

Police Response Times to Calls for Service: Fragmentation, Community Characteristics, and Efficiency

Daniel S. Bennett*

November, 2018

Abstract

I assembled a unique dataset of every event recorded by the Computer Aided Dispatch systems of 40 different police agencies from December of 2015 to January of 2016 to measure police efficiency by their response times to public calls for service. Using the detailed geographic information provided by these systems, I geocode the calls, match them to the census block group from which they originate, and calculate a predicted response time based on the optimal placement of police response units using a Maximum Covering Model with capacity constraints. I find that minority communities can expect slower response times on average for lower priority calls, but there is considerable heterogeneity across jurisdictions. In areas with less fragmented law enforcement, response times are significantly faster for calls requiring an immediate response.

*Email: dsbennet@stanford.edu. I would like to thank Caroline Hoxby, Michael Boskin, and Mark Duggan for their excellent advice and mentorship. Special thanks are due to the many records specialists in police departments and dispatching organizations who went above and beyond the call of duty in providing me with the data for this research. Thank you to the participants of the Stanford Public Economics and Environmental Economics Seminar for helpful comments. I gratefully acknowledge support for this research from the Bradley Graduate and Postgraduate Fellowship through a grant to the Stanford Institute for Economic Policy Research. The contents of this work are solely my responsibility.

Law enforcement services in the United States have always been fragmented or primarily provided at the local level by municipalities and counties. The system is purposefully decentralized, but how the degree of decentralization impacts the ability of police to operate efficiently is a question that remains open. A consistent difficulty faced by researchers studying this issue and attempting to evaluate police performance in general is the lack of convincing measures that can be compared across and within jurisdictions, neighborhoods, and population groups. Crime rates and the rate of reported crimes cleared by arrest have historically been the most commonly used metrics, but as researchers, and even the agencies collecting such data, acknowledge, crime rates and clearance rates have serious limitations. Many crimes go unreported, and the decision to not report a crime almost certainly depends on the effectiveness of police. Low crime rates could indicate that an agency is efficiently deterring and preventing crimes, or that people have so little faith in the police they don't bother reporting crimes in the first place. Similarly, low clearance rates could indicate that agencies do poor investigative work, or that they efficiently allocate resources to crimes that are relatively difficult to clear.

To contribute to a better understanding of police efficiency, I made a series of Freedom of Information Act (FOIA) Requests, and assembled a unique dataset of every event recorded by the Computer Aided Dispatch (CAD) systems of 40 different law enforcement agencies from January 2015 to December of 2016. Using this data I was able to construct a measure of police effectiveness with clear advantages over the traditional measures. Specifically, response times to public calls for service. Unlike crime rates and clearance rates, response times unambiguously differentiate between better and worse outcomes because, given that a call for service has been made, a faster response is always preferred to a slower response.

This paper uses the CAD data I assembled to gain a better understanding of how fragmentation impacts police efficiency and how the characteristics of different communities within a jurisdiction impact the efficiency of service those communities receive. Both of these questions

have a long standing historical legacy dating back to the first professional police forces. The effect of fragmentation is ambiguous from a theoretical perspective, and empirical studies have been necessarily limited in scope. Furthermore, although the effects of demographic characteristics on police behavior in individual interactions have been studied extensively in a range of different academic fields, the effect of community characteristics on the service those communities receive are less well understood, and it is not obvious a priori what the sign or magnitude of the effects are likely to be.

Complicating any analysis using response times, however, is the fact that different police agencies face very different circumstances, both in the severity of the calls they respond to and in the geographic realities of the areas they serve. To make response times comparable across jurisdictions I make use of the detailed information provided by the CAD systems. Specifically, I use the physical locations to which police were dispatched to calculate an optimal placement of police response units using a maximum covering model (MCM) with capacity constraints. I generate a predicted response time for each call from the results of the MCM and subtract the predicted response time from the actual response time to get a “residual” response time that controls for the differences in the distribution of call locations, physical distances, road networks, and average traffic conditions that different agencies face.

This paper uses the residual response times to answer two specific questions. How does the degree to which police services are fragmented impact police efficiency, and how do the characteristics of the different communities police serve impact the efficiency of the service those communities receive. Using a two stage least squares estimator, I find that residual response times are significantly faster for calls requiring an immediate response in Metropolitan Statistical Areas (MSAs) with less fragmented police. However, residual response times are higher for lower priority calls in such MSAs. Many community characteristics have statistically significant impacts on residual response times. A characteristic of particular interest is the racial makeup of communities. I find communities with a higher share of white

residents can expect longer residual response times for calls requiring an immediate response, but shorter residual response times for lower priority calls. Using ordinary least squares regressions I evaluate the degree to which the estimated impacts of community characteristics vary among police departments. There is a significant amount of heterogeneity across all characteristics, and in particular, in only 8 departments is race a statistically significant predictor of residual response times for calls requiring an immediate response. In all except three cases, a ten percentage point increase in the share of community that is white would change residual response times by less than a minute for these calls.

The remainder of the paper will proceed as follows. Section I gives some background information, and a brief review of the theoretical and empirical literature analyzing the effects of fragmentation on police. I include an overview of police research using response times. In Section II I describe in detail the CAD data that I assembled and the methods I used to collect it. I briefly discuss the other data sources used in this research, which are generally well known. Section III describes the process of calculating an optimal placement for police response units, including the specifics of the Maximum Covering Model and the assumptions used to find a solution. I also describe the estimation procedure and the instrument I use to estimate a two stage least squares regression. Section IV presents the main results, and Section V concludes.

I Literature Review

The fragmented system of law enforcement in the United States exists primarily for historical reasons. Americans of the 19th century, when the first professional police departments were formed, had a strong preference for the local provision of public services as a safeguard against abuses by officials at higher levels of government. These preferences received an important theoretical justification in 1956 when Charles Tiebout published his seminal paper on the provision of local public goods. Tiebout (1956) argued that municipalities experience a kind of competitive pressure when seeking to attract residents that leads them to provide the efficient levels of local public goods. Following work largely bore out some of the empirical predictions of Tiebout’s Hypothesis (Howell-Moroney (2008)), and in many contexts competition between local governments has been shown to improve outcomes. Just a few examples of the effects of competition include; increasing student test scores in Hoxby (2000), protecting against deforestation in Wright et al. (2016), and reducing the price of state lottery tickets in Knight and Schiff (2012).

While the widely varying preferences for police services among the population provide scope for Tiebout style competition to improve welfare, there are also large fixed costs associated with operating police departments. Too much fragmentation could cause agencies to fail to take advantage of economies of scale in providing police services. Complicating the issue even further are the positive and negative externalities of policing that are not present in other public goods. Since criminals can commit crimes in multiple jurisdictions, arrests in one jurisdiction can reduce crime in all of the neighboring ones.¹ Similarly, effective policing can displace criminals into neighboring jurisdictions where they believe they are less likely to be caught. For example, Gonzalez-Navarro (2013) found that approximately one fifth of the reduction in crime that states in Mexico enjoyed after the introduction of an anti car theft device was displaced into neighboring states that did not adopt the new

¹Much of the literature calls this an incapacitation effect.

technology. Theoretical models in which multiple municipalities simultaneously choose a level of police services to provide and take account of these externalities, such as Pinto (2007), Lee and Pinto (2009), and Bandyopadhyay, Pinto and Wheeler (2011) find that local provision dominates centralized provision only in very specific circumstances.

Despite the theoretical ambiguity, empirical studies of fragmentation have found predominantly positive effects of more decentralized provision. Ostrom, Parks and Whitaker (1973) and Ostrom and Whitaker (1973) find in case studies of several police departments in the Indianapolis metropolitan area that citizens living in the smaller police departments had more favorable opinions of the police and rated the service they received more highly. Wheaton (2006) comes the closest to the analysis I perform in this paper, trying to estimate the impact of fragmentation on crime rates and clearance rates in Metropolitan Statistical Areas (MSAs) across the United States. He estimates a structural model of municipalities choosing a level of police services and criminals simultaneously choosing a location in which to commit crimes, finding that Metropolitan Statistical Areas (MSAs) with higher number of police departments have both lower expenditures on police and lower crime rates. He attributes this finding to an X-efficiency gain from the competitive pressure of Tiebout choice, but it is difficult to directly interpret his results given the weaknesses of crime rates discussed above.

Response times to calls for service have been used to evaluate emergency services in general and police in particular for quite some time. Recent work in economics has even reinforced the importance of response times for police. Vidal and Kirchmaier (2018) find, contrary to previous research, that response times can significantly impact the probability of clearing a crime by arrest. Faster response times increase the likelihood of apprehending a suspect at the scene and increase the likelihood of complainants identifying a suspect. Research concerning police response times in the United States goes back to at least the 1970's. Pate et al. (1976) is an early example and one of several papers studying response times using data

from a study of the Kansas City Missouri Police Department. They examine the relationship between response times and citizen satisfaction, finding a positive impact of fast response times on citizen opinions. More recent examples including Cihan, Zhang and Hoover (2012) and Lee, Lee and Hoover (2016), who study response times of the Huston Texas police department, and Cihan (2014) who compares the response times of the Huston and Dallas police departments. These studies find similar results, namely that the circumstances of the calls impact response times and that economically disadvantaged census tracts can expect faster response times.

