model framework to estimate the amount of rebalancing. The research is conducted using bike sharing data from Barcelona and Seville, Spain. The resulting modeling framework provides a template for examining bicycle rebalancing in different contexts, and a tool to improve system management of bicycle sharing systems.

Keywords: bike sharing, rebalancing, linear mixed model, Points of Interest

1. Introduction

Bike sharing systems are an emerging mode of transportation that provide the temporary rental of publicly available bicycles. These programs have the potential to reduce car usage in dense neighborhoods, hence reducing congestion; additionally they promote healthy living and are environmentally friendly. Over 400 cities have operating bike sharing programs worldwide, including North and South America, Europe and Asia (see <http://bike-sharing.blogspot.com>). In the process, many cities and planners have conducted feasibility studies of existing and proposed bike sharing systems. These studies often include demand estimation and corresponding methodology. For example, Philadelphia (DVPR, 2011), New York (NYC, 2011), London (TFL,

2011), and many others have published bike sharing feasibility studies. In these reports, many assumptions and hypotheses are presented for demand estimation. The feasibility studies make hypotheses about various socio-demographic, land use, economic and infrastructure factors. The work of Krykewycz et al. (2011) is representative of demand estimation methodology for bike sharing systems. They hypothesize about factors which contribute to trip generation and trip attraction. These factors are grouped into three main categories: socio-demographic, land-use and infrastructure spatial attributes. Our approach also uses these three categories while employing system usage data.

The goal of this paper is to explain the factors influencing trip generation and trip attraction from station occupancy snapshot data. Earlier studies have shown that snapshot data can provide a useful representation of bike share usage (de Chardon and Caruso, 2015). We transform station occupancy snapshot data into two categories: (i) customer arrivals and departures, and (ii) system operator rebalancing (removal and refill), in order to perform this analysis. Augmenting the system data from Barcelona and Seville in Spain with census level socio-demographic data, and points of interest data, we test the hypotheses about the various factors affecting bike sharing usage. We select Barcelona as the primary city for our study because of its popularity and large number of bike trips. Additionally, Barcelona has been previously studied in the literature, see Froehlich et al. (2008 and 2009). The goal of these two papers is to predict the availability of bikes at each station. In contrast, our goal is not prediction, but explaining the factors that contribute to trip generation and attraction. We also select Seville as a case study because of the urban redevelopment and innovative transport planning that has taken place over the last ten years (Cycling Mobility, 2011).

The proposed research effort evaluates two hypotheses and their corresponding questions. First, we hypothesize that bike-sharing demand is influenced by bicycle infrastructure (bicycle station numbers and capacity), land use (population density, employment density and points of interest) and temporal variables (such as temperature and humidity). We answer the question of quantifying the impact of these various factors on bike-sharing demand, in a multivariate setting. Second, we hypothesize that rebalancing requirements at each station can be partitioned into the quantity of rebalancing at a station and the frequency of rebalancing. While these metrics are likely to be substantially affected by the same set of variables that influence bike-sharing demand, we expect the influence of these variables to be different for rebalancing. We answer the question of how the influences of these factors differ for rebalancing, as compared to demand, in magnitude and sign. For instance, while mixed land use areas increase bicycle sharing demand they might also reduce demand for rebalancing as bicycle flows occur across all parts of a dense neighborhood.

The outline of this paper is as follows. In Section 2, we review relevant literature. In Section 3, we discuss the unique data sources used in the empirical estimation strategy. We present the linear mixed model and discuss the most salient challenges and solutions to its estimation in Section 4. In Section 5, we discuss the estimation results for Barcelona and Seville, including the explanations for user trips, rebalancing operations caused by imbalance of bikes in the system, as well as policy implications for system design. We conclude in Section 6.

1.1 Contributions

The major contributions of this paper are as follows. First, we generate a unique dataset through a combination of three data sets – (1) station occupancy snapshots collected from the

operators' websites; (2) socio-demographic, economic and housing data from census data (for this paper, from Eurostat); and (3) Points of Interest data that describe land-use from TeleAtlas. Second, we separate user arrivals and departures from operator rebalancing (removal or refill of bikes) using a heuristic approach. Through empirical analysis we test assumptions about the factors that influence customer arrivals and departures and rebalancing refill and removal. We apply a methodology for analyzing such systems using behavioural models, specifically, linear mixed models. Finally, we present the first empirical analysis of system rebalancing by the operator focused on understanding the factors creating such imbalances, using an approach consisting of a binary logit model (for identifying stations that need rebalancing) and a linear regression model for the amount of rebalancing. This analysis can help in creating plans for rebalancing well in advance, as well as in creating incentive mechanisms for customers to rebalance bikes.

2. Literature Review

Demand estimation for non-motorized travel modes have been fairly well-studied (see Rietveld et al., 2001, Cao et al., 2006, Chatman, 2005, Handy et al., 2006, Kitamura et al., 1997 and Schwanen and Mokhtarian, 2005). The FHWA report (FHWA, 1999) describes key factors for trip generation for non-motorized modes. Other studies have attempted the demand analysis of bicycle usage and trip rates (Chatman, 2005), usage and mode choice (Baltes, 1996, Beck and Immers, 1996, Cervero and Duncan, 2003 and Hunt and Abraham, 2007) and usage and travel mileage (Ewing et al., 2005). There are relatively fewer studies about bicycle ownership and its relation to bicycle use. Also, measures of bicycle usage are subject to inaccuracy in travel surveys and are therefore often poorly documented (BTS, 2000).

Bike sharing systems, however, are expected to have some similar and some differing characteristics compared to other non-motorized modes. For example, a characteristic differentiating bike sharing systems from other non-motorized systems is that they do not necessitate ownership of bikes and therefore facilitate increased complementarity between biking and transit. A characteristic common to bike sharing and other non-motorized systems is the age group population that is expected to favor their use.

There is emerging literature on bike sharing systems. Studies such as Shaheen et al. (2010) and deMaio (2009) have described the history of bike sharing systems while Carballeda et al. (2010) survey public bike systems in Spain. The majority of quantitative studies focus on state prediction of the system using time series models. Borgnat et al. (2009a, 2009b, 2010), Kaltenbrunner et al. (2010) and Vogel & Mattfeld (2010) present time series models of bike sharing. Jensen et al. (2010) also infer the travel speeds of bikes in the bike sharing program in Lyon. Froehlich et al. (2008, 2009) use Bayesian networks and clustering to predict bike availability.

Recently, there have been several quantitative studies examining bike sharing systems from different dimensions. For example, several studies demonstrate that increasing bike sharing systems infrastructure (number of stations and capacity) or increasing bicycle routes around stations increases bike sharing systems usage (Buck and Buehler, 2012, Faghih-Imani et al., 2014, and Wang et al., 2015). Faghih-Imani and Eluru (2014) found evidence for the self-selection hypothesis indicating that ignoring the installation decision process in modeling usage tends to over-estimate the impact of bike sharing system infrastructure. Studies found that stations in areas with higher job or population density or stations with higher number of point of interests (such as

restaurants, retail stores and universities) in the vicinity experience higher arrivals and departures (Rixey, 2013, and Faghih-Imani et al., 2014). Furthermore, the relationship between bike sharing systems and other public transportation systems such as subway or bus transit system are also examined by several research efforts (Nair et al., 2013, Faghih-Imani et al., 2014, Faghih-Imani and Eluru, 2015 and González et al., 2015). Analyses on temporal attributes of bike sharing systems show that the peak usage is observed during the evening peak hours while weekdays tend to have higher rates of usage compared to weekends, indicating that bike sharing systems are used on weekdays for commuting purposes (O'Brien, 2014, Faghih-Imani et al., 2014, and Murphy and Usher, 2015). Several studies analyze the impact of weather characteristics (such as temperature and humidity) on the usage of the bike sharing systems (Gebhart and Noland, 2014, and Faghih-Imani et al., 2014).

Another stream of literature focuses on operational efficiency of bike sharing systems. Nair et al. (2013) characterize the spatial-temporal supply and demand asymmetries inherent in bike sharing systems. Complementarily, Raviv et al. (2013) and Lin et al. (2011) seek to address these asymmetries by optimizing bike repositioning operations.

