DISCUSSION PAPER

August 2015 ■ RFF DP 15-39

Bicycle Infrastructure and Traffic Congestion

Evidence from DC's Capital Bikeshare

Timothy L. Hamilton and Casey J. Wichman

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES
FOR THE FUTURE

Electronic copy available at: http://ssrn.com/abstract=2649978

Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare*

Timothy L. Hamilton† Casey J. Wichman‡

August 19, 2015

Abstract

We explore the impact of bicycle-sharing infrastructure on urban transportation. Accounting for selection bias in a matching framework, we estimate a causal effect of the Capital Bikeshare on traffic congestion in the metropolitan Washington, DC, area. We exploit a unique traffic dataset that is finely defined on a spatial and temporal scale. Our approach examines within-city commuting decisions as opposed to traffic patterns on major thruways. Empirical results suggest that the availability of a bikeshare reduces traffic congestion by 2 to 3% within a neighborhood. We also identify geographic spillovers that may counteract benefits from reductions in local pollution.

JEL Codes: L91, H40, Q53, R41, R53
Keywords: Traffic congestion, public transportation, bicycle-sharing, pollution, environmental externalities

*We would like to thank Lesley Turner, Ashley Langer, and Ashley Chaifetz, as well as conference participants at the Southern Economics Association annual meeting, for insightful comments. We would also like to thank the CATT Lab at the University of Maryland for providing data for this study. Much of this research was performed while Wichman was a graduate student at the University of Maryland, College Park.

†Corresponding Author. E-mail: thamilt2@richmond.edu, Tel.: +1-804-287-1815. Department of Economics, University of Richmond.

‡E-mail: Wichman@rff.org, Tel.: +1-202-328-5055. Resources for the Future, Washington, DC.

Electronic copy available at: http://ssrn.com/abstract=2649978

1 Introduction

Tailpipe emissions from transportation constitute 27% of greenhouse gas emissions in the United States.[1] The effect of automobile pollution is amplified further by increases in congestion in urban areas, which exacerbate both private and public damages. Schrank et al. (2012) estimate national congestion costs arising from time loss and wasted fuel at more than $120 billion in 2011, while annual CO_2 emissions attributable strictly to congestion are 56 billion pounds.[2]

In response to these concerns, government agencies have imposed highway tolls, built high-occupancy vehicle lanes, invested in public transit infrastructure, imposed fuel economy standards, and relied on voluntary information campaigns in an effort to reduce vehicle miles traveled (VMTs), alleviate congestion, and mitigate the associated environmental damages. A new mechanism to reduce urban traffic congestion that is currently gaining traction for its purported cost-effectiveness, environmental-friendliness, and positive health impacts is the adoption of citywide bicycle-sharing systems. This infrastructure provides an alternative to driving for short trips and extends the existing network of public transit within a metropolitan area. Further, bicycling infrastructure augments the environmental bona fides of densely populated urban areas (Kahn, 2010). If bikeshares reduce traffic congestion, they may provide a low-cost policy lever to reduce automobile externalities in urban areas.

Bicycle-sharing programs have seen substantial uptake in European cities such as Amsterdam, Paris, Copenhagen, and London, but only recently have US cities adopted these transportation systems (Nair et al., 2013).[3] Of note, Washington, DC, Minneapolis-St. Paul, Boston, and New York City have in-

[1] http://www.epa.gov/climatechange/ghgemissions/sources.html.

[2] These estimates are the emissions arising solely from congested travel, as opposed to free-flowing traffic.

[3] A typical bikeshare system works as follows. A user registers for an annual, multiday, or (infra-)day membership at one of many bikeshare "docks," which house the bicycles not in use. The user is then given a key, physical or numerical, to unlock a bicycle for transportation and she is allowed to return it to any other dock within the bikeshare system. The user pays according to a rate structure based on the elapsed time of the trip.

stalled city-wide bikeshares. It is thus worth examining the policy importance of environmental benefits of a bicycle-sharing program in urban areas. Specifically, we focus on metropolitan Washington, DC's, Capital Bikeshare, which was introduced in 2010. Schrank et al. (2012) show that the Washington area ranks first in pounds per automotive commuter for CO_2 emissions produced during congested travel, at 631 pounds per commuter annually. This estimate, of course, ignores any local pollutants that contribute to ambient air quality. Further, a journalist at the *Washington Post* notes, "Capital Bikeshare ... funded the original bikes and the docking stations with federal grants earmarked for local programs that mitigate congestion and improve air quality."[4] In particular, Capital Bikeshare leveraged federal money through the Congestion Mitigation and Air Quality Improvement Program (CMAQ), which funds programs in air quality nonattainment areas for ozone, carbon monoxide, and particulate matter that reduce congestion related emissions.[5] In order to assess these environmental benefits, however, a causal link between bikeshares and traffic congestion must be established; identifying this effect is the focus of our paper.

Specifically, we motivate a conceptual framework for examining the impacts of introducing additional transportation options into an existing transit network within a large metropolitan area. Using the introduction of the Capital Bikeshare program as a source of quasi-experimental variation, we examine the effects of bikeshare dock locations on traffic congestion. The expansion of the bikeshare program between 2010 and 2012 allows for identification of changes in congestion within a neighborhood over time. Bikeshare dock locations and ridership history from the start of the program are matched with micro-level traffic data on a finely defined spatial and temporal scale within the city. These data provide an advantageous approach to examining within-

[4]Badger, Emily. "Why DC's bikeshare is flourishing while New York's is financially struggling." *The Washington Post*, 1 April 2014. http://www.washingtonpost.com/blogs/wonkblog/wp/2014/04/01/why-dcs-bikeshare-is-flourishing-while-new-yorks-is-financially-struggling/. Last Accessed: October 13th, 2014.

[5]Fact Sheets on Highway Provisions, http://www.fhwa.dot.gov/safetealu/factsheets/cmaq.htm. Last Accessed: October 13, 2014.

city commuting decisions as opposed to examining changes in traffic patterns on major thruways and arterial highways. In the economics literature, this is the first paper to examine the causal effect of large-scale bicycle-sharing infrastructure on motor vehicle traffic, with implications for environmental and health benefits.

Empirically, we develop a framework to capture the effect of bikeshare systems on traffic congestion in multiple ways. Fixed effects models controlling for time-invariant unobservables in a neighborhood allow us to explore the effect of the bikeshare systems on traffic congestion, while highlighting the bias from endogenous selection of bikeshare locations. Our preferred causal specifications, using the presence of a bikeshare dock as a treatment, utilize a series of matched samples on observable socioeconomic characteristics and pre-treatment traffic congestion to mitigate the effect of selection bias. Estimates from our preferred models range from a 2 to 3% reduction in congestion due to the presence of a bikeshare. For context, a 1% reduction in congestion results in a roughly 1% increase in vehicle miles per hour driven for a representative road segment in our sample. Additionally, we model the effect of bikeshare docks in adjacent neighborhoods. We find some evidence that congestion increases in neighborhoods adjacent to those that possess a bikeshare dock. These spatial spillovers are potentially driven by motor vehicle drivers avoiding bicycle traffic.

In the next section, we outline the relevant literature as it relates to traffic congestion and environmental quality. We develop in Section 3 a stylized model of transportation choice to motivate our empirical analysis. In Section 4, we describe the data used. We present our quasi-experimental strategy and empirical models in Section 5 and discuss results in Section 6. Policy implications are discussed in Section 7, followed by robustness checks in Section 8 and conclusions in the final section.

2 The Economics of Traffic Congestion

Recently, the economics literature on transportation has focused on motor vehicle drivers' behavior. Particularly, research has characterized consumer responses to changes in the price of gasoline as well as the incidence and distributional implications of taxing gasoline to curb its negative externalities (Bento et al., 2009; West and Williams III, 2005, 2007). This research informs the policy-relevant debate concerning the optimal mechanism (e.g., Pigouvian taxation of gasoline) to reduce environmental damages that arise from motor vehicles. In addition, researchers have examined how variability in gasoline prices affects speed and congestion on freeways (Burger and Kaffine, 2009), carpooling behavior (Bento et al., 2013), and the substitutability of other modes of public transportation (Currie and Phung, 2007; Spiller et al., 2012). While this research sheds light on consumers' decision-making process within the existing transportation infrastructure for motor vehicles, it leaves open the question of how consumer behavior changes with the addition of a new, purportedly environmentally friendly, transportation option.

The effect of adding a bikeshare network to an existing transit system can be examined similarly to the introduction of a rail line, for example, as it effectively reduces the relative cost of public transportation. This additional transportation option is particularly relevant in urban areas where public transit is more cost-effective and may, in the short run, circumvent the fundamental law of road congestion (Duranton and Turner, 2011), which posits that VMTs increase proportionally with additional vehicle lanes and exhibit no response from additional public transit service. Among research at the intersection of motor vehicle traffic and public transit, no studies have provided an estimate of the impact of bicycle-sharing programs on traffic congestion nor their ability to augment existing public transit in an urban area.

A parallel literature explores the investment in public transit infrastructure and finds that despite garnering a large fraction of public support, a small number of commuters actually use public transportation. As such, several studies have concluded that investment in public transit does little to reduce traffic

congestion and thus fails to reap the corresponding environmental benefits (Rubin et al., 1999; Winston and Langer, 2006; Winston and Maheshri, 2007; Duranton and Turner, 2011). Beaudoin et al. (2014) provide evidence, however, that a 10% increase in public transit capacity reduces traffic congestion by 0.8%, though this effect is stronger in densely populated cities. Further, Anderson (2014) shows that commuters most likely to support public transit are those who would otherwise commute along highly congested motorways. Using a transit strike in Southern California as a natural experiment, he finds that average highway travel delays increase by about 0.2 minutes per mile, a 47% increase. The evaluation of a bikeshare system's effect on congestion is particularly relevant within this context as it provides a low-cost alternative to larger capital investment and could increase the efficiency of existing transit options by improving accessibility in a metropolitan area.

The literature on the local environmental effects of congestion is sparse in comparison to the research on public investment. However, Parry et al. (2007) lay out an extended review of the ways in which motor-vehicle trips produce environmental and public externalities, including the cost-effectiveness of policies designed to curb these externalities. Overall, the analyses summarized in Parry et al. (2007) indicate that the local environmental effects of traffic congestion are substantial.