I contribute to this literature in several important ways. I abstract away from the general problem of how many resources to dedicate to policing while individuals simultaneously choose how many crimes to commit, and instead focus on a specific issue. Given the resources dedicated to policing, how do fragmentation and community characteristics influence the efficient utilization of those resources. Instead of a case study of one or a few very closely related police departments I gather data from across the country to compare the efficiency of different agencies. I analyze all public calls for service received by these departments, rather than focus on a specific type of call, and I use Geographic Information Systems (GIS) analysis to calculate residual response times. This allows me to control for the distribution of calls locations and other physical geographic differences different agencies face in a way other studies have not been able to. The GIS analysis also allows me to examine, at a uniquely fine geographic level, the degree to which community characteristics impact police service.

II Data

I use three different datasets for this research. Two of the datasets are well known, and I created the third, namely the CAD dataset that was used to calculate response times. I describe the method by which I obtained the CAD data and then briefly discuss the other two datasets. I assembled the CAD dataset from a series of FOIA requests to individual

police departments across the United States. I requested an export of every event recorded by their CAD system from January 2015 to December 2016.² These CAD events include calls made to 911, calls to non-emergency lines, and events that were initiated by officers such as traffic stop to issue a citation. I selected the sample from the set of agencies that submitted data to the FBI's Uniform Crime Reports (UCR). Although a small number of the smallest agencies in America are not included in the UCR, it is the most comprehensive list available,³ and the FBI estimates that 97.8% of the population live in the jurisdictions' of agencies that report to the UCR. I designed the sampling method to ensure sufficient variation in the measure of police fragmentation or competition since this is the explanatory variable of interest. I measure competition using a Hirsch-Herfindahl Index, where a market is defined as a Metropolitan Statistical Area, and an agency's market share is defined as the percent of the MSA's population that lives in the agency's jurisdiction. Since the HHI is measured at the level of the MSA I randomly selected MSAs first, and then randomly selected an agency within that MSA to send a request to. To ensure sufficient variation in the HHI, I partitioned MSA's into quintiles of the HHI and selected 60% of my sample from the top and bottom quintiles with the remaining 40% from the middle three quintiles. When choosing an agency from an MSA, I selected only those with at least 20,000 people living in its jurisdiction to exclude atypical cases such as university police and cities like Vernon California that have very few residents, but a large number of people who commute into the community to work.⁴

In some places, agencies depend on county level organizations that dispatch for multiple law enforcement agencies. In these cases I made the request to the dispatching organization, and typically obtained data for every agency for which the organization provided dispatch

²Four agencies in sample switched CAD durring this time period and as a result gave me only 12 months or 18 months worth of data.

³The Justice Department conducts a census of State and Local Law Enforcement Agencies, but the most recent publicly available data from the CSLLE was collected in 2008.

⁴Vernon, with a population of 112, has a technically accurate but highly misleading crime rate of about 4 crimes per person per year, which superficially makes it a crime capital of the United States.

services. I contacted a total of 249 agencies with FOIA requests, and while almost every agency responded to my initial request, only 57 agencies provided me with data including New Orleans and Detroit whose CAD data is publicly available. Although police call logs and CAD data are typically covered by FOIA laws, the agencies that declined to fulfill my FOIA requests typically did so due to technical limitations. Specifically their CAD systems were not designed to allow users to extract customized reports, and the default reports contained personal information, such as names, that would have to be redacted by hand. These datasets would have been prohibitively expensive to assemble. This raises the possibility that my sampling method is biased by systematically selecting agencies with the technical ability to provide the data I am analyzing. Furthermore an additional 17 agencies had to be dropped from the sample because the agencies provided me with incomplete data. In these cases the data was either missing a description of the calls' priority or an accurate address to which police were dispatched. The final sample consists of 40 agencies representing 7.5 million CAD events, of which, 3.4 million were public calls for service. Table 1, shows a comparison of all agencies that satisfied my sampling restrictions and the 40 agencies in my sample. There are significant differences in the summary statistics of the agencies in sample and the universe. As we might expect, the agencies in sample have larger populations and come from more concentrated MSAs. They also have higher crime rates and fewer officers per resident. These differences should be taken into account when interpreting the results of this research. Future work may be able to improve on this limitation by collecting a more representative sample. Figure 1, shows the location of the complying agencies.

For each agency and for each call in the sample I was able to obtain the time that four key events occurred and descriptive information about the characteristics of the call. The timing variables are the time that the call was created by the CAD system, the time the first officer was dispatched, the time the first officer arrived on scene, and the time the call was cleared from the system. Except where explicitly stated otherwise, I define response time as

Table 1: Summary Statistics Comparing Agencies in Universe with the Agencies in Sample.

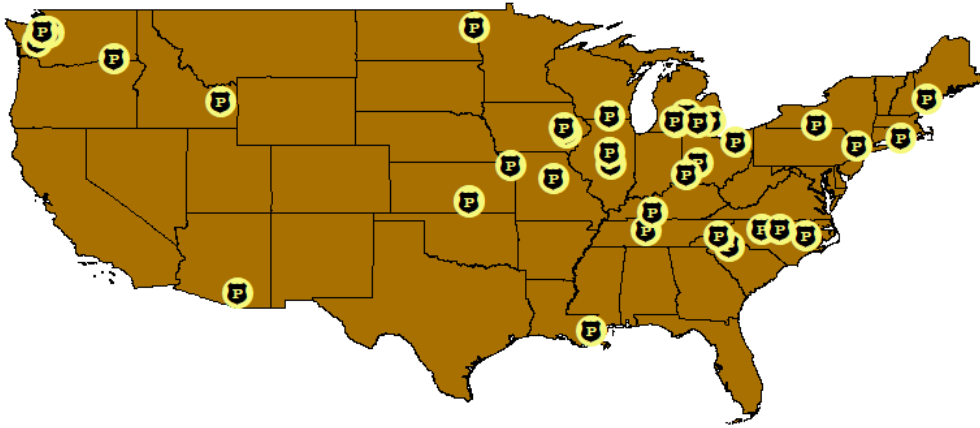
	Universe			Sample		
	Mean	Median	Sd	Mean	Median	Sd
Population	64,041	28,673	195,974	121,192	72,694	143,687
HHI	1,786	1,510	1,386	2,880	3,000	1,463
Crime Rate	3,155	2,621	2,204	5,209	4,757	2,410
Murder Rate	3.4	0.0	6.4	6.3	0.0	10.0
Clear Rate	0.31	0.31	0.16	0.25	0.26	0.12
Officer	17.5	16.0	15.7	18.0	16.0	6.4
Civilian	6.9	4.0	14.5	4.1	4.0	2.1
Injury Rate	0.39	0.0	1.74	0.54	0.0	1.50

Notes: The agencies “in Universe” are those agencies which could have been sampled. The atypical departments I excluded from my sampling method are also excluded from the table. Crime rate is measured as crimes per 100,000 people. Clearance rate is the share of reported crimes cleared by arrest. Officer and Civilian are the number of sworn officers and civilian employees per 1,000 people, respectively.

the time between the call being created and the first officer arriving on scene, since this is the delay the person making the call would most likely perceive.

The descriptive information consists of four variables. They are the nature of the call, a priority code, the address to which police were dispatched, and a clearance code. The nature of the call is the reason police were dispatched. This variable is typically fairly specific. Some examples include, “domestic disturbance,” “robbery,” and “burglary in progress.” Priority codes are an indicator of how quickly the officer should respond to a call. They are not a description of the order in which an officer should perform tasks. A shooting in progress will always have the same code, even if an officer is ordered to respond to one shooting first. The nature and the priority code are information given to the officers when they are dispatched. What the police actually find when they arrive may be significantly different, but since this “true” nature is unknown to the officer before she arrives it should not affect her response time. Clearance codes describe the actions officers took after arriving on scene, usually in a very broad sense. For example, a significant portion of calls have the clearance code, “took

Figure 1: Map of Agencies In Sample



notes: Yellow police shields indicate the location of an agency included in the sample.

necessary action.” The last piece of descriptive information is the address to which officers were dispatched. Most of the agencies I requested data from were able to provide me with the exact addresses. Some agencies, with stricter public privacy laws, were only able to provide me with the block address. In other words, the address “656 18th Avenue” would be recorded as “600-699 18th Avenue” in the data. In almost all jurisdictions, certain calls, particularly those in which a child was the victim of a crime, have their address obscured to the block level to protect the privacy of the victim. Despite these cases in which addresses are partially obscured, the data is still fine enough to use in my GIS analysis. Section III describes the process of GIS analysis in detail.