2.1 Trip Generation and Attraction Factors

Many cities have released bike sharing feasibility and demand forecasting studies which use similar methodologies (see New York City (NYC, 2011), London (TFL, 2011), and Philadelphia (DVRP, 2011). Most of these studies posit three main user groups for bike sharing: commuters, students and tourists. The feasibility studies typically use stated preference surveys and census tract level data to estimate uptake rates for each user group. Additionally, the surveys attempt to ask questions which lead to estimates of trip substitution and mode share changes. Our approach is different in that we are estimating trip generation and attraction factors from *revealed* preference (usage) data. Krykewycz et al. (2011) is the most directly related work to this paper.

Krykewycz et al. (2011) presents a systematic framework for estimating demand for a bicycle sharing program in Philadelphia, Pennsylvania. In addition, they hypothesize about factors which contribute to trip origin and trip attraction. Origin factors include population density and group quarter population density. Attraction factors include job density, location of tourist attractions, and proximity to parks and recreation. They also consider network facilities and infrastructure features like rail stations, bike lanes and bus stops. We complement the work of Krykewycz et al. (2011) by using observed usage data to uncover the factors that impact trip generation and attraction. While results based on census data are common in bike sharing feasibility studies -- see Philadelphia (DVRP, 2011), New York (NYC, 2011), London (TFL, 2011) -- we present an empirical analysis to test the factors that influence bike sharing usage.

3. Description of Data Sources

Our goal is to explain select factors influencing bike sharing trip generation and attraction in Barcelona and Seville, Spain. The dependent variables of interest in our analysis are the public bicycle usage arrival and departure rates and operator rebalancing refill and removal rates in Sub-City Districts (SCDs) of the city. SCDs are spatially homogenous regions (in terms of social structure and built environment) formed for ease of planning. The independent variables include census level data at the SCD-level on socio-demographics, economics, and housing, from

EuroStat. Points of interest (POI) data such as the location of businesses, metro stations, leisure activities, restaurants, etc., are used as proxies for land-use characteristics and trip purpose. We describe the compilation exercise for each of these data sources in detail below.

3.1 Bike Sharing Usage Data

We developed an information systems infrastructure that includes a web crawler, to capture bike sharing system state (snapshot) data in real-time via the websites of these programs. The dataset, obtained from the websites of the bike sharing programs, spans from May 1 - September 20, 2009. Through this process, the state information of all bike stations in the city is captured every 5 minutes (due to restrictions on the crawler). However, due to intermittent errors in the information systems infrastructure, several stations and time points have missing data. We clean this data to result in 34 days and 21 days of 5-minute state data for each station in Barcelona and Seville, respectively. Trip rate information in this study is derived from the collected state information. We also record the latitude and longitude of each bike station in the city, and the total number of bike stations in each SCD. Through this process, we have compiled a unique longitudinal dataset on usage at each individual station and SCD.

To transform this data into the dependent variables used in our models, we first compute the total arrival and departure rates at each station at a 5-minute level. Note that total arrivals and departures of bikes can be influenced by both customer usage as well as rebalancing operations by the operator. We split the apparent total arrival rate and total departure rate into four components - (i) arrival rate, (ii) departure rate, (iii) refilling rate, and (iv) removal rate, using a heuristic approach. Rates (i) and (ii) are due to customer usage and (iii) and (iv) due to rebalancing by the operator. The assumption behind our separation is the following – when the operator rebalances bikes at a station, usually there will be a significant change in the total number of bikes at the station (refilling or removal) in a short span of time, as compared to users borrowing and returning bikes. Therefore, when we observe a 5-minute total arrival (total departure) rate that is greater than the 99th percentile of arrival (departure) rate for that station, we assume that a rebalancing operation (refilling or removal) is performed by the operator. Specifically, our heuristic assumes that when the total arrival (total departure) rate exceeds 99th percentile of the arrival (departure) rate for that station, the arrival (departure) rate due to public demand is the average rate of the last two 5-minute arrivals (departures) for that station, and the remaining is due to refilling (removal) by the operator (Note that a different threshold, for example, the 95th percentile, could also be used depending on the observations in the data)¹. The 5-minute level data of (i) arrival *rate*, (ii)

¹ We did not have the rebalancing information from the operator to validate our rebalancing heuristic; hence, we employed the exact same approach using Boston bicycle-sharing system data to “test” our heuristic approach. The Boston data was available online with aggregate daily rebalancing information (at the system level for each day). A comparison of the estimated rebalancing (using our approach) and actual aggregate numbers provided in the Boston data yielded a maximum daily error margin of about 20%. Given that there might be some intrinsic differences between how the Barcelona and Seville systems and the Boston system are operated we considered this an acceptable error margin for our heuristic method. Moreover, we have tested different values (99th, 98th and 95th) for threshold of our heuristic approach and the 99th percentile threshold provided lowest error in the comparison of estimated rebalancing and actual numbers. Moreover, we believe that it is very unlikely that two rebalancing operations happen for one station in two consecutive hours. The comparison of 99th, 98th and 95th percentile showed us that the 99th percentile appropriately reduce the likelihood of such occurrences.

departure *rate*, (iii) refilling *rate*, and (iv) removal *rate* are further aggregated temporally and spatially to create their corresponding hourly metrics at the SCD-hour level.

3.2 Eurostat Urban Audit

The European Union’s statistical agency, Eurostat, and member state national statistical agencies compile data at the intra-city level for a select number of cities in the Urban Audit (Eurostat, 2007). The Eurostat urban audit data provides variables in the categories of: sociodemographic, economic and housing. The data is available at the city level and Sub-City District (SCD) level. For this study we use data at the SCD level. This study uses data collected in the 2006-2007 urban audit. The number of variables at the SCD level in 2006-2007 with completed entries is limited. We were able to extract the following variables at the SCD level: population density, female population, one person households and labor market participation rate.

3.3 Tele Atlas Points of Interest

We use Points of Interest (POI) data from Tele Atlas (see www.teleatlas.com), provider of geographic databases. The Tele Atlas data consists of the latitude and longitude of Points of Interest (POI) in a city. Tele Atlas places each POI into one of 68 categories. Following the bike sharing literature, Krykewycz et al. (2011), we combine the 68 categories into super-categories of POIs for our analysis. The eight categories are businesses, transportation, leisure, worship, hotels, hospitals, restaurants and universities.

We combine the arrival and departure rates at SCD-hour level with SCD-level Eurostat data and the TeleAtlas Point of Interest (POI) data to create the data samples for Barcelona and Seville. We now describe the data sets for Barcelona and Seville generated from this procedure.

