Clustered into Control: Heterogeneous Causal Impacts of Water Infrastructure Failure*

Brandon Cunningham[4], Jacob LaRiviere[1,4,5], and Casey J. Wichman[2,3]

[1]Microsoft
[2]Georgia Institute of Technology
[3]Resources for the Future
[4]University of Washington
[5]University of Tennessee

September 12, 2020

Abstract

We estimate economic impacts from decaying water infrastructure in the U.S. Using water main breaks in Washington, DC, and a yearlong panel of hourly traffic speeds, we estimate causal effects of water main failures on traffic congestion. We use k-means clustering to create clusters of streets that are similar to each other: treated observations are compared to other units in the cluster. We identify heterogeneous treatment effects algorithmically while retaining straightforward standard error calculations. We find strong evidence of heterogeneous treatment effects across clusters but small welfare impacts of water main breaks on traffic patterns overall.

Key Words: k-means clustering, program evaluation, water main breaks, traffic congestion, water infrastructure
JEL Codes: H41, H70, C10, L95, Q51, R42

*LaRiviere: jlariv@microsoft.com. Cunningham: brandon.cunningham.20@gmail.com. Wichman: wichman@gatech.edu. The authors thank Victoria Fleming at DC Water for her prompt help in facilitating onerous FOIA requests and Kay Kiratsuka for great data munging and visualization. Matt Harding, Marty Smith, participants at the 2016 Association of Environmental and Resource Economists annual conference, and 2017 Economics of Water and Energy Workshop provided helpful comments and suggestions. The authors are grateful for research support from the US Department of Homeland Security (DHS) and feedback from Angela Blair and Tony Turner. Findings expressed here do not represent official positions of the DHS.

1 Introduction

Water infrastructure is widely acknowledged to be in need of repair and reinvestment in the United States and many other Organization for Economic Co-operation and Development (OECD) countries.[1] Water infrastructure in particular gets a lot of media attention. The 2014 Flint, MI, water crisis is resulting in criminal charges, a main break on the UCLA campus in 2014 received national media attention, and the traffic impact of water main breaks are routinely reported by news outlets.[2]

Because water infrastructure is a public good, determining efficient reinvestment is a public policy issue sitting with city, county, state and federal governments. Trade groups estimate the cost of water instrastructure repair to be on the order of $1 trillion dollars over the next 25 years (AWWA, 2012). To develop efficient policy, its important for policy makers to understand the expected benefits and expected costs of each investment decision.

Unfortunately, there are no causal estimates for how water supply disruptions impact any measure of economic welfare for OECD countries to our knowledge.[3] As a result, even if OECD utilities were allocated funds to invest in water infrastructure, there is no empirical evidence for the economic benefits of different investment options. Although the American Water Works Association notes, "The need to rebuild... pipe networks must come on top of other water investment needs, such as the need to replace water treatment plants and storage tanks, and investments needed to comply with standards for drinking water quality" (AWWA, 2012), there is no clear causal evidence on the direct or indirect economic returns of such an

[1] The *Harvard Business Review* and *The Economist* routinely run stories on crumbling water (and road) infrastructures. See https://hbr.org/2015/05/what-it-will-take-to-fix-americas-crumbling-infrastructure and http://www.economist.com/news/united-states/21605932-country-where-everyone-drives-america-has-shoddy-roads-bridging-gap.

[2] See https://www.washingtonpost.com/news/dr-gridlock/wp/2016/06/23/water-main-break-in-alexandria-likely-to-cause-traffic-delays/.

[3] There is virtually no work in the economics literature that addresses the causal economic impacts of well-functioning water infrastructure in OECD countries. Most of the work addressing the causal impacts of water infrastructure focuses on less-developed countries (e.g., Galiani et al., 2005; Gamper-Rabindran et al., 2010; Devoto et al., 2012; Bel et al., 2010). Some closer research focuses on estimating a dose-response function of water pollution on infant health (Currie et al., 2013) and bottled water purchases in response to water quality violations (Graff Zivin et al., 2011). The majority of the work on water infrastructure in OECD countries, however, uses computable general equilibrium (CGE) models with parameters taken from the literature (Rose and Liao, 2005). Although a valuable modeling technique, because the parameter values used are often not causal it is unclear how much policymakers should prioritize water infrastructure improvements based upon CGE output.

investment.

This paper estimates the causal effect of water infrastructure failure on an important economic outcome—traffic—in a dense, urban OECD city. Water main breaks occur relatively frequently in an unpredictable fashion and are an ideal example of water infrastructure failure of pipe networks. When a water main breaks, it is typically repaired by the local utility or another vendor immediately. The repair often shuts down streets and impacts traffic because construction crews have to cut through cement and asphalt to repair the broken water main. We estimate the causal effect of water main breaks on traffic speeds. While there is a healthy literature examining the impacts of various market events and regulations on traffic and driving behavior (e.g., Burger and Kaffine, 2009; Anderson, 2014; Bento et al., 2014; Wolff, 2014a,b), there has been no work on the traffic impacts of water main breaks.

We study the universe of all water main breaks over a 12-month period in Washington, DC, from July 2014 through June 2015. Our data on water main breaks include location of the break, the severity of the break, and the time at which a break is reported and when repairs are completed. We merge in high-frequency and spatially detailed traffic speed data for 2,182 urban road segments in DC. We use a generalized difference-in-difference (DD) research design by comparing observed traffic speeds on "treated" road segments near a break to "comparison" road segments further away from the break, in addition to "spillover" road segments in the middle that may violate our stable-unit treatment value assumption. The DD research design is important because we find breaks are more likely to occur during lower traffic speed days (e.g., when it is colder and mains are more likely to break).

Because there are substantial differences in road types in our data set, it is unclear which road segments serve as good counterfactuals for a treated road segment when there is a water main break. In our data, individual road segments have unique IDs, traffic direction and traffic speeds. This can create problems in selecting comparison road segments for segments that are "treated" with a water main break. For example, we have interstate road segments, within-city arterial roads, and non-arterial roads in our data. The location and direction of each road segment determines which road segments are susceptible to morning commute versus afternoon commute traffic. With non-descriptive identifiers we run the risk of having dissimilar road segments serve as control for treated segments.[4]

There are some methods available to applied econometricians and statisticians to use a data-driven approach to pick appropriate counterfactuals. Researchers sometimes hand-select comparison groups to include "like" observations (Bento et al., 2014) or samples are trimmed so that units have similar support (Ferraro and Miranda, 2017). Hand trimming, though, is *ad hoc* and not algorithmic; it requires the researcher to decide what appropriate variable cutoffs are. Synthetic control techniques are increasingly used to estimate unit-specific treatment effects using panel data algorithmically (Abadie et al., 2010; Quistorff, 2017; Xu, 2017). While appealing, synthetic controls have both complicated standard error calculations and are computationally intensive, especially as the number of units increases.

In this paper we leverage k-means clustering as a simple technique to select comparison groups algorithmically as in Bonhomme et al. (2017) and Aliprantis et al. (2017). k-means clustering involves pre-processing panel data using an unsupervised machine-learning algorithm to classify similar units based upon levels and summary statistics of observed variables in the dataset. While k-means clustering has been used in the economics literature previously for classification (e.g., Crone, 2005; Caballero, 2016; Castledine et al., 2014) it has not been used to create matched clusters in a program evaluation context before our work and Bonhomme et al. (2017) and Aliprantis et al. (2017). Observations within the same cluster are most similar to one another on observable margins, and are thus likely balanced on unobservables in expectation. Pre-processing the data in this way permits estimation of cluster-specific treatment effects, particularly for a treatment that varies over space and time as in our setting.

One desirable feature of k-means clustering over other techniques is that it can be easily

[4] This type of data is sometimes called "unstructured panel data" in computer science. Examples of unstructured large panel datasets include website browsing data, product use data, and anonymized healthcare use data where the types of outcomes common in economics datasets isn't present. With web browsing data a company uses cookies to identify specific users and then logs the universe of their click behavior on their website. For product use data, internet connected devices like thermostats monitor electricity consumption of households in near real time. Anonymized healthcare analytics track how individuals on different healthcare programs use healthcare services. One challenge of conducting researching on these types of datasets is that when a particular subject is treated, it is unclear what the proper counterfactual ought to be. Put another way, an infrequent healthcare service user is probably a poor control for an intense health care user, in particular if treatment could have heterogeneous treatment effects.

applied within quasi-experimental research designs like difference-in-differences estimators as we do here, in addition to experimental variation. Thus, our use of k-means clustering adopts an algorithmic approach to traditional econometric pre-processing that allows for estimating cluster-specific treatment effects. Our method is distinct from alternative algorithmic approaches in Athey and Imbens (2016) and Wager and Athey (2017). Those approaches leverage supervised ML algorithms to hunt for heterogeneous treatment effects along well-defined covariates in the data. By contrast, our approach creates categorical variables called "clusters" for which we explicitly estimate unique treatment effects.

We pre-process our data set using k-means clustering to classify similar road types into clusters based upon observed traffic speed levels, variance, changes, and directions of traffic on each segment for different hours of the day. We present evidence that the clustering algorithm removes interstate and main thoroughfares as comparison units for small surface streets. We describe how roads within the same cluster are similar to one another and therefore provide a more plausible counterfactual outcome. Our identifying assumption is that unobservable features of segments within a cluster are assumed to be uncorrelated with the timing and location of water main breaks, so that the error term is uncorrelated with the treatment variable within clusters. We verify the difference-in-differences pre-trends requirement holds within a cluster. We also perform a placebo test to provide evidence that conditional on a break occurring it occurs exogenously within a cluster.

Clustering is important for internal validity in our application. Using a DD design without clusters, we find that water main breaks are associated with a statistically significant decrease in traffic speeds (1.4%) in road segment clusters where they occurred. By measuring treatment effects at the cluster level, however, we find that average traffic speed impacts range between not statistically different from zero to a significant 5% decrease for different clusters. Hence, there are meaningful statistical differences across road segment clusters. We also find evidence of spillovers: traffic speed impacts decrease as distance from the break increases, radiating a half-mile from the location of a water main break. These results are robust to a variety of alternative specifications, including: changing the number of clusters, temporal and spatial controls, serial correlation of the standard errors, and falsification tests. Lastly, we find evidence that using road segments within the same cluster but far from main breaks as a control increases statistical power relative to including the fully pooled set of road segments as a control.

There is a clear temporal pattern in the traffic impacts of water main breaks across clusters as well: impacts range from over 5% decreases at morning rush hour down to statistically insignificant effects during off-peak hours within impacted segments. We take this as evidence, consistent with Anderson (2014), that accounting for temporal heterogeneity is an important part of traffic studies. Further, we find that the cross-cluster heterogeneity (e.g., spatial heterogeneity) is just as important as the intertemporal heterogeneity for water main break traffic delays. To our knowledge, this is the first algorithmic evidence in the economics literature for heterogeneity in traffic delays over space as being the same order of magnitude as heterogeneity over time-of-day for a given type of traffic disruption (e.g., water main breaks).

Finally, while statistically significant and robust across specifications, the aggregate magnitude of these effects is economically small, even for the road segments and times of day where it matters most. For the average break in our sample, a central estimate of the private congestion costs is approximately $1,350 per break. Total costs to drivers over our 12-month sample were $695,275, or approximately $1 per resident of Washington, DC. To our knowledge, this is the first causal estimate of water infrastructure supply disruptions on any economic outcome in an OECD country. Despite widespread media attention to water main breaks, our results imply that economic losses from traffic congestion due to breaks are not a reasonable justification for large-scale infrastructure repair during, for example, low traffic periods at night. Of course, there are other important attributes to consider in a full cost-benefit analysis, including indirect economic costs due to public, commercial, and residential buildings being without water; lost revenue from leaked water; health risks due to water quality degradation; and direct repair costs. To that end, our paper is a starting point rather than a decision point for policymakers considering optimal water infrastructure investment policy.

2 Background

A water main is a pipe that supplies water to residential, commercial, and industrial buildings in a water supply system. When cities are constructed, water mains are often placed under city streets with smaller pipes leading into individual buildings. The water in a water main is pressurized to ensure access for utility customers.

Water main breaks occur due to the combination of pressurized water and pipe failure. Failure is related to pipe age, but also to sharp changes in temperature that cause the material making up the pipe to expand and contract. When a break occurs, "downstream" users may lose water and there is sometimes an "urban geyser" where the pressurized water breaks through the ground much like an opened fire hydrant. Much talk about crumbling infrastructure occurs due to increased likelihood of failure. Additional recent concerns also deal with securing infrastructure from human threats. In both cases, the infrastructure's age plays a critical role (AWWA, 2012).