Additionally, several papers utilize novel identification strategies to uncover the forces that map transportation policies to environmental and health outcomes. Specifically, Cutter and Neidell (2009) examine the effect of "Spare the Air" information campaigns, in which commuters are asked to voluntarily forgo motor vehicle trips on days when local ambient pollution levels are dangerous. This voluntary mechanism, designed to mitigate the incidence of exposure to ozone in central California, was found to reduce traffic volume and increase public transit usage. Sexton (2012), however, provides an empirical counterpoint to Cutter and Neidell (2009) by examining general equilibrium effects induced by free-fare days for public transportation. Sexton shows that free-fare days for public transit increase motor-vehicle traffic, as well as the corresponding local pollution, due to an unintended reduction in relative costs

of driving. Contrastingly, Chen and Whalley (2012) show that the introduction of an urban rail line in Taipei induced a 5 to 10% reduction in carbon monoxide, suggesting a substantial decrease in tailpipe emissions. Moreover, Currie and Walker (2011) study the effect of reduced congestion induced by the introduction of E-ZPass on infant mortality and birthweight. They provide convincing evidence of the positive public health spillovers of reduced congestion on local environmental quality.

Overall, research in this vein suggests that a reduction in congestion improves local air quality; however, there may be perverse incentives for commuters that work against the desired social optimality of policy designs. In this paper, we build on this literature by examining an altogether different type of transportation policy with implications for environmental and health outcomes through its effect on traffic congestion. While support for bicycle-sharing programs often touts environmental benefits, ex ante predictions of the effect of bicycle-sharing programs on traffic congestion are mixed. We formalize this intuition in a stylized model of transportation mode choice in the follow section.

3 Model of Transportation Choice

To motivate the empirical strategy, we construct a utility-maximization framework that demonstrates consumers' transportation decisions similar to Cutter and Neidell (2009) and Sexton (2012). For each travel opportunity, a representative consumer chooses among car (C), metro rail (M), and bikeshare (B). Utility for individual i is represented by

$$U_i = max\{U_i^C, U_i^M, U_i^B\}, \tag{1}$$

where U_i^j indicates indirect utility for transportation mode $j \in C, M, B$. Car and metro trips are both subject to the impact of congestion, which arises from aggregate demand for each mode of transportation. Let Q^j denote the number of individuals that choose transportation mode j.

The utility gained from each mode of transit is defined by the following:

$$U_i^C = \theta_i^C - P^C U_i^0 - f^C(Q^C, Q^B) \tag{2}$$

$$U_i^M = \theta_i^M - P^M U_i^0 - f^M(Q^M) \tag{3}$$

$$U_i^B = \theta_i^C - P^C U_i^0 - f^B(Q^B, Q^C). \tag{4}$$

Each transit mode involves some level of intrinsic utility θ^j. This may include, among other things, the convenience of a particular mode or the private benefits of reducing pollution by not driving. The utility obtained from a numeraire composite consumption good is denoted as U_i^0. Multiplication of this term by the per unit cost of each transit mode, P^j, gives utility net of private consumption.

The final term in each of the above equations implies the substitutability of the transportation modes and highlights the behavioral responses that are key to the analysis. As mentioned earlier, congestion is a factor for transportation choices so the aggregate quantity for each mode directly impacts utility for that mode. In addition, car and bike travel must physically compete for space on the road. Therefore, the quantity of cars on the road impacts congestion for bike travel, while the quantity of bikes on the road impacts congestion for car travel. Thus, f^C represents the disutility of congestion when driving that is determined by both cars and bikes, f^M represents the disutility of congested metro trains, and f^B represents the disutility of congestion when biking, also determined by both cars and bikes. Each of the $f(\cdot)$ functions is increasing in all arguments.

In addition, the aggregate quantity of car trips, Q^C, directly impacts utility from choosing the bikeshare option. This could be the result of two factors. First, cars and bikes may have to physically compete for space on the road. Many roads (including portions of the study area) have dedicated bike lanes, though this is not true of all locations. Even with bike lanes, an increase in the number of cars on the road can increase the risk of accidents involving bikes and create an increasingly congested area that slows bike travel. Second, an

increase in the number of cars on the road will increase the level of automobile emissions. An increase in emissions causes damages for the entire population, but individuals on bikes may be more susceptible to local pollutants.

Given that environmental concerns serve as a major justifications for bikeshare programs, it should be noted that air pollution as it impacts the entire community is not absent from this analysis. It is simply subsumed by the ordinal nature of utility. While the external costs from car travel are likely significant, this study is focused primarily on measuring transportation choices. Thus, we consider such costs from an aggregate perspective and model only their source.

Given the above utility framework, we can analyze the impact of introducing a bikeshare program. Access to the bike option may attract riders who would have otherwise driven, used the metro rail, or taken no trip at all. If bikeshare riders are substituting away from driving, additional bike trips will offer the external benefits of reduced traffic congestion and emissions. Of course there then exists the impact of bike trips on road congestion. The components of the utility functions help to illustrate the role that congestion plays. An increase in bike trips may also reduce traffic congestion generated by cars. A reduction in congestion, however, increases the utility of driving, mitigating the impact that bikeshares would otherwise have on aggregate demand for car trips. At the same time, the reduction in congestion also increases utility from using the bikeshare, creating a feedback effect in the opposite direction. If bikeshare riders are substituting away from using the metro rail, the benefits of reduced traffic congestion and air pollution do not exist. The only external benefits in this case would be those related to congestion on the metro.

The conceptual model outlines the interactions among different modes of transportation in a consumer's choice set. While there exists a multidirectional relationship between the different modes of transportation, a full general equilibrium model is necessary to identify the individual forces at work. Rather, we are currently concerned with the net effects of the bikeshare program on alternative modes of transportation. More specifically, we are interested in the degree to which the existence of the bikeshare program has

impacted the demand for car trips. This suggests the use of an aggregate measure of car trips as a key variable. Due to data availability and the mechanism by which quantities of alternative modes impact decisions, however, we use a measure of traffic congestion to capture the relationships discussed above. The following sections outline the empirical strategy used to estimate these impacts.

4 Data

4.1 Transportation Choice Data

The Capital Bikeshare, which began in September 2010, serves the metropolitan Washington, D.C., area. It is funded publicly by the District of Columbia, City of Alexandria, and Arlington County and operated by a private company. In the fall of 2013, Capital Bikeshare expanded to Montgomery County in Maryland. Uptake in ridership and the number of station locations increased substantially over its first three years of operation. Figure 1 shows the number of rides initiated at Capital Bikeshare docks over our study period, demonstrating an obvious increase in ridership.

We are first concerned with the locations of bikeshare docks, analyzing the impact arising from the existence of docks. The geographic locations of these docks are publicly available from Capital Bikeshare. Our data include 158 docks located throughout the metropolitan area. It is important to note that some bikeshare docks were established after the first period of our dataset. Thus observations of traffic congestion in the same block group before and after a dock is established serve as the foundation of our identification strategy.

A key component of our analysis is access to traffic data that is finely disaggregated on both a spatial and temporal scale. Furthermore, we use observations of city streets and arterial roads, rather than only major highways. Traffic data were obtained through partnership with the CATT Lab at the University of Maryland, College Park. Archived real-time information on traffic speed by road segments was obtained through the Vehicle Probe Project

(VPP) Suite within the Regional Integrated Transportation 5 Information System (RITIS). The unit of observation for speed data is a road segment that is identified by its Traffic Message Channel (TMC) code and exhibits a much richer geographic span than standard traffic monitors that capture flow at various points along major highways. Each TMC road segment is characterized by the latitudinal and longitudinal coordinates of its start and end points and tends to be less than one-half of a mile, with a mean and median length of 0.37 and 0.24 miles, respectively, in our sample. These road segments cover a more comprehensive range of within-city traffic patterns than was covered in previous research. To manage the spatial nature of our dataset, we construct road midpoints as the geographic midpoint between start and end points. These midpoints define a road observation. For each TMC road segment, real-time speed information is transmitted via radio frequency from vehicle GPS navigation systems. Over the multiple years of the study, archived speed data is accessible for every TMC road segment for every minute.

Recall that the primary variable of interest is traffic congestion. We construct a normalized metric of traffic congestion by comparing observed speed to a reference speed (defined by CATT Lab) that is a typical historical speed for a road segment. Congestion for a particular segment is defined as follows:

$$\text{CONG}_j = \frac{\text{Speed}_j^R}{\text{Speed}_j^O}, \tag{5}$$

where Speed_j^R is a constant reference speed for free-flowing traffic on road segment j determined by the prevailing speed limit and historical speed patterns and Speed_j^O is the observed speed at any point in time. Given this definition, congestion is decreasing in observed speed. We measure congestion using speeds aggregated to 30-minute intervals, focusing on morning rush-hour commuting times from 6:00 am through 10:00 am. In addition, we look at April, May, September, and October from 2010 to 2012, resulting in 3,120 time periods across 2,790 road segments. Specific months are chosen to capture times of the year in which cycling is a reasonable commuting option and commuters have typical work schedules. Summary statistics for our dependent variable

congestion are shown in Table 1.

4.2 Census Block Groups

An important component of our analysis is the spatial nature of the data, as bikeshare docks and speed (and thus congestion) are observed at particular geographic coordinates. It is therefore necessary to establish a geographic link between the variables.

We use US Census Block Groups as the geographic unit of observation. Census block groups make up the second smallest geographic area defined by the US Census. The smallest designation, census blocks, is unsuitable for the analysis because of our treatment of road segments. Since we use the midpoint of each road as the designation of its location, census blocks are small enough that a road may pass through several blocks but of course its midpoint is in only one. This problem is largely alleviated by using a slightly larger geographic designation. In addition, the size of census blocks makes it likely that individuals move across blocks for their commuting choice (e.g., an individual might walk to a bikeshare dock in a bordering block or bike through a block that borders that of the usual car commute), so that such a small area may not accurately capture the relationship of transportation mode choices. Our study area includes 488 block groups.

