ROUTE PREFERENCES IN BIKE-SHARING SYSTEMS

ABSTRACT. Over 300 cities have implemented large scale bike-share systems, including London, Paris and New York city. A typical bike-share system includes a communal stock of sturdy, low-maintenance bicycles distributed over a network of docking stations. Using transaction-level trip and bike-availability data from London’s bike sharing system, we construct a structural model of commuters’ preferences over different routes that captures the effects of two components of route-length: the biking and the walking distances, along with the service-levels experienced at both ends. Our data has a large proportion of 0-use routes, we propose a novel 2-stage method to efficiently handle these routes, in addition to a station clustering scheme. We illustrate the use of our estimates for designing and expanding the station network— for example to choose between adding stations to increase station density in existing coverage areas or to expand the coverage area. We also compare a docking station model with a dock-less model in terms of overall system usage.

INTRODUCTION

The cities of Paris, Barcelona, London, Wuhan, Hangzhou, Shanghai, New York, and Chicago (among many others) have set up large-scale bike-share systems that facilitate the use of bicycles in cities. A typical bike-share system includes a communal stock of sturdy, low-maintenance bicycles distributed over a network of parking stations. A registered user can “check out” any available bicycle from a station and, at the end of her commute, can return the bicycle to any station in the network.

Consumer desirability for a route in a bike-share system depends on three aspects of the route: First is the biking-distance, that is the distance between the originating and end station. Routes that involve too much biking are unlikely to be popular, as other means of transport such as the subway may be preferred. On the other hand, if the distance is too short, using the system might not be worth the trouble and commuters may prefer walking. Second, a commuter has to walk from the start point of the route, such as the residence, school or office to a nearby station and again walk from the ending station to her destination after returning the bike. There is a disutility associated with walking on both ends. Finally, another distinctive feature of bike sharing systems is that two service levels or availabilities matter: bike availability at starting stations and dock availability at ending stations. High service-levels on both ends allow users to rely on bike sharing systems as a reliable way of daily commuting.
We build a consumer-choice model that includes preferences for the biking distance, the walking distances and the service-levels, while controlling for demographics and locational factors which determine the baseline interest in the route. We estimate the model using data that covers all trips taken on the bike-share system in the city of London in the year 2014 covering over 10 million trips and half a million distinct routes. We pair this data with bike-availability data, Google places data on the location of a million or so points of interest in the city which are classified in 90 categories such as cafes, schools, bus stations etc., and demographic data that includes population density, median housing prices and median household incomes.

**Model and Estimation.** One characteristic of our data set is that around 40% of routes have 0 usage. This “many 0-share problem” is also prevalent in demand estimation in other contexts, most notably online retail, where retailers can offer very large assortments, yet the sales of many individual products are 0 for a given period. Traditional choice or demand estimation methodologies cannot deal with products with 0 share and the 0s are typically ignored, which could bias the estimates. Counterfactual analyses based on models that do not address this issue will also overstate the impact of adding new stations; these models presume all routes have non-zero usage.

We propose a new 2-stage procedure to deal with the problem. We first build and estimate a predictive model for classifying a 0-trip route vs. a route with non-zero usage using a random subset of our full dataset. The second stage of our model captures commuter choices between different non-zero routes and outside options as a function of the biking distance, the disutility of walking, and the service levels experienced. The model follows the celebrated BLP-model but with the cross-sectional heterogeneity in demand for the origin-destination pair playing the role of random coefficients. We control for the interest in different routes by including the demographics and the number of places of interest at the origin and at the destination.

To deal with the issue of identification and endogeneity we use the GMM approach of Berry et al. [1995] and adopt the MPEC optimization formulation proposed by Dubé et al. [2012]. We use two sets of instruments. One type of instruments is similar to BLP instruments; we use demographic and google place features that are faraway to one station as instruments for the availability measures of the focal station. The idea is that the unobserved utility error of each route will be uncorrelated with demographic and places features of areas that are far away from starting and ending station of the route. Secondly, we use a novel set of instruments that exploit the trip flow into the station network. For example, for a route starting from station $i$ to station $j$, we count the outgoing trip from the the starting station $i$ to stations that are far
away from station $j$, which is correlated with bike availability at the station $i$ but uncorrelated with unobserved route utility $\xi_{ij}$.

Our first stage model provides over 70% out-of-sample prediction accuracy. Overall, out of sample tests shows that our demand model provides a good fit and a much better prediction of demand than reduced-form models.

**Results and Counterfactuals.** Our initial results show that routes of distances between 400 to 800m are the most popular routes. For routes shorter than 500m, an additional 100m of biking increases the usage by 1.85%, while after 500m, an additional 100m of biking reduces the usage by 4.28%.

We illustrate how our analyses support the choice between deploying (or growing a system) to have a larger coverage but with stations that are far apart, or a system that has smaller coverage but closer stations. While the first system can increase the potential user base and the network of destinations that can be accessed by bike-share, on the other hand, the second offers more routes of the preferred distance and makes it easier to access stations.

We also illustrate the application of our estimated consumer preference model by comparing the performance of two different system designs. The first with stationary docks (as in London and most other major cities) and a second dock-less system that allows for free placement of bikes. In dock-less systems, each bike is geo-located and electronically locked. Users can pick up and drop off bikes at any available street-parking spot. Most systems in China are of this type, while systems in Europe and North America generally have docking stations. Dock-less systems are more convenient for commuters on the destination side since there is no need to find a station nearby where she wants to go and worry about dock availability. However on the origination side, there could be a cost associated with having to look for a bike and disutility of the uncertainty of nearby bike locations. Our estimates help advise the trade-off and identify conditions where each of the systems is preferred.

**References**