In our study area the distributed water infrastructure is somewhat old. Table 1 shows the composition of mains by material and a coarse measure of main age for all water mains that had a break in the data we were provided by DC Water, the water utility for Washington, DC, through a Freedom of Information Act (FOIA) request. Almost all DC water mains are cast iron although 3% of breaks occurred in pipes of "unknown" material.[5]

More surprising is the age of mains that broke in our sample. We observe 515 breaks between July 1, 2014, and June 30, 2015; however, we focus on 278 breaks that occurred near a road in our data set. Of these, roughly 46% of breaks occurred in water mains that were over 100 years old, and the oldest break was from a main installed before the Civil War, in 1859. Unfortunately, we were not able to obtain the age distribution of the entire water main system with our FOIA request due to security concerns, so we cannot compare the age of broken mains relative to the entire water supply system. DC Water reports on its website, however, that the median age of all water pipes is 79 years, which is similar to the median

[5]Staff at DC Water noted that there is some incompleteness in the materials records. Rather than the material actually being unknown, these are likely instances of incomplete recording.

age in our sample (90 years).[6] In the full set of 515 breaks, the median age is 81 years—only two years older than the population median.

While local utilities are responsible for upkeep of their distributed water infrastructure, such as water mains and sewage lines, they are also responsible for maintenance and expansion of centralized water infrastructure. Centralized water infrastructure takes the form of water intake pipes from water sources, water treatment facilities, and pump houses. In allocating public money for an optimal portfolio of infrastructure improvements, it is unclear how to allocate funds across centralized and decentralized projects. There is a separate question of the impact of disruptions on centralized water infrastructure. In this paper, we do not address disruptions to centralized water infrastructure or any other type of infrastructure (e.g., transportation or electricity) that local and regional governments must address.

In the interests of tractability and precision, we focus on a single outcome that is affected by distributed water infrastructure: the effect of water main breaks on traffic. For ba water main break there are other important outcomes that should be addressed in any complete cost-benefit analysis. For example, in many cities commercial buildings must be closed if they do not have access to drinking water. However, knowing precisely which buildings were impacted by a water main shutdown requires more detailed information than we have. Our research design and results, though, could be extended to this important economic impact in future work. As a result, we focus on estimating an accurate effect of water main breaks on traffic speeds as a first step in informing the larger policy question of optimal water infrastructure investment.

3 Empirical methodology

To estimate the causal impacts of water main breaks on traffic congestion we combine unique data sets covering the Washington, DC, area. We then use a machine learning algorithm to cluster road segments into groups that are observationally similar. Finally, we use a flexible difference-in-difference design to test whether traffic is affected by main breaks and whether

[6] https://www.dcwater.com/about/rates/default.cfm.

this effect diffuses over space. This section summarizes each of the steps in detail.

3.1 Data

We purchased traffic data from INRIX, a company that aggregates high frequency and fine granularity traffic speed data, for Washington, DC, covering July 1, 2014–June 30, 2015.[7] INRIX collects data by partnering with commercial and government agencies to place GPS based speed sensors in cars then aggregates readings using their own proprietary algorithm to provide segment by minute speeds. We have speed data in miles per hour (MPH) at one-minute intervals on each day in our study period, for 2,182 individual road segments in Washington, DC. Similar data are commonly used in the economics literature for a wide variety of traffic topics (Burger and Kaffine, 2009; Anderson, 2014; Bento et al., 2014; Wolff, 2014a,b; Hamilton and Wichman, 2018). Included in the set of road segment characteristics are the latitude and longitude points to identify road segment location, the direction of traffic flow, and the reference speed for the road. A road segment is typically around 0.25 miles in length and ranges from a small city street to an interstate highway; these segments, geographically indexed by their midpoint, serve as the unit of observation in our application. For tractability in our analysis, we use hourly averages of speed for each road segment and we drop observations on weekends and those outside of the 5AM–11PM time frame. As such, we have 8,956,589 individual hour-by-road-segment observations. Unlike Bento et al. (2013) and Anderson (2014), for example, who use traffic flow and delay data from the California Freeway Performance Measurement System (PeMS), we require data that is more finely disaggregated on a spatial scale to identify the impact of water main breaks within urban areas. The primary limitation of these data, however, is that the sole time-varying metric we have on traffic patterns is speed, which does not capture important characteristics such as the number of vehicles on the road.

DC Water provided us with a list of water main breaks in response to a FOIA request, including the intersections or addresses of the breaks that occurred during the time period of our traffic data. These data include the date of reporting the water main break and the time

[7] See `http://inrix.com/`.

of completion of work. We geo-referenced the locations (i.e., the street intersection or street address) of main breaks using Google Maps' API.[8]

We merge the two data sets—INRIX and DC Water—using latitude and longitude coordinates. The merged data are shown in Figure 1 with points representing water main breaks and lines representing streets with observed speed data. Because the geographic locations of water main breaks do not overlap perfectly with road segment midpoints, we assign a water main break to each road segment within a fixed distance from the break. We then let this distance vary by econometric specification as discussed below.

The main limitation of our water main break data is that there is no information on when work actually began for each water main break. We observe when DC Water reports a repair completed and we observe when a problem is reported. If, however, there is a lapse in work during which there is no construction, that lapse will count as a "treated" period even though traffic could be flowing normally, leading to a lower bound for our estimated average treatment effects. We solve this errors-in-variables problem in several ways. First, we interpolate repair times for breaks that are implausibly long by replacing their repair times with the median repair times of breaks denoted as "most severe." This approach is motivated by severe breaks being prioritized so that their repairs garner the most immediate use of resources. Additionally, we include specifications that define a repair time as the lesser of (a) the difference between the time of a reported break and its completion and (b) one week from reported completion to provide a lower bound for traffic speed impacts.

3.2 k-means clustering

Our INRIX traffic data contain speeds for both surface streets and highways in DC. In our sample, there are several types of surface streets, including arteries and smaller residential streets that have commuter traffic and those without, and so forth. With 2,182 individual road segments, we adopt a method for classifying observationally similar streets together to provide the best possible counterfactual outcome for a road segment that is affected by a

[8] In Appendix Table A.1, we list water main break summary statistics. There are 278 total breaks in the data. Recall, though, that one break often impacts multiple road segments so there are 515 times a water main break impacts a road segment our data.

water main break.

In order to construct a measure of observationally similar streets from a time series of speed data for each road segment, there are two tasks. The first is to use the time series data to summarize the important characteristics of traffic patterns. The second is to use a method of classification based upon these summary statistics. We create a set of 52 summary statistics to characterize streets using the year's worth of data. These include mean speed by hour, standard deviation of speed by hour, difference between maximum observed hourly mean and mean speeds during commuting hours (to measure congestion), and categorical variables for traffic direction.[9]

Classifying road segments is a unique challenge for this paper. Our approach is similar in spirit to using propensity-score matching to construct a comparison group from observable characteristics (Rosenbaum and Rubin, 1985). Economists traditionally approach classification in this context by matching treatment and control units on the probability of being treated (Rosenbaum and Rubin, 1983) or based on similarities in covariates (Ferraro and Miranda, 2017; Wichman and Ferraro, 2017). Our situation is fundamentally different because we do not have a constant treatment and control group throughout the study. That is, we need to construct a cluster of road segments that will "turn on" as controls when a road segment in that cluster is affected by a main break, and turn off when traffic is flowing normally within the cluster. Otherwise we could be comparing interstate speeds to surface street speeds, because both are present in our data. As a result, we require a tool to classify roads with no a priori information about the correct groups. Similar challenges exist for in other settings, like stores classifying customer types to construct optimal price discrimination menus (e.g., what features should they stratify customers along, how to weight those features and where to draw boundaries between groups based upon the features).

[9]Specifically, from our hour-by-segment level traffic speed data, we drop all observations that occurred before 5AM, after 10PM, or on Saturday or Sunday. We then aggregate the data to a segment level and generate variables giving the mean and standard deviation of speed over the entire year of data, with one variable for each hour of the day (i.e., annual mean and standard deviation of speed for the hour beginning at 5AM, 6AM, ..., 10PM). We also construct the difference in means for several peak hours relative to a baseline hour with minimal traffic (5AM–6AM). Lastly, using the road segment characteristics provided to us by INRIX, we create dummy variables for cardinal directions (NB, SB, EB, WB, clockwise, and counterclockwise) and highways (one variable indicating whether a road is an interstate, another for US routes). There are 52 total variables in the clustering matrix.

Fortunately there is a set of tools used in machine learning for exactly this problem: unsupervised learning algorithms.[10] We use a simple unsupervised learning algorithm—k-means clustering—which is a statistical method used to group a set of objects based on characteristic variables. This approach classifies N objects in an I-dimensional space into K clusters, choosing to minimize the Euclidean distance between an object's vector and a cluster center (the mean of all vectors in the group) (MacQueen, 1967). K, the number of clusters, and I, the set of clustering variables, are chosen by the researcher.

k-means clustering minimizes the within-cluster sum of squares, using the Euclidean distance within a cluster weighting each of the I dimensions equally,

$$\sum_{k \in K} \sum_{i \in I} ||x_i - \bar{x}_i^k||^2, \tag{1}$$

where x_i is a vector of the ith variable and $\bar{x}_i^k$ is the mean of the ith variable in cluster k. As with all machine learning classification algorithms, the precise form of the algorithm defines what k-means clustering is. The algorithm begins by assigning K group centers to random points.[11,12] Then, it iterates as follows:

1. Assignment step: Each data point is assigned to the nearest group center.

2. Update step: Group centers are adjusted to match the sample means (i.e., centroid) of the data points.

[10] Unsupervised learning is a term used in data science to put structure on data when there is no left-hand-side variable of interest.

[11] k-means clustering has several limitations. One is that the random assignment of starting points can lead to very different clusters based on where the initial placement is (i.e., multiple local maxima). One solution is to repeat the process many times and pick the result with the smallest squared error or, in the case of several with the same squared error, use some sort of average. Bernhardt and Robinson (2007) use multiple iterations and note the importance of doing this for clustering a large number of objects together. Another limitation is that k-means clustering does not consider the shape and distribution of the data. As a result, it is up to the researcher to provide the appropriate summary statistics to use for classification. A third limitation is the "hard" design of k-means clustering. Points are assigned to exactly one cluster, including border points that influence (and are influenced by) points in nearby clusters. This limitation spawned a second type of k-means algorithm known as "soft" or "fuzzy" clustering. This returns a membership degree for each cluster-object pair (Rezankova, 2014). While these aspects of clustering are largely beyond the scope of our application, our results are remarkably robust to various sensitivity tests in clustering.

[12] As a sensitivity test, we also apply k-median clustering to our data. k-median clustering is similar to k-means, but uses the 1-norm distance instead of Euclidean distance to assign objects to clusters (Anderson et al., 2006). Primary results for this approach are included in the Appendix Table A.6

3. Repeat (1) and (2) until the assignments do not change.

The k-means algorithm will continue to run until each observation is located in a cluster with other observations that have similar elements to the clustering variables I. Because simple Euclidean distance will overweight variables with larger nominal values, we standardize our clustering variables to weight each variable equally. We adopt the method recommended by Milligan and Cooper (1988), which is to create $\hat{x}_j = x_j/(\max(x_j) - \min(x_j))$ where x_j is the jth variable.

Figure 2 shows the results of the k-means clustering procedure with $K = 10$. We choose 10 road clusters to allow for two (incoming and outflowing traffic) interstate roads, main surface streets, small surface streets, peripheral streets, and "other." In Figure 2, each road segment included in a panel is part of the cluster in that panel. The algorithm does well at matching similar road segments from visual inspection. $K = 10$ is our preferred number of clusters, but results are robust to other number of clusters. Primary results for $K = 8, 15$ are presented in Appendix Tables A.7 and A.8.

Table 2 shows summary statistics of clusters selected by the k-means algorithm. There are four clusters (5, 7, 8, and 10) containing many road segments and six smaller clusters. The larger clusters have lower average traffic speeds, suggesting that we have more observations on roads with more traffic. Additionally, the k-means clustering effectively groups streets by direction of traffic and along surface-highway delineations. Similarities in the variables within a row and differences across rows imply that the algorithm did an adequate job of clustering.

While we are one of the first papers to use k-means clustering to estimate heterogeneous treatment effects, there is a growing need for tools like these in other contexts due to increased availability of unstructured large panel datasets. Four examples are website browsing data, product use data, anonymized healthcare use data, and traffic data. With web browsing data, a company uses cookies to identify specific users and then logs the universe of their click behavior on their website. For product use data, internet-connected devices like thermostats monitor electricity consumption of households in near real time. Anonymized healthcare analytics track how individuals on different healthcare programs use healthcare services. Traffic data records traffic speeds and flows are different locations over time. Thus we expect this technique and others like it to become more prevalent moving forward as new research develops k-means clustering methodology for estimating heterogeneous treatment effects (Bonhomme et al., 2017; Aliprantis et al., 2017).

3.3 Treatment effects over space and time

Now that we have grouped road segments into similar clusters, we can use this classification to inform our identification strategy. Let Y_{it} be the outcome variable of interest—traffic speed on road segment i at time t. The unit of observation for speed is the road-segment level, as defined by our INRIX data. Road segments are assigned to a cluster j based on our k-means algorithm.