The question becomes one of the relationship between aggregate demand for each mode of transportation in a particular block group. Through the use of GIS software, bike trips are easily linked to the block group in which any particular dock is located. For congestion data, we aggregate to the block group level by taking mean congestion for all road segment midpoints located within a particular block group. Figure 2 shows block group boundaries of the study area with road segment midpoints overlaid. Bikeshare docks (as of the final time period of our sample) and Metrorail stations are displayed in the same geographic boundaries in Figure 3. Our sample includes 889,800 observations of block group-time combinations. As shown Table 1, 12.9% of our sample observations are block groups with bikeshare docks.

4.3 Adjacent Block Groups

As a final component in the empirical analysis, we consider the impact of bike docks that are close to road segments, but perhaps are in different census geographies. Individuals are not confined to a particular census block. An individual may, for example, forgo a car trip and use a bikeshare dock in an adjacent block group. It is therefore possible that the impact we are trying to identify is not confined solely to transportation demand within the same block group. In addition, geographic space is continuous and census block groups are somewhat arbitrary to individuals. Thus we are concerned about docks that may be located in different block groups, though they may be very close to the area of measured road congestion. For each block group, we measure the docks that are in neighboring block groups (those that share a border). We also measure the docks that are in neighboring blocks, obtaining a better measure of those docks that are close to measured congestion, but perhaps just across a block group border. Across our sample of block groups and time periods, 40.4% of our observations have a bikeshare dock in a bordering block group, and 21.6% have a bikeshare dock in a bordering block.

Table 2 displays summary statistics for our bikeshare dock count measures. Summary statistics show that capturing bikeshare docks in adjacent locations could be considerably important. Of the 488 block groups, 19.9% have a bikeshare dock at some point in time. Of these, only 32% (14.4%) have more than 1 (2) dock. Alternatively, 57.4% of block groups have a dock within a neighboring block group at some point in time, and 75.7% (57.5%) of these have more than 1 (2) dock within the neighboring block groups.

5 Empirical Strategy

Our general empirical approach seeks to identify a causal effect of bikeshare programs on traffic congestion. In this section, we first develop a series of models controlling for unobservable effects at the block group level to assess the relationship between traffic congestion and various metrics of bikeshare

existence and use. Within these models, we note the potential bias induced by nonrandom siting of bikeshare docks. To correct for this, we develop a treatment effect model that mitigates this bias using propensity score matching. Within the latter framework, we estimate an average treatment effect of the bikeshare on congestion.

5.1 Panel Data Model

To estimate the impact of the bikeshare program on traffic congestion, we take a reduced form approach and estimate several different linear equations. The first general model examines the relationship between the existence of bikeshare docks and the average level of traffic congestion in a block group. We estimate the natural log of traffic congestion as

$$\ln CONG_{jhmt} = \alpha + \delta_j + \nu_h + \mu_m + \tau_t + \beta_D Dock_{jhmt} + \varepsilon_{jhmt}, \qquad (6)$$

where $CONG_{jhmt}$ denotes average congestion among all road segments with midpoints in block group j during half-hour period h, month m, and year y. The parameters δ_j, ν_h, μ_m, and τ_t represent block group, half-hour period, month, and year dummy variables, respectively. The variable $Dock_{jhmt}$ indicates whether a dock exists in block group j in time period $\{h, m, t\}$. We are therefore primarily interested in the coefficient β_D.

Two additional specifications each include an additional variable to indicate the presence of a dock in an adjacent area. These specifications use the same definition of average congestion as (6) and are otherwise identical except for a single addition variable. In the second specification we include $Dock_Adj_{jhmt} = 1$ if there is a bikeshare dock in an adjacent census block group. In the third specification, $Dock_KAdj_{jhmt} = 1$ if there is a bikeshare dock in an adjacent census block.

Given that a block group may have more than one dock, and of more consequence, that a block group may be surrounded by a large number of bike docks, an additional set of equations estimates the impact on congestion of the number of docks in or adjacent to a block group. Therefore, three

specifications employ a count of docks as independent regressors,

$$\ln CONG_{jhmt} = \alpha + \delta_j + \nu_h + \mu_m + \tau_t + \beta_D Dock_Count_{jhmt} + \varepsilon_{jhmt}. \quad (7)$$

The variable $Dock_Count_{jhmt}$ is equal to the number of bikeshare docks in a block group. In addition to Equation (7), we again consider the impact of nearby docks with two additional specifications that include $Dock_Count_Adj_{jhmt}$ and $Dock_Count_KAdj_{jhmt}$, respectively. These variables are similarly defined as the number of docks in bordering block groups and blocks, respectively.

We refer to the econometric framework of Equations (6) and (7), including specifications with variables to indicate nearby docks, as the fixed-effects panel data (FEPD) model.

5.2 Treatment Effect Model

In our empirical framework we recognize the potential selection bias when examining the impact of the existence of bikeshare docks. It is likely the case that docks were established in areas of high congestion for two reasons: these are the locations that are most in need of bikeshare docks from the city's perspective and these are likely the locations with the highest demand for bikeshares. Alternatively, docks require ample sidewalk space and may be best suited to areas of residential and commercial concentration, rather than commuting corridors. In either case, it is well known that endogenous selection of treatment (i.e., bikeshare docks) of this nature will generate biased estimates. Therefore, we use propensity score matching techniques to properly identify the impact of bikeshare docks. The following section outlines our approach to and justification for matching.

Define treatment status T to be equal to 1 (treatment) if a block group contains a dock and equal to 0 otherwise (untreated). Define the potential outcome in the treatment condition as $CONG^1$ and the potential outcome in the untreated condition as $CONG^0$. The goal of the analysis is to identify the

average treatment effect on the treated (ATT),

$$ATT = E[CONG^1 - CONG^0 \mid T = 1]. \tag{8}$$

The econometric strategy discussed in the previous section essentially estimates

$$E[CONG^1 \mid T = 1] - E[CONG^0 \mid T = 0]. \tag{9}$$

Equation (9) is only an unbiased estimator of the ATT if $E[CONG^0 \mid T = 1] = E[CONG^0 \mid T = 0]$, which is unlikely to be true given the decision to locate bikeshare docks in areas of high congestion. To the extent that this is true, we expect our FEPD estimates to be biased upwards.

The matching technique we employ is based on the more plausible assumption that

$$E[CONG^0 \mid X, T = 1] = E[CONG^0 \mid X, T = 0]. \tag{10}$$

Equation (10) states that observed $E[CONG^0 \mid X, T = 0]$ offers a proper counterfactual for the unobserved $E[CONG^0 \mid X, T = 1]$. By conditioning on the set of variables X, we can create a counterfactual sample of block groups with no docks (untreated) that are observably similar to those that do have a dock (treated). Rather than matching on the multidimensional set of covariates X, we use propensity scores, $P(X)$, which denotes the predicted probability that a particular block group will have a bikeshare dock conditional on X.

Note that we combine the identifying power of the FEPD estimator with that of matching, following the approach of Ho et al. (2007) and Ferraro and Miranda (2014). The fixed effect specification alone controls for any congestion shocks that are time-invariant or specific to a particular block group. Our matching strategy becomes important in the context of the implicit assumptions of the model. The econometric framework assumes homogeneous treatment effects and similar time trends. By matching observations based on block group descriptors and pre-treatment congestion, we make these assumptions more plausible. Thus we use propensity score matching to process the

data and generate a set of control observations that are similar to the treated block groups. We then apply the FEPD estimator to the matched samples to obtain unbiased estimates. A further discussion of our matching approach and its empirical content follows a brief review of results from a naive FEPD model that fails to account for any selection bias.

6 Results

6.1 FEPD Estimates

Results from estimation in the fixed effects panel data framework are reported in Tables 3 and 4.[6] All estimates contain block group and time fixed effects as discussed above. The following estimates should be interpreted in light of the previously discussed potential for biased estimates. The FEPD model does not account for any endogenous selection into treatment and the nature of dock siting suggests an upward biased in the estimated coefficient on the *Dock* variable.

Estimates in Table 3 offer mixed evidence for the effect of the bikeshare program. The first and third columns show positive significant impacts of the presence of a dock. The second specification in Column 2 indicates a significant reduction in the average traffic congestion of a block group due to the presence of a bikeshare dock. The coefficient in Column 1 is interpreted as a 0.426% increase in a block group's congestion. Similarly, the estimated congestion reduction in the second specification is only 0.21%. In Table 4, the measured impact of bikeshare docks on traffic congestion when considering the number of docks is similar to the previous specification. Again, only a small percentage of block groups have multiple docks, so this result is expected. A congestion-increasing impact of bikeshare docks is evident in specifications 1 and 3, though a small congestion-reducing impact is found in specification 2. Recall that there is reason to expect these estimates to be biased upwards due

[6] Due to computation constraints, estimation proceeds on a random sample of 50% of our observations.

to the siting of bikeshare docks. We will return to this analysis as it relates to our matching approach to correct for any bias.

The FEPD model is presented as a baseline to highlight the expected bias generated by endogenous selection of dock locations. Therefore, we place little confidence in the marginal impacts discussed in this section. In the following section, estimates from a treatment effects model that uses propensity score matching demonstrate the importance of controlling for the endogenous selection of dock locations.

6.2 Propensity Score Matching

Based on the potential endogeneity problem in previous specifications, we employ matching techniques in an attempt to obtain unbiased coefficient estimates and identify a causal impact. We focus on the results of the specification in Equation (6), examining the impact of the presence of docks. The choice to use this model moving forward is based on a treatment effect defined as the presence of a bikeshare dock, rather than multiple treatments that account for the number of docks. In the context of a quasi-experimental approach, recall that selection into treatment may occur if the decision to establish a bikeshare dock in a particular location is dependent on the level of congestion in that location. Our matching technique therefore seeks to establish a set of control block groups with no bikeshare docks that are similar to those block groups that do have bikeshare docks, so that our analysis has a sound counterfactual. We then estimate linear regressions identical to the specification in Equation (6) on our constructed sample of matched observations.