It is important to re-emphasize here that no two law enforcement agencies are the same. In particular, virtually every agency uses a different system of natures and priority codes. Comparing response times across jurisdictions requires controlling for the severity of the call since some agencies systematically respond to more calls that require an immediate response. A natural inclination might be to control for severity with the call’s nature. But two calls with the same nature, for example “assault,” may require very different responses. Police try

to respond immediately to an assault that is in progress, but would assign a lower priority to an assault that occurred a few days ago. So instead of using the call's nature, I use priority codes to control for the severity of the call. Most priority code systems are a set of numbers. Most typically a code of 1 indicates the highest priority while successively bigger numbers indicate successively lower priority statuses. While the modal agency uses a system with nine numbers, the number of codes used by a department ranges pretty widely, from Holly Hills Florida, which does not use priority codes at all and had to be excluded from the sample, to New Orleans Louisiana, which uses 35 different codes. To make the priority codes comparable, I mapped each individual agency's priority codes into a universal system consisting of three levels. Priority 1 calls are calls that require an immediate response, which includes crimes in progress and some medical emergencies. There are 667,172 priority 1 calls in the sample making up 19.4% of calls for service. Priority 2 calls are calls requiring an expedited response, which includes crimes that occurred recently, some traffic accidents, etc... There are 1,608,127 priority 2 calls, representing 46.9% of calls for service in the sample. Finally priority 3 calls are calls that require a routine response, which includes crimes without a suspect, noise complaints, reports and patrols, etc... The Priority 3 calls make up the remaining 33.7% or 1,156,145 calls for service in the sample.

Some agencies gave me detailed descriptions of the process by which they assign priority codes to calls and what the codes were meant to indicate. In these cases, mapping the individual priority codes into the universal system was trivial. Other departments only told me the ordering of highest priority to lowest priority, which made determining the mapping more difficult. In the difficult cases I determined the mapping using what I called "cut off" calls, or calls that are almost never in certain priority codes. To distinguish between codes that should map to priority 2 and priority 3, I mostly used nature codes relating to fraud, since it is almost always reported after the fact and without a suspect. To distinguish between priority 2 and priority 1 calls I used nature codes that had "in progress" as part

of the description, since crimes in progress almost always require an immediate response.⁵ I never assign different calls with the same original priority code assigned by the reporting police department to two different priority codes in the universal system. It is important to emphasize here that, two agencies receiving the same call may classify it differently. For example, one department receiving a call complaining of a parked car blocking a driveway may classify the call as a priority 2 call while another would classify it as a priority 3 call. Therefore the universal priority code system controls for the severity of the call **as perceived by the responding agency**.

Table 2 shows summary statistics of the raw response time broken out by the priority of the call. As expected higher priority calls have shorter mean and median response times, but notice that the mean is quite a bit larger than the median across all priority groups and that the standard deviation is also quite large. Table 2 shows the summary statistics with the extreme outliers, that are likely errors, removed the sample. The large means are driven by long right tails in the distribution of response times across all priority codes. To get a better idea of the entire distribution, Figure 2 shows the CDF of the raw response times broken out by priority code. Reassuringly, the CDF of priority 1 calls lies almost entirely to the left of priority 2 calls which lie almost entirely to the left of priority 3 calls. Meaning the probability of a response time being less than m minutes is greater for a priority 1 call than a priority 2 call and greater for a priority 2 call than for a priority 3 call. Surprisingly, for very small values of m this is not the case, and the probability of a response time being less than 1 minute is higher for priority 3 calls than for priority 1 calls. It's hard to know why this would be the case, but there are several possible reasons. Crimes in progress may often pose a threat to the responding officers' safety. An officer responding to a priority 1 call should take a moment to evaluate the situation and ensure their own safety before rushing to intervene. Similarly, it is stressful to operate in dangerous conditions, and officers may compete with each other to respond to lower priority calls that may be more pleasant to

⁵The exception being some misdemeanors in some areas.

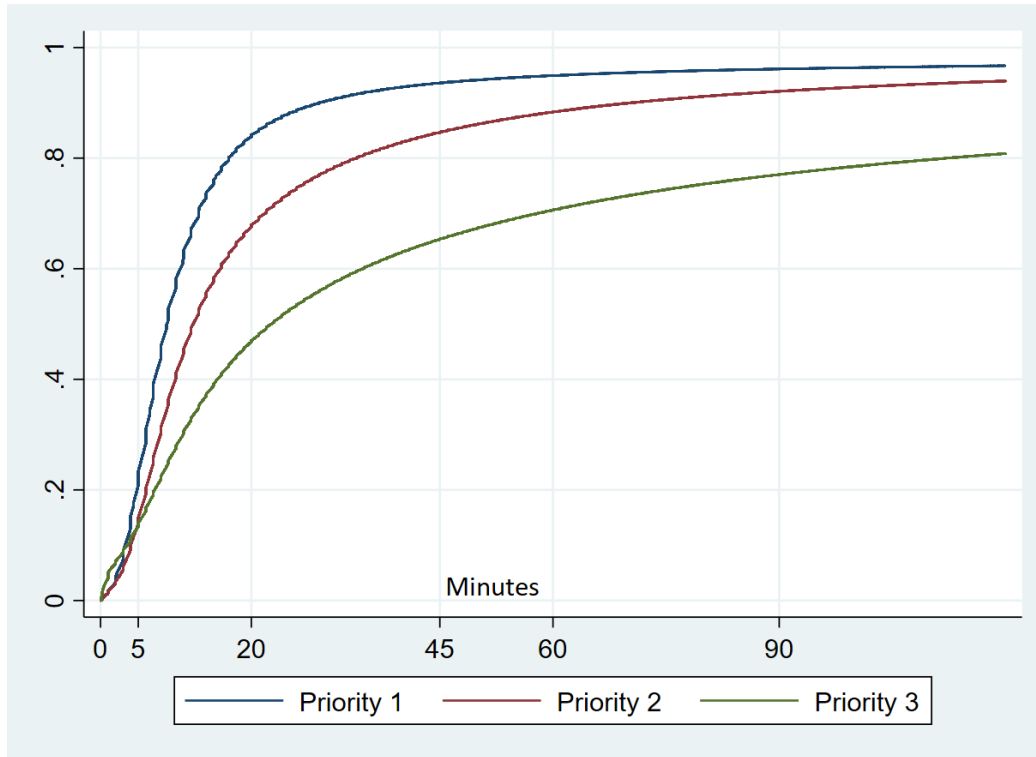
Table 2: Summary Statistics for Call Data

	Priority 1	Priority 2	Priority 3
Mean	24.8	37.8	94.3
Median	8.8	12.3	22.7
Std. Dv.	81.5	93.2	175.9

notes: Response time is measured as the time between the call being received and the first officer arriving on scene in minutes.

handle. That competition could result in shorter response times for lower priority calls in some situations.

Figure 2: Cumulative Distribution Function of Raw Response Time in Minutes by Priority



The remaining two well known sources of data are the UCR, and the American Community Survey (ACS). The UCR is a monthly survey administered by the FBI. Agencies self report crimes reported to them each month using the uniform definitions of crimes provided by the FBI. The UCR includes some information about each reporting agency, including an estimate of the population living in each agency's jurisdiction and the number of sworn

officers and civilians employees. The HHI that I use to measure competition is calculated from the population estimates, and the number of officers and civilian employees per 1000 residents are included as controls in my regression analysis.

The ACS is an ongoing yearly survey conducted by the Census Bureau. It provides detailed demographic and economic characteristics of households at multiple different geographic levels. I utilized GIS software to spatially match each call in the CAD data to the census block group containing the address to which police were dispatched. This allowed me to match the characteristics of the census block group from the ACS to each call. Census block groups are geographies in between census blocks and census tracts, and are the smallest geographic unit for which the Census Bureau publishes publicly available estimates. Typically they contain about 600 to 3,000 residents. For small geographic units like block groups, the census only makes available estimates that are derived from five years worth of survey data to protect the privacy of the respondents. I use the 2016 five year estimates, which are appropriate in this context, since my research does not attempt to measure changes in demographics. Table 3 contains summary statistics of the variables from the ACS I use in the regression analysis. These variables are measured as a share of the census block group’s population.

Table 3: Summary Statistics of Demographic and Economic Variables of Census Block Groups in Sample

	Mean	Std. Dev.
Share White	0.59	0.34
Share Married	0.36	0.21
Share Below Poverty Line	0.21	0.18
Share Unemployed	0.07	0.06
Share of Housing Vacant	0.15	0.14
Share of Householders Renting	0.44	0.26

notes: “White” is the share of the census block group reporting “white only” as their race. “Married” is the share of the population that is living in a married household. “Poverty” is the share of the population living below the federal poverty line. “Unemployed” is the share of the working age population who are unemployed. “Vacant” is the share of all housing units that are unoccupied. “Renting” is the share of households living in rented units as opposed to living in a unit they own.