We are interested in identifying the effect of a series of exogenous water main breaks on nearby traffic patterns. Since water main breaks vary over space and time throughout our sample, we assign treatment status, $T_{it} \in \{0, 1\}$, to any road segment within $\omega_1 = 0.15$ mile of a water main break during the time period when our data indicate the presence of a water main break. Although this distance choice is admittedly arbitrary, we choose 0.15 mile as a distance that will capture the immediate effect of a water main break and also provide sufficient power to identify effects on congestion. Results are robust to varying this threshold, though estimated effects decrease in absolute value monotonically as this bandwidth increases, consistent with attenuation bias. We also define a cluster indicator, $C_i = j$, for road segments that are beyond $\omega_1 + \omega_2 = 0.5$ mile from the water main break, but are within the same jth cluster as any treated road segment. After grouping segments by k-means clustering, we contend that the control road segments are observationally similar to the treated road segments, conditional on segment (α_i) and time (τ_t) fixed effects. We can write this formally in the potential outcomes framework,

$$E[Y_{it}^0 | \alpha_i, \tau_t, C_i = j, T_{it} = 1] = E[Y_{it}^0 | \alpha_i, \tau_t, C_i = j, T_{it} = 0], \tag{2}$$

where Y_{it}^0 is the potential outcome in the absence of treatment. The previous equation asserts that the potential outcomes for observations *in the same cluster as a treated segment* ($C_i = j$) provide a proper counterfactual for the unobserved term, $E[Y_{it}^0|\alpha_i, \tau_t, C_i = j, T_{it} = 1]$.

Using the clustered control group in a generalized difference-in-difference framework, we can then estimate the average treatment effect on the treated (ATT), defined as

$$\text{ATT} = E[Y_{it}^1 - Y_{it}^0|\alpha_i, \tau_t, T_{it} = 1]. \tag{3}$$

Because water main breaks are conditionally exogenous to our outcome variable, and thus our cluster indicator, we contend that the marginal effect of a water main break on affected road segments, relative to prevailing traffic patterns in the same cluster, is causal. Equation 3 identifies a global treatment effect, although heterogeneous treatment effects are facilitated simply by conditioning on the cluster. Thus, the cluster-j ATT is simply

$$\text{ATT}_j = E[Y_{it}^1 - Y_{it}^0|\alpha_i, \tau_t, C_i = j, T_{it} = 1]. \tag{4}$$

There are a few important assumptions required for our heterogeneous treatment effects to be valid and the treatment effects themselves to be causal. First, for our heterogeneous treatment effects to be valid we use only features characterizing the distribution of road segment traffic speeds during hours when there is no reported water main break to perform clustering. This type of "sample splitting" implies that there should be no bias in the treatment effects due to the clustering algorithm itself.[13] Second, the k-means clustering serves as a form of sample trimming. The trimming, however, is performed in a transparent algorithmic way so that like units specified by the algorithm are compared to other like units. All units not contained within a given cluster are excluded fully as in other sample trimming techniques. Here, the features used for clustering are generated based on the characteristics of relatively unstructured traffic data rather than well-defined sociodemographic data used traditionally by economists.

[13] A more general sample splitting approach would use only half the data to perform the clustering, then the other half of the data to estimate the treatment effects such as the Causal Forest technique and others common in the ML causal inference literature. That technique is a very straightforward extension of this approach.

For causality we require conditional exogeneity, as stated above. Conditional exogeneity is likely to hold in our context: deviations from average traffic speeds is unlikely to cause water main breaks. Water main breaks are a function of water main ages, temperatures, and random (from the econometrician's perspective) mechanical failures. We also note that the stable unit treatment value assumption (SUTVA) plays an important role in our analysis. Given the natural spatial correlation of traffic patterns in a dense, urban road network, it is likely that the effect of a water main break at a given point may spill over into nearby road segments and contaminate the control group. As a result, we engage in a trimming procedure discussed at length below to remove contaminated control road segments in addition to explicitly modeling spillovers using a pragmatic semi-parametric technique and thus preserve the SUTVA assumption even with an interconnection road network.

In Figure 3, we present a simplified diagram of our treatment assignment to highlight the spatial dimension of our analysis. If a water main break occurs at the point in the center of the diagram, we treat all road segments in the circle A (within ω_1 miles from the water main break) as treated. As shown, the markers # and + in A represent treated road segments (i.e., $C = j, T = 1$). All other # and + segments in B and C represent potential comparison road segments that are in the same cluster as the treated segment (i.e., $C = j, T = 0$). In the example shown, the marker $\star$ is not treated in A and hence none of its cluster-segments are considered treated (i.e., $C = -j, T = 0$).

Using Figure 3 as a reference point, we conduct three complementary econometric analyses to explore the potential bias arising from treatment spillovers. Specifically, SUTVA is violated if treatment in A affects the outcome in B. If the correlation between treatment in A and outcomes in B is positive, as is likely when considering traffic patterns, then the causal effect of the water main break is likely biased downward. To combat this, we explore this potential bias directly. First, we estimate a naive model using segments lying in A as treated ($C = j, T = 1$), while clustered segments in B and C serve as "controls" ($C = j, T = 0$). Second, we estimate a model using road segments in A as treated, and road segments in C (i.e., greater than $\omega_1 + \omega_2$ miles from the water main break) as controls. The treated clusters that lie in B are excluded from the set of controls ($C = -j, T = 0$). Last, we estimate the spillover effect directly by defining an indicator that corresponds to treated segments that lie in A and another that corresponds to spillover segments that lie in B, and all segments in C corresponding to a treated cluster are controls.[14]

4 Empirical results and discussion

This section reports our main results. We begin with an econometric model that does not leverage clustering but otherwise shares the same research design. As a result, we first estimate the average treatment effect of water main breaks on traffic speeds across the entire sample. We then estimate the same model but allowing for heterogeneous treatment effects across the algorithmically determined clusters. We proceed in this way to be as explicit and transparent as possible in how the informational gains from algorithmic clustering can be used to identify heterogeneous treatment effects. In the Appendix we include a short section which shows evidence that the DD pre-trends assumption holds within clusters and also discuss it in more detail below.

4.1 Diff-in-Diff with uniform ATT

We use a research design which allows for water main breaks to have a direct impact on road segments within 0.15 miles of the main break. Further, we allow a "spillover" impact on segments between 0.15 and 0.5 miles from the break. Finally, we include an indicator variable taking the value of one for all hours in which any main break is active. In this specification, as in subsequent ones, the dependent variable is log of traffic speeds for a given road segment:

$$\ln(speed_{it}) = \alpha_i + \lambda_t + \beta \cdot 1\{Any\ break_t\} + \gamma \cdot 1\{Break_{it}\} + \gamma_S \cdot 1\{Spillover_{it}\} + \varepsilon_{it} \quad (5)$$

[14]Table A.2 in the Appendix shows that there is some variation in the number of impacted segments per break per impacted cluster for a 0.15 mile impact radius. Over 278 breaks there are a total of 12,355 unique impacted segments within a 0.15 mile radius. The average number of segments impacted per break range from just over one up to over four. The variability is correlated with the number of segments in a cluster: clusters with more segments understandably tend to have more impacted segments per break.

where α_i is a road-segment fixed effect and λ_t is a time fixed effect. In this specification we define $1\{Any\ break_t\} = 1$ for each time period within the 12 hours before any break in the DC Water database is repaired, and zero otherwise. We choose 12 hours because it is the median repair time for the common type of break in our sample.[15] As discussed above and shown visually in Figure 3, we define a treated segment, $Break_{it}$, as any segment within 0.15 mile of the address of a reported break. Thus, for any road segment i within 0.15 miles of a main break during a period i, $1\{Break_{it}\}$ takes the value of 1. The variable $1\{Spillover_{it}\}$ is defined similarly for road segments between 0.15 and 0.5 miles of an active break. The coefficient γ is thus the coefficient of interest in this specification.

Equation (5) and all subsequent regressions are estimated using Cochrane-Orcutt standard errors (Cochrane and Orcutt, 1949) in an autoregressive panel framework. We leverage this approach because idiosyncratic variation in hourly traffic speeds within a road segment are likely correlated. To ensure this solves the serial correlation problem, we test for serial correlation in the error term using the Bhargava et al. (1982) modified Durbin-Watson error term and the Baltagi-Wu LBI statistic (Baltagi and Wu, 1999) in all models.

Table 3 shows the results of the econometric model in equation (5). It starts with an overly simple model with only the indicator variable for hours with main breaks, then adds road segment and time period fixed effects before reporting results for the full specification in column (3). Columns (1) and (2) show that, on average, water main breaks occur on days and hours with slightly higher speeds (approximately 0.76% in column (2)). The key result is from column (3): the coefficient on road segments from water main breaks is -0.0179 and it is statistically significant. Spillover segments between 0.15 and 0.5 miles are significantly impacted at roughly 1/4 the rate of road segments within 0.15 miles (-0.0179 versus -0.0046). In words, there is a roughly 1.8 log-point decrease in traffic speeds on road segments impacted by water main breaks when averaged across all road segments.

[15]There were five types of breaks in the utility's data and they are sorted ordinally based upon break severity. Specifically breaks are denoted by 1–5, with 1 being the least severe and 5 the most severe, as well as an "unreported" category. As shown, the median repair time decreases with the severity of the break which is not surprising given that utilities don't arbitrarily resource breaks and often fix most severe breaks first. DC Water notes, "A simple water main repair can be completed in six to eight hours, but large or complicated repairs may take several days to a week" (source: `https://www.dcwater.com/wastewater/watermain_break.cfm`).

4.2 Diff-in-Diff with cluster-specific ATT

In our next specification, we generalize the model in equation (5) to allow for differential impact of water main breaks across road segments from different clusters. As discussed above and shown visually in Figure 3, we define a treated segment, $Break_{it}$, as any segment within 0.15 mile of the address of a reported break. We define control segments, $Cluster_{it}$, as any segment that is in the same cluster as a treated segment and more than 0.5 mile from a break during the time of a break. Thus, $Cluster_{it}$ is the k-means analog to the population level $1\{Any\ break_t\}$ in the previous specification. Put another way, when a segment is treated only segments in its same cluster serve as controls. We also allow for segments in the same cluster between 0.15 and 0.5 mile from a break to be spillover segments, $Spillover_{it}$, during time periods where a break occurs. We estimate a treatment effect of these segments to determine any possible diffusion of congestion radiating from a break.

The simplest form of the estimating equation is

$$\ln(speed_{it}) = \alpha_i + \beta \cdot 1\{Break_{it}\} + \gamma \cdot 1\{Cluster_{it}\} + \gamma_S \cdot 1\{Spillover_{it}\} + \lambda_t + \varepsilon_{it} \qquad (6)$$

Summarizing the intuition for the regression, for time periods when there are no main breaks in the data, none of the indicator variables takes the value of one. During time periods where there is a break within 0.15 miles of a segment i, $1\{Break_{it}\} = 1$ for that segment. All segments in the same cluster as segment i have $1\{Cluster_{it}\} = 1$ as would segment i. Similarly, all segments in the same cluster as segment i and between 0.15 and 0.5 miles of the break have $1\{Spillover_{it}\} = 1$. Thus, the coefficient γ describes the average difference in traffic speeds when a break occurs relative to baseline for impacted clusters. By assumption, this simple specification imposes that the average speed differential when a water main break occurs is uniform across clusters. We relax this assumption below. The coefficient γ_S is the spillover effect of traffic from a road segment where a break occurs. Our identifying assumption for causality is that a break occurs exogenously within a cluster, since γ controls for average speed differences during break hours.

The coefficient of interest is β, which is the causal impact of a break on traffic speeds, corresponding to the ATT in Equation 3 relative to road segments in the same cluster but more than 0.5 miles from a water main break. This simple specification imposes that the ATT is the same across all clusters; we estimate a more general specification below which allows heterogeneity across clusters. Thus the key difference between this estimating equation and the previous one is the set of controls changes to the clusters of any road segments impacted by a water main break.

Since Equation (6) is a generalized difference-in-difference specification, we briefly discuss pre-trends in our outcome variable before turning to results. Although the timing of water main breaks is exogenous, there may be underlying characteristics of the road segments that are correlated with the likelihood that a water main break occurs. We present a simple pre-trend analysis in the appendix (Figure A.2) by summarizing the hourly speeds on the day previous to a water main break in our sample for treatment and control segments. For six out of our ten clusters, where we observe the majority of our data, pre-trends look like nearly parallel level shifts and the difference in levels is small. Still, we estimate our main specifications with road segment fixed effects to eliminate level differences between treated and control segments.