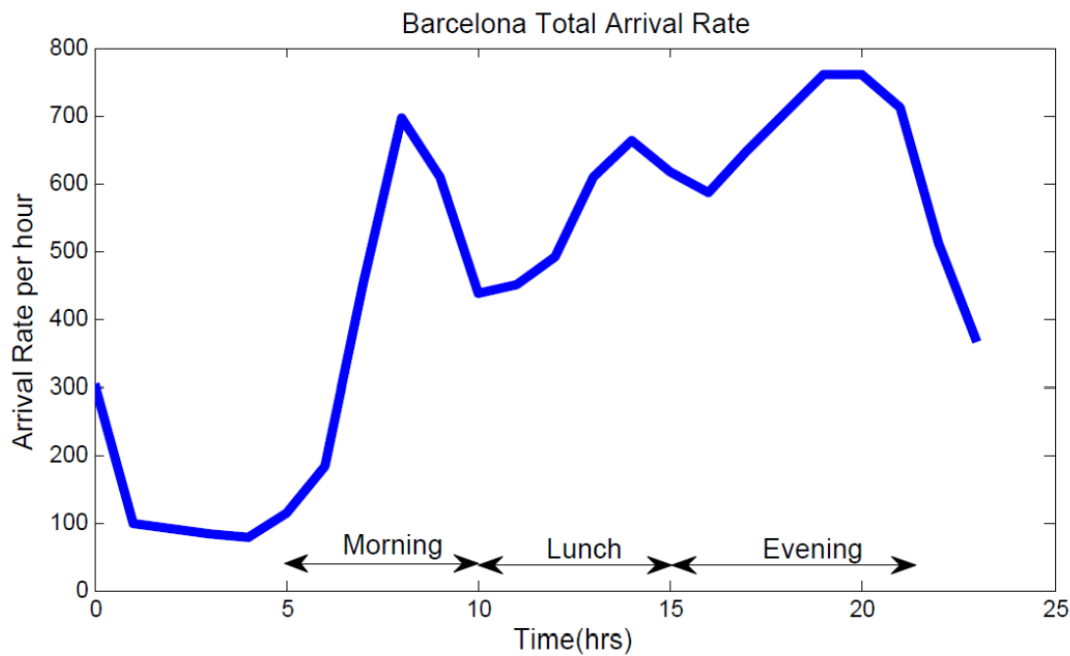


Fig. 1 Barcelona Average Total Arrival Rate for 24 hours

3.4 Barcelona Dataset Description

The bike sharing program in Barcelona, Spain, named Bicing, started operation on March 3, 2008, operated by the company Clear Channel. During our study period, it consisted of 402 fixed location bike stations and approximately 6000 bikes. Barcelona has 86 train and metro stops with a bike station nearby (OBIS, obisproject.com). Barcelona has 56 SCDs in the Eurostat urban audit. The Tele Atlas data contains 6,893 points of interest for Barcelona. During the observation period, Barcelona has 168 points of interest that are categorized as transport, 2,809 categorized as businesses, 401 as leisure and 60 as universities. We use data from 34 days of observations, with arrival rate and departure rate averaged over each hour of the day. Thus we have 816 observations per SCD. Only weekdays are selected for analysis in this study. In our dataset, there are 28,632 average trips per day in Barcelona. Figure 1 shows the city-wide total arrival rate for Barcelona for each hour in a 24 hour period. It shows the morning, lunch and evening behavior of the arrival rate. The total city arrival rate has three peaks, corresponding to the morning, lunch and evening periods.

The total arrival rate attains a maximum during the evening period around 7pm. A descriptive summary of arrival, departure, refilling and removal rates are provided in Table 1. Specifically, the average rates for the dependent variables across the different time periods are provided. The values clearly indicate that the lunch and evening time periods have higher arrivals and departures as well as increased rebalancing operations. Table 2 provides a description of the sample characteristics of the independent variables employed in the model estimation. The statistics indicate that there are about 7 stations in every SCD while the average capacity at the SCD level is about 67 bicycles. Further, it is clear that businesses and restaurants form the majority of the POIs in the Barcelona region.

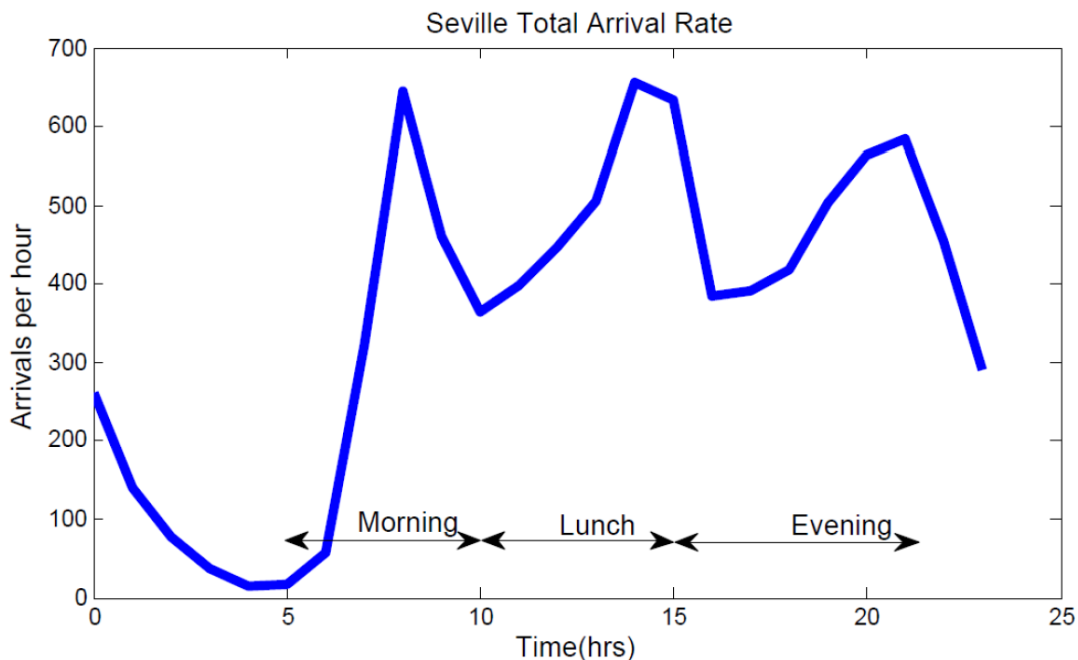


Fig. 2 Seville Average Total Arrival Rate for 24 hours

3.5 Seville Dataset Description

The bike sharing program in Seville, named Sevici, started operation on July 24, 2007. It is operated by the company JCDecaux. During our study period, it consisted of 271 fixed location bike stations and approximately 2000 bikes. According to the EU-sponsored Optimizing Bike Sharing in Europe (OBIS) working group, Seville has 19 train and metro stops with a bike station nearby. In contrast to Barcelona, Seville has a flat elevation profile. There are 38 sub-city districts in the city. The Tele Atlas data contains 1,700 points of interest for the city, in which 168 are categorized as transport, 808 as business, 278 as leisure and 30 as universities. We use data from 21 days of observations. Similar to Barcelona, only weekdays are selected for the analysis in this study. From these days, we construct 504 hourly observations of the arrivals and departures per SCD. In the dataset, there are an average of 8,173 trips per day in the city.

Figure 2 shows the total arrival rate for Seville for each hour in a 24-hour period. It shows morning, lunch and evening peaks in the arrival rate, similar to that in Barcelona. One key observation is that in Seville, the activity in the evening period is less prominent compared to the morning and lunch periods, as compared to Barcelona. Additionally, the peak during the lunch period is slightly later in the day compared to Barcelona. Tables 1 and 2 describe the sample characteristics of the dependent variables (arrival, departure, refilling and removal rates) and independent variables used in the model estimation.

4. Research Methodology

4.1 Linear Mixed Model Approach for Arrival and Departure Rates

In this section, we describe the methodology employed for model estimation in our paper. The data compiled contains repeated measures for the same SCD across multiple days and hours. A traditional cross-sectional linear regression model would neglect the inherent correlation across the multiple repeated measures and the resulting models would be econometrically inefficient and the parameter estimates likely to be biased. Moreover, any quantitative computation from such models will result in erroneous predictions. Toward estimating the accurate impact of exogenous factors we estimated a parsimonious linear mixed model that allows us to simultaneously incorporate different correlation structures. In our analysis, we found that the auto regressive moving average (ARMA) correlation structure offered superior statistical fit. In this model structure, we consider that observations within a day for each SCD are correlated in the ARMA structure. The assumption allows us to estimate a model that captures the influence of bicycle infrastructure attributes and socio-demographic characteristics at the SCD level. The same model structure is employed for customer arrivals and departures and operator refilling and removal for Barcelona and Seville.

Let $q = 1, 2, \dots, Q$ be an index to represent a sub-city district (SCD), $d = 1, 2, \dots, D$ be an index to represent the various days on which data was collected and $t = 1, 2, \dots, 24$ be an index for hourly data collection period at SCD q and day d . The logarithm of the dependent variable is modeled using linear regression equation which in its most general form takes the following structure:

$$y_{qdt} = \beta X + \xi$$

where y_{qdt} is the logarithm of the observed dependent variable, X is an $L \times 1$ column vector of attributes including a constant that influences the dependent variable including SCD q level variables, such as number of stations, population density, day d related variables, such as week of the day, and time period t related variables, such as hour of the day. The model coefficients, form a $L \times 1$ -column vector. An idiosyncratic random error term, ξ , is assumed to be normally distributed across the data records.

In the context of data collected in our study, the idiosyncratic error term could be potentially be partitioned into three components. The first component represents the common unobserved factors related to the SCD in influencing the dependent variable across different days and times. The second component represents common unobserved factors related to a particular day that influence the dependent variable across the region, usually related to spatial effects such as occurrence of rain or unusually hot weather. The third component considers the influence of unobserved factors on the dependent variable values across different time periods in a day for each SCD. An analysis that incorporates the three components requires the partitioning of the ξ term into three categories: (1) SCD component, (2) Day component and (3) Time of day component. However, incorporating these three components in the regression model simultaneously is far from trivial and will result in a covariance matrix of the order of the size of the dataset (i.e. $45,696 \times 45,696$ in our empirical context for Barcelona). Given the considerably large size of data and the number of dependent variables considered in our study, it is quite prohibitive in terms of run-times to estimate the combined influence of the three components simultaneously (see Bhat et al. (2010) for discussion on complexity involving spatial models). So, we adopt the approach of identifying an appropriate correlation structure through a systematic process of model estimation while focusing on a parsimonious specification. Specifically, we resort to examining the influence of SCD-specific common unobserved effects and SCD-time of day related common unobserved effects.