Our matching approach rests on the policy guidelines that determined the location of bikeshare docks. Without access to an explicit set of rules or methods to site docks, there is reason to believe that the siting process depended on both socioeconomic variables and current traffic congestion. In an August 2010 application for funding from the Transportation Investments Generating Economic Recovery II (TIGER II) Competitive Grant Program administered by the US Department of Transportation, the Metropolitan Washington Coun-

cil of Governments indicated a criterion for expanding the Capital Bikeshare conditional on improving transportation options for underserved populations. In particular, the report reads, "[Capital Bikeshare] will bring an affordable, convenient, and healthy travel option directly to 30% of the region's households and population and provide access to 45% of the region's jobs, particularly in areas where low-income and/or transit-dependent populations are concentrated."[7] However, a policy report in Georgetown's Public Policy Review shows that bikeshare docks are located predominantly in wealthy areas of metropolitan Washington with docks located sparsely throughout impoverished neighborhoods.[8] Since dock locations are not independent of socioeconomic neighborhood characteristics nor transit accessibility, this suggests that a matching approach along socioeconomic and pre-treatment traffic patterns is a viable strategy to reduce bias from any nonrandomness in the siting of a bikeshare dock. Finally, we consider the possibility that bikeshare docks may serve as complements or substitutes to the Washington DC, Metrorail, the rapid transit system that serves the metropolitan area. Individuals may use the bikeshare for partial trips from home to a Metrorail station or from a Metrorail station to their final destination. Using similar bikeshare data, Gebhart and Noland (2013) demonstrate that bikeshare users exhibit different ridership behavior at docks within close proximity to Metrorail stations. This result establishes the important role that Metrorail stations play in the individual's transportation decision. We therefore use distance from a block group to the nearest Metrorail station, measured as a simple geographic distance, as an additional matching variable.

While our dataset includes observations across block groups and time periods, our matching approach concerns only block groups. This is based on

[7] A Regional Bike-sharing System for the National Capital Region, Metropolitan Washington Council of Governments, August 23, 2010, p. 12. http://www.mwcog.org/uploads/committee-documents/bV5YWlxe20100820155649.pdf. Last accessed, October 13, 2014.

[8] Johnson, Kristine. "Capital Bikeshare in low-income areas: The question no one is asking." Georgetown Public Policy Review, Domestic Policy, Energy & Environment. September 2, 2014. http://gppreview.com/2014/09/02/capital-bikeshare-low-income-areas-question-one-asking/. Last accessed, October 13, 2014.

the following assumption: given that a dock is to be established in a particular block group, the timing of its establishment is random. Note that the decision to build a dock is not a decision each period, but instead a onetime decision to locate a semipermanent dock in a block group. There are instances of removal of a dock in our dataset, but no instances of removal followed by reestablishment.

An important caveat to note is that we do not incorporate dedicated bicycle lanes into the analysis. Such bicycle lanes are obvious complements to the bikeshare program that not only increase the ease of commuting by bike, but may also have an impact on traffic congestion through their existence and location. As of 2009, Washington DC, had 44.7 miles of bicycle lanes, up from only 24.7 miles in 2006. However, over our study period, miles of bicycle lanes increased only slightly to 50.3 in 2010 and to 51.3 in 2011.[9]

We construct a subsample of the full dataset of treated block groups and untreated block groups, in which the untreated block groups are chosen as the most observationally similar to the treated block groups. To accomplish this we use a block group's propensity score $P(X)$, the probability that a block group has a dock conditional on the set of observed covariates X. We estimate a probit model to predict $\hat{P}(X)$. Due to the desire to match on covariates before treatment, we use publicly available US Census data from 2009 to construct X.

The set of covariates X used to estimate and predict propensity scores is shown in Table 5. We use block group socioeconomic data to help predict the probability that a dock is sited in a particular block group. We calculate mean congestion and the standard deviation of congestion in each of the three months prior to the bikeshare program being developed, in an effort to match the trend in the dependent variable. To define a match, we use both nearest neighbor and caliper matching techniques to find a valid control block group. The caliper approach differs only slightly from nearest neighbor matching. With caliper matching, a treated observation may have more than one corresponding

[9]DC Department of Transportation: http://ddot.dc.gov/publication/2014-bike-program-fact-sheet.

control block group. This allows us to take advantage of multiple control observations that may all be very good counterfactuals, rather than having to choose only the closest. Caliper matching also removes outliers and inliers from the dataset. Treatment observations for which there is no untreated observation with a propensity score within the caliper range are dropped from the sample. In contrast, nearest neighbor matching forces the best match, regardless of how close of a match it may be. The choice regarding the size of the caliper is left primarily to the researcher. As its objective is to remove poor matches, we match with a caliper equal to .05, as well as an alternative match with caliper equal to the standard deviation of the predicted propensity scores, ranging from approximately 0.18 to 0.21. To avoid issues that arise from having to choose a particular order of matching (Rosenbaum, 1995), we match with replacement in all specifications.

6.2.1 Matched Samples

We estimate propensity scores using three matching techniques. The use of congestion as a covariate, however, forces us to change the time periods allowed in our sample. To avoid matching on our dependent variable, only observations from 2011 onward are used in the estimation sample. In this way, our treatment and control groups are matched based on observed characteristics of the block groups, as well as congestion before the possibility of treatment. The drawback to this approach is that we lose identification power derived from observing the same block group with and without a dock. We drop block groups that received treatment in 2010 (3 block groups) from the sample of potential matches, as they lack 2010 pre-treatment observations of congestion.

Coefficients from the probit regressions indicate that among the socioeconomic variables, Med_Inc, Per_Educ, and $Dist_Metro$ were significant predictors.[10] When using only the pre-treatment congestion variables, mean congestion in September and October, as well as $Dist_Metro$, had statistically significant impacts in predicting dock locations. With the full set of covari-

[10]Full estimation results are available from the authors upon request.

ates, Per_Educ, $Dist_Metro$, and mean congestion in September and October were again statistically significant.

In assessing the performance of our matching technique we analyze block group descriptors in the treatment and control samples. The objective of matching is to obtain balanced samples in which covariates are similarly distributed across samples. We refer to the set of unmatched observations from 2011 onward as the full sample.

6.2.2 Full Sample

Looking first at the full sample, Table 6 shows mean covariate values for block groups with bikeshare docks (treatment group) along with mean covariate values for block groups without docks, those that would act as counterfactuals using the full sample. For each covariate, we conduct a hypothesis test with the null hypothesis that both means are the same. The p-value from each of these tests is reported in the third column. In addition, we perform a Kolmogorov-Smirnov test (K-S test) for each covariate. The K-S test examines the empirical distribution functions of the treatment and control observations to test whether the samples were drawn from the same distribution. Again, the p-values correspond to a test with a null hypothesis that the treatment and control samples were drawn from the same distribution. Finally, we also report the ratio of sample variances of the two samples. From Table 6, the p-value and K-S p-value suggest that sample selection bias likely exists in the full sample. With the exception of Sq_Mi a test of means leads us to reject the hypothesis that sociodemographic covariates have identical means across the treatment and control samples at the 5% significance level. For Med_Inc, the test is rejected at the 10% significance level. Similar results are clear for the K-S test of the covariate distributions. Pre-treatment congestion is already fairly well balanced for May and September but the hypothesis of equal means is rejected for October, the closest month to possible treatment.

Table 7 displays similar statistics for covariate balance in the matched samples. The three panels correspond respectively to the three methods of defining a match: nearest neighbor, caliper = σ, and caliper = 0.05. The

full sample includes 82 treated block groups. The nearest neighbor matched sample thus has 82 treated and 82 control block groups. Using a caliper of (σ), the number of treated block groups decreases to 61, as block groups are dropped from the analysis for not having an appropriate counterfactual. Given the allowance of multiple matches with calipers, these block groups are matched to 71 control block groups from the full sample control of 337. With the smaller caliper equal to 0.05, the number of treated block groups decreases to 52, matched to 62 control block groups. While a comparison of means suggests fairly well-balanced samples, formal hypothesis tests support considerable improvement from matching. With the exception of $House_Age$ in the nearest neighbor match and larger caliper match, mean comparison p-values are insignificant for all socioeconomic covariates. The K-S tests show similar results and offer additional support for balanced samples. Variance ratios show modest improvement beyond the full sample. For distance to the nearest Metrorail, we continue to reject the hypothesis of equal means across all three specifications. While formal hypothesis tests on distance to the nearest Metrorail station continue to offer weak evidence of perfectly matched samples, treatment and control means and variances are much more closely aligned, relative to the unmatched sample. In general, results suggest that our matching approach is effective at balancing covariates across treatment and control samples, thus reducing any bias that may be present in estimation.

6.3 Event Study

To further motivate our regression analysis, we present a graphical representation that examines the impact of treatment in the form of a bikeshare dock, using an analysis similar to that of Davis et al. (2014). We run the following regression with the objective of observing the trend of block group congestion before and after treatment,

$$\ln CONG_{jhmt} = \alpha + \sum_{k=-17}^{25} \beta_k 1[p_{jhmt} = k] + \delta_j + \nu_h + \mu_m + \tau_t + \varepsilon_{(jhmt)}, \quad (11)$$

on our sample matched on a caliper of .05, using only the treated block groups. The term p_{jhmt} represents the 30-minute time period in our sample. We then group the 30-minute periods into intervals of 50 consecutive periods to generate 17 aggregated intervals before treatment and 25 aggregated intervals after treatment. The term $1[p_{jhmt} = k]$ is an indicator variable that equals 1 if the number of 30-minute intervals before or after treatment falls into interval k. Thus k suggests the length of time before or after treatment and $k = 0$ denotes the exact period during which the dock was established. The regression also contains fixed effects for block group, year, month, and time, similarly to our main empirical analysis. Designating the first period of treatment as the omitted category, the β_k coefficients indicate any congestion trend leading up to and following treatment. Figure 4 shows a fitted trend line with 95% confidence intervals for the β_k coefficients against periods relative to treatment, using a locally weighted polynomial regression. The top panel shows regression results for the treated block groups. In the bottom panel, we find a similar trend line for coefficient estimates when Equation (11) is estimated on observations in control block groups. Control block group observations are assigned a hypothetical time before or after treatment based on their corresponding treated block group match and time period.

Results suggest preliminary evidence of the impact of bikeshare docks. For treated block groups, we see a slight upward trend in congestion that levels off and decreases slightly at the point of congestion, implying a congestion-reducing effect of treatment. Such an effect, however, is not evident for our control observations, as congestion remains relatively constant over the sample period. With indications of a causal impact of the presence of bikeshare docks, we now turn to estimates from our regression analysis.