The results from estimating equation (6) are shown in Table 4. Each column of the table adds more controls and independent variables until column (4), which has the full model with controls in equation (6). In each case, the coefficient on $1\{Break_{it}\}$ is the coefficient of interest and measures the impact of a water main break on road segments. There are two initialy takeaways. First, the point estimate for the causal impact of water main breaks on traffic speeds in the complete model in column (4) is a -0.0142 log point reduction, or about 25% smaller than the -0.0179 log point reduction in the equation (5) specification (column 4). This difference, however, is not statistically different, although a 25% change in a treatment effect can be quite large in other settings. Second, we also find no evidence of statistically significant spillover effects. Each of these findings is robust to varying the number of clusters, as shown in Appendix Tables A.7 and A.8.

We also estimate equation (6) with cluster-specific treatment effects so that there is a unique cluster identifier for all 10 clusters interacted with the $1\{Break_{it}\}$ indicator. Table 5 shows those results.[16] For clusters 1, 7, and 10 there are statistically significant treatment effects ranging from 1.3% to 3.6% in the final specification. Estimating different point estimates across segments highlights the value of using clustering to identify heterogeneous impacts. Recalling that clusters 5, 7, 8, and 10 are the largest clusters in the sample, the lack of significance in clusters other than 1 is plausibly attributable to power issues rather than a true zero effect. Clusters 5 and 8 have the expected sign and magnitudes, but are significant only in columns (1) and (2). In this specification we do not find evidence of nonzero spillover effects.

4.3 Diff-in-Diff with cluster-specific ATT and cluster-specific controls

Lastly, we estimate equation (6) with both cluster-specific treatment effects and cluster-specific control indicators. We view this specification as the one that best leverages the k-means clustering algorithm for identifying heterogeneous treatment effects. That specification is:

$$\ln(speed_{ict}) = \alpha_i + \Sigma_c \beta_c \cdot 1\{Break_{ict}\} + \Sigma_c \gamma_C \cdot 1\{Cluster_{ict}\} + \gamma_S 1\{Spillover_{it}\} + \lambda_t + \varepsilon_{it} \quad (7)$$

In equation (7) we don't allow the spillover effect to vary across clusters since we view spillovers as second order and, having estimated a cluster-specific spillover model, their inclusion does not alter the main findings in a meaningful way. The key difference between this model and the previous one in Table 5 is that we allow the timing of breaks to be arbitrarily correlated with sampling variation in the average speed of different clusters (e.g., average speed differences within a cluster during a treatment event are controlled for). Given that we observe only 515 unique segments impacted by a water main break (sometimes within the same cluster) sampling variation could be non-trivial. Hence, we view this model as both our most flexible

[16] Note that Table 5 includes only 2,180 road segments because 2 road segments in our sample have no identifying characteristics to be used in our clustering algorithm.

model and our preferred specification.

We report results from equation (7) in Table 6. Compared to results in Table 5, the two statistically significant and largest in magnitude point estimates decrease with cluster specific controls. Further, in the full specification (column 4) the number of statistically significant road segments (at the 10% level) increases from two to six. This finding appears to be driven by statistically significant heterogeneity in the cluster control variable. We take this as evidence of increased precision in estimating the treatment effect due to decreased noise in cluster specific untreated road segment speeds making up the control relative to untreated road segment speeds across all clusters. Put another way, the power needed to identify significant treatment effects with lower variation in control outcomes decreases. We view this as a key benefit to leveraging the clustering algorithm in this type of an applied setting where there is meaningful heterogeneity across a population.

In both Tables 5 and 6, we observe some cluster types with significantly impacted speeds due to water main breaks but for some we do not. Applied econometricians historically use their own priors to determine groups or clusters whereas in this approach we've used an algorithm to do so. Further, the algorithm creates clusters before any causal effects are estimated. While it might be unsettling to not have priors drive cluster selection for some practioners, it is far more transparent and data driven. Since it is done separately from estimation of ATTs, it is also unbiased.

We can also refer back to Table 2 and evaluate the summary statistics of the type of road segments which have significant negative speed impacts attributable to water main breaks. They tend to be road segments with the lowest mean speeds (i.e., clusters 1, 7, 8 and 10). These findings can lead to data-driven theories about why some road segments would be impacted by water main breaks while other segments would not. For example, road segments that are already more congested could be impacted the most. Alternatively, small roads with low speeds could become impassible with a water main break being repaired. We view algorithmic data driven theory based upon unbiased causal inference to be a good thing for economics.

In this section, we found three main econometric artifacts of leveraging k-means to cluster road segments. First, when we allow clusters to serve as controls for road segments as opposed to all road segments, our point estimate decreased by 25%. Although not an order of magnitude difference and not statistically different, this effect is still meaningful. Second, including cluster-specific controls in addition to cluster specific treatment effects increases statistical power. We attribute this to a reduction in sampling variation when leveling cluster level identifiers in control road segments. Third, estimating heterogeneous treatment effects explicitly along with heterogeneous control groups shows significant heterogeneity in causal effects. Point estimates relative to the simple model not leveraging k-means clustering varies from more than doubling of the causal effect (cluster 4 at -0.043) to some segments having no statistically significant impact (clusters 2, 3, 5 and 6). Of course, we only examine a single economic costs and there are likely to be others to break proximate businesses and residences. It stands to reason that policymakers in this and other contexts could construct optimal policy which leverages large heterogeneous impacts like these (e.g., prioritization of repairs for the types of clusters which have a larger economic cost).

4.4 Robustness checks

Although our results are fairly consistent across specifications, in order to ensure that our estimates can be attributed to water main breaks we perform several robustness checks.

One challenge of this study is possible measurement error in our treatment identifier. For severe breaks, which receive the highest priority, the median time between when a break is reported and when it is repaired is 12 hours (rounded down; see Table A.1). However, the least severe breaks have median repair times of over 200 hours. This is likely due to lower-resourced and less timely repair schedules for less "important" breaks. This concern initially led us to define treatment as the 12 hours before a repair is completed in order to mitigate the errors-in-variables problem.

As a robustness check, we estimate our main specification using the lesser of (a) the difference between the time of a reported break and its completion and (b) one week from reported completion as the treatment window. Results are in Appendix Table A.3. The alternate treatment window finds estimated results of -1.9%, relative to that of our primary treatment definition of -1.4% in the analogous specification above. These estimates, however, are not statistically different. Given the robustness of our primary result to this alternative treatment definition, we view this as evidence that our preferred specification is likely to provide an accurate point estimate.

We estimated the same regression as in equation 6 with severity-level treatment effects, rather than pooled, to account for the prioritization of DC Water directly. We present this table in the Appendix (Table A.4) and we find significant point estimates between 1% and 5% reductions for severity level 1, 2, 3, and 5 breaks. Notably, the treatment effect for severity 5 breaks is statistically similar to our preferred treatment estimate and does not suffer from small sample problems. Further, it is these breaks that are prioritized to be fixed immediately, so this result suggests that our preferred estimates are robust to congestion mitigation efforts by the constructions crews (such as waiting until nighttime, when there is less traffic, to repair the main).

Lastly, we estimate a placebo test of randomly generated water main breaks. We generate 515 random water main breaks in our sample. We then construct treated and control segments using the exact same procedure as with reported breaks. Table A.5 reports the results from estimating our main specification on the placebo data. We repeat the procedure several times, but report the results from only a single run. In no case do we find a statistically significant impact of breaks on traffic speeds.[17]

4.5 Heterogeneous impacts by time of day

Anderson (2014) shows that the impacts of transit infrastructure disruptions vary by time of day. Intuitively, a disruption is more problematic during high traffic volume periods when the marginal impact of another commuter is larger. As a result, we estimate both aggregated and cluster-specific versions of the econometric model restricting the sample to time-of-day

[17] We also include alternatives to our main specification using k-median clustering in Table A.6, rather than k-means clustering. In our primary results, we set the number of clusters to 10. In a robustness check, we also estimate our primary model with $k = 8$ (Table A.7) and $k = 15$ (Table A.8).

bins. Specifically, we break the day into five parts: 7AM–10AM, 10AM–1PM, 1PM–4PM, 4PM–7PM, and 7PM–10PM.

Table 7 shows results for our time-of-day regressions. We find several important patterns in the data that are robust to alternative specifications. First, the causal impact of breaks varies throughout the day. Largest impacts are during the morning commute (-3.97%) and the magnitude of these impacts weaken throughout the day. This result is consistent with repairs having a higher probability of being fixed by later in the day. To that end, we find a positive and insignificant impact of treatment on speeds during the afternoon rush hour.

Second, spillovers are much more pronounced when breaking out results by time of day. In all but one case, the spillover effect is smaller in magnitude than the direct treatment impact. During the time period when the spillover effect is larger than the treatment effect (7PM–10PM) the two estimated coefficients are not significantly different. This finding is consistent with a spatial diffusion of delays with strongest impacts at the point of the water main break.

Third, having the appropriate control group takes on extra importance in the time-of-day results. Table A.9 in the Appendix shows results including cluster-specific controls. As before the coefficients on *Break* and *Spillover* are defined as marginal impacts on top of speeds in the control streets in the same cluster. The table shows statistically significant heterogeneity in the control cluster speeds by time of day. These results reveal increases in precision and magnitude of treatment effects by hour of day. We note that spillover impacts remain unchanged relative to the specification where the average impact of a control period is assumed to be uniform across clusters.

5 Policy Implications

We find small but statistically significant impacts of water main breaks on traffic speeds. The impacts range from $0-5\%$ decreases in traffic speeds in road segments proximate to the break. These results are robust to a variety of specifications and classification criteria. Our falsification tests show the estimated effects are driven by main breaks. While traffic patterns certainly are correlated to population densities, we show additional heterogeneity conditional on urban density levels (e.g., across clusters within Washington, DC). The impacts also vary with time of day and range similarly between $0 - 5\%$. To our knowledge, this is the first algorithmic evidence in the economics literature providing evidence that traffic effects which cause delays can have the same level of heterogeneity over space as they do over time of day. This insight is useful because it is consistent with managers prioritizing different types of road segments over others for repair to avoid traffic delays.

The direction of our results is sensible but the magnitudes are somewhat surprising for two reasons. First, water main breaks are frequently reported by local and national media outlets. Second, there is a growing acknowledgment that water and other public infrastructure is deteriorating. Our evidence is consistent with these stylized facts. In our study, however, we find that the costs of a single type of public infrastructure break is not large for the single outcome we examine. Changing water infrastructure investment strategies because of concerns about the effects of water main breaks on indirect economic outcomes (e.g., traffic delays) seems not to be justified.

To put our estimated treatment effects in context, we approximate welfare impacts of traffic disruptions attributable to water main breaks using both average impacts over the entire sample and the heterogeneous time-of-day impacts in the spirit of Anderson (2014). To do so we download daily traffic count data from Washington, DC. The average city street has roughly 12,500 unique cars travel on it per day.[18] Consistent with the Department of Transportation guidelines, we use half the hourly wage rate in the Washington, DC, metropolitan statistical area (MSA) as reported by the Bureau of Economic Analysis website to value time: \$18.80/hour. Table A.10 in the Appendix shows mean speed by hour of day over all city streets.[19] By using average speed by hour of day, we construct the number of minutes taken to travel one mile. We can compare average speeds and expected speeds during treated hours

[18]See `http://rtdc.mwcog.opendata.arcgis.com/datasets/fd3a40a7e317420faff13864c7b82bc7_0?uiTab=table`.

[19]This table also includes mean and standard deviation of the INRIX "score" for the speed data. Score measures the data quality averaged over an hour. 30 is an actual reading and perfect data, 10 is an interpolated speed reading. The overall average data quality according to this metric is 26 and data quality is roughly consistent across our sample.

to infer the time cost attributable to water main breaks.

We make two simplifying assumptions to make the welfare calculation tractable. First, we have to determine the total number of miles of street that are subject to the treatment effect and the spillover effect. To do so, we assume there is a unique street every 0.1 mile since city blocks are commonly 0.1 mile. We also ignore diagonal arterial streets in DC. This is shown in Figure 4. Shaded area is considered the treated area and the unshaded counted as spillover. Each gray line is a single street. In the welfare calculations, we assume the total length of all streets in the shaded circle of radius 0.15 mile is the length of all treated streets during a "treated" period. The total street length in the doughnut surrounding the shaded region is the length of spillover streets. We calculated street lengths using the Pythagorean Theorem since streets are assumed to be spaced at exactly 0.1 miles and circles are symmetric.

Second, we have to determine how many cars travel on each road segment over a day and, in the time-of-day calculation, each time period of the day. To do so, we assume each street has a total of 12,500 cars traveling on it each day. We both assume cars are uniformly distributed throughout the day and that volumes more than double during rush hours in different specifications. Because we have no data on volumes by road segment type, we focus exclusively on temporal heterogeneity since temporal traffic patterns are more well known than spatial patterns. We take parameter estimates from Tables 5, 8, and 9 to calculate the time costs. The magnitude of the time costs is similar to that using other parameter estimates.