4.1.1 SCD structure

In this structure, the multiple responses for each SCD are considered to be correlated. This means that in the Barcelona dataset, this results in 56 SCDs with 816 repeated measures. Estimating a full covariance matrix is computationally intensive while providing very little intuition. Hence, we parameterize the covariance matrix (Ω):

$$\Omega = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \dots & \sigma_1^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^2 & \dots & \dots & \sigma^2 + \sigma_1^2 \end{pmatrix}$$

The parameters estimated in this correlation structure are (σ^2 and σ_1^2).

4.1.2 SCD-day-time of day structure

In this structure, the data can be visualized as 24 records for each SCD-day combination for a total of 1804 observations. In this structure, the 24 records, corresponding to 24 hours, from the same SCD and same day are assumed to be correlated. Again, for estimating a parsimonious

specification, we assume a first-order autoregressive moving average (ARMA) correlation structure with three parameters σ ; φ and ρ as follows:

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \varphi\rho & \dots & \varphi\rho^{23} \\ \varphi\rho & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \varphi\rho^{23} & \dots & \dots & 1 \end{pmatrix}$$

4.2 Rebalancing Operation Models

For modeling the rebalancing operation, we employ two models: 1) a binary choice model for the identification of a rebalancing action 2) a linear regression model for the rate of rebalancing actions. We examine the refilling and removal operations separately. To elaborate, two binary choice models are estimated to analyze the contributing factors that influence the refill and removal operations. In these two models, the dependent variables are dummy variables indicating if a refill or removal operation is performed or not. Next, given a refill (removal) operation is performed, a linear regression model is employed to examine what factors affect the refill (removal) rates in each rebalancing operation. The sample size for linear regression models reduces to the records where a refill (removal) is identified.

The model for identification of a rebalancing action takes the well-known binary logit formulation as the following:

$$u_r = \beta_r X_r + \eta_r$$

where u_r is the utility obtained for rebalancing occurrence, X_r is the vector of attributes that influences the rebalancing operation and β_r is the model coefficients to be estimated. The random error term, η_r , is assumed to be independent and identically Gumbel-distributed across the dataset. Given this notation, the probability expression takes the typical binary logit form as follows:

$$P_r = \frac{1}{1 + \exp(-u_r)}$$

by maximizing the log-likelihood of this probability function, the model parameters β_r are estimated. The model for the rate of rebalancing actions take the simplest form of linear regression models. The linear mixed model described above collapses to the model for the rebalancing rate when the φ and ρ are assumed to be zero.

4.3 Model Estimation

The approach employed for arrival and departure models in the estimation is based on the Restricted Maximum Likelihood Approach (REML) that is slightly different from the maximum likelihood (ML) approach. The REML approach estimates the parameters by computing the likelihood function on a transformed dataset. The approach is commonly used for linear mixed models (Harville, 1977). For operator rebalancing models, the binary choice model is estimated using a maximum likelihood approach while linear regression model is estimated using an ordinary least squares approach. The models were estimated using the SPSS software.

The results from the estimation are discussed separately for usage and rebalancing (operator-enforced movement) measures. The former analysis will allow us to evaluate bicycle demand generated at an SCD level while the latter analysis will allow us to provide information to operators on when and where to focus on rebalancing operations. In particular, the current analysis

is focused on examining four rate components: (i) arrival rates, (ii) departure rates, (iii) refilling rate and (iv) removal rate.

Several variables were considered in the analysis. These variables can be broadly classified as: (1) Bicycle infrastructure attributes including number of stations in a SCD, stations per unit area, capacity of the SCD, total capacity in a SCD, total capacity per unit area and (2) Land use characteristics of the SCD including population density, proportion of one person households, proportion of females to males, elevation measures in the SCD and percentage point of interests (POIs) for business, recreation, transport, restaurants, places of worship and universities. In the model estimation, several forms of the variables from the two groups were considered. These variables were also interacted with temporal dimensions to allow for differential sensitivities across different time periods. The final model selection was based on the restricted log-likelihood and Bayesian Information Criterion metrics. Our model estimation process was guided by parsimony and intuitiveness considerations. The model estimation process required us to estimate SCD level models and SCD-day combination models. The models that incorporated temporal correlations using the ARMA model frameworks offered improved fit measures. Hence, in Section 5, we focus our discussion on the ARMA model results.

5. Estimation Results

5.1 Usage

In this section, we focus our attention on understanding how bicycle infrastructure attributes and land use characteristics influence public bicycle usage (arrival and departure rates) in Barcelona and Seville. The model estimation process began with simple linear regression model. Employing the results from the linear regression a linear mixed model was estimated. The data fit as measured by log-likelihood clearly highlights the model fit improvement offered by the mixed models relative to simple linear regression model in both cities. Specifically, the log-likelihood ratio test statistic for comparing the mixed linear model to simple linear model is 6650.0 for Barcelona arrivals model, 10580.4 for Barcelona departures model, 2918.4 for Seville arrivals model, and 4552.8 for Seville departures model. The test statistic is substantially higher than the corresponding chi-squared table value with only 2 degrees of freedom at any practical level of significance for all the models. The estimation results for the log-linear mixed model estimations are presented in Tables 3 and 4.

5.1.1 Bicycle Infrastructure attributes

The results indicate that the arrivals and departures in a SCD are strongly influenced by bicycle infrastructure variables. In the model specifications, different functional forms of the number of stations and capacity are considered. The station density variable and total capacity per unit area variable offered the most intuitive fit while providing statistically significant results. For Barcelona, as is expected, as the station density increases the number of arrivals and departures in an SCD increase. It is interesting to note that the impact of the station density varies for different time period of the day. The result indicates that the impact of station density is highest on the 'Evening' and 'Late' time period i.e. the arrivals or departures in the evening and late night period are most strongly related to station density. This is expected because during evening and late time

periods the demand is predominantly a function of the number of stations where as during other time periods, demand is dependent on other amenities. For Seville, on the contrary, the impact of station density is negative. However, it is important to note that the effect of station density variable and total capacity per unit area variable must be considered together; i.e. increase in one parameter would simultaneously increase the other parameter. Hence, the estimates obtained are the overall effect of station and capacity density variables. The capacity variable exhibits intuitive impacts. For both Barcelona and Seville, as capacity in an SCD increases, arrivals and departures are positively influenced. Further, we see that capacity has an overall positive impact on the different time periods of the day (Morning, Lunch and Evening) with the largest impact during the lunch period for Barcelona and the morning period for Seville.

5.1.2 Land use characteristics

A host of land use characteristics affect the arrival and departure rates in our models for Barcelona. The mean and standard deviation of the SCD elevation levels offer interesting results. We observe that SCDs with higher average elevation tend to have lower arrival and departure rate increase indicating that bicycle demand is lower in regions with higher elevation. However, as the standard deviation of elevation of the stations within the SCD increases, the arrival and departure rates increase, indicating that within SCDs with same mean elevation, bicyclists prefer stations that are on lower elevation levels for arrivals and departures. The Points of Interest variables by different categories offer plausible results. All categories of POIs positively influence bicycle arrival and departure rates (except place of worship). The reader would note that the POIs are indicative of the presence of potential activity centers in the various parts of the city that strongly influence station location and capacity decisions. Hence they are more likely to have a positive association on arrival and departure rates. Nonetheless, controlling for the variation of POI proportion is important in the context of our modeling effort. The other land-use variables including proportion of females to males, proportion of one person households and population density serve as controls for land-use effects on arrival and departure rates.

The impact of land-use effects in the Seville region broadly follow the same pattern. The mean of elevation of the SCD influences usage similarly, with increased elevation decreasing usage. The standard deviation of elevation was not significant in explaining the usage (as was the case for Barcelona). The POI variables for different categories such as businesses, recreation and transportation all increase usage (both arrivals and departures), similar to the case of Barcelona. As expected, the population density variable has positive impact on the arrivals and departures. However, the impact of population density on arrivals in the morning time period is less than other time periods. It is intuitive, since we expect more departures from higher population density areas in the morning period than arrivals, which would result in higher arrivals in the areas with lower population density.