III Methodology

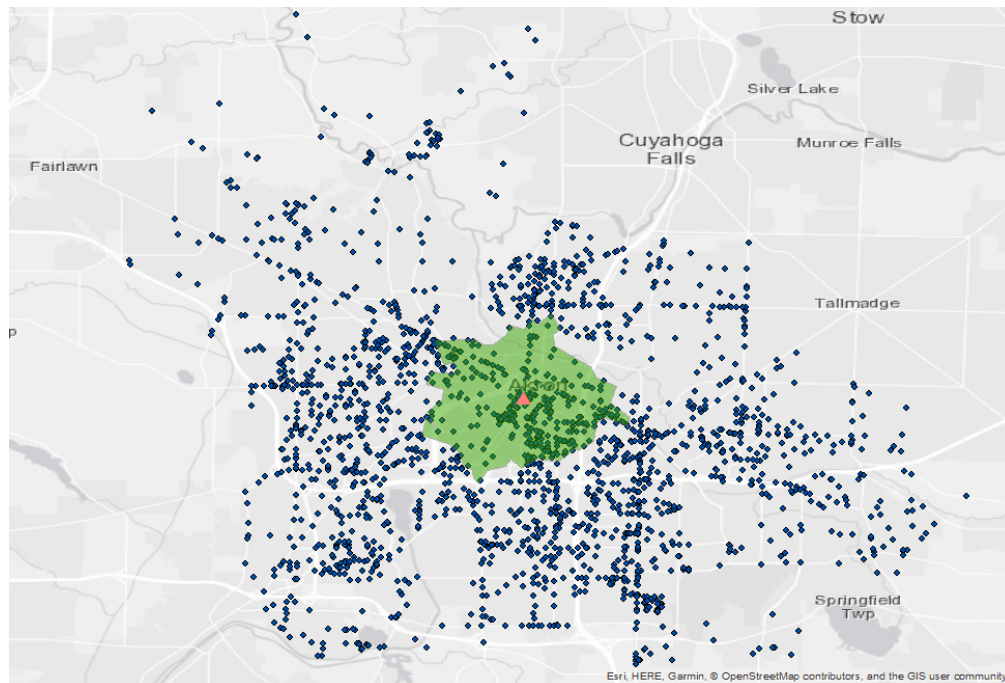
A GIS Analysis

One of the key challenges in comparing response times across jurisdictions is the differences in geographic characteristics different agencies face. Not only are there differences in physical distances, road networks, and traffic conditions, but also in the distribution of call locations. Two jurisdictions may cover similar geographic areas, but if the call locations are concentrated in one area in one jurisdiction and spread out evenly in the other jurisdiction, comparisons of raw response times could be misleading. To control for the geographic differences I calculate an optimal placement of police response units given the distribution of call locations using a Maximum Covering Model (MCM) with capacity constraints. From the optimal placement, I calculate a predicted response time, which represents the travel time faced by an optimally placed response unit. I subtract the prediction from the actual response time to get a residual response time which controls for these confounding factors and is plausibly comparable across jurisdictions.

The MCM is a special case of a large class of problems called location allocation problems. In my specification, the problem is to choose locations for police response units to maximize the number of calls within a specified travel time of at least one response unit, subject to the constraint that each response unit can only answer a limited number of calls. Figure 3 shows an example from January of 2016 in Akron Ohio. Blue dots in the figure represent CAD events, while the red triangle shows a potential location for a response unit. The green polygon surrounding the triangle shows the locations that are within a 3 minute drive time of the potential response unit. If Akron PD had only one response unit to place, the objective would be to position the red triangle such that the maximum number of points are within the green polygon. The maximum covering model has been studied extensively, and it has been used to plan emergency services in many different contexts, including the placement of

emergency sirens in Current and O’Kelley (1992) and the placement of ambulances in rural areas in Adenso-Diaz and Rodriguez (1997). Curtin, Hayslett-McCall and Qiu (n.d.) use a varriant of the MCM to redesign the patrol areas of the Dallas Police Department, arguing that the MCM gives police a quantitative and objective tool for designing and evaluating effective patrol areas.

Figure 3: Akron, Ohio. CAD Events Created Between 12:00AM and 7:59AM in January of 2016.



notes: Blue dots indicate the location of CAD events. The red triangle is a potential place a response unit could locate. The green polygon surrounding the red triangle indicates the locations that are within a 3 minute drive time of the potential response unit.

It is important to note here that the MCM is an imperfect description of the police’s “true” objective function. Police might put more importance on placing units to provide each other with backup. The police might weight certain calls differently so that covering a priority 1 call is more valuable than covering a priority 3 call. A department might try to minimize the total, average, or median travel time to call locations, and there is no obvious a priori reason to prefer one of these objective functions over the others. In addition to these issues, some police functions have nothing to do with response times at all. For example, efficient

police departments keep good records, but dedicating officer time to record keeping and filing reports prevents those officers from using that time to answer calls for service. Despite these limitations, The MCM is a useful benchmark that is actually used in determining the disposition of emergency services, even if it is not a perfect description of police objectives. Using a richer model that could better address the multiple different facets of police objectives or generate more accurate predicted response times are potentially fruitful ways of extending this research.

I assume that police re-optimize the location of their response units every six months using one month’s worth of data. Allowing the police to re-optimize is critical, because not only do police seek to locate optimally to respond to calls for service, but criminals may also try to locate their activities to avoid detection. This endogeneity of the call locations can’t be perfectly controlled for with the available data, but the re-optimization mitigates the impact. Moreover, while some types of crimes and criminals are highly mobile, a significant share of calls, such as domestic disputes, are unlikely to change location in response to police deployments.

In addition to the biannual re-optimization, the police also optimize for two different “shifts.” A graveyard shift from midnight to 8 am, and a day shift from 8am to midnight. There are two main reasons for optimizing over different times of day. The first is that it’s a closer approximation to the actual operations of the departments. The hours between midnight and 8 am experience the lowest volume of calls for service, and most departments respond by staffing fewer units during those hours. Unsurprisingly, every department is different, and there is a large amount of variation in how shifts are scheduled and the staffing model the agencies use, but this is a practical approximation that captures what is typically the largest factor in time varying staffing. The second reason for splitting the data into multiple shifts is a practical limitation on the number observations that can be handled by ArcGIS. Splitting the data into shifts greatly speeds up the optimization, and the four

largest departments in my sample; New Orleans, Detroit, Tucson, and Wichita, have to be broken up into 4 six hour shifts to accommodate the large number of CAD events.

All of the GIS analysis was performed using ESRI's ArcGIS software. The analysis proceeded in three steps. 1; geocoding the call locations from their text addresses. 2; calculating an optimal placement using the MCM. 3; generating predicted response times from the solution to the MCM.

Geocoding Geocoding is the process of assigning XY coordinates to text based addresses. Addresses in the CAD data were geocoded using ESRI's US Address Locator, based on their 2013 North American Street Map data. Police data is typically formatted in ways that are difficult for the locator to analyze. I reformatted the addresses to increase the locator's accuracy. This mostly consisted of replacing the delimiters for intersections and repositioning some directional indicators on street names. Even with high quality data the share of calls successfully located, or the match rate is virtually always less than 100%. There are a number of reasons the locator may fail to find an address. Alternative spellings, changing street names, and a lack of precise location indicators for some streets, particularly highways, all contribute to imperfect match rates. In my sample the match rate varies among agencies. The highest match rate I obtained for any department was 95%, while the lowest was 85%. The combined match rate of all the jurisdictions is 86%. Low match rates are a potential cause for concern because the matched calls may not accurately represent the true distribution of locations. Ratcliffe (2004) however, uses Monte Carlo simulations to estimate that for police data 85% is an acceptable match rate when geocoding crime locations.

Calculating an Optimal Placement The MCM requires the researcher to specify four parameters. They are the number of response units to locate, the potential places in which response units can locate, the capacity constraint of the response units, and finally the travel time cutoffs determining if a call is covered.

I estimated the number police response units to locate directly from the CAD data. Although the UCR contains information on the staffing levels of police departments, the number of officers and even the number of patrol cars can be bad indicators of the actual resources available to respond to calls for service. Not all officers are assigned to patrol duties, and even official schedules do not always reflect the actual number of officers as people take time off, get injured, or attend trainings. Most officers ride by themselves, but some departments put multiple officers in some or all of their cars, and many calls require two or more officers to respond.