Table 8 shows the results of the time costs attributable to water main breaks that occurred over the 12 months we study. Accounting for temporal heterogeneity rather than simple average impacts, we find time costs increase by roughly 400%. This result is consistent with Anderson (2014) who finds the impacts of transit infrastructure disruptions vary by time of day in a similar way.

Our preferred cost calculation is the bottom one in which we use our estimated time-of-day effects and assume more traffic occurs during rush hours. In doing so we estimate a time cost per water main break of roughly $1,350. This works out to roughly $700,000 over the entire year. While half or twice this number is possible, we are reasonably confident this is the correct order of magnitude. Given that the total population of Washington, DC, is roughly 700,000, this works out to roughly $1 per person. In this case $1 per person is almost surely an overestimate: the time-weighted population of Washington, DC, is much larger than 700,000, as many people commute into the city from more suburban areas. We do not view this as a large cost.

The use of these estimates for other urban areas is somewhat plausible, but they probably do not transfer to less urban areas. Washington is a dense urban area with various alternative transport options. The metropolitan DC area consistently ranks as one of the most congested cities, ranking first in annual hours of delay per commuter (Schrank et al., 2015). As a result, the effect of a water main break on traffic patterns in DC may be small relative to a city with fewer alternative commuting options, whether those are alternative routes or different modes of transport. This logic would imply that our results are externally valid for dense road networks in urban cities and likely a lower bound when fewer substitutes are present. Despite this, urban areas on average tend to contain older infrastructure that is of critical policy importance.

6 Policy and Methodological Discussion

Our results do not suggest that infrastructure investment is not important. In fact, the number of water main breaks, and the corresponding age of the mains, for a single urban area within our year-long study period is alarming. Rather, we provide evidence that a single indirect economic cost (increased congestion) from distributed water infrastructure failure is small. Other direct and indirect effects could be large. That said, if other indirect costs of failure were small, then centralized water infrastructure improvements could provide more value than improvements to distributed infrastructure.

It could be that observed failures are not the right measure in this space. Water infrastructure investment might be best framed in terms of forgoing the worst possible outcome, much as electric utilities plan to mitigate blackouts. Low frequency but massive water main failures or infrequent large public health disasters are certainly more salient, and they also may be more important from a benefit cost perspective. In that case, though, we are not aware of a good economic framework for estimating the impacts of those large, and in some cases never observed, events.[20]

More generally, our paper is a starting point rather than a decision point for policymakers in this space. There is a gap in the literature in identifying causal impacts of water infrastructure failure on economic outcomes. While there is a larger literature on dose-response functions that could be used to perform back-of-the-envelope calculations on the costs of deterioration, there is a need to inform policymakers so that they can plan their infrastructure investments efficiently.

In addition to policy insights, this paper also offers lessons for applied microeconometricians interested in leveraging machine learning to algorithmically classify subjects in panel data. We found three implications of using a clustering algorithm to inform heterogeneous treatment effects relative to population average treatment effects. First, when we restricted the set of control road segments for treated road segments to be in the same cluster, our point estimate decreased by 25%. Second, including cluster-specific controls in addition to cluster specific treatment effects increases statistical power. We attribute this to reduced sampling variation when leveling cluster level identifiers in control road segments. Third, estimating heterogeneous treatment effects explicitly along with heterogeneous control groups shows significant heterogeneity in causal effects. Point estimates relative to the simple model not leveraging k-means clustering varies from more than doubling of the causal effect to some segments having no statistically significant impact. It stands to reason that policy makers in this and other contexts could construct optimal policy which leverages large heterogeneous impacts like these. Thus, the econometric insights enabled using k-means clustering could have direct impacts for optimal policy construction in different settings. Much like for our policy implications, we view using unsupervised learning algorithms to form clusters as a starting point which is both a useful and scalable way to complement the existing expertise

[20]One example is trying to identifying the causal impact of a never before observed human threat to water supplies.

of applied econometricians working on important policy topics.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.

Aliprantis, D, A Bergman, and G Caetano, "Identifying Sorting Into Echo Chambers," *Working Paper*, 2017.

Anderson, Benjamin J., Deborah S. Gross, David R. Musicant, Anna M. Ritz, Thomas G. Smith, and Leah E. Steinberg, "Adapting K-Medians to Generate Normalized Cluster Centers.," in "Proceedings of the Sixth SIAM International Conference on Data Mining," 2006, pp. 165–175.

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

Athey, Susan and Guido Imbens, "Recursive partitioning for heterogeneous causal effects," *Proceedings of the National Academy of Sciences*, 2016, *113* (27), 7353–7360.

AWWA, *Buried No Longer: Confronting America's Water Infrastructure Challenge*, American Water Works Association, 2012.

Baltagi, Badi H. and Ping X. Wu, "Unequally spaced panel data regressions with AR (1) disturbances," *Econometric Theory*, 1999, *15* (6), 814–823.

Bel, Germà, Xavier Fageda, and Mildred E. Warner, "Is private production of public services cheaper than public production? A meta-regression analysis of solid waste and water services," *Journal of Policy Analysis and Management*, 2010, *29* (3), 553–577.

Bento, Antonio, Daniel Kaffine, Kevin Roth, and Matthew Zaragoza-Watkins, "The effects of regulation in the presence of multiple unpriced externalities: Evidence from the transportation sector," *American Economic Journal: Economic Policy*, 2014, *6* (3), 1–29.

Bento, Antonio M., 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*, 2013, *77*, 41–56.

Bernhardt, Daniel and Peter Robinson, "Detecting affect from non-stylised body motions," *Affective Computing and Intelligent Interaction*, 2007, pp. 59–70.

Bhargava, Alok, Luisa Franzini, and Wiji Narendranathan, "Serial correlation and the fixed effects model," *Review of Economic Studies*, 1982, *49* (4), 533–549.

Bonhomme, S, T Lamadon, and E Manresa, "Discretizing Unobserved Heterogeneity," *Institute for Fiscal Studies, Working Paper*, 2017.

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

Caballero, Julián A., "Do surges in international capital inflows influence the likelihood of banking crises?," *The Economic Journal*, 2016, *126* (591), 281–316.

Castledine, Anita, Klaus Moeltner, Michael K. Price, and Shawn Stoddard, "Free to choose: Promoting conservation by relaxing outdoor watering restrictions," *Journal of Economic Behavior & Organization*, 2014, *107*, 324–343.

Cochrane, Donald and Guy H. Orcutt, "Application of least squares regression to relationships containing auto-correlated error terms," *Journal of the American Statistical Association*, 1949, *44* (245), 32–61.

Crone, Theodore, "An alternative definition of economic regions in the United States based on similarities in state business cycles," *Review of Economics and Statistics*, 2005, *87* (4), 617–626.

Currie, Janet, Josh Graff-Zivin, K. Meckel, Matt Neidell, and Wolfram Schlenker, "Something in the water: Contaminated drinking water and infant health," *Canadian Journal of Economics/Revue canadienne d'économique*, 2013, *46* (3), 791–810.

Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Parienté, and Vincent Pons, "Happiness on tap: Piped water adoption in urban Morocco," *American Economic Journal: Economic Policy*, 2012, *4* (4), 68–99.

Ferraro, Paul J. and Juan José Miranda, "Panel data designs and estimators as substitutes for randomized controlled trials in the evaluation of public programs," *Journal of the Association of Environmental and Resource Economists*, 2017, *4* (1), 281–317.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky, "Water for life: The impact of the privatization of water services on child mortality," *Journal of Political Economy*, 2005, *113* (1), 188–200.

Gamper-Rabindran, Shanthi, Shakeed Khan, and Christopher Timmins, "The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach," *Journal of Development Economics*, 2010, *92* (2), 188–200.

Graff Zivin, Joshua, Matthew Neidell, and Wolfram Schlenker, "Water quality violations and avoidance behavior: Evidence from bottled water consumption," *The American Economic Review*, 2011, *101* (3), 448–453.

Hamilton, Timothy L. and Casey J. Wichman, "Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare," *Journal of Environmental Economics and Management*, 2018, *87*, 72–93.

MacQueen, James, "Some methods for classification and analysis of multivariate observations," in "Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability" number 1. In '14.' Oakland, CA, USA. 1967, pp. 281–297.

Milligan, Glenn W. and Martha C. Cooper, "A study of standardization of variables in cluster analysis," *Journal of Classification*, 1988, *5* (2), 181–204.

Quistorff, Brian, "The Effects of Energy Prices on Manufacturing Employment," *Microsoft, Working Paper*, 2017.

Rezankova, Hana, "Nominal variable clustering and its evaluation," in "Proceedings of the 8th International Days of Statistics and Economics," Melandrium, Slanỳ 2014, pp. 1293–1302.

Rose, Adam and S. Liao, "Modeling regional economic resilience to disasters: A computable general equilibrium analysis of water service disruptions," *Journal of Regional Science*, 2005, *45* (1), 75–112.

Rosenbaum, Paul R. and Donald B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, *70* (1), 41–55.

_ **and** _ , "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *The American Statistician*, 1985, *39* (1), 33–38.

Schrank, David, Bill Eisele, Tim Lomax, and Jim Bak, "2015 Urban Mobility Scorecard," 2015. Texas A&M Transportation Institute.

Wager, Stefan and Susan Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 2017, (just-accepted).

Wichman, Casey J. and Paul J. Ferraro, "A cautionary tale on using panel data estimators to measure program impacts," *Economics Letters*, 2017, *151*, 82–90.

Wolff, Hendrik, "Keep your clunker in the suburb: Low-emission zones and adoption of green vehicles," *The Economic Journal*, 2014, *124* (578), F481–F512.

_ , "Value of time: Speeding behavior and gasoline prices," *Journal of Environmental Economics and Management*, 2014, *67* (1), 71–88.

Xu, Yiqing, "Generalized synthetic control method: Causal inference with interactive fixed effects models," *Political Analysis*, 2017, *25* (1), 57–76.

Tables

Table 1: Age and material of DC water mains

	Count	Percentage
Total no. of water main breaks (July 1, 2014–June 30, 2015)	278	
Total no. main breaks with installation year recorded	268	
Mean year	1921	
Median year	1926	
Before 1916	122	45.86
Before 1900	76	28.57
Before 1865	5	1.88
Total no. of water main breaks with material info	266	
Cast iron	260	97.01
Ductile iron	5	1.87
PCCP-LCP	1	0.37
Steel	2	0.75

Notes: We analyze 278 water main breaks that are near roads for which we have traffic information, which is a subset of the total number of water main breaks that occurred in this time period. DC Water reported 515 total water main breaks for this time period.

Table 2: Summary statistics for each cluster

Cluster ID	No. segments	Speed (Mean)	Speed (SD)	Max. diff. (MPH)	NB (Pr.)	SB (Pr.)	EB (Pr.)	WB (Pr.)	IS (Pr.)	US (Pr.)
1	121	22.823	6.392	4.913	0.975	0	0	0	0.008	0.008
2	41	41.358	8.534	7.311	0	0	1	0	0.073	0
3	121	42.501	9.639	10.127	0	0.678	0	0.298	0.025	0.033
4	38	29.237	13.261	13.916	0.079	0.737	0.079	0.105	0.053	0.026
5	442	17.77	5.57	4.382	0	1	0	0	0	0
6	90	41.816	10.205	11.313	1	0	0	0	0.044	0.011
7	484	18.252	6.301	3.796	0	0	0	1	0	0.004
8	355	15.829	4.117	3.049	1	0	0.003	0	0	0.006
9	130	23.54	6.297	5.49	0	0	1	0	0	0
10	358	16.049	4.442	3.769	0	0	1	0	0	0

Notes: NB = northbound, SB = southbound, EB = eastbound, WB = westbound, IS = interstate, and US = US highway. Max. diff. is the maximum difference in mean speeds during each hour of the day relative to speeds at 5AM within each cluster

Table 3: The effect of a water main break on aggregate traffic speeds in Washington, DC

	(1)	(2)	(3)
	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$
$Any\ break_t$	0.400***	0.00762***	0.00766***
	(0.000361)	(0.000369)	(0.000369)
$Break_{it}$			-0.0179***
			(0.00326)
$Spillover_{it}$			-0.00458***
			(0.00136)
Observations	8,954,407	8,954,407	8,954,407
R-squared	0.0001	0.147	0.151
Number of segments	2,182	2,182	2,182
Fixed effects:			
Hour FE	NO	YES	YES
Weekday FE	NO	YES	YES
Month FE	NO	YES	YES

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Average treatment and cluster break effects

	(1)	(2)	(3)	(4)
	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$
$Break_{it}$	-0.0187***	-0.0174***	-0.0141***	-0.0142***
	(0.00350)	(0.00326)	(0.00327)	(0.00327)
$Cluster_{it}$			-0.00491***	-0.00489***
			(0.000311)	(0.000316)
$Spillover_{it}$				-0.00055
				(0.00138)
Observations	8,954,407	8,954,407	8,954,407	8,954,407
Number of segments	2,182	2,182	2,182	2,182
Fixed effects:				
Hour FE	NO	YES	YES	YES
Weekday FE	NO	YES	YES	YES
Month FE	NO	YES	YES	YES
Modified Bhargava et al. Durbin-Watson	0.6976	0.7823	0.7826	0.7826
Baltagi-Wu LBI	0.6997	0.7847	0.7849	0.7849