5.1.2 Temporal Parameters

In our modeling efforts, we also control for temporal variables. For Barcelona, we observe lowest demand during late night time period as expected. For Seville, AM time period has the lowest arrival demand while for departures, the lowest demand is for AM and late night time periods.

5.2 Rebalancing Operation Analysis

In the rebalancing framework the influence of bicycle infrastructure attributes and land-use characteristics are examined on the decision of rebalancing operation and the rates of refills and removals. The estimates for binary choice and linear regression models for Barcelona and Seville are presented in Tables 5 to 8. In Barcelona, the results indicate that areas with the need for more rebalancing operations require lower rebalancing rates. The results pertaining to number of stations and average capacity offer interesting insights on infrastructure operation. Specifically, increasing the number of stations results in the increase in number of rebalancing needs for both refills and removals and reduction in the rate of refills and removal. Increasing capacity also results in increased rebalancing needs both in terms of occurrence and rates. The result indicates that a reduction in station density in an SCD would result in higher rates of refills and removal but with lower rebalancing operation needs.

Among land-use characteristics, increase in mean elevation and the standard deviation of the elevation reduces rebalancing needs because the usage is marginally lower in these SCDs. However, when a rebalancing occurred, it is more likely associated with higher refill or removal rates. Population density also reduces rebalancing needs for both refills and removals. Moreover, SCDs with higher population density are more likely to have rebalancing operations both in terms of occurrence and rates during night time period. The points of interest variables also have similar trends. The SCDs with more points of interest have higher needs for rebalancing but with fewer amount of rebalancing per instance. Temporal parameters indicate more performance of rebalancing operations during the night time period.

In Seville, increasing capacity results in increased rebalancing needs but reduces the amount of refills and removals. The number of stations estimate is not significant for the removal model. However, the SCDs with higher station density have higher needs, for refills. Contrary to Barcelona, the effect of elevation parameters are not significant.

We now introduce a notion of *spatial flexibility* for shared mobility systems. This notion is closely related to *mixed-use development*. We claim that SCDs that contain a heterogeneous set of points of interest support many different usage patterns and are spatially flexible. By the term ‘usage pattern’, we are referring to temporal usage patterns associated with a given point of interest. Heterogeneous usage patterns, if complementary, lead to higher operational efficiency via economies of scope, or multiplexing.

SCDs with less varied types of points of interest support fewer usage patterns and are less spatially flexible. For example, stations in an SCD that contain only business POIs experience an influx of bikes due to the morning commute causing the stations to become full. There would be fewer full stations if this SCD also contained POIs with a demand inversely related to business demand. The operational benefit of heterogeneous usage patterns was introduced in previous literature on car sharing, see Barth and Shaheen (2002). We provide the first empirical support for spatial flexibility for shared mobility systems.

In the rebalancing estimation, we measure spatial flexibility by the POI variance across the SCDs by computing the average of the square of the deviation from the mean for all POI categories i.e. a larger value indicates significant differences from overall mean values in the region. The results for this variable in both Seville and Barcelona indicate that as the overall variance computed increases, the need for rebalancing reduces. However, when a rebalancing occurred, SCDs with higher POI variance are associated with higher rates of refills and removals. Again, the result provides credence to the hypothesis that increased spatial flexibility reduces requirement to

rebalance because the usage patterns for these POIs are complementary and lead to an inherent rebalancing by the bicyclists without the operator's intervention. For both Seville and Barcelona, the POI variables have negative impact in binary choice model while they have positive impact in linear regression model of refills and removals.

5.3 Elasticity Effects

The parameter estimates in Section 5.1 and 5.2 provide an intuitive understanding of the exogenous factors that are important. However, these estimates do not provide us with tangible comparison measures across various exogenous variables. Towards this end, we compute elasticity effects based on parameter estimates. The elasticity effects are computed for customer arrivals and departures. For the sake of brevity, we restrict ourselves to elasticity effects for the Barcelona region.

5.3.1 Arrivals and Departures

In this section, we focus on examining the influence of various variables identified in the discussion of results. To elaborate, we compute the elasticity effect of changing a particular variable. For instance, we illustrate the contribution of number of stations in an SCD by increasing their count by 1 and examining its influence on the overall arrivals and departures. In this exercise, we focus on the following variables: 1) increasing number of stations without increasing SCD level capacity, i.e., we redistribute capacity to create a new station, 2) increasing number of stations allowing for SCD capacity to increase (the increase is same as the average capacity), and 3) increasing average SCD capacity (by adding bicycles at each station). For each of these variables, the percentage change in the predicted outcome is computed. The elasticity measures are also computed for different time periods. The results of the elasticity computation for Barcelona are presented in Table 9.

The following observations can be made from the elasticities. First, increasing bicycle infrastructure facilities (stations and/or capacity) results in an increase in arrivals and departures. Second, the positive impact of increasing number of stations and capacity is substantially higher than the positive impact of increasing either capacity or stations for arrivals and departures. An important finding is that the impact of adding a new station is similar to increasing the capacity by 5 units to existing stations in the SCD for arrivals. On the other hand, for departures, adding a station has a stronger impact than capacity increase of 5 units.

6. Conclusion

The current study contributes to literature on bicycle sharing program usage and operation. The paper documents research undertaken toward answering two questions related to quantifying and comparing the influence of bicycle infrastructure attributes and land-use characteristics on: (a) demand, consisting of customer arrivals and departures, and (b) rebalancing, consisting of the frequency and quantity of operator refills and removals. The research questions also seek to understand if specific bicycle infrastructure attributes or land-use characteristics influence demand and rebalancing operations differently, and if so, in what manner.

The study employs usage data compiled through scripts that record station level bicycle availability every few minutes in urban regions of Barcelona and Seville. The resulting data is composed of a multi-day 24 hour data for station level bicycle availability. The station level bicycle availability data was aggregated to a Sub-City District (SCD) level and the temporal dimension is aggregated to an hourly value. The aggregation of snapshot data allows us to derive bicycle arrivals and departure rates and rebalancing rates at a regional level. The compiled bicycle usage data is augmented with intra-city level data on sociodemographic, economic and housing characteristics at the SCD level for the year 2006-2007. Specifically, the following variables at the SCD level were extracted: population density, female population, one person households and labor market participation rate. We also compiled Points Of Interest (POI) data from Tele Atlas in eight categories: business, transport, leisure, restaurants, worship, hotels, hospitals and universities. The POI variables are used as an indicator of land-use variability at the SCD level.

A total of 8 models were estimated. The model results highlight the importance of bicycle infrastructure attributes and land-use characteristics on bicycle usage and operator rebalancing. The results for bicycle usage from Barcelona and Seville follow a similar pattern with minor differences across the two cities. The model results for the Barcelona region highlight bicycle station density, average capacity, and percentage of POIs from business, recreation, and restaurants as important contributors of arrival and departure rates. The results for the Seville region are similar except for the insignificance of station density attribute. The results for operator rebalancing for both Barcelona and Seville region have fewer parameters. In particular, for the Barcelona region we find that increasing the number of stations actually results in a reduction for operator rebalancing. The result has significant implications for shared bicycle infrastructure operators in Barcelona. In the Seville region, the number of stations variable was insignificant for operator rebalancing indicating that increasing capacity directly results in increasing rebalancing requirements. In addition, we also see that the presence of heterogeneous POIs in an SCD leads to lower requirements of rebalancing. This confirms empirically the hypothesis that the presence of a variety of POIs indicates several purposes for bikes, incoming and outgoing from the SCD, and therefore the likelihood of customers automatically rebalancing bikes increases.

In addition to the parameter estimates, to directly compare the impact of various exogenous variables, we also compute elasticity effects. The elasticity effects were computed for various bicycle infrastructure attributes and land-use characteristics for Barcelona. The results provide quantitative estimates of how increasing the number of bicycle stations or station capacity affect bicycle usage. The estimates suggest very important policy implications from the perspective of the bicycle infrastructure operation. The results indicate that redistributing existing capacity into smaller capacity based stations thus increasing total number of stations has a positive effect on usage. The authors believe that the quantitative computation provided in this paper provides a template for examining bicycle rebalancing across the world. A prediction framework for SCD level rebalancing will offer improved system management for bicycle sharing system operators.