6.4 Matched Samples Estimates

Given strong evidence of successful matching and graphical treatment that suggests an impact of treatment, we now turn to our regression results. We report additional baseline estimates that correspond to the full sample of post-

2010 unmatched observations, followed by estimates from matched samples. Baseline estimates, reported in Table 8, are close to those reported earlier in Table 3, though we see slightly more evidence of a congestion-reducing impact of bikeshare docks. For specifications 1 and 3, the impact of a dock is insignificant. Specification 2 indicates that the presence of a dock reduces congestion, though by a very small amount.

To put congestion reduction in context, consider a 1.0% reduction in traffic congestion. Relative to mean congestion in our sample of 1.23, for a road segment with a reference speed of 40 mph, this change in congestion corresponds to a change in actual speed 0.98%. Alternatively, a 1% change corresponds to a reduction from median congestion to the 47th percentile.

From the coefficients *Dock_Adj* and *Dock_KAdj*, our results indicate a congestion-increasing effect of docks in neighboring geographic areas. A potential explanation for this is a spillover effect, in which there is substitution from adjacent block groups among drivers that seek to avoid bike traffic. Thus the increase in congestion could be the result of car traffic that would otherwise have been in a different block group. Note that the estimated impact of a dock in a neighboring block group is considerably larger than that of a dock in a neighboring block. This may be the result of using only an indicator variable for the presence of docks when the distribution of the number of docks in an adjacent block group is more positively skewed than the number of docks in an adjacent block. Still, these results suggest that the geographic spillover of traffic due to the presence of bikeshare docks could be a significant factor in analyzing traffic congestion impacts.

Alternatively, a possible explanation for the positive coefficient on adjacent docks is an additional endogeneity problem. Bikeshare docks are highly centralized in transportation corridors of the study area, so treated block groups are more likely to be located close to other block groups with docks. Given this correlation, endogenously treated block groups with high congestion have a higher probability of adjacent docks, leading to upward bias in the coefficients on *Dock_Adj* and *Dock_KAdj*. Though similar to earlier results, the magnitude of coefficients in matching specifications is greater than that of

estimation on the full sample.

Recall that each specification achieved strong balance across treatment and control observations. Average treatment on the treated (ATT) estimates, reported in Table 9, are negative and statistically significant across models and specifications. Results using nearest neighbor matching indicate the presence of a bikeshare dock reduces congestion by upwards of 2.5%. With the larger caliper, the impact of the presence of a dock is a decrease in congestion ranging from 2.81 to 3.46%. For the smaller caliper, the estimated decrease in congestion ranges from 1.65 to 2.61%. For context, a coefficient of -0.025 implies an increase in speed of 2.55%. Differences from the full sample estimates are significant at the 1% level for all three specifications within each of the matching methods. There is less consistency in the impact of bikeshare docks in adjacent locations, relative to the basline model or across specifications. From the coefficients on $Dock_Adj$ and $Dock_KAdj$, our results again indicate a possible spillover effect in which bikeshare docks increase congestion in neighboring geographic areas, though the range of estimates makes this conjecture less conclusive.

With the exception of the first specification in the nearest neighbor matching model, a pattern in coefficients emerges as we examine the three specifications in each of our models. The first specification leads to the smallest estimated congestion-decreasing effect. When we move to the third specification, controlling for docks that are in neighboring areas but still geographically close, the magnitude of this effect increases. In the second specification, controlling for docks in broader neighboring areas, the magnitude of the effect increases. One explanation for this is the spatial concentration of bikeshare docks, and thus the possible confounding of effects in the first specification. The coefficient on $Dock$ in column 1 may be partially picking up the impact of (econometrically) unaccounted for docks in nearby locations. Likewise, the coefficient on $Dock$ in column 3 may be partially picking up the impact of docks that are in neighboring block groups, but beyond the boundaries of adjacent blocks.

6.5 Heterogenous Treatment Effects

A final empirical model examines whether the impact on congestion is uniform across road types. The matched samples are divided into quartiles of reference speed to distinguish between roads with low reference speeds and those with higher reference speeds that may better represent traffic commuting corridors. We run our previous specifications on the partitioned subsamples, estimating the impact of a bikeshare dock for each subsample. We report coefficients for each of the quartiles from samples generated by matching with the smaller caliper, as results are similar across specifications. Speed quartiles for the sample are 21.50, 25.63, 31.25, and 56.00. While the upper bound of the 4th quartile (sample maximum) is relatively large, note that this subsample has a mean of 35.09 and a median of 33.25, showing considerable skew in the overall distribution.

Figures 5 and 6 display coefficient estimates and 95% confidence intervals for the *Dock* coefficient in the second and third specifications, controlling for docks in adjacent locations. Results indicate a much larger impact for the third and fourth quartiles, those observations on block groups with higher average reference speeds. For the second specification, in which we control for bikeshare docks in adjacent blocks groups, it is clear from Figure 5 that the impact is much larger in block groups with higher average reference speeds, as the *Dock* coefficient is -0.032. These results are somewhat expected, as the congestion-reducing impact of the bikeshare program should be strongest on those roads that are part of the morning commute, rather than roads in slower residential areas. Earlier results suggest an impact that is in between that seen here for the lower quartiles and the upper quartiles. Thus our average impact is picking up a very strong congestion-reducing force on some roads, while being partially mitigated by little impact on other roads. When controlling for bikeshare docks in bordering blocks, in Figure 6, estimates are fairly consistent across quantiles.

7 Discussion

In general, coefficient estimates from matching models imply a causal link between the presence of bikeshare docks and congestion reduction. Given evidence of improved covariate balance in several matching models, our regression results suggest that there is self-selection of block groups into treatment. The FEPD estimates are biased by observations of bikeshare docks that are placed in high-congestion areas, thus generating a positive correlation between the presence of a dock and congestion levels. Matching prevents low-congestion block groups with no docks from serving as the counterfactual for high-congestion block groups that select into treatment, i.e. have a dock established. Thus coefficients estimated from matched samples present a more accurate estimate of the causal impact of bikeshare docks, and we see considerable evidence that bikeshare docks are effective at reducing congestion in their immediate areas.

An examination of the adjacency coefficients relative to those on *Dock* illustrates the possibility of other important effects of the presence of bikeshare docks. First, results indicate an impact of docks in adjacent locations (block group or block) with a magnitude as large as the impact of a dock in the location. Thus the traffic spillover effect is substantial relative to the direct effect of an alternative transportation mode. Such a result raises the possibility that the impact of bikeshare docks is simply to push traffic to different locations, with little or no aggregate impact on pollution. Second, the impact of a dock in an adjacent block group appears to be significantly larger than the impact of a dock in an adjacent block. While this could simply be the result of failing to control for the larger number of docks that exist in a geographically larger block group, it may also be explained by the continuous nature of transportation in our geographically discretized model. Since a bikeshare dock in a neighboring block more often tends to be an easy alternative for a commuter, relative to the designation of a dock in a neighboring block group, such a dock should provide some congestion reduction. This effect as it appears in our estimates could therefore be mitigated to some degree by a congestion-

reducing force of nearby docks. Of course, it should also be noted that these estimates could simply be the result of endogenous siting of docks, in which there is some spatial correlation in siting that leads to a higher probability of high-congestion areas having docks in adjacent locations.

Finally, a pattern of coefficients across the three specifications controlling for nearby bikeshare docks illustrates another important point related to the spatial nature of transportation. While block group boundaries are an essential part of the empirical analysis, such boundaries mean very little to commuters. It is likely that commuters treat the landscape as continuous rather than being made of discrete geographies. Controlling for docks in nearby blocks, we see a larger congestion-reducing (smaller congestion-increasing) impact of a bikeshare dock at the point of measurement for all models and specifications. The congestion-reducing impact becomes even larger when we control for docks in a larger nearby geographic area, neighboring block groups. At the same time, the positive spillover effect of nearby docks becomes significantly larger. We interpret this as a confounding effect that appears in the *Dock* coefficient. When we fail to properly account for the positive spillover from nearby docks, this congestion-increasing impact obscures the congestion-reducing impact of the dock in the measured location.

8 Robustness

We conduct additional empirical analysis to examine the robustness of our results. First, we estimate clustered standard errors that are robust to heteroskedasticity and correlation. Two models cluster on time (half-hour period) and block group, respectively. Results are presented in the Appendix. Table A1 displays coefficient estimates with standard errors clustered on time in the second row of each panel. This allows for heteroskedasticity across time periods and correlation among residuals across block groups during the same time period. Standard errors are only slightly larger than those reported earlier and all previously significant coefficients remain so. An alternative approach is to cluster standard errors on block groups. This allows the unobserved variance

in congestion over time to vary across block groups and controls for autocorrelation over time. However, time periods are sequential only for several half hours in a given day, but then jump to different days and different months. Still, heteroskedasticity across different block groups may be closer to the true correlation structure than heteroskedasticity across time. Results for this error specification are shown in third row of each panel in Table A1. Results from the samples generated by nearest neighbor matching exhibit a loss in statistical significance on the *Dock* coefficient. In the caliper matched samples, however, significance is maintained on the *Dock* coefficient in specifications that control for adjacent bikeshare docks.

A second concern is that of the caliper size and the degree to which the chosen caliper size influences our estimates. Recall that the above results used two different calipers, sized 0.05 and 0.26, respectively, where the value of the larger caliper reflects the standard deviation of fitted propensity scores. We report sensitivity analysis for both specifications that control for docks in adjacent localities. We then reestimate the model with matched samples based on a caliper ranging from 0.04 to 0.30. Coefficient estimates on *Dock* and the adjacency variable, along with standard errors, are shown in Tables A2 and A3. In the second specification, estimates of the congestion impact of a bikeshare dock range from -2.40%, when using a relatively narrow caliper equal to 0.10, to -3.46% when the largest caliper of 0.30 is employed. Also note that the magnitude of the coefficient on *Dock_Adj* tends to behave in an opposite manner, reflecting the earlier discussion of a pattern in coefficients. For the third specification, there is a larger coefficient range, from 1.66 to 2.94, for calipers of size 0.10 and 0.30, respectively. A similar pattern emerges for the coefficient on $Dock_KAdj$.