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Cluster-specific treatment effects

	(1)	(2)	(3)	(4)
	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$
$Break_{it} \times Cluster1$	-0.0504***	-0.0393***	-0.0359**	-0.0360**
	(0.0154)	(0.0143)	(0.0143)	(0.0143)
$Break_{it} \times Cluster2$	0.00721	0.0134	0.0176	0.0176
	(0.0402)	(0.0375)	(0.0375)	(0.0375)
$Break_{it} \times Cluster3$	0.00617	0.00818	0.0118	0.0118
	(0.0269)	(0.0251)	(0.0251)	(0.0251)
$Break_{it} \times Cluster4$	-0.0539**	-0.0370	-0.0340	-0.0340
	(0.0253)	(0.0236)	(0.0236)	(0.0236)
$Break_{it} \times Cluster5$	-0.0138*	-0.0125*	-0.00956	-0.00958
	(0.00730)	(0.00681)	(0.00681)	(0.00681)
$Break_{it} \times Cluster6$	0.0287	0.0353	0.0387	0.0387
	(0.0287)	(0.0269)	(0.0269)	(0.0269)
$Break_{it} \times Cluster7$	-0.0230***	-0.0250***	-0.0219***	-0.0219***
	(0.00729)	(0.00678)	(0.00678)	(0.00678)
$Break_{it} \times Cluster8$	-0.0181**	-0.0138*	-0.0106	-0.0106
	(0.00811)	(0.00756)	(0.00757)	(0.00757)
$Break_{it} \times Cluster9$	-0.0311*	-0.0220	-0.0184	-0.0184
	(0.0172)	(0.0161)	(0.0161)	(0.0161)
$Break_{it} \times Cluster10$	-0.0122	-0.0161**	-0.0125*	-0.0126*
	(0.00784)	(0.00728)	(0.00729)	(0.00729)
$Cluster_{it}$			-0.00491***	-0.00489***
			(0.000311)	(0.000316)
$Spillover_{it}$				-0.00055
				(0.00138)
Observations	8,952,305	8,952,305	8,952,305	8,952,305
Number of segments	2,180	2,180	2,180	2,180
Fixed effects:				
Hour FE	NO	YES	YES	YES
Wkday FE	NO	YES	YES	YES
Month FE	NO	YES	YES	YES
Modified Bhargava et al. Durbin-Watson	0.6976	0.7823	0.7826	0.7826
Baltagi-Wu LBI	0.6997	0.7847	0.7849	0.7849

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Cluster-specific treatment effects with cluster-specific controls

	(1)	(2)
	$\ln(speed_{it})$	$\ln(speed_{it})$
$Break_{it} \times Cluster1$	-0.0302**	-0.0302**
	(0.0144)	(0.0144)
$Break_{it} \times Cluster2$	-0.00160	-0.00165
	(0.0380)	(0.0380)
$Break_{it} \times Cluster3$	0.00881	0.00878
	(0.0252)	(0.0252)
$Break_{it} \times Cluster4$	-0.0434*	-0.0434*
	(0.0241)	(0.0241)
$Break_{it} \times Cluster5$	-0.00673	-0.00677
	(0.00682)	(0.00682)
$Break_{it} \times Cluster6$	0.0298	0.0297
	(0.0270)	(0.0270)
$Break_{it} \times Cluster7$	-0.0206***	-0.0207***
	(0.00679)	(0.00679)
$Break_{it} \times Cluster8$	-0.0153**	-0.0154**
	(0.00758)	(0.00758)
$Break_{it} \times Cluster9$	-0.0268*	-0.0269*
	(0.0162)	(0.0162)
$Break_{it} \times Cluster10$	-0.0136*	-0.0136*
	(0.00730)	(0.00730)
$Break_{it} \times Cluster\ Control1$	-0.0128***	-0.0127***
	(0.00184)	(0.00184)
$Break_{it} \times Cluster\ Control2$	0.0164**	0.0164**
	(0.00676)	(0.00676)
$Break_{it} \times Cluster\ Control3$	-0.00108	-0.00104
	(0.00296)	(0.00296)
$Break_{it} \times Cluster\ Control4$	0.00869	0.00880
	(0.00698)	(0.00698)
$Break_{it} \times Cluster\ Control5$	-0.00933***	-0.00930***
	(0.000611)	(0.000613)
$Break_{it} \times Cluster\ Control6$	0.00732*	0.00738*
	(0.00379)	(0.00379)
$Break_{it} \times Cluster\ Control7$	-0.00686***	-0.00683***
	(0.000559)	(0.000561)
$Break_{it} \times Cluster\ Control8$	0.00197***	0.00202***
	(0.000750)	(0.000753)
$Break_{it} \times Cluster\ Control9$	0.00607***	0.00610***
	(0.00184)	(0.00184)
$Break_{it} \times Cluster\ Control10$	-0.00352***	-0.00348***
	(0.000680)	(0.000682)
$Spillover_{it}$		-0.00100
		(0.00138)
Observations	8,952,305	8,952,305
Number of segments	2,180	2,180
Fixed effects:		
Hour FE	YES	YES
Weekday FE	YES	YES
Month FE	YES	YES
Modified Bhargava et al. Durbin-Watson	0.7827	0.7827
Baltagi-Wu LBI	0.7850	0.7850

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Average treatment and cluster break effects: by time of day

	(1)	(2)	(3)	(4)	(5)
	ln($speed_{it}$)	ln($speed_{it}$)	ln($speed_{it}$)	ln($speed_{it}$)	ln($speed_{it}$)
$Break_{it}$	-0.0397***	-0.0182***	-0.0180**	0.0130	-0.0126*
	(0.00863)	(0.00703)	(0.00727)	(0.00799)	(0.00695)
$Cluster_{it}$	-0.00828***	-0.0141***	-0.0321***	-0.0310***	-0.0340***
	(0.000783)	(0.000668)	(0.000706)	(0.000781)	(0.000692)
$Spillover_{it}$	-0.0190***	-0.00482	-0.00829***	0.00458	-0.0194***
	(0.00385)	(0.00301)	(0.00304)	(0.00329)	(0.00287)
Observations	1,680,016	1,680,019	1,673,519	1,677,910	1,673,462
Number of segments	2,182	2,182	2,182	2,182	2,182
Hours	7AM-10AM	10AM-1PM	1PM-4PM	4PM-7PM	7PM-10PM
Fixed effects:					
Hour FE	YES	YES	YES	YES	YES
Weekday FE	YES	YES	YES	YES	YES
Month FE	YES	YES	YES	YES	YES
Modified Bhargava et al.					
Durbin-Watson	0.9390	0.9013	0.8963	0.8548	0.9037
Baltagi-Wu LBI	1.2807	1.2330	1.2734	1.2021	1.2063

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Annual traffic time costs of Washington, DC, water main breaks

Method	Coefficients table	Rush hour car volume	Normal car volume	Total cost
Average (without spillover)	5	n/a	n/a	\$125,988
Average (with spillover)	5	n/a	n/a	\$159,222
Average (with spillover)	8	n/a	n/a	\$444,490
Time of day	9	2,500	2,500	\$648,279
Time of day	9	4,000	1,500	\$695,275

Notes: Assume 12,500 total volume per road/day and value of time of \$18.80/hour. A total of water main breaks occurred between July 1, 2014, and June 30, 2015. For the time-of-day, non-uniform calculation, we find the time cost per break is roughly \$1,350.

Figures

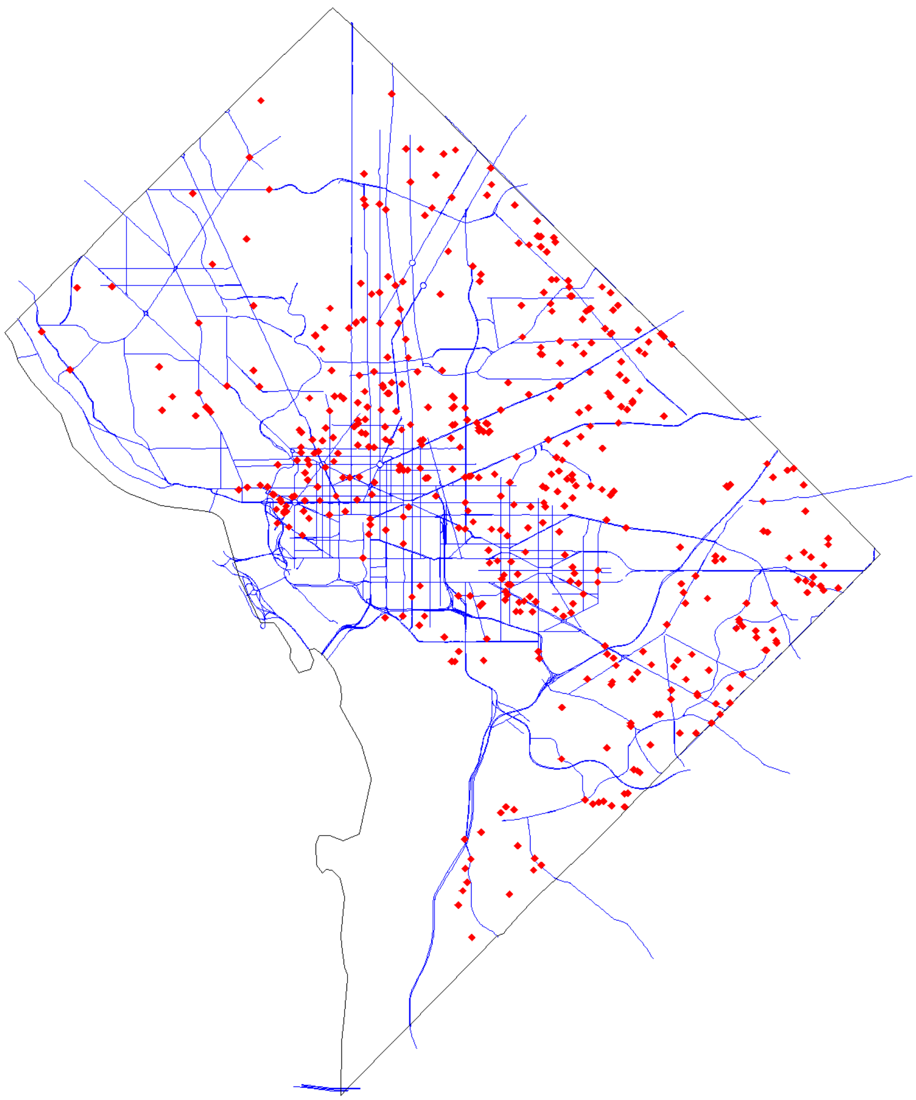

Figure 1: Merged INRIX road segment and DC Water main break data from July 1, 2014, through June 30, 2015

