Pre-Analysis Plan for "Affluence and Influence? Examining the Extent of Non-Policy Representation in American Cities"

Brian T. Hamel

June 11, 2018

Recent research on representation has considered the extent to which policymakers are more responsive to the preferences of some income groups over others. Bartels (2008), for example, shows that legislator ideology in the Senate is more closely associated with the ideological position of high-income constituents, relative to middle and low-income earners. Gilens (2012) offers a similar take, demonstrating that aggregate-level policy change follows preferences for change among the affluent, especially when those preferences diverge from those of lower income groups. Still others suggest that the rich do not "win" more often than the middle or the poor (Bashir 2015; Branham, Soroka, and Wlezien 2017; Enns 2015; Tausanovitch 2016), often noting that detecting differential representation should is made difficult by the strong association in preferences across income groups (Soroka and Wlezien 2008).

Policy representation, however, is not the only way in which policymakers, or government more broadly, may prioritize one income group over others. And indeed, given the challenge of separating and distinguishing the policy preferences of high-income earners from low-income earners, it may be the case that examining policy output is a theoretically unimportant way to consider representation. Constituent service—a form of distributive politics—is another mode of representation. Here, we may expect to find that governments provide fewer services to poorer neighborhoods, for example, in part because poorer neighborhoods may be less likely—due to a lack of resources, for example—to officially request services (such as snow removal, or garbage pickup) from government. As indicated, examining representation from this angle may be more reasonable, as preferences—or more aptly, the propensity to lobby for government assistance— may actually diverge across income groups. As a result, differences are likely to be much more discernible for government officials. To the extent that it exists, then, reliably detecting differential representation might be more likely.

My proposals asks: do city governments provide fewer services to the poor, relative to the rich? My hypotheses are that city governments will (1) be more likely to respond to and complete service requests as the median income of a neighborhood increases; (2) conditional on response, respond to and complete service requests faster—as measured by the number of days between request and completion—as the median income of a neighborhood increases; and (3) that these differences between high and lower income areas with respect to the time to completion will be smaller as an election approaches. I will test my expectations using 311 data from Boston (to begin) . The below pre-analysis plan outlines the data and my empirical strategy for testing all three of these hypotheses using data from Boston between 2011 and 2018.

Data

Most major cities in the U.S.—and indeed, around the world—make data on constituent requests for government services readily available online. These are typically called 311 service requests. Boston, for example, updates the log of 311 service requests daily. 311 service requests can be made to report any number of non-emergencies: a broken street sign, a broken traffic signal, an out street lamp, a pothole, graffiti, etc. Requests can be made over the phone, online, or through the city's 311 smartphone app. These requests are then logged and made publicly available. As noted, my pre-analysis plan will focus on data from Boston. Boston's dataset is readily available on the city's website, and includes an exhaustive amount of information on the nature of each request for service. This includes information on the date of the request, the address where the service was requested, the type of service requested and the city department to which the request was assigned to, the neighborhood and city council district where the request was made, whether the service request was completed, and if so, how long it took for the service request to be completed. These data are available beginning in 2011, and for the purposes of my analysis, will be updated as of April 15, 2018.

Critically, the 311 data also include the estimated latitude and longitude at which the service request was made. These two variables allow me to geolocate each service request within a Census block, and to calculate the distance in nautical miles between the service request and City Hall (as needed for hypothesis 3). The geolocated data will then be merged with data from the 2011-2016 American Community (ACS). The ACS is a monthly survey conducted by the U.S. Census Bureau that captures a wealth of information about education attainment, income, employment, and housing. Data are made publicly available each year, but aggregating data to small geographic units, such as Census blocks, while still maintaining a large enough sample within each unit requires aggregating over a large time period, such as 5 years. The 2011-2016 ACS is particularly useful for this research problem as it overlaps almost entirely with the time series of Boston's 311 data. Using the 2011-2016 ACS, I will calculate my primary explanatory variable: the median income within each Census block. The variable will then be logged to account for likely skew in the data.

From here, two outcome variables will be generated using the raw 311 data. First, I will create a binary variable equal to 1 if the service request was completed (as of April 15, 2018). Second, among those requests marked as closed, I will calculate the difference—in days—between the time at which service was requested and the time at which the service request was completed. These two variables, respectively, will allow me to test the expectations outlined above: (1)

that city governments are more likely to respond to and complete service requests as the median income of a neighborhood increases; and (2) conditional on response, respond to and complete service requests faster—as measured by the number of days between request and completion—as the median income of a neighborhood increases.