9 Conclusion

In this analysis we present causal evidence of the impact of bikeshare programs on traffic congestion. Using a unique dataset of city roads, we construct a finely spatially defined measure of congestion that allows us to examine congestion

effects at a disaggregated geographic level. A panel dataset of half-hourly traffic observations at the census block group level suggests that the existence of a bikeshare dock in a block group reduces traffic congestion. To control for non-randomness of dock locations, we use a propensity score matching approach to identify the causal effect of the presence of a dock. A comparison of estimates from matched samples to those from the full sample of observations suggests a selection bias in the placement of bikeshare docks.

Our empirical results indicate that the average treatment effect of the presence of bikeshare docks is a 2 to 3% reduction in traffic congestion. In addition to these results, we also find evidence of a potential spillover effect, in which docks increase congestion in neighboring locations, perhaps as they lead drivers to find alternative routes to avoid bicycle traffic. The magnitude of this impact is substantial relative to the effect in which docks offer transportation alternatives and reduce traffic congestion. Our results are robust to various caliper values in propensity score matching. Estimates are also robust to heteroskedasticity and autocorrelation when standard errors are clustered on time, though we lose much of our statistical significance when we cluster on location. This suggests the need for further work related to the spatial dimension of congestion.

Our model assumes a fairly simplistic relationship between local congestion and local transportation choices. Further research is necessary to fully capture the spatial interactions in our observed and unobserved variables. The identification of mechanisms and structural impacts requires a more involved model of consumer choice. We also stop short of estimating any complementary effect among transportation modes as, for example, bikeshare docks may extend the network of Metrorail travel. Similarly, we ignore the impact of dedicated bicycle lanes that have an obvious impact on congestion and are likely correlated with the location of bikeshare docks. Still, our model is able to identify causal treatment effects and provides evidence for the efficacy of bikeshare programs in reducing traffic congestion.

References

Michael L. Anderson. Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *The American Economic Review*, 104(9):2763–2796, 2014.

Justic Beaudoin, Y. Hossein Farzin, and C.-Y. Cynthia Lin. Evaluating public transit investment in congested cities. Working Paper, University of California, Davis, 2014.

Antonio M. Bento, Lawrence H. Goulder, Mark R. Jacobsen, and Roger H. von Haefen. Distributional and efficiency impacts of increased US gasoline taxes. *The American Economic Review*, 99(3):667–699, June 2009.

Antonio M. Bento, Jonathan E. Hughes, and Daniel Kaffine. Carpooling and driver responses to fuel price changes: Evidence from traffic flows in Los Angeles. *Journal of Urban Economics*, 77:41–56, September 2013.

Nicholas E. Burger and Daniel T. Kaffine. Gas prices, traffic, and freeway speeds in Los Angeles. *The Review of Economics and Statistics*, 91(3):652–657, August 2009.

Yihsu Chen and Alexander Whalley. Green infrastructure: The effects of urban rail on air quality. *American Economic Journal: Economic Policy*, 4 (1):58–97, 2012.

Graham Currie and Justin Phung. Transit ridership, auto gas prices, and world events: New drivers of change? *Transportation Research Record*, 1992(1): 3–10, January 2007.

Janet Currie and Reed Walker. Traffic congestion and infant health: Evidence from E-Z Pass. *American Economic Journal: Applied Economics*, 3(1):65–90, January 2011.

W. Bowman Cutter and Matthew Neidell. Voluntary information programs and environmental regulation: Evidence from "Spare the Air". *Journal of Environmental Economics and Management*, 58(3):253–265, November 2009.

Lucas W. Davis, Alan Fuchs, and Paul Gertler. Cash for coolers: Evaluating a large-scale appliance replacement program in Mexico. *American Economic Journal: Economic Policy*, 6(4):207–238, 2014.

Gilles Duranton and Matthew A. Turner. The fundamental law of road congestion: Evidence from US cities. *The American Economic Review*, 101(6): 2616–2652, October 2011.

Paul J. Ferraro and Juan Jose Miranda. Panel data designs and estimators as alternatives for randomized controlled trials in the evaluation of social programs. Working Paper, 2014.

Kyle Gebhart and Robert B. Noland. The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, (41):1205–1225, 2013.

Daniel E. Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15:199–236, 2007.

Matthew E. Kahn. Urban policy effects on carbon mitigation. NBER Working Paper 16131, 2010.

Rahul Nair, Elise Miller-Hooks, Robert C. Hampshire, and Ana Busic. Large-scale vehicle sharing systems: Analysis of Velib. *International Journal of Sustainable Transportation*, 7(1):85–106, January 2013.

Ian W. H. Parry, Margaret Walls, and Winston Harrington. Automobile externalities and policies. *Journal of Economic Literature*, 45:373–399, June 2007.

Paul Rosenbaum. *Observation Studies*. Springer Verlag, New York, 1995.

Thomas Rubin, James E. Moore II, and Shin Lee. Ten myths about US urban rail systems. *Transport Policy*, 6(1):57–73, 1999.

David Schrank, Bill Eisele, and Tim Lomas. Texas A&M Transportation Institute's 2012 urban mobility report. Technical report, Texas A&M University, 2012.

Steven E. Sexton. Paying for pollution? How general equilibrium effects undermine the "Spare the Air" program. *Environmental and Resource Economics*, 53(4):553–575, August 2012.

Elisheba Spiller, Heather Stephens, Christopher D. Timmins, and Allison Smith. Does the substitutability of public transit affect commuters' response to gasoline price changes? *Resources for the Future Discussion Paper*, 2012.

Sarah E. West and Roberton C. Williams III. The cost of reducing gasoline consumption. *The American Economic Review*, 95(2):294–299, May 2005.

Sarah E. West and Roberton C. Williams III. Optimal taxation and cross-price effects on labor supply: Estimates of the optimal gas tax. *Journal of Public Economics*, 91(3):593–617, April 2007.

Clifford Winston and Ashley Langer. The effect of government highway spending on road users' congestion costs. *Journal of Urban Economics*, 60:463–483, 2006.

Clifford Winston and Vikram Maheshri. On the social desirability of urban rail transit systems. *Journal of Urban Economics*, 62:362–382, 2007.

Tables

Table 1: Summary Statistics

	mean	std. dev.	10th percentile	90th percentile
Congestion	1.247	0.476	0.968	1.586
Congestion (by block group)	1.230	0.271	1.025	1.472
Dock	0.129	-	-	-
Dock_Adj	0.404	-	-	-
Dock_KAdj	0.216	-	-	-

Table 2: Frequency Distribution of Dock Counts

	0	1	2	3	4	5	6	7	8	9	10+
Docks (max = 10)	774,743	80,592	22,436	6,908	955	3,173	326	184	339	144	
Docks_Adj (max = 41)	530,630	104,167	70,383	60,740	32,688	19,650	12,236	10,609	8,873	7,110	32,714
Docks_KAdj (max = 10)	698,040	127,080	36,332	14,083	8,721	4,170	8	699	341	326	

Table 3: Presence of Bikeshare Docks

	1	2	3
Dock	0.00426***	-0.00208*	0.00198*
	(0.0011)	(0.0012)	(0.0012)
Dock_Adj		0.01313***	
		(0.0008)	
Dock_KAdj			0.00798***
			(0.0010)
Adjusted R^2	0.4681	0.4693	0.4683
Observations	889,800	889,800	889,800

numbers in parentheses indicate standard errors

*,**, and *** indicate statistical significance at the .1, .05, and .01 levels, respectively

Table 4: Count of Bikeshare Docks

	1	2	3
Dock	0.00193***	-0.00206***	0.00063
	(0.0006)	(0.0007)	(0.0006)
Dock_Count_Adj		0.00149***	
		(0.0001)	
Dock_Count_KAdj			0.00227***
			(0.0005)
Adjusted R^2	0.4678	0.4679	0.4678
Observations	889,800	889,800	889,800

numbers in parentheses indicate standard errors

*,**, and *** indicate statistical significance at the .1, .05, and .01 levels, respectively

Table 5: Variables for Propensity Score Matching

Variable	Description
Med_Inc	median household income
Price	median price of housing (rented and owned units)
Pop	block group population
Per_Educ	percentage of population +25 years with bachelor's degree or further education
Per_Own	percentage of housing units owned by tenant
House_Age	median age of housing units
Sq_Mi	square mile area
Dist_Metro	distance (in meters) to nearest Metrorail station
Cong_Mean_May	mean congestion in May 2010
Cong_SD_May	standard deviation of congestion in May 2010
Cong_Mean_Sep	mean congestion in September 2010
Cong_SD_Sep	standard deviation of congestion in September 2010
Cong_Mean_Oct	mean congestion in October 2010
Cong_SD_Oct	standard deviation of congestion in October 2010
Dist_Metro	distance from block group centroid to nearest Metrorail station (in meters)

Table 6: Covariate Balance: Post-2010 Sample, Unmatched

Variable	Treatment Mean	Control Mean	Variance Ratio	p-value	K-S p-value
Med_Inc	82,269.720	90,888.147	1.814	0.050	0.052
Price	3,500.425	4,410.788	1.542	0.001	0.001
Pop	1,751.110	1,432.763	0.692	0.008	0.000
Per_Educ	0.662	0.530	1.275	0.000	0.001
Per_Own	0.412	0.571	1.371	0.000	0.000
House_Age	48.683	53.952	0.465	0.017	0.006
Sq_Mi	0.166	0.201	2.528	0.109	0.141
Cong_Mean_May	1.266	1.265	1.436	0.474	0.540
Cong_SD_May	0.168	0.162	1.672	0.361	0.085
Cong_Mean_Sep	1.252	1.265	1.801	0.248	0.587
Cong_SD_Sep	0.144	0.164	2.769	0.074	0.090
Cong_Mean_Oct	1.201	1.269	2.167	0.000	0.012
Cong_SD_Oct	0.091	0.151	3.477	0.000	0.000
Dist_Metro	570.970	1,607.028	9.748	0.000	0.000

Table 7: Covariate Balance: Matching on Socioeconomic Variables and Pre-Treatment Congestion