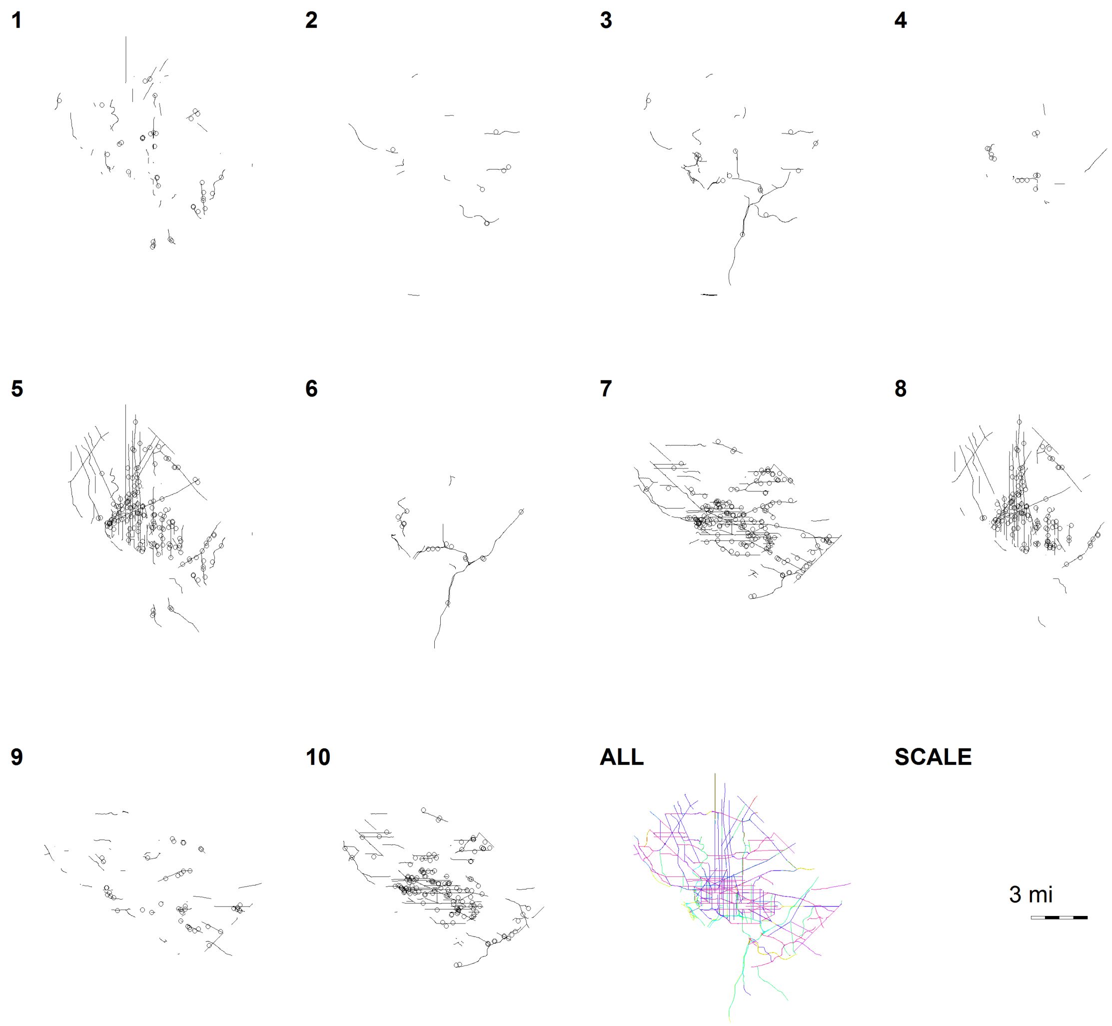

Figure 2: Map of individual road segment clusters and water main breaks

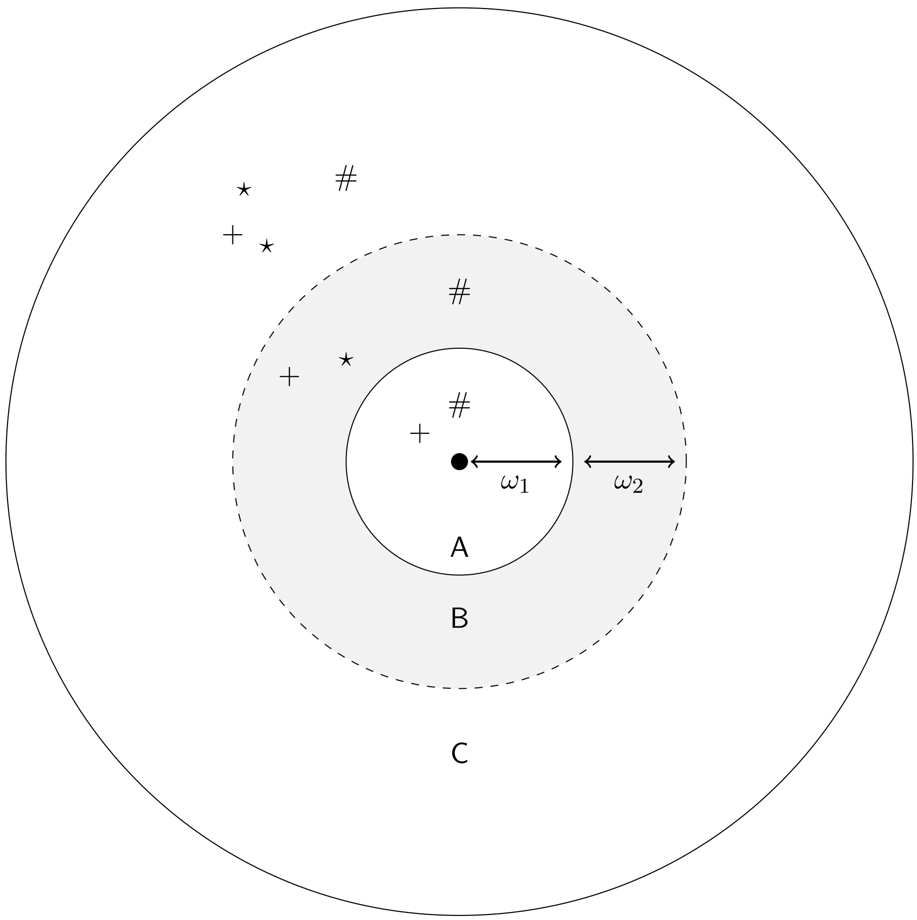

Figure 3: Simplified spatial treatment diagram

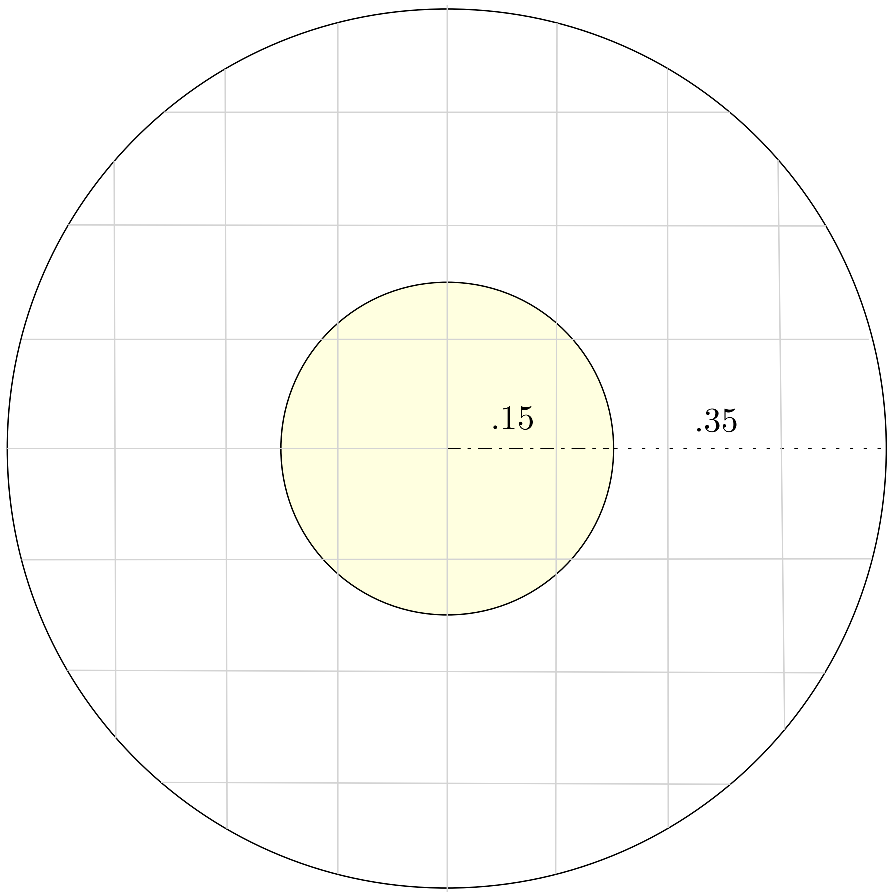

Figure 4: Schematic of assumed street layout

Online appendix – Not for publication

Table A.1: Difference between reported and completion time (in hours) by severity level

Severity Level	Count	1Q	Median	3Q	Mean
1	2	201.6	269.5	337.5	269.5
2	7	329.9	382.5	542.4	423.5
3	41	99.3	189.4	363.1	338.3
4	79	22.6	46.3	96.5	110.0
5	144	9.5	12.8	19.7	22.5
Unreported	5	7.6	18.9	50.4	45.6

Table A.2: Summary statistics cluster-specific breaks (.15 miles)

Cluster ID	No. segments	Speed (Mean)	Speed (SD)	Unique Segment Breaks	Ave Segments-Breaks per break
1	121	22.823	6.392	961	4.56
2	41	41.358	8.534	146	2.55
3	121	42.501	9.639	431	1.26
4	38	29.237	13.261	277	2.02
5	442	17.77	5.57	2348	3.94
6	90	41.816	10.205	434	1.23
7	484	18.252	6.301	2394	3.88
8	355	15.829	4.117	1960	3.76
9	130	23.54	6.297	1217	1.48
10	358	16.049	4.442	2187	3.48

Notes: Table counts and shows the average number of impacts segments for a break that impacts at least one segment in a cluster. Conditional on them being impacted, Cluster 1 is slightly more impacted by main breaks than others accounting for the clusters being larger or smaller. For example cluster 1 averages 4.56 impacted segments per impacting break which is more than segment 7 which has four times the road segments. However, in general the there is a rough ratio of 1 treated segment per 100 road segments conditional on any road segment within a cluster being treated.

Table A.3: Average treatment and cluster break effects: reported time = start time (maximum 1 week)

	(1)	(2)	(3)	(4)
	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$
$Break_{it}$	-0.0119***	-0.0170***	-0.0195***	-0.0199***
	(0.00226)	(0.00202)	(0.00202)	(0.00202)
$Cluster_{it}$			0.00781***	0.00807***
			(0.000311)	(0.000315)
$Spillover_{it}$				-0.00462***
				(0.00081)
Observations	8,954,407	8,954,407	8,954,407	8,954,407
Number of segments	2,182	2,182	2,182	2,182
Fixed effects:				
Hour FE	NO	YES	YES	YES
Weekday FE	NO	YES	YES	YES
Month FE	NO	YES	YES	YES
Modified Bhargava et al. Durbin-Watson	0.6976	0.7824	0.7826	0.7826
Baltagi-Wu LBI	0.6997	0.7847	0.7849	0.7849

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Average treatment and cluster break effects by break severity

	(1)	(2)	(3)	(4)
	ln($speed_{it}$)	ln($speed_{it}$)	ln($speed_{it}$)	ln($speed_{it}$)
$Break_{it} \times Severity1$	-0.159***	-0.103***	-0.101***	-0.101***
	(0.0289)	(0.0267)	(0.0267)	(0.0267)
$Break_{it} \times Severity2$	-0.0686***	-0.0683***	-0.0644***	-0.0644***
	(0.0262)	(0.0244)	(0.0244)	(0.0244)
$Break_{it} \times Severity3$	-0.0541***	-0.0519***	-0.0488***	-0.0488***
	(0.00830)	(0.00773)	(0.00774)	(0.00774)
$Break_{it} \times Severity4$	0.00485	0.00210	0.00531	0.00528
	(0.00621)	(0.00575)	(0.00575)	(0.00575)
$Break_{it} \times Severity5$	-0.0162***	-0.0147***	-0.0114**	-0.0114**
	(0.00514)	(0.00481)	(0.00482)	(0.00482)
$Break_{it} \times Severity\ Unreported$	0.0607	0.0815**	0.0855**	0.0855**
	(0.0381)	(0.0355)	(0.0355)	(0.0355)
$Cluster_{it}$			(0.000311)	(0.000316)
			(0.000314)	(0.000319)
$Spillover_{it}$				-0.000561
				(0.00138)
Observations	8,954,407	8,954,407	8,954,407	8,954,407
Number of segments	2,182	2,182	2,182	2,182
Fixed effects:				
Hour FE	NO	YES	YES	YES
Weekday FE	NO	YES	YES	YES
Month FE	NO	YES	YES	YES
Modified Bhargava et al. Durbin-Watson	0.6976	0.7823	0.7826	0.7826
Baltagi-Wu LBI	0.6998	0.7847	0.7849	0.7849

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Placebo clusters and breaks

	(1)	(2)	(3)	(4)
	ln($speed_{it}$)	ln($speed_{it}$)	ln($speed_{it}$)	ln($speed_{it}$)
$Break_{it}$	0.00177	2.05e-06	-7.51e-05	-7.30e-05
	(0.00654)	(0.00617)	(0.00619)	(0.00619)
$Cluster_{it}$			9.97e-05	9.70e-05
			(0.000550)	(0.000555)
$Spillover_{it}$				0.000133
				(0.00369)
Observations	8,954,407	8,954,407	8,954,407	8,954,407
Number of segments	2,182	2,182	2,182	2,182
Fixed effects:				
Hour FE	NO	YES	YES	YES
Weekday FE	NO	YES	YES	YES
Month FE	NO	YES	YES	YES
Modified Bhargava et al. Durbin-Watson	0.6976	0.7823	0.7824	0.7824
Baltagi-Wu LBI	0.6997	0.7846	0.7847	0.7847

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: k-medians clustering

	(1) $\ln(speed_{it})$	(2) $\ln(speed_{it})$
$Break_{it}$	-0.0130***	-0.0130***
	(0.00327)	(0.00327)
$Cluster_{it}$	-0.00644***	-0.00644***
	(0.000339)	(0.000345)
$Spillover_{it}$		-0.000136
		(0.00146)
Observations	8,954,407	8,954,407
Number of segments	2,182	2,182
Fixed effects:		
Hour FE	YES	YES
Weekday FE	YES	YES
Month FE	YES	YES
Modified Bhargava et al. Durbin-Watson	0.7826	0.7826
Baltagi-Wu LBI	0.7849	0.7849