For each department and shift, I estimated the number of response units to locate as the number of calls a department can respond to simultaneously. It's easy to think of these response units as cars to be placed, but since many calls require multiple officers to respond, what I am estimating is more accurately described as response units rather than cars. First, I defined "wait time" as the time between a call being received by the CAD system and an officer being dispatched. I calculated the distribution of wait times and then identified the calls that had abnormally long ones. I defined abnormal as the 75th percentile. Among these calls with abnormally long wait times I then identified the priority 1 calls. There are a number of possible explanations for why a call would have a long wait time. Sometimes there are language barriers or the dispatcher is unable to locate the caller, but the most likely reason that calls requiring an immediate response had abnormally long wait times is that there were no units available to respond. In other words, the department was at capacity.

For each high priority call with an abnormally long wait time, I calculated the number of "open calls" at the time the call was received. Specifically, the number of calls for which an officer had been dispatched, but the call had not yet been cleared from the CAD system. There is a distribution of open calls so there are a number of statistics that could potentially be used as an estimate. Since abnormally long wait times are more likely to occur when a particular shift is understaffed, the median number of open calls is probably too small of an

estimate. The maximum is typically also a bad indicator, since the highest numbers of open calls tend to correspond to natural disasters or large public events for which all or almost all of the department's officers are put on duty. I use the 75th percentile of open calls as the number of response units to place.

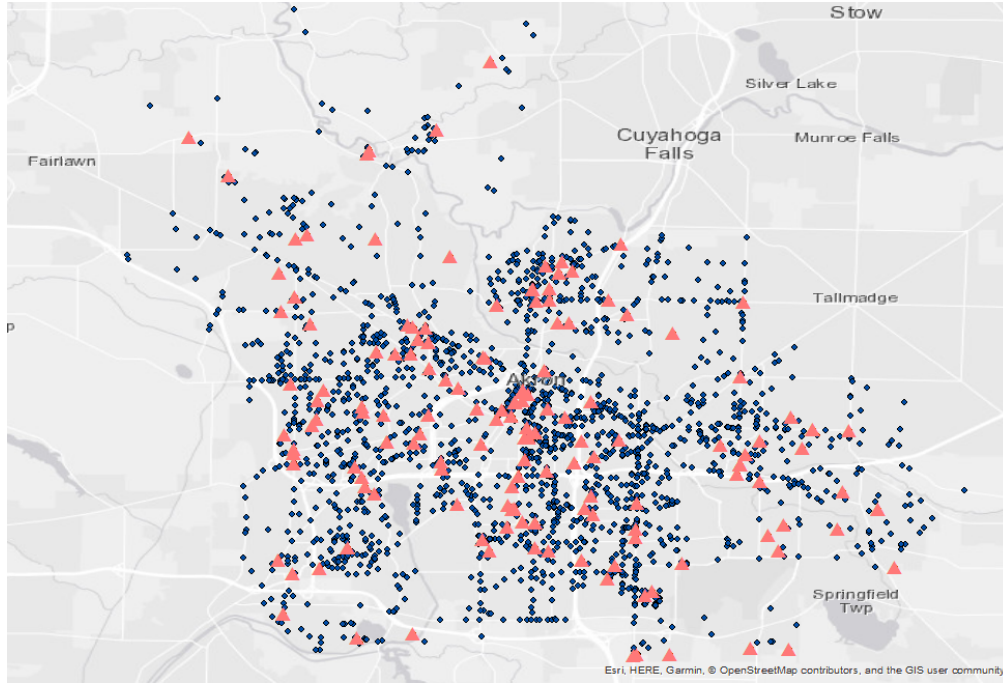
Next the researcher has to determine the potential places in which the response units can locate. The difficulty in selecting potential places to locate is the trade off between tractability and coverage. Technically the police could locate on any inch of the road, but the practical limits of the software make analyzing such a set of potential locations impossible. On the other hand, limiting the number of places where police can locate may cause the researcher to inadvertently remove the optimal location. With this trade off in mind, I choose to use a 5%⁶ random sample of actual call locations as potential places to locate units. Since the optimal solution is likely close to where the calls are located, the actual locations of the calls is a good place to begin looking. This allowed me to limit the number of potential places to locate and still maintain a set of points that are likely to contain good solutions. Figure 4, shows the potential places to locate for Akron Ohio in our example.

I estimate the capacity constraint from the average amount of time between a call being received and a call being cleared. I calculate the number of calls that could be answered per hour with the calculated average call duration, and then multiply by the number of hours the response unit could spend answering calls.⁷ I use this as the capacity constraint for each response unit. The capacity constraints are important because without them the response units would tend to be more spread out than they otherwise should be.

⁶For larger jurisdictions a 5% random sampled still resulted in too many potential places for the software to be able to run in a reasonable mount of time. In these cases a 2.5% random sample was used. Similarly, in some of the smaller jurisdictions a 10% random sample was used ensure enough coverage of potential places to locate.

⁷Kitsap county in Washington did not provide me with the time calls were cleared from the CAD software. I used the estimate from a set of Agencies in King County for the capacity constraints in Kitsap.

Figure 4: Akron Ohio, Graveyard Shift of January 2016, Potential Places to Locate.



notes: Blue dots indicate CAD event locations. Red triangles indicate potential places for police to place response units.

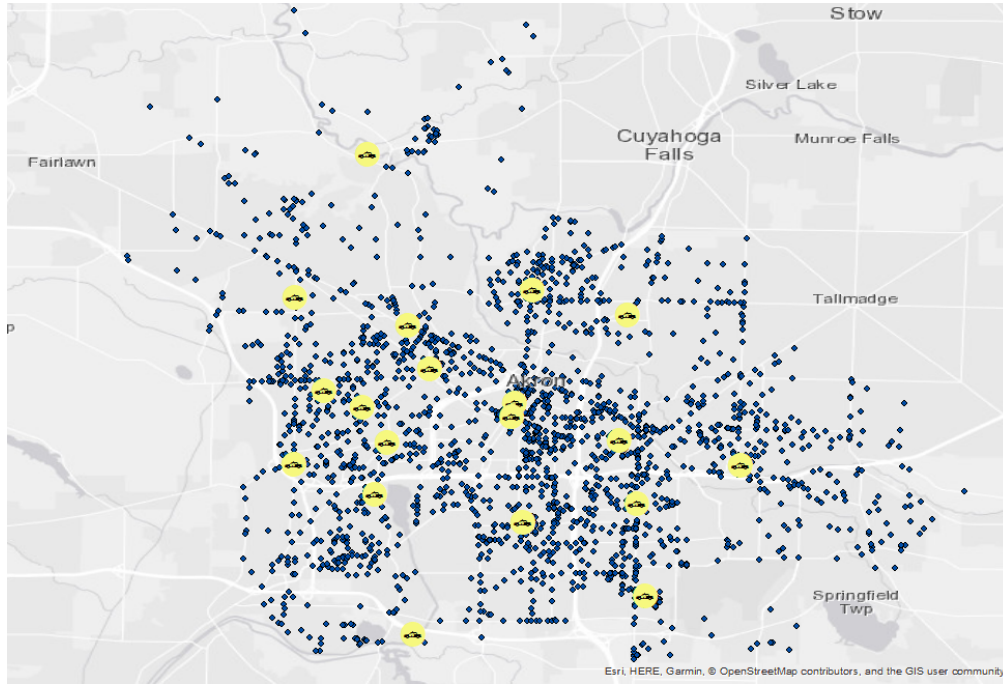
The most difficult parameters to specify were the cutoff times for covered calls. There are no national guidelines for acceptable response times to emergency calls, let alone routine calls. Furthermore there are no federal, state, or local statutes requiring police to meet any standards for response times. In the MCM model the resources available to the department are treated as given. A hypothetical police planner would specify the cutoffs based on the benefit of arriving within a certain time frame, and since that benefit should be the same whether the call happens in Rhode Island or Washington I use the same cutoff for all departments. To determine the cutoffs I examined published guidelines from a few departments that made them available and consulted with some of the analysts who provided me with CAD data. I chose cutoffs that reflected the approximate average of the objectives of departments who were willing to discuss them with me. I use cutoffs of 5 minutes for priority 1 calls, 20 minutes for priority 2, and 45 minutes for priority 3.

There are certainly alternative values for the cut offs that are also plausible. An analysis of the sensitivity of the results to the cutoff times is appropriate in this circumstance, but unfortunately the extremely time consuming nature of the optimization process prevents extensive sensitivity analyses. A positive side effect of the time constraint is that specification searching is also infeasible, but the sensitivity analysis I perform is necessarily limited. I provide an example in one jurisdiction to give an indication of how much the results may change when varying the cutoffs. Continuing with the graveyard shift of January 2016 in Akron Ohio, I solved the MCM using alternative cutoff times for covered calls, specifically I reduced them by 40% to 3, 12, and 27 minutes for priority codes 1, 2, and 3 calls respectively. I then calculated the predicted response times for CAD events and compared them to the base case. The sum of all predicted response times increases by 14% when the cutoffs are reduced. This is predominantly caused by the MCM choosing more central locations in higher call density areas, which causes fewer calls on the outskirts of the jurisdiction to be covered. The median absolute difference in predicted response times between the two sets of cutoffs was 32 seconds. The fact that a relatively large change in the cutoff times leads to a relatively small change in the predicted response times should give us some confidence that the results of the analysis may not be especially sensitive.