To be sure, the study is not without its limitations. The data collected allows the arrival rate and departure rate to be computed at intervals of 5 minutes. It is possible that multiple transactions occur within this time period and result in an erroneous estimate of the bicycle usage. However, we believe that the resulting error from our assumption is minimal². The quantitative models

² To address the magnitude of error from our 5-minute data, we used another dataset. Specifically, we explored data from Montreal that was collected every 1 minute to check for the underestimation of demand if the data collection interval was 5 minutes. We found that the hourly arrival and departures rates were under estimated by about 20% if

developed will benefit substantially if additional data on commuting patterns and general urban trends in these cities were available. Finally, in this paper, we focus on SCD level models. In future research, other types of spatial aggregation might be considered.

References

- Baltes, M., Factors influencing nondiscretionary work trips by bicycle determined from 1990 U.S. census metropolitan statistical area data. *Transportation Research Record: Journal of the Transportation Research Board* 1538, 96–101 (1996).
- Barth, M., Shaheen, S., Shared-use vehicle systems: Framework for classifying carsharing, station cars, and combined approaches. *Transportation Research Record: Journal of the Transportation Research Board* 1791, 105–112 (2002).
- Beck, M., Immers, L.H. Bicycle ownership and use in Amsterdam. *Transportation Research Record: Journal of the Transportation Research Board* 1441, 141–148 (1996).
- Bhat, C.R., Sener, I.N., Eluru, N., A flexible spatially dependent discrete choice model: Formulation and application to teenagers weekday recreational activity participation. *Transportation Research Part B* 44(8-9), 903–921 (2010).
- Borgnat, P., Abry, P., Flandrin, P., Rouquier, J.B. Studying Lyon's Vlo'v: A statistical cyclic model. In: *Dans European Conference on Complex Systems 2009 - ECCS'09* (2009a).
- Borgnat, P., Fleury, E., Robardet, C., Scherrer, A., Spatial analysis of dynamic movements of Velo'v, Lyon's shared bicycle program (2009b). URL <http://prunel.ccsd.cnrs.fr/ensl-00408150/en/>
- Borgnat, P., Robardet, C., Rouquier, J.B., Abry, P., Flandrin, P., Fleury, E., Shared bicycles in a city: A signal processing and data analysis perspective (2010). URL <http://prunel.ccsd.cnrs.fr/ensl-00490325/en/>
- BTS, Bicycle and pedestrian data: Sources, needs, & gaps. BTS 00-02. Tech. rep., U.S. Department of Transportation and Bureau of Transportation Statistics (2000).
- Buck, D., Buehler, R., Bike lanes and other determinants of capital bikeshare trips. Paper presented at the 91st Transportation Research Board Annual Meeting, Washington, DC (2012).
- Cao, X., Handy, S., Mokhtarian, P., The influences of the built environment and residential self-selection on pedestrian behavior: Evidence from Austin, TX. *Transportation* pp. 1–20 (2006).
- Carballeda, L., Velasco, L., Rojo, M., Gonzalo-Orden, H. Public bike systems in Spain. Association for European Transport and contributors, University of Burgos (2010)
- Cervero, R., Duncan, M., Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay Area. Tech. rep., University of California Transportation Center, University of California, Berkeley (2003).
- Chatman, D., How the Built Environment Influences Non-work Travel: Theoretical and Empirical Essays. Ph.D. Dissertation. Department of Urban Planning, University of California, Los Angeles (2005).
- de Chardon, C.M. and Caruso, G., Estimating bike-share trips using station level data. *Transportation Research Part B: Methodological*, Vol. 78, pp.260-279 (2015).

the data collected interval was 5 minutes. So while we realize reducing the time interval would have improved model accuracy, the restrictions on these websites forced us to adopt a larger time interval with error.

- deMaio, P., Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation* pp. 41–56 (2009).
- DVRP, Philadelphia bikeshare concept study. Tech. rep., Bonnette Consulting and Delaware Valley Regional Planning Commission (2011). URL www.bicyclecoalition.org/files/PhiladelphiaBikeshareConceptStudy.pdf
- Eurostat, (2007). URL <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>
- Ewing, R., Schroeder, W., Greene, W., School location and student travel analysis of factors affecting mode choice. *Transportation Research Record: Journal of the Transportation Research Board* 1895, 55–63 (2005).
- Faghih-Imani A., N. Eluru, A. El-Geneidy, M. Rabbat, U. Haq, How does land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal, *Journal of Transport Geography*, Vol. 41, pp. 306-314 (2014).
- Faghih-Imani A., N. Eluru, Analyzing Bicycle Sharing System User Destination Choice Preferences: An Investigation of Chicago's Divvy System, *Journal of Transport Geography*, Vol. 44, pp. 53-64 (2015).
- Faghih-Imani A., N. Eluru, Role of Bicycle Sharing System Infrastructure on Usage: Evidence from Montreal, Presented at the 5th Innovations in Travel Modeling Conference, Baltimore, MD, (2014).
- FHWA, F.H.A., Guidebook on methods to estimate non-motorized travel: Overview of methods. Technical report, 1999.
- Froehlich, J., Neumann, J., Oliver, N., Measuring the pulse of the city through shared bicycle programs. In: *International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems- UrbanSense08* (2008).
- Froehlich, J., Neumann, J., Oliver, N., Sensing and predicting the pulse of the city through shared bicycling. In: *Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)* (2009).
- Gebhart, K., Noland, R., The Impact of Weather Conditions on Bikeshare Trips in Washington, DC. *Transportation*, Vol. 41, pp. 1205-1225 (2014).
- González, F., C. Melo-Riquelme, and L. de Grange, A combined destination and route choice model for a bicycle sharing system. *Transportation*, pp. 1-17 (2015).
- Handy, S., Cao, X., P.L., M., Self-selection in the relationship between built environment and walking? Evidence from Northern California. *Journal of the American Planning Association* pp. 55–74 (2006).
- Harville, D.A., Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72(358), 320-338 (1977).
- Hunt, J., Abraham, J., Influences on bicycle use. *Transportation* 34, 453–470 (2007).
- Jensen, Pablo, Jean-Baptiste Rouquier, Nicolas Ovtracht, and Céline Robardet. Characterizing the speed and paths of shared bicycle use in Lyon. *Transportation research part D: transport and environment* 15, no. 8 522-524 (2010).
- Kaltenbrunner, A., Meza, R., C, J.G., Codina, J., Banchs, R., Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing* 6 (2010).
- Kitamura, R., Mokhtarian, P., Laidet., L., A micro-analysis of land use and travel in five neighborhoods in the San Francisco bay area. *Transportation* pp. 125–158 (1997).
- Krykewycz, G.R., Puchalsky, C.M., Rocks, J., Bonnette, B., Jaskiewicz, F., Defining a primary market and estimating demand for major bicycle-sharing program in Philadelphia,

- Pennsylvania. Transportation Research Record: Journal of the Transportation Research Board 2143, 117–124 (2011).
- Lin, J.R., Yang, T.H., Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E* pp. 284–294 (2011).
- Murphy, E., and J., Usher, The Role of Bicycle-sharing in the City: Analysis of the Irish Experience, *International Journal of Sustainable Transportation*, Vol. 9, pp. 116-125 (2015).
- Nair, R., Miller-Hooks, E., Hampshire, R., A., Busic, Large-Scale Vehicle Sharing Systems: Analysis of Velib. *International Journal of Sustainable Transportation*, Vol. 7, pp. 85-106 (2013).
- NYC, Bikeshare opportunities in New York City. Technical report, NYC Department of Planning (2011). URL www.nyc.gov/html/dcp/pdf/transportation/bike_share_complete.pdf
- O'Brien, O., Cheshire, J., and M. Batty, Mining bicycle sharing data for generating insights into sustainable transport systems, *Journal of Transport Geography*, Vol. 34, pp. 262-273 (2014).
- Optimising bike sharing systems in Europe. URL: www.obisproject.com
- Raviv, T., Tzur, M., & Forma, I. A. Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics*, 2(3), 187-229 (2013).
- Rietveld, P., Hensher, K., Button, D., Biking and walking: The position of non-motorized transport modes in transport systems. *Handbook of Transport Systems and Traffic Control* pp. 299–319 (2001).
- Rixey, R., Station-Level Forecasting of Bikesharing Ridership: Station Network Effects in Three U.S. Systems. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2387, pp. 46-55 (2013).
- Schwanen, T., Mokhtarian, P., What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled. *Transportation Research Part D* pp. 127–151 (2005).
- Shaheen, S.A., Guzman, S., Zhang, H., Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future. *Transportation Research Record: Journal of the Transportation Research Board* 2143, 159–167(2010).
- Tele Atlas. URL www.teleatlas.com
- TFL, Feasibility study for a central London cycle hire scheme. Tech. rep., Transport of London (2011). URL www.tfl.gov.uk/assets/downloads/businessandpartners/cycle-hire-scheme-feasibility-fullreport-nov2008.pdf
- The bike-sharing blog. URL <http://bike-sharing.blogspot.com/>
- The circle of life. *Cycling Mobility* pp. 72–79 (2011).
- Vogel, P., Christian, D., Sensing and predicting the pulse of the city through shared bicycling. In: 12th World Conference on Transport Research (WCTR) (2010).
- Wang, X., G. Lindsey, J. Schoner and A. Harrison, Modeling Bike Share Station Activity: Effects of Nearby Businesses and Jobs on Trips to and from Stations. *Forthcoming Journal of Urban Planning and Development* (2015).