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Average treatment and cluster break effects, k=8

	(1) $\ln(speed_{it})$	(2) $\ln(speed_{it})$
$Break_{it}$	-0.0141***	-0.0141***
	(0.00327)	(0.00327)
$Cluster_{it}$	-0.00515***	-0.00515***
	(0.000301)	(0.000305)
$Spillover_{it}$		-1.40e-05
		(0.00135)
Observations	8,954,407	8,954,407
Number of segments	2,182	2,182
Fixed effects:		
Hour FE	YES	YES
Weekday FE	YES	YES
Month FE	YES	YES
Modified Bhargava et al. Durbin-Watson	0.7826	0.7826
Baltagi-Wu LBI	0.7849	0.7849

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table A.8: Average treatment and cluster break effects, k=15

	(1)	(2)
	ln($speed_{it}$)	ln($speed_{it}$)
$Break_{it}$	-0.0143***	-0.0144***
	(0.00327)	(0.00327)
$Cluster_{it}$	-0.00443***	-0.00442***
	(0.000339)	(0.000344)
$Spillover_{it}$		-0.0024
		(0.00144)
Observations	8,954,407	8,954,407
Number of segments	2,182	2,182
Fixed effects:		
Hour FE	YES	YES
Weekday FE	YES	YES
Month FE	YES	YES
Modified Bhargava et al. Durbin-Watson	0.7825	0.7825
Baltagi-Wu LBI	0.7848	0.7848

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Cluster-specific treatment effects with cluster-specific controls by time

	(1)	(2)	(3)	(4)	(5)
	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$	$\ln(speed_{it})$
$Break_{it} \times Cluster1$	-0.0488	-0.00579	-0.0508	0.0249	-0.0684**
	(0.0405)	(0.0290)	(0.0323)	(0.0312)	(0.0298)
$Break_{it} \times Cluster2$	-0.0230	-0.0344	-0.00681	-0.0740	0.0716
	(0.121)	(0.0710)	(0.0864)	(0.0896)	(0.0802)
$Break_{it} \times Cluster3$	0.126*	0.0589	0.0112	-0.0110	-0.0485
	(0.0647)	(0.0525)	(0.0552)	(0.0595)	(0.0586)
$Break_{it} \times Cluster4$	-0.0987	-0.0793	-0.108**	0.0251	-0.122***
	(0.0722)	(0.0617)	(0.0538)	(0.0469)	(0.0453)
$Break_{it} \times Cluster5$	-0.0312*	-0.0305**	-0.0214	0.0164	-0.0158
	(0.0178)	(0.0148)	(0.0152)	(0.0166)	(0.0143)
$Break_{it} \times Cluster6$	-0.0149	0.0709	0.00300	0.111*	0.0680
	(0.0706)	(0.0580)	(0.0690)	(0.0655)	(0.0611)
$Break_{it} \times Cluster7$	-0.0532***	0.00501	-0.0376**	-0.0181	-0.0273*
	(0.0181)	(0.0145)	(0.0150)	(0.0170)	(0.0147)
$Break_{it} \times Cluster8$	-0.0485**	-0.0454***	-0.00767	0.0173	-0.000380
	(0.0202)	(0.0167)	(0.0168)	(0.0186)	(0.0158)
$Break_{it} \times Cluster9$	-0.0452	-0.00224	-0.00626	-0.0590	-0.0646**
	(0.0395)	(0.0348)	(0.0365)	(0.0406)	(0.0322)
$Break_{it} \times Cluster10$	-0.0467**	-0.0283*	-0.0193	0.0237	0.00594
	(0.0189)	(0.0155)	(0.0160)	(0.0185)	(0.0161)
$Break_{it} \times Cluster\ Control1$	0.00404	0.000199	0.0191***	0.00617	-0.00197
	(0.00459)	(0.00398)	(0.00433)	(0.00416)	(0.00405)
$Break_{it} \times Cluster\ Control2$	0.200***	-0.0204*	0.175***	0.0844***	0.0391***
	(0.0188)	(0.0124)	(0.0148)	(0.0178)	(0.0150)
$Break_{it} \times Cluster\ Control3$	0.0176**	0.0336***	0.0865***	0.138***	0.142***
	(0.00848)	(0.00567)	(0.00577)	(0.00698)	(0.00741)
$Break_{it} \times Cluster\ Control4$	-0.152***	-0.0538***	0.100***	0.00870	0.0927***
	(0.0227)	(0.0180)	(0.0155)	(0.0135)	(0.0135)
$Break_{it} \times Cluster\ Control5$	-0.00419***	-0.0193***	-0.0366***	-0.0371***	-0.0397***
	(0.00148)	(0.00131)	(0.00134)	(0.00141)	(0.00127)
$Break_{it} \times Cluster\ Control6$	0.129***	0.0915***	0.170***	0.0458***	0.0164*
	(0.00857)	(0.00784)	(0.0101)	(0.00927)	(0.00875)
$Break_{it} \times Cluster\ Control7$	-0.0131***	-0.00991***	-0.0290***	-0.0284***	-0.0335***
	(0.00139)	(0.00116)	(0.00123)	(0.00140)	(0.00122)
$Break_{it} \times Cluster\ Control8$	-0.00327*	-0.0237***	-0.0473***	-0.0366***	-0.0416***
	(0.00189)	(0.00161)	(0.00160)	(0.00177)	(0.00151)
$Break_{it} \times Cluster\ Control9$	0.0302***	0.0259***	0.0392***	0.0355***	0.0340***
	(0.00464)	(0.00365)	(0.00436)	(0.00460)	(0.00414)
$Break_{it} \times Cluster\ Control10$	-0.0247***	-0.0212***	-0.0467***	-0.0497***	-0.0425***
	(0.00171)	(0.00147)	(0.00148)	(0.00171)	(0.00152)
$Spillover_{it}$	-0.0194***	-0.00472	-0.00917***	0.00415	-0.0188***
	(0.00385)	(0.00301)	(0.00304)	(0.00329)	(0.00287)
Observations	1,679,622	1,679,625	1,673,125	1,677,518	1,673,074
Number of segments	2,180	2,180	2,180	2,180	2,180
Hours	7AM-10AM	10AM-1PM	1PM-4PM	4PM-7PM	7PM-10PM
Fixed effects:					
Hour FE	YES	YES	YES	YES	YES
Weekday FE	YES	YES	YES	YES	YES
Month FE	YES	YES	YES	YES	YES
Modified Bhargava et al. Durbin-Watson	0.9392	0.9014	0.8965	0.8552	0.9040
Baltagi-Wu LBI	1.2808	1.2331	1.2736	1.2024	1.2066

Notes: All models control for road-segment fixed effects. All models adjusted for autocorrelation. Robust standard errors in parentheses clustered at the road segment level. *** p<0.01, ** p<0.05, * p<0.1

Table A.10: Mean speed and data quality score by time of day

Time	Mean speed	SD speed	Mean score	SD score	Observations
7A-10A	20.51	9.93	26.17	4.14	1,682,760
10A-1P	21.2	10.6	26.21	4	1,682,760
1P-4P	20.79	10.3	26.14	4	1,676,262
4P-7P	19.36	9.47	25.73	4.09	1,680,658
7P-10P	21.61	10.22	24.17	4.18	1,676,496
All	20.7	10.14	25.69	4.16	8,398,936

A Pre-trends discussion

This subsection characterizes breaks at the cluster level. It also assesses the quality of the clustering exercise in satisfying the pre-trends assumption at the root of our diff-in-diff design.

To show pre-trends within a cluster of treated and untreated segments within a cluster we plot hourly speeds by treatment status the day before a break impacts a cluster. Figure A.2 shows "day before" pre-trend results by cluster with 95% confidence intervals. Blue is treated and red is control.

As expected, clusters with the poorest match in Figure A.2, clusters 3, 4, and 6 have the least number of road segments and the smallest number of treated "segment-days". Cluster 9 has a similarly small set of road segments. Part of this is by design: Clusters 3, 4, 6 and 9 have relatively fast speeds indicating they are main roads. This is borne out in Figure 2. As a result, the pre-trends are worst where we expect them to be worse.

For the remaining clusters, accounting for the vast majority of road segments, the pre-trends looks quite similar for treated and control segments. Clusters 2 and 10 lie mostly on top of each other. Clusters 1, 5, 7, and 8 look exactly like level shifts. That said, the level differences are very small in percentage terms, generally around 10%. Still, we estimate our main specifications with road segment fixed effects to eliminate level differences between treated and control segments. We take Figure A.2 as strong evidence that the pre-trends assumption of the difference-in-differences research design is satisfied at the cluster level. In the subsequent analysis, we'll see that these pre-trend matching clusters, 1, 7, 8 and 10 is where we estimate the most important impacts.

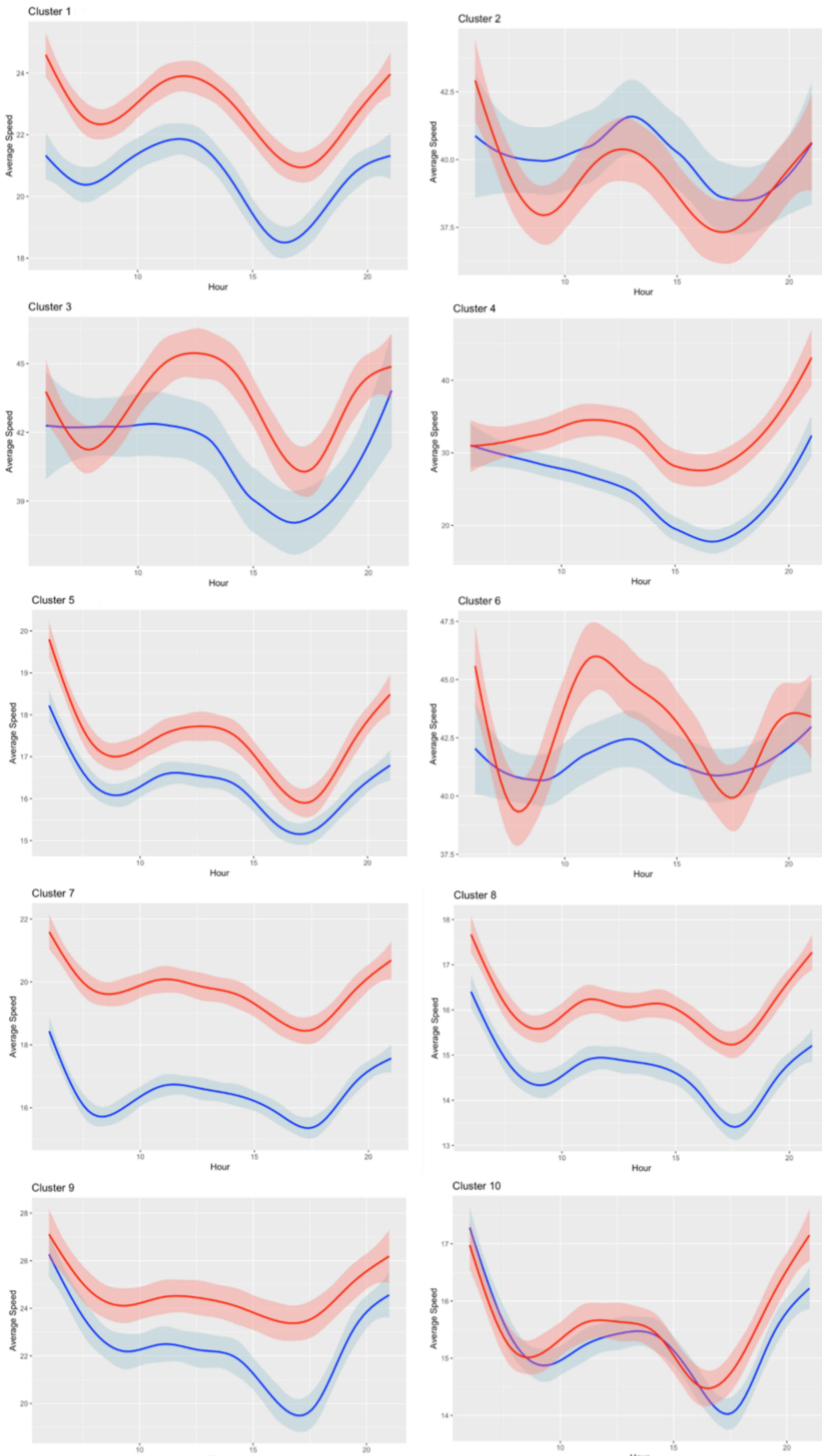

Figure A.2: Each line summarizes average hourly speeds by treated road segments (blue) and control road segments (red) in the day before a main break impacts a subset of segments within a cluster. Blue and red hues show 95% confidence intervals for hourly speeds.