Since the set of potential places to locate is finite, the optimum could hypothetically be found by brute force, however the combinatorial complexity of the problem necessitates the use of heuristics. First an origin-destination matrix, calculating the shortest time it takes to travel between each potential response unit location and call location, is created. The matrix undergoes a process of editing due to Hillsman (1984) after which a semirandomized set of solutions is generated and a Teitz-Bart vertex substitution heuristic is used to refine these solutions. A metaheuristic combines solutions to obtain a set of better solutions, iterating this process until convergence.⁸ The results are a set of locations for police to place their

⁸These are the default algorithms utilized by ArcGIS Network Analyst Extension. "Algorithms used by the ArcGIS Network Analyst extension," ArcGIS handbook.

Figure 5: Akron Ohio, Graveyard Shift of January 2016, Chosen Locations for Police Response Units



notes: Blue dots indicate CAD event locations. Yellow cars indicate optimal placement of police response units.

response units.

Calculating Predicted Response Times The predicted response times are calculated as the travel time in minutes between the nearest response unit and the call's location. The prediction is subtracted from the observed response time to get a residual response time. Note that the predicted response times are calculated as if the covering unit could be dispatched immediately, and as if the responding unit was always being dispatched from that optimal location. This is a simplification since police patrol rather than park in one spot waiting for a call. The predicted response time should be thought of as an average travel time based on the center of the patrol area covered by the response unit. While the residual response times control for the differences in geography they do not control for how busy the department was at the time the call was received. This has to be controlled for in the two stage least squares regressions. Table 4 shows the summary statistics of the residual

Table 4: Summary Statistics of Residual Response Times in Minutes

	Priority 1	Priority 2	Priority 3
Mean	22.9	35.5	90.1
Median	6.8	10.4	20.4
Std. Dv.	82.0	92.8	173.2

notes: Observed response time less predicted travel time in minutes.

response times in minutes, by priority.

B Estimation

I estimate the impact of fragmentation and community characteristics on police efficiency by regressing the residual response times on the HHI and the demographic characteristics of the census block group to which police were dispatched. I use a two stage least squares estimator due to the possibility of endogeneity in the HHI. In a 2006 Justice Department survey⁹ of 165 newly formed police departments, 68% cited wanting to improve response times as a reason for establishing their own police departments. If the respondents were reporting truthfully, then not only does the degree of fragmentation impact response times, but response times also help determine the degree of fragmentation. This implies that Ordinary Least Squares estimates will be biased and inconsistent.

Following in the spirit of Levitt (2002), who uses the number of firemen in a municipality to instrument for the number of policemen in estimating the impact of police on crime, I use the number of fire departments in an MSA, calculated from the National Fire Department Registry, to instrument for the market concentration of police. The logic of the instrument is very simple. Many of the factors that impact the fragmentation of police also impact the fragmentation of fire departments. Local government officials tend to prefer more control over less control of public goods. This is especially true with regards to police because in areas without police departments county sheriffs offices are responsible for providing law

⁹Spence, Webster and Connors (2006)

enforcement. Sheriffs are elected officials, while police chiefs are typically appointed by a mayor or city council, so creating a police department greatly increases the amount of control these officials can exercise. Additionally, since police jurisdictions have to follow the boundaries of the municipalities, county, or state that gives them authority, the accidents of history and geography that have led to the fragmentation of local governments also contribute to the fragmentation of police departments. These same preferences for local control and exogenous variation in municipalities impacts fire departments in the same way they impact police departments. Empirically, the number of fire departments and police departments have a high degree of correlation giving strong evidence that fire departments fulfill the inclusion restriction for identification.

The exclusion restriction is satisfied because in general municipalities do not start a new fire department when the police have poor response times. It is not immediately obvious that the exclusion restriction holds in every case. In many municipalities emergency services share or coordinate dispatchers, which makes it likely that the response times of fire departments and police departments are correlated. If the number of fire departments is also endogenous to the fire department's response times the exclusion restriction might not be satisfied. However, as long as the correlated part of fire and police response times, namely the efficiency of the dispatchers, does not cause municipalities to establish or dissolve police and fire departments the exclusion restriction will be satisfied. Since modern technology makes it easy to distinguish between the operations of the dispatchers, the police, and the firemen, the exclusion restriction is likely to hold.

IV Results

Table 5 shows the main results of the two stage least squares regressions. I run a separate regression for each priority code. The effect of fragmentation, which is measured by police market concentration, on residual response times is reported in the first row of the table.

In each regression the effect of concentration on residual response times is statistically significant at the 1% level. Surprisingly, the effect has a different sign across the priority codes. Police departments in more concentrated MSA's have faster residual response times for priority 1 calls, but slower residual response times for priority 2 and 3 calls. This contrasts with Wheaton (2006), whose research suggests that increased competition has only positive impacts on the efficiency of police. One possible explanation for the different signs is a potential trade off between efficiently responding to priority 1 calls and lower priority calls. Tiebout competitive pressure on police departments in fragmented areas could cause departments to focus and expend more effort on "quality of life" calls that tend to have lower priority codes. Alternatively, the difference may be caused by different advantages of centralized and decentralized policing affecting different kinds of calls. For example, suppose that there are economies of scale that more concentrated areas can take advantage of, and that those economies of scale mostly improve the ability of police to respond effectively to high priority calls. Furthermore suppose that, as Ostrom and Whitaker (1973) suggest, more fragmented departments are better able to engage with communities in a way that improves their ability to respond effectively to low priority calls. This would generate the same pattern of concentration improving response times for high priority calls and worsening response times for low priority calls. The exact mechanism that produces these results is an interesting question for future research.

The size of the HHI's effect is quite large across all call priorities. A one standard deviation increase in the HHI leads to an almost 4 minute improvement in the residual response times for priority one calls. Relative to the mean shown in Table 4, that translates to about a 17% improvement. Relative to the median it's an almost 45% improvement! The magnitude of the effect for priority 2 and priority 3 calls is similar although in the opposite direction. Since these calls have longer residual response times however, as a percentage they represent a smaller change. Approximately 8% and 11% of the means for priority 2 and priority 3 calls

Table 5: Two Stage Least Squares Regressions of Residual Response Time, by Priority

	Priority 1	Priority 2	Priority 3
HHI	−0.003** <i>0.001</i>	0.002** <i>0.001</i>	0.007** <i>0.001</i>
White	10.580** <i>1.726</i>	−15.856** <i>1.995</i>	−15.889** <i>4.861</i>
Married	−3.073 <i>4.170</i>	−2.792 <i>5.507</i>	2.866 <i>11.176</i>
Poverty	11.736** <i>2.627</i>	8.408** <i>3.520</i>	23.516** <i>7.179</i>
Unemployed	−0.502 <i>7.123</i>	−28.185** <i>9.790</i>	−6.164 <i>22.371</i>
Vacant	−2.955 <i>3.144</i>	−7.276** <i>5.280</i>	−24.480** <i>10.050</i>
Renting	−7.622** <i>2.204</i>	−12.312** <i>2.302</i>	−27.330** <i>5.272</i>
Officer Emp	−2.325** <i>0.114</i>	−0.083 <i>0.099</i>	−3.162** <i>0.235</i>
Civilian Emp	2.158** <i>0.299</i>	−0.505 <i>0.554</i>	8.360** <i>1.263</i>
Call Volume	1.271** <i>0.026</i>	1.793** <i>0.032</i>	3.408** <i>0.083</i>
Hourly Max Temp	0.189** <i>0.028</i>	0.0170 <i>0.055</i>	0.627 <i>0.094</i>

notes: Point estimates are listed first with standard errors in italics beneath them. * and ** indicate statistical significance at the 5% and 1% levels respectively. The demographic variables are measured as a share of the census block group. The notes of Table 3 provide more detailed descriptions. “Officer Emp” and “Civilian Emp” are the number of sworn officers and civilians employed by the police department per 1000 residents in their jurisdiction respectively. “Call Volume” is the number of calls received in the same hour as the call, and “Hourly Max Temp” is the maximum dry bulb temperature measured from the nearest NOAA station. A constant and monthly fixed effects are included in the estimation, but not reported.

respectively.

The effects of demographic variables shown in Table 5 are also statistically significant in many instances. However, since these variables are measured as a share the impact on response times is smaller than a cursory glance would suggest. For example, a 10 percentage point increase in the share of households living in renter occupied housing decreases residual response times by about 45 seconds. Cihan, Zhang and Hoover (2012) find broadly similar results in the sense that our estimates of community characteristic effects tend to have the same sign, although direct comparisons are difficult due to the differences in methods we employ.