LIST OF TABLES

Table 1: Descriptive Summary of Dependent Variables

Table 2: Descriptive Summary of Independent Variables

Table 3: Model Estimation Results for Arrival and Departure Rates - Barcelona

Table 4: Model Estimation Results for Arrival and Departure Rates - Seville

Table 5: Model Estimation Results for Binary Choice of Refilling and Removal - Barcelona

Table 6: Model Estimation Results for Binary Choice of Refilling and Removal - Seville

Table 7: Model Estimation Results for Refilling and Removal Rates- Barcelona

Table 8: Model Estimation Results for Refilling and Removal Rates- Seville

Table 9: Elasticity Effects for Arrival and Departure Rates - Barcelona

Table 1: Descriptive Summary of Dependent Variables

Dependent Variable	Morning	Mid-day	Evening	Late	Total
	<i>Barcelona</i>				
Mean Arrival Rate	4.05	5.12	5.53	3.95	4.61
Mean Departure Rate	4.55	5.11	5.65	3.38	4.60
Mean Refill Rate	0.46	0.29	0.15	0.50	0.36
Mean Removal Rate	0.45	0.28	0.15	0.52	0.36
	<i>Seville</i>				
Mean Arrival Rate	.97	2.08	2.13	1.33	1.59
Mean Departure Rate	1.42	1.95	2.11	1.09	1.61
Mean Refill Rate	.05	0.04	0.04	0.02	0.03
Mean Removal Rate	.03	0.02	0.03	0.01	0.02

Table 2: Descriptive Summary of Independent Variables

Independent Variable	Barcelona		Seville	
	Mean	Standard Deviation	Mean	Standard Deviation
Number of Stations	7.2	4.9	8.3	5.1
Average SCD Capacity	66.9	82.7	147.7	105.0
Mean of station elevation in the SCD	58.0	48.5	14.4	4.1
Standard Deviation of station elevation in the SCD	13.6	18.7	3.0	0.9
Proportion Females to Males (for every 100 males)	112.6	9.9	110.8	7.0
Proportion of One Person Households	26.6	5.8	18.8	6.9
Population Density (1000 residents per sqkm)	30.2	14.6	15.8	10.7
Percentage of Business POI	43.9	12.4	45.6	18.8
Percentage of Recreational POI	4.4	4.4	11.0	9.6
Percentage of University POI	0.9	2.4	1.6	3.1
Percentage of Transport POI	17.4	7.3	16.9	18.4
Percentage of Hospital POI	0.7	1.3	0.7	1.5
Percentage of Restaurant POI	27.0	10.5	14.9	8.7
Percentage of Worship POI	2.8	2.4	5.3	5.3
Percentage of Hotel POI	2.9	3.9	3.9	5.3

Table 3: Model Estimation Results for Arrival and Departure Rates - Barcelona

Variable	Arrival Rate		Departure Rate	
	Estimate	t-statistic	Estimate	t-statistic
<i>Intercept</i>	-1.0144	-5.160	-1.4988	-6.514
<u>Bicycle Infrastructure Attributes</u>				
Number of Stations per unit area	0.0139	3.177	0.0217	4.122
Number of Stations per unit area * Morning	-0.0111	-2.278	-0.0258	-4.898
Number of Stations per unit area * Midday	-0.0205	-4.067	-0.0220	-4.053
Total capacity per unit area (10^{-3})	0.1861	10.359	0.1906	9.791
Total capacity per unit area * Morning (10^{-3})	0.0823	3.685	0.0547	2.349
Total capacity per unit area * Midday (10^{-3})	0.1404	5.061	0.1390	5.071
Total capacity per unit area * Evening (10^{-3})	0.0446	1.975	-	-
<u>Land use characteristics</u>				
Mean elevation of the SCD	-0.0131	-20.784	-0.0136	-17.384
Standard Deviation of elevation of the SCD	0.0198	12.060	0.0226	11.351
Percentage of recreation POIs	0.0297	6.765	0.0314	6.004
Percentage of transportation POIs	-	-	0.0060	2.336
Percentage of Restaurants POIs	0.0066	3.904	0.0059	2.990
Percentage of place of Worship POIs	-0.0208	-3.700	-	-
Percentage of hotel POIs	0.0346	6.314	0.0373	5.738
Proportion of Females to Males in the SCD	0.0184	9.641	0.0180	8.099
Proportion of One Person Households in the SCD	-0.0343	-7.676	-0.0326	-5.893
Population Density (10^{-3})	0.0114	10.727	0.0135	10.809
<u>Temporal Parameters</u>				
AM	0.2771	7.133	0.7667	18.396
Mid-day	0.6128	14.857	0.7200	16.466
PM	0.5478	22.956	0.6687	30.981
<u>ARMA correlation Parameters</u>				
σ	1.7861	118.476	2.0738	99.917
ρ	0.7512	89.902	0.7697	118.192
φ	0.3707	66.693	0.4896	89.436
<u>Restricted Log-Likelihood</u>		-74263.2		-74800.4

Table 4: Model Estimation Results for Arrival and Departure Rates - Seville

Variable	Arrival Rate		Departure Rate	
	Estimate	t-statistic	Estimate	t-statistic
<i>Intercept</i>	-9.6038	-11.084	-9.0148	-10.936
<u>Bicycle Infrastructure Attributes</u>				
Number of Stations per unit area	-0.0433	-2.289	-	-
Number of Stations per unit area * Morning	-	-	-0.0696	-7.837
Number of Stations per unit area * Midday	-	-	-0.0367	-4.004
Number of Stations per unit area * Evening	0.0148	1.745	-	-
Total capacity per unit area (10^{-3})	0.4990	6.843	0.5265	15.975
Total capacity per unit area * Morning (10^{-3})	0.1552	5.008	-	-
Total capacity per unit area * Evening (10^{-3})	-	-	-0.1745	-4.836
<u>Land use characteristics</u>				
Mean elevation of the SCD	-0.1360	-16.400	-0.1315	-15.381
Percentage of business POIs	0.0406	7.775	0.0389	7.733
Percentage of recreation POIs	0.0322	6.825	0.0399	8.045
Percentage of transportation POIs	0.0603	9.973	0.0596	10.998
Percentage of Restaurants POIs	0.0186	3.216	0.0219	3.586
Percentage of hotel POIs	0.0367	5.887	0.0409	5.889
Proportion of Females to Males in the SCD	0.0622	10.532	0.0529	8.634
Population Density (10^{-3})	0.0290	5.582	0.0100	3.073
Population Density (10^{-3}) * Morning	-0.0248	-7.034	-	-
Population Density (10^{-3}) * Mid-day	-0.0157	-4.040	-	-
Population Density (10^{-3}) * Evening	-0.0159	-4.059	-	-
<u>Temporal Parameters</u>				
AM	-0.4375	-6.063	-	-
Mid-day	0.5219	7.037	0.7360	10.741
PM	0.5130	7.176	0.8061	14.798
<u>ARMA correlation Parameters</u>				
σ	2.0349	66.817	2.4206	64.222
ρ	0.6142	41.958	0.5906	49.682
φ	0.4602	53.802	0.5376	72.289
<u>Restricted Log-Likelihood</u>		-26820.1		-27448.5