Nearest Neighbor

Variable	Treatment Mean	Control Mean	Variance Ratio	p-value	K-S p-value
Med_Inc	82,269.720	85,317.179	1.279	0.358	0.973
Price	3,500.425	3,752.341	1.291	0.301	0.477
Pop	1,751.110	1,758.128	1.166	0.487	0.751
Per_Educ	0.662	0.626	1.173	0.238	0.416
Per_Own	0.412	0.445	1.182	0.257	0.521
House_Age	48.683	55.026	0.396	0.023	0.013
Sq_Mi	0.166	0.197	3.534	0.314	0.514
Cong_Mean_May	1.266	1.242	0.853	0.221	0.708
Cong_SD_May	0.168	0.167	1.721	0.493	0.392
Cong_Mean_Sep	1.252	1.233	1.032	0.261	0.948
Cong_SD_Sep	0.144	0.132	0.463	0.215	0.420
Cong_Mean_Oct	1.201	1.232	1.250	0.134	0.184
Cong_SD_Oct	0.091	0.115	0.470	0.042	0.007
Dist_Metro	570.970	697.313	1.323	0.050	0.335

Caliper (1 sd of prop. score)

Variable	Treatment Mean	Control Mean	Variance Ratio	p-value	K-S p-value
Med_Inc	86,479.639	84,198.735	1.195	0.392	0.532
Price	3,678.093	3,707.847	1.065	0.475	0.655
Pop	1,707.262	1,641.939	1.396	0.369	0.442
Per_Educ	0.663	0.604	1.062	0.121	0.133
Per_Own	0.437	0.465	1.190	0.294	0.756
House_Age	49.230	54.796	0.383	0.048	0.023
Sq_Mi	0.157	0.186	5.324	0.295	0.720
Cong_Mean_May	1.237	1.243	1.321	0.414	0.416
Cong_SD_May	0.153	0.157	1.764	0.440	0.515
Cong_Mean_Sep	1.234	1.234	1.430	0.500	0.559
Cong_SD_Sep	0.132	0.129	0.684	0.414	0.534
Cong_Mean_Oct	1.228	1.235	1.233	0.393	0.379
Cong_SD_Oct	0.112	0.113	0.462	0.468	0.548
Dist_Metro	629.664	836.133	1.617	0.007	0.015

Caliper (.05)

Variable	Treatment Mean	Control Mean	Variance Ratio	p-value	K-S p-value
Med_Inc	87,312.538	84,198.735	1.073	0.361	0.797
Price	3,836.419	3,707.847	0.957	0.401	0.820
Pop	1,560.769	1,641.939	1.655	0.339	0.878
Per_Educ	0.641	0.604	1.019	0.239	0.328
Per_Own	0.459	0.465	1.099	0.452	0.889
House_Age	51.904	54.796	0.441	0.194	0.090
Sq_Mi	0.168	0.186	4.704	0.375	0.504
Cong_Mean_May	1.239	1.243	1.159	0.456	0.524
Cong_SD_May	0.157	0.157	1.586	0.496	0.728
Cong_Mean_Sep	1.241	1.234	1.271	0.402	0.861
Cong_SD_Sep	0.136	0.129	0.607	0.326	0.525
Cong_Mean_Oct	1.238	1.235	1.100	0.471	0.774
Cong_SD_Oct	0.118	0.113	0.420	0.383	0.663
Dist_Metro	670.177	836.133	1.536	0.028	0.051

Table 8: Presence of Bikeshare Docks: Full Sample

	1	2	3
Dock	-0.0013	-0.00518***	-0.00260
	(0.0017)	(0.0017)	(0.0017)
Dock_Adj		0.01187***	
		(0.0012)	
Dock_KAdj			0.00566***
			(0.0014)
Adjusted R^2	0.4640	0.4641	0.4640
Observations	652,052	652,052	652,052

numbers in parentheses indicate standard errors

*,**, and *** indicate statistical significance at the .1, .05, and .01 levels, respectively

Table 9: Presence of Bikeshare Docks: Matched Samples

Nearest Neighbor			
	1	2	3
Dock	-0.02167***	-0.02509***	-0.01683***
	(0.0011)	(0.0012)	(0.0011)
Dock_Adj		0.00522***	
		(0.0012)	
Dock_KAdj			0.00173***
			(0.0014)
Adjusted R^2	0.5733	0.5733	0.5735
Observations	221,873	221,873	221,873
Caliper (1 sd of prop. score)			
	1	2	3
Dock	-0.02813***	-0.03464***	-0.02940***
	(0.0010)	(0.0011)	(0.0011)
Dock_Adj		0.01442***	
		(0.0012)	
Dock_KAdj			0.00614
			(0.0014)
Adjusted R^2	0.5040	0.5044	0.5040
Observations	192,078	192,078	192,078
Caliper (.05)			
	1	2	3
Dock	-0.01650***	-0.02608***	-0.01953***
	(0.0011)	(0.0012)	(0.0011)
Dock_Adj		0.03219***	
		(0.0012)	
Dock_KAdj			0.01911***
			(0.0014)
Adjusted R2	0.5265	0.5284	0.5270
Observations	167,875	167,875	167,875

numbers in parentheses indicate standard errors

*,**, and *** indicate statistical significance at the .1, .05, and .01 levels, respectively

Figures

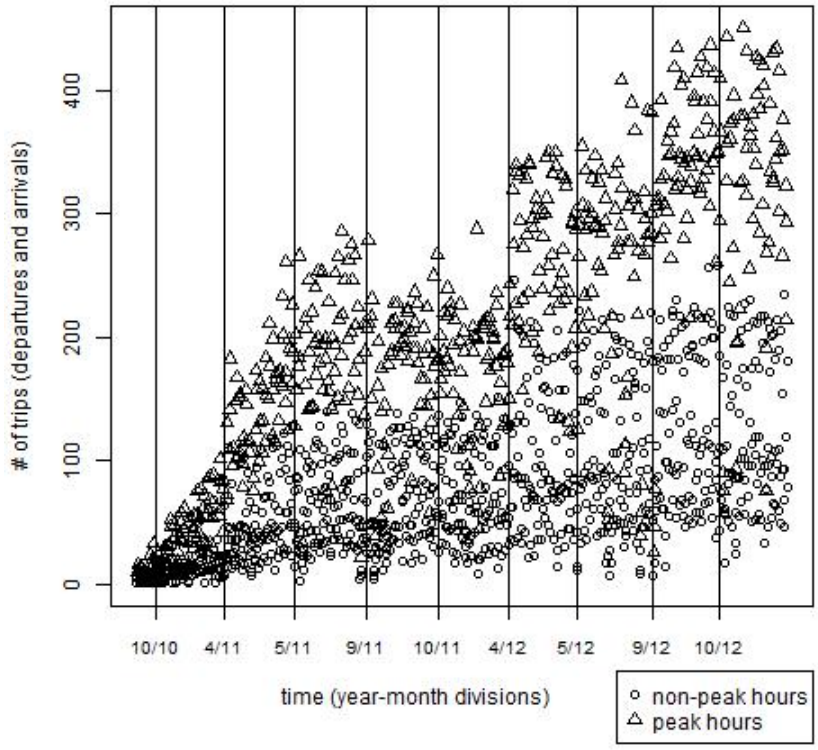

Figure 1: Bikeshare Trips (departures and arrivals)