Interestingly, across all priority codes the share of households living in renter occupied housing improves residual response times, and the effect is particularly large for priority 3 calls. In other words communities with more renters can expect faster response times across all priority codes. If Tiebout style competition for residents is a mechanism for determining response times, then police may choose to dedicate more resources to relatively more mobile residents like those who rent instead of own their housing.

The effect of the racial composition is of particular interest. The effect is statistically significant across all priority codes and while for priority 1 calls the effect is positive, for priority 2 and priority 3 calls the effect is negative. A 10 percentage point increase in the share of a block group that is white increases residual response times for priority 1 calls by about one minute and decreases residual response times by about one and half minutes for both priority 2 and 3 calls. The implication that minority communities receive better service in some circumstances, is in some sense surprising. Although Lee, Lee and Hoover (2016) also find that for domestic violence calls black complainants can expect police to respond faster, in many other contexts discretionary police actions have been shown to harm minorities. In just one example, a recent meta-analysis by Chochel, Wilson and Mastrofski

(2011) of 40 published and unpublished works of research found that non white suspects had an almost 30% higher chance of being arrested than white suspects. The fact that domestic violence calls can span all three priority codes depending on other circumstances of the call further emphasizes the need for careful nuance in evaluating police performance.

While the average effects of demographic characteristics are important there is considerable heterogeneity across jurisdictions in the size of those effects. I estimate ordinary least squares regressions of residual response time on community characteristics for each jurisdiction to explore the degree of heterogeneity. Table 6 shows summary statistics of the estimated coefficients from these regressions. Notice that in every case the standard deviation is much larger than the mean, in most cases by a factor of three or more. While the sign of the means mostly correspond to the estimates from the combined regression, in every case there are estimates on both sides of zero.

Table 6: Summary Statistics of Estimated Coefficients

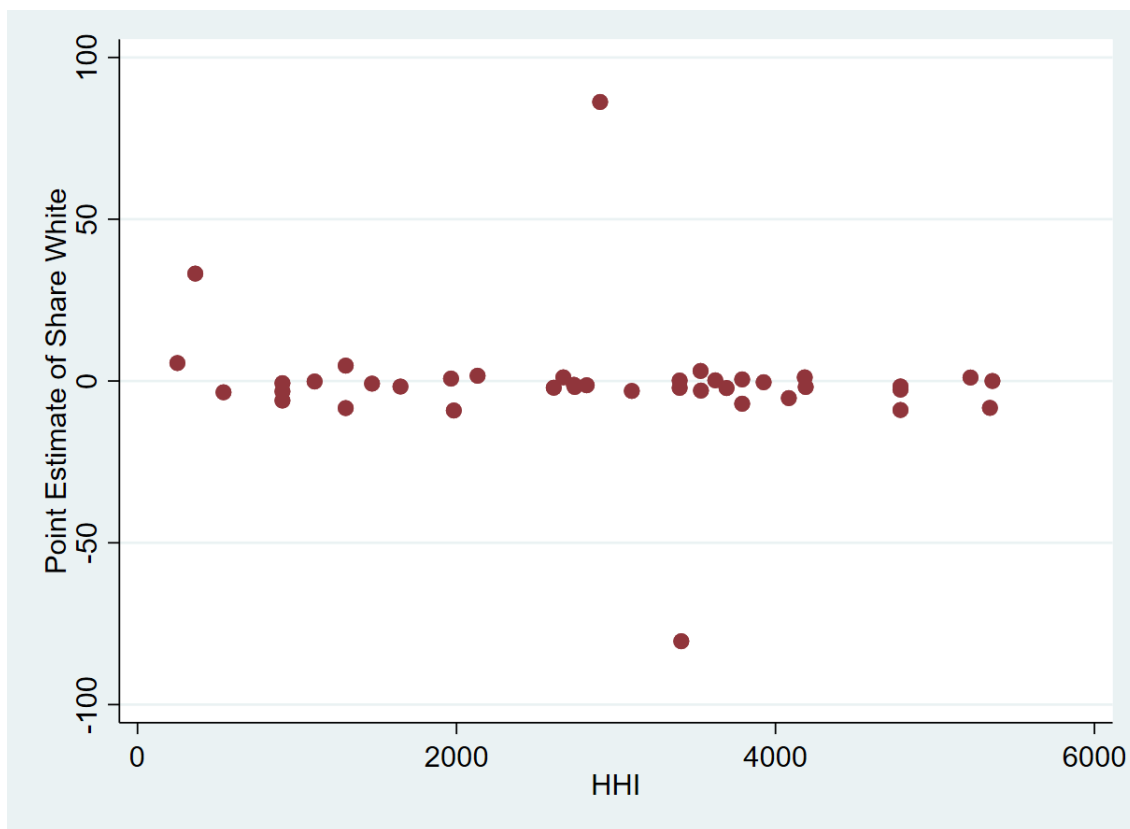
	Priority 1		Priority 2		Priority 3	
	Mean	Sd	Mean	Mean	Mean	Sd
White	-0.68	19.97	-6.69	22.52	-7.28	21.20
Married	3.03	16.33	1.38	8.18	6.64	32.88
Poverty	-2.44	11.93	-1.40	8.24	-1.29	29.09
Unemployed	-5.75	27.88	-5.64	28.34	-26.33	91.20
Vacant	-4.98	16.87	-0.88	8.03	-4.16	19.77
Renting	1.37	15.24	-2.27	6.75	-3.89	21.55
Call Volume	0.64	1.11	1.10	1.14	1.85	1.55
Hourly Max Temp	-0.01	0.10	-0.01	0.11	-0.01	0.21

notes: The mean is the simple average over jurisdictions of the point estimates or coefficients. "Sd" is the standard deviation of the distribution of means. It is **not** the mean of, or related to, the estimated standard errors of the coefficients.

To further illustrate the degree of heterogeneity consider the effect of race. Figure 6 shows for priority 1 calls the estimated coefficient of the share of a block group that is white graphed against market concentration for all 40 police departments in the sample. Notice that for all but three of the estimates, a 10 percentage point increase in the share of white residents corresponds to changes in residual response times of less than one minute. In addition to

the small magnitudes, among these estimates only 8 are statistically significant at the 95% level. Figure 7¹⁰ shows these estimates. Even among these estimates there is a good amount of variation, and both positive and negative effects are observed. This strongly suggests that the effects of demographics may be very specific to particular jurisdictions. Similar patterns are observed when examining priority 2 and priority 3 calls, although a larger, but still minority, share of jurisdictions have statistically significant estimates.

Figure 6: Estimated Coefficient on Share White for Priority 1 Calls: All Jurisdictions

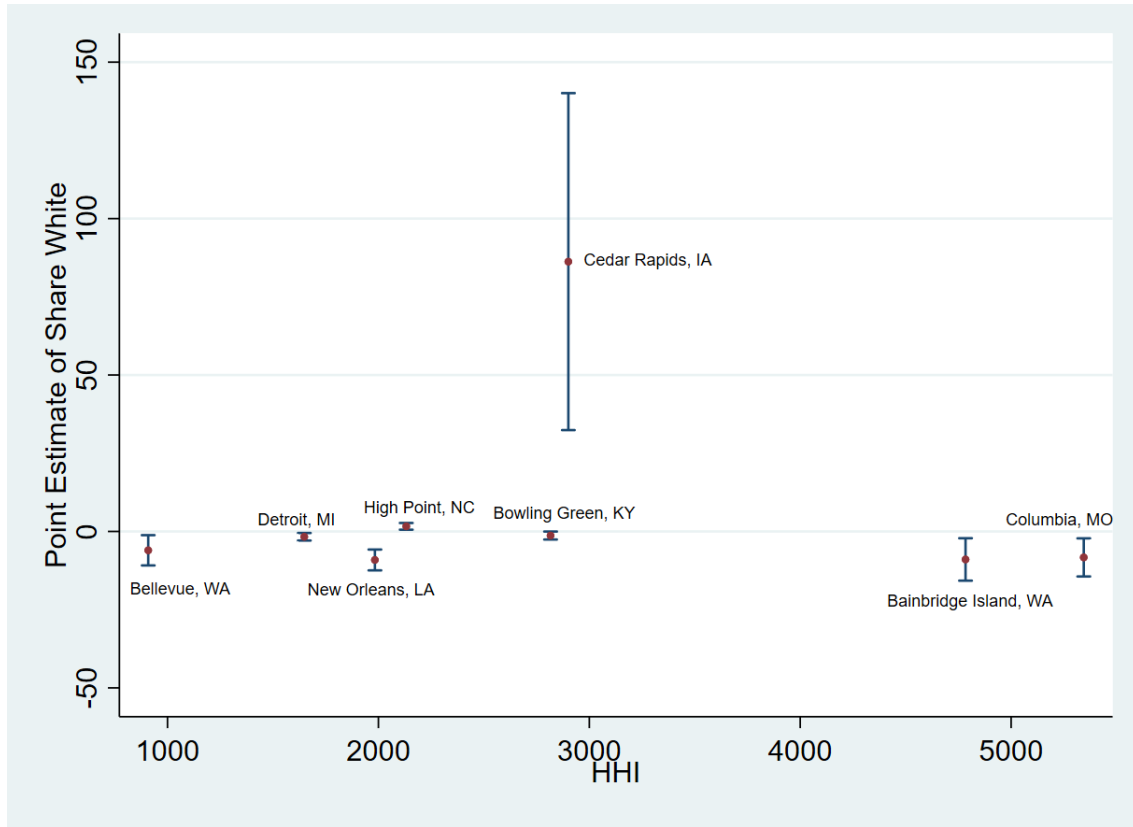


notes: The coefficient is estimated from an ordinary least squares regression of residual response times on community characteristics, call volume, hourly temperature, and month fixed effects.