Table 5: Model Estimation Results for Binary Choice of Refilling and Removal - Barcelona

Variable	Refill		Removal	
	Estimate	t-statistic	Estimate	t-statistic
<i>Intercept</i>	-6.0329	-7.636	-3.0901	-29.831
<i>Bicycle Infrastructure Attributes</i>				
Number of Stations per unit area	0.0285	3.414	0.0187	2.830
Number of Stations per unit area * Morning	-0.0289	-2.854	-	-
Number of Stations per unit area * Midday	0.0402	3.053	0.0602	4.887
Number of Stations per unit area * Evening	0.0822	6.875	0.0778	7.023
Total capacity per unit area (10^{-3})	0.3830	22.521	0.3671	21.648
Total capacity per unit area * Midday (10^{-3})	0.2277	5.860	0.2143	5.446
<i>Land use characteristics</i>				
Mean elevation of the SCD	-0.0116	-9.771	-0.0074	-7.765
Standard Deviation of elevation of the SCD	0.0165	5.678	0.0100	3.867
Population Density (10^{-3})	0.0067	2.759	0.0112	4.726
Population Density (10^{-3}) * Morning	-0.0100	-3.128	-0.0188	-6.093
Population Density (10^{-3}) * Mid-day	-0.0378	-9.798	-0.0537	-13.490
Population Density (10^{-3}) * Evening	-0.0294	-7.556	-0.0318	-8.005
Percentage of recreation POIs	0.0441	5.181	0.0274	5.762
Percentage of transportation POIs	0.0175	2.092	-	-
Percentage of hotels POIs	0.0297	3.050	-	-
Percentage of business POIs	0.0159	1.996	-	-
Percentage of Restaurants POIs	0.0193	2.454	0.0091	4.428
POI Variance Measure (10^{-3})	-0.1757	-2.913	-0.3603	-6.763
Proportion of Females to Males in the SCD	0.0146	5.956	-	-
<i>Temporal Parameters</i>				
Morning	0.5937	4.767	0.7033	6.776
Mid-day	0.4036	2.952	0.6729	5.177
Evening	-0.4700	-3.126	-0.3888	-2.626
<u>Log-Likelihood</u>	-12027.0		-11941.0	

Table 6: Model Estimation Results for Binary Choice of Refilling and Removal - Seville

Variable	Refill		Removal	
	Estimate	t-statistic	Estimate	t-statistic
<i>Intercept</i>	-8.9647	-10.927	-10.5854	-11.504
<u><i>Bicycle Infrastructure Attributes</i></u>				
Number of Stations per unit area	0.0403	1.814	-	-
Number of Stations per unit area * Evening	-	-	-0.0811	-3.861
Total capacity per unit area (10^{-3})	0.3239	4.412	0.6301	12.026
Total capacity per unit area * Morning (10^{-3})	-	-	-0.4404	-5.129
<u><i>Land use characteristics</i></u>				
Population Density (10^{-3})			-0.0157	-2.673
Population Density (10^{-3}) * Morning	-0.0858	-7.929	-	-
Population Density (10^{-3}) * Mid-day	-0.0357	-3.171	-0.0752	-5.130
Population Density (10^{-3}) * Evening	-0.0214	-2.142	-	-
POI Variance Measure (10^{-3})	-0.2879	-3.893	-0.2773	-3.391
Proportion of Females to Males in the SCD	0.0430	5.981	0.0565	7.155
<u><i>Temporal Parameters</i></u>				
Morning	1.9843	10.654	1.3355	6.908
Mid-day	1.1079	5.220	1.4095	6.112
Evening	0.9547	4.800	1.3818	6.371
<u>Log-Likelihood</u>	-2713.9		-2404.8	

Table 7: Model Estimation Results for Refilling and Removal Rates - Barcelona

Variable	Refill		Removal	
	Estimate	t-statistic	Estimate	t-statistic
<i>Intercept</i>	11.2218	7.181	11.0151	7.085
<u>Bicycle Infrastructure Attributes</u>				
Number of Stations per unit area	-0.4510	-20.314	-0.4750	-19.994
Number of Stations per unit area * Midday	0.2023	5.358	0.2130	5.607
Number of Stations per unit area * Evening	0.1154	2.774	0.1129	2.739
Total capacity per unit area (10^{-3})	-	-	0.1159	2.196
<u>Land use characteristics</u>				
Mean elevation of the SCD	0.0126	3.364	0.0108	2.950
Standard Deviation of elevation of the SCD	-0.0321	-3.225	-0.0278	-2.866
Population Density (10^{-3})	0.1206	19.115	0.1215	17.934
Population Density (10^{-3}) * Mid-day	-0.0666	-6.449	-0.0609	-4.992
Population Density (10^{-3}) * Evening	-0.0526	-4.867	-0.0304	-2.257
Percentage of recreation POIs	-0.1401	-5.068	-0.1179	-4.294
Percentage of transportation POIs	-0.0652	-3.795	-0.0572	-3.335
Percentage of business POIs	-0.0542	-3.369	-0.0505	-3.145
Percentage of Restaurants POIs	-0.0692	-4.222	-0.0613	-3.742
POI Variance Measure (10^{-3})	1.6137	8.339	1.5618	7.864
<u>Temporal Parameters</u>				
Morning	-0.6587	-4.123	-1.1684	-7.380
Mid-day	-1.5150	-3.361	-2.1933	-4.891
Evening	-1.5844	-2.969	-2.3532	-4.533
<u>R Square</u>	0.249		0.257	

Table 8: Model Estimation Results for Refilling and Removal Rates- Seville

Variable	Refill		Removal	
	Estimate	t-statistic	Estimate	t-statistic
<i>Intercept</i>	-0.6009	-2.513	0.1494	1.407
<u>Bicycle Infrastructure Attributes</u>				
Number of Stations per unit area	0.0476	3.513	-	-
Number of Stations per unit area * Morning	0.0625	3.571	0.0802	2.162
Number of Stations per unit area * Midday	-0.0278	-2.279	-	-
Total capacity per unit area (10^{-3})	-0.1790	-3.385	-0.0860	-4.190
Total capacity per unit area * Morning (10^{-3})	-0.1779	-2.316	-0.1038	-2.877
<u>Land use characteristics</u>				
Standard Deviation of elevation of the SCD	0.1881	4.438	-	-
Population Density (10^{-3})	-	-	0.0115	4.578
Population Density (10^{-3}) * Mid-day	-	-	-0.0104	-1.660
Percentage of transportation POIs	0.0077	2.909	0.0046	2.515
Percentage of hotels POIs	-0.0169	-2.450	-	-
Percentage of business POIs	0.0061	2.171	0.0049	2.487
Percentage of business POIs * Midday	0.0201	4.392	-	-
Percentage of worship POIs	0.0367	4.768	-	-
POI Variance Measure (10^{-3})	0.3397	7.025	0.1972	4.983
<u>Temporal Parameters</u>				
Morning	-	-	0.3060	4.206
Mid-day	-0.5996	-2.744	0.2649	2.976
<u>R Square</u>	0.266		0.249	

Table 9: Elasticity Effects for Arrival and Departure Rates - Barcelona

<u>Variable Change Considered</u>	Arrivals					Departures				
	Morning	Mid-day	Evening	Late	Overall	Morning	Mid-day	Evening	Late	Overall
<u>Increasing Number of Stations</u> <i>(Total Capacity remains same)</i>										
1	0.37	-0.86	1.84	1.84	0.91	-0.54	-0.04	2.89	2.89	1.42
2	0.74	-1.71	3.73	3.73	1.85	-1.08	-0.08	5.88	5.88	2.90
<u>Increasing Number of Stations</u> <i>(Total Capacity increases by average station capacity)</i>										
1	2.48	1.37	3.41	3.23	2.70	1.37	2.24	4.19	4.33	3.12
2	5.09	2.82	6.97	6.61	5.53	2.82	4.60	8.61	8.91	6.42
<u>Increase Average SCD capacity</u> <i>(Number of Stations remains same)</i>										
5	1.05	1.28	0.90	0.73	0.97	0.96	1.29	0.75	0.75	0.91
10	2.12	2.58	1.82	1.46	1.95	1.94	2.61	1.50	1.50	1.84