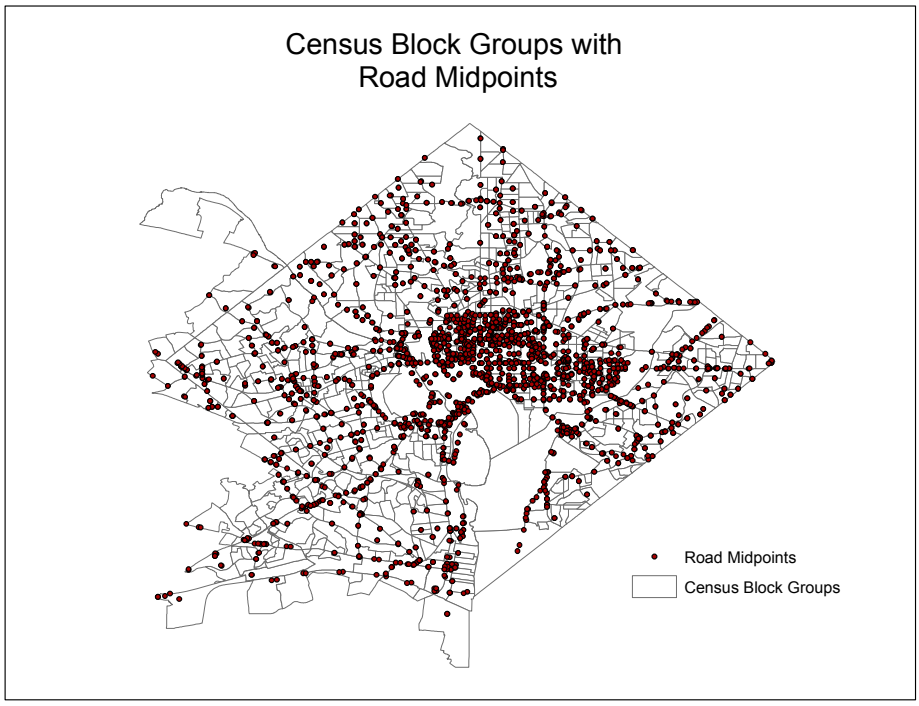

Figure 2: Census Block Groups with Road Midpoints

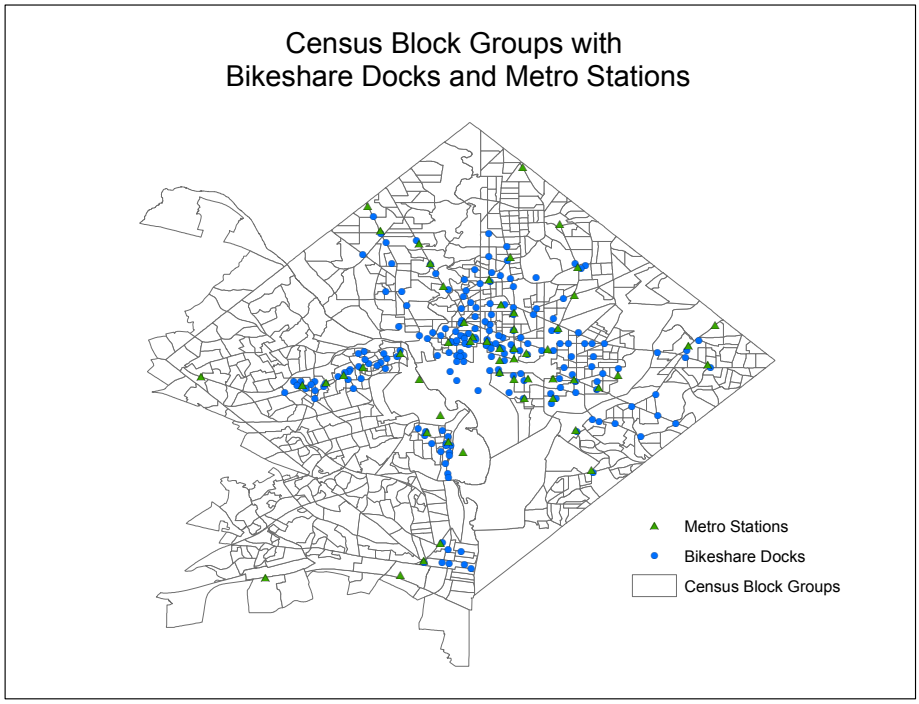

Figure 3: Census Block Groups with Bikeshare Docks and Metrorail Stations

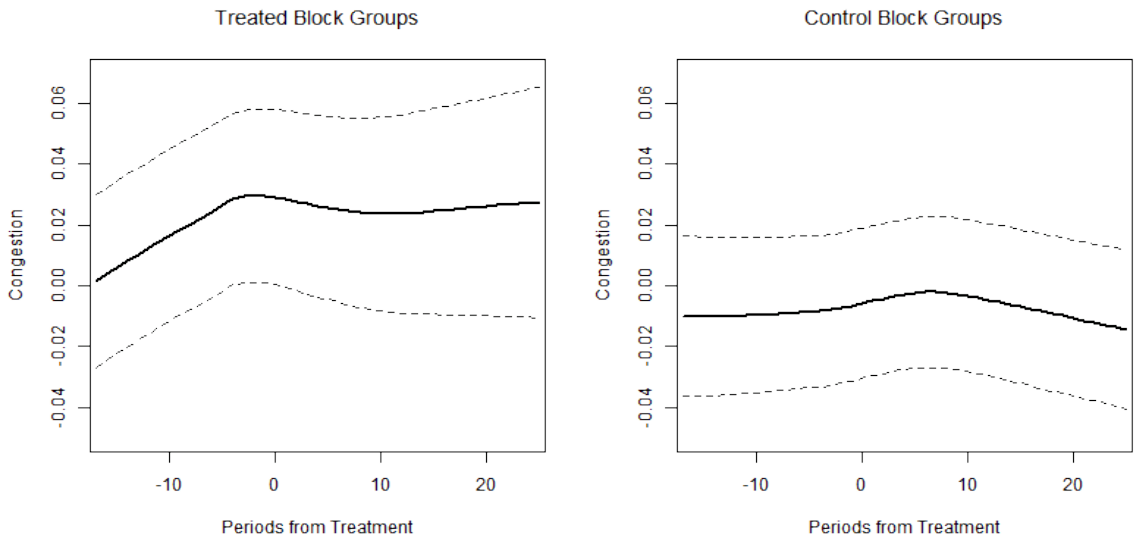

Figure 4: The Effect of Bike Dock Treatment on Block Group Congestion

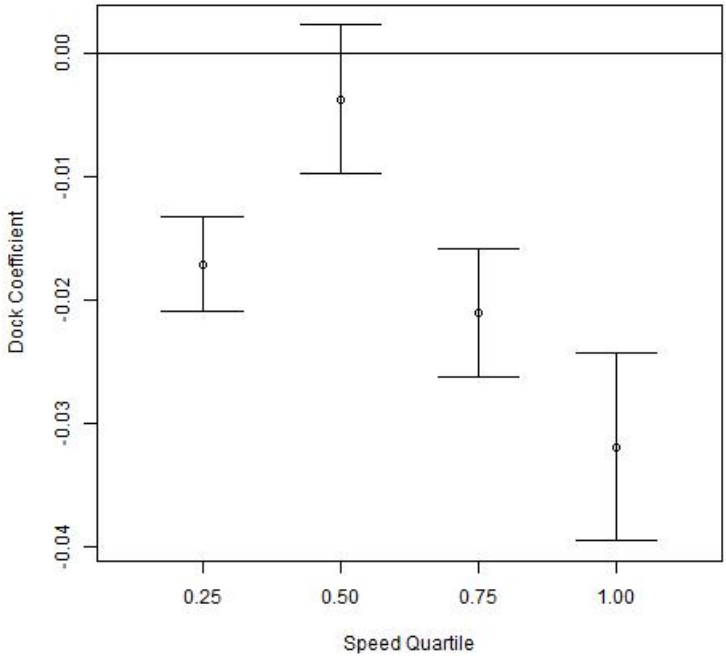

Figure 5: Dock Coefficients with Partitioned Subsamples: Control for Neighboring Block Groups

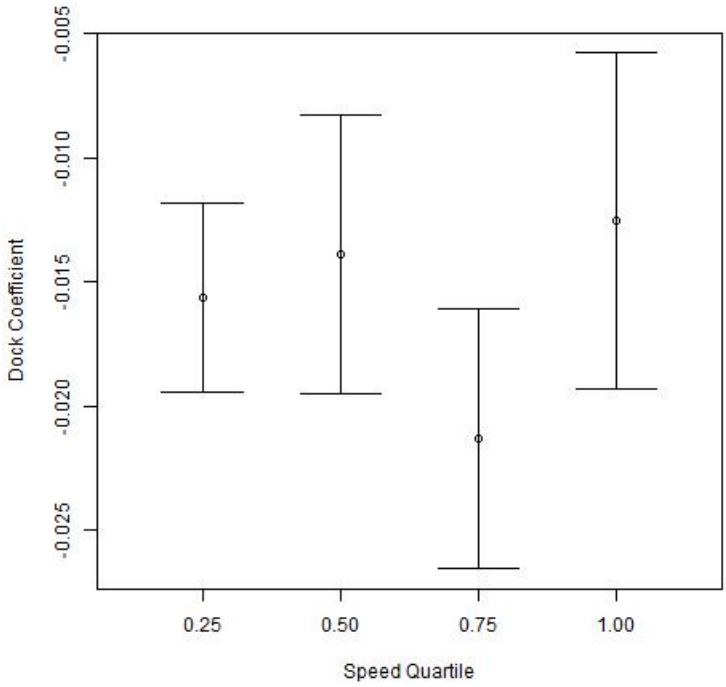

Figure 6: Dock Coefficients with Partitioned Subsamples: Control for Neighboring Blocks

Appendix

Table A1: Presence of Bikeshare Docks: Matched Samples, Clustered Standard Errors

Nearest Neighbor			
	1	2	3
Coef	-0.02167	-0.02509	-0.01683
SE_Time	(0.0035)***	(0.0036)***	(0.0038)***
SE_BG	(0.0288)	(0.0337)	(0.0263)
Caliper (1 sd of prop. score)			
	1	2	3
Coef	-0.02813	-0.03464	-0.0294
SE_Time	(0.0024)***	(0.0022)***	(0.0023)***
SE_BG	(0.0118)**	(0.0127)***	(0.0112)***
Caliper (.05)			
	1	2	3
Coef	-0.0165	-0.02608	-0.01953
SE_Time	(0.0020)***	(0.0017)***	(0.0019)***
SE_BG	(0.0114)	(0.0129)**	(0.0110)*

numbers in parentheses indicate standard errors

*,**, and *** indicate statistical significance at the .1, .05, and .01 levels, respectively

Table A2: Caliper Robustness: Specification 2

Caliper Size	Dock	SE	Dock_Adj	SE
0.04	-0.0254	(0.0012)	0.0286	(0.0013)
0.06	-0.0248	(0.0012)	0.0320	(0.0013)
0.08	-0.0240	(0.0012)	0.0346	(0.0013)
0.10	-0.0240	(0.0012)	0.0346	(0.0013)
0.12	-0.0273	(0.0012)	0.0259	(0.0012)
0.14	-0.0311	(0.0012)	0.0249	(0.0013)
0.16	-0.0311	(0.0012)	0.0249	(0.0013)
0.18	-0.0334	(0.0012)	0.0220	(0.0013)
0.20	-0.0334	(0.0012)	0.0220	(0.0013)
0.22	-0.0334	(0.0012)	0.0220	(0.0013)
0.24	-0.0346	(0.0011)	0.0144	(0.0012)
0.26	-0.0346	(0.0011)	0.0144	(0.0012)
0.28	-0.0346	(0.0011)	0.0144	(0.0012)
0.30	-0.0346	(0.0011)	0.0144	(0.0012)

Table A3: Caliper Robustness: Specification 3

Caliper Size	Dock	SE	Dock_Adj	SE
0.04	-0.0203	(0.0011)	0.0212	(0.0014)
0.06	-0.0187	(0.0011)	0.0195	(0.0015)
0.08	-0.0166	(0.0011)	0.0189	(0.0015)
0.10	-0.0166	(0.0011)	0.0189	(0.0015)
0.12	-0.0189	(0.0012)	0.0107	(0.0014)
0.14	-0.0242	(0.0011)	0.0120	(0.0014)
0.16	-0.0242	(0.0011)	0.0120	(0.0014)
0.18	-0.0266	(0.0011)	0.0087	(0.0014)
0.20	-0.0266	(0.0011)	0.0087	(0.0014)
0.22	-0.0266	(0.0011)	0.0087	(0.0014)
0.24	-0.0294	(0.0011)	0.0061	(0.0014)
0.26	-0.0294	(0.0011)	0.0061	(0.0014)
0.28	-0.0294	(0.0011)	0.0061	(0.0014)
0.30	-0.0294	(0.0011)	0.0061	(0.0014)