In addition to the dispersion of the point estimates, the figures also show that there is a slight negative correlation between the point estimates and the level of police concentration. Communities with a larger minority population can expect longer residual response times

¹⁰The reader should note that the axes of Figure 6 and Figure 7 do not use the same scale.

Figure 7: Estimated Coefficient on Share White for Priority 1 Calls, Statistically Significant



notes: The coefficient is estimated from an ordinary least squares regression of residual response times on community characteristics, call volume, hourly temperature, and month fixed effects.

in more concentrated areas. This contrasts with some of the common wisdom that more centralized law enforcement would be less prone to racial discrimination. To engage in some speculation, it could be the case that competition among jurisdictions encourages police to operate more equitably towards minority communities. This observation however, must be interpreted with extreme caution, since the correlation is very small, nowhere near statistically significant, and possibly not a causal relationship. It does however reinforce that there is little evidence that centralizing police departments would reduce discrimination.

V Conclusion

This paper represents a step towards answering two fundamental questions about police operations. How does the decentralized provision of police services effect their efficiency, and how does the quality of police service differ based on the demographic characteristics of different communities? Answering these questions requires a measure of police performance that is objective and comparable across jurisdictions. To make a measure that plausibly satisfies these requirements, I collected the CAD data of 40 different agencies and used a maximum covering model with capacity constraints to calculate residual response times for approximately 3.4 million calls for service received by these agencies in 2015 and 2016.

Using a two stage least squares estimator to account for the possible endogeneity of market concentration, I find that concentration has relatively large and statistically significant impacts on residual response times across all priority codes. A 1 standard deviation increase in market concentration leads to an almost four minute decrease in residual response times for calls requiring an immediate response. In some cases 4 minutes could quite literally be the difference between life and death. While increased concentration improves residual response times for priority 1 calls, it simultaneously worsens residual response times for lower priority calls.

The effects of demographic characteristics are also significant determinants of residual response times. While the signs of the effects tend to be consistent across priority codes, the effect of race is a major exception. Communities with a higher share of white residents can expect longer residual response times for priority 1 calls, but shorter residual response times for lower priority calls. While it is impossible with the available data in this study to determine exactly why this is the case, observed differences in the quality of service that correlate with the racial make up of communities are a serious concern. However, analysis of individual departments show that there is a large amount of variation in the effects

of demographic characteristics on residual response times among different jurisdictions. Although on average police respond faster to communities with a higher minority population for priority 1 calls, in the majority of jurisdictions the effect has the opposite sign. Furthermore, in only 8 of the 40 jurisdictions is race a statistically significant predictor of residual response times. Similar observations are true across all priorities.

Perhaps the most important finding of this research is that many characteristics like the degree of centralization can have heterogeneous impacts across different kinds of police activities. There is a tendency in some research to oversimplify the relationships among various factors and outcomes, but law enforcement is a complicated combination of multiple public goods. The findings in this paper will have to be further explored to understand the mechanisms behind the effects, and derive more concrete policy implications.

References

- Adenso-Diaz, Belarmino, and F. Rodriguez.** 1997. "A Simple Search Heuristic for the MCLP: Application to the Location of Ambulance Bases in a Rural Region." *Omega*, 25.
- Bandyopadhyay, Subhayu, Santiago M. Pinto, and Christopher H. Wheeler.** 2011. "Urban Crime and Labor Mobility." *Journal of Public Economic Theory*, 13(3).
- Chochel, Tammy R., David B. Wilson, and Stephen D. Mastrofski.** 2011. "Effect of Suspect Race on Officers' Arrest Decisions." *Criminology*, 49.
- Cihan, Abdullah.** 2014. "Social Disorganization and Police Performance to Burglary Calls: A Tale of Two Cities." *Policing: An International Journal of Police Strategies & Management*, 37.
- Cihan, Abdullah, Yan Zhang, and Larry Hoover.** 2012. "Police Response Time to In-Progress Burglary: A Multilevel Analysis." *Police Quarterly*, 15.

- Current, John, and Morton O’Kelley.** 1992. “Locating Emergency Warning Sirens.” *Decision Sciences*, 23.
- Curtin, Kevin M., Karen Hayslett-McCall, and Fang Qiu.** n.d.. “Determining Optimal Police Patrol Areas with Maximal Covering and Backup Covering Location Models.”
- Gonzalez-Navarro, Marco.** 2013. “Deterrence and Geographical Externalities in Auto Theft.” *American Economic Journal: Applied Economics*, 5(4).
- Hillsman, Edward L.** 1984. “The p-Median Structure as a Unified Linear Model for Location—Allocation Analysis.” *Environment and Planning A*, 16.
- Howell-Moroney, Michael.** 2008. “The Tiebout Hypothesis 50 Years Later: Lessons and Lingering Challenges for Metropolitan Governance in the 21st Century.” *Public Administration Review*, 68(1).
- Hoxby, Caroline.** 2000. “Does Competition among Public Schools Benefit Students and Taxpayers?” *The American Economic Review*, 90(5).
- Knight, Brian, and Nathan Schiff.** 2012. “Spatial Competition and Cross-Border Shopping: Evidence from State Lotteries.” *American Economic Journal: Economic Policy*, 4(4).
- Lee, JaeSeung, Jonathan Lee, and Larry T. Hoover.** 2016. “What Conditions Affect Police Response Time? Examining Situational and Neighborhood Factors.” *Police Quarterly*, 20.
- Lee, Kanogh, and Santiago M. Pinto.** 2009. “Crime in a Multi-Jurisdictional Model with Private and Public Prevention.” *Journal of Regional Science*, 49(5).

- Levitt, Steven D.** 1998. "The Relationship Between Crime Reporting and Police: Implications for the Use of Uniform Crime Reports." *Journal of Quantitative Criminology*, 14(1).
- Levitt, Steven D.** 2002. "Using Electoral Cycles in Police Hiring to Estimate the Effects of Police on Crime: Reply." *The American Economic Review*, 92.
- Ostrom, Elinor, and Gordon Whitaker.** 1973. "Does Local Community Control of Police make a Difference? Some Preliminary Findings." *American Journal of Political Science*, 17(1).
- Ostrom, Elinor, Roger B. Parks, and Gordon P. Whitaker.** 1973. "Do We Really Want to Consolidate Urban Police Forces? A Reappraisal of Some Old Assertions." *Public Administration Review*, 33(5).
- Pate, Tony, Amanda Ferrara, Robert A. Bowers, and Jon Lorence.** 1976. *Police Response Time*. The Police Foundation.
- Pinto, Santiago M.** 2007. "Tax Competition in the Presence of Interjurisdictional Externalities: the Case of Crime Prevention." *Journal of Regional Science*, 47(5).
- Ratcliffe, Jerry H.** 2004. "Geocoding Crime and a First Estimate of a Minimum Acceptable Hit Rate." *International Journal of Geographical Information Science*, 18.
- Spence, Deborah, Barbara Webster, and Edward Connors.** 2006. *Guidelines for Starting and Operating a New Police Department*. U.S. Department of Justice.
- Tiebout, Charles M.** 1956. "A Pure Theory of Local Public Expenditure." *Journal of Political Economy*, 64(5).
- Vidal, Jordi B. I., and Tom Kirchmaier.** 2018. "The Effect of Police Response Time on Crime Clearance Rates." *The Review of Economic Studies*, 85.

Wheaton, William C. 2006. “Metropolitan Fragmentation, Law Enforcement Effort and Urban Crime.” *Journal of Urban Economics*, 60(1).

Wright, Glenn D., Krister P. Andersson, Clark C. Gibson, and Tom P. Evans. 2016. “Decentralization Can Help Reduce Deforestation When User Groups Engage with Local Government.” *Proceedings of the National Academy of Sciences*, 113(52): 14958–14963.