Empirical Strategy

Simply regressing whether a request was completed or not on the median income of the locale, however, is challenging. Imagine that we find a negative relationship between income and the length of time that it takes to complete a service request, such that requests from higher income areas are resolved faster relative to lower income areas. We may wish to interpret this as reflecting differential representation on the basis of income, a la Bartels (2008) and Gilens (2012). But it may also be an artifact of the likelihood to request services, an artifact of differential needs for particular services, and perhaps even of the size and scope of various elements of government. For example, perhaps higher income areas are more likely than lower income areas to request services that are easier to fix. This possibility makes intuitive sense, too: we may expect lower income areas to suffer from more structural issues, such as potholes, etc., and so as a request, lower income areas may be much more likely to request services that require a large team and a significant amount of time, effort, and resources to solve. These concerns become amplified if it is also the case that lower income areas are more likely to make requests to divisions within government that are understaffed and underresourced, making their ability to respond to and complete services requests in a timely manner that much more difficult. In this scenario, the negative relationship uncovered before is hardly a function of a bias toward helping high income earners, but a function of the differing ways in which high income areas and lower income areas use 311 and of government's capacity to respond to service requests.

Given this, a stronger empirical strategy is to make comparisons only *within* service domains. As noted, Boston's 311 data categorizes each request by the city department or division it was assigned to upon receipt and by the nature of the request. For example, requests could have been assigned to the Public Works Department, Boston Public Schools, the Women's Commission, etc. Requests are then categorized by their "reason:" street lights, sanitation, street cleaning, cemetery, valet, water issues, pothole, animal issues, etc. To account for the possibility above—that high income areas and lower income areas simply face different challenges and as such use 311 for different purposes—we will hold constant both the department and request type through a set of fixed effects. In doing so, we are able to hold constant these unobservable differences that may relate to income and service requests, and instead examine the effect of within-department, within-task variation in income on the probability of a given request being completed.

We will estimate the first model—considering the relationship between income and completion—using logit as follows, where λ_i represents department fixed effects and τ_i represents service task fixed effects:

$$\textit{Request completed}_i = \alpha + \beta_1 \textit{Median income}_i + \lambda_i + \tau_i + \epsilon_{it} \tag{1}$$

We will estimate the second model—conditional on completion, the relationship between income and the length of time it took for the service request to be completed—using either Poisson or negative binomial. We will make these determination *after* examining the data. If the mean and variance of the outcome variable are the same, or within 1 day of one another, we will proceed with Poisson. If the mean and variance differ, we will proceed with negative binomial. This is the only estimation choice that we will make following examining the data. As before, λ_i will represent department fixed effects and τ_i will represent service task fixed effects:

$$Days\ to\ completion_i = \alpha + \beta_1 Median\ income_i + \lambda_i + \tau_i + \epsilon_{it} \tag{2}$$

The third model—one estimating effect of income on time completion, conditional on the number of months that the service was requested prior to the next November election—will again be estimated using Poisson or negative binomial (as outlined above). Here, the key variable is an interaction between income and the number of months prior to the next November election that the request was made. Support for the hypothesis would mean a positive relationship, such that high income areas are more likely to see their service requests completed faster as the number of months to the next election increases (i.e., as the election is further away). More simply, it would suggest that the bias in service completion decreases as elections get closer. Again, λ_i will represent department fixed effects and τ_i will represent service task fixed effects:

$$\begin{aligned} Days\ to\ completion_i = \alpha + \beta_1 Median\ income_i + \beta_2 Months\ till\ the\ next\ election_i \\ + \beta_3 Median\ income_i * Months\ till\ the\ next\ election_i + \lambda_i + \tau_i + \epsilon_{it} \end{aligned} \tag{3}$$

I will estimate these models in R, as follows, for each hypothesis. For H2 and H2, we assume that we will proceed with negative binomial.

```
library(MASS)

hypoth1 <- glm(completed ~ median_income + as.factor(department) +
                 as.factor(type),
               data = bos_311, family = "binomial")
```

```r
hypoth2 <- glm.nb(days ~ median_income + as.factor(department) +
                as.factor(type),
              data = bos_311)
hypoth3 <- glm.nb(days ~ median_income*months_elections +
                as.factor(department) + as.factor(type),
              data = bos_311)
```

It is likely the case that there are other factors that predict service completion, and the time to service completion, outside of Census block income levels, the department to which the request was assigned, and the kind of request made. One such variable is likely to be government capacity. We hope that these are picked up by the department fixed effects, but we may also want to estimate the effect of government resources more precisely. Indeed, as noted earlier, it seems highly plausible that any negative association between income and time to completion could also be unrelated to income and instead a function of the ability of government to complete the task. As such, I will re-estimate the models associated with H2 and H3 and include a variable measuring nautical miles between the location of the service request and City Hall. The idea here is quite simple: we might expect higher income areas to be closer to City Hall, and that requests are simply easier to complete in a timely manner the closer it is to City Hall. As a result, we would want to make sure that any estimate associated with the main income variable is not biased by not including the "ease" with which government may be able to resolve service requests across high and low income areas. I will estimate these models as follows in R:

```r
hypoth2_capac <- glm.nb(days ~ median_income + hall_distance +
                          as.factor(department) + as.factor(type),
              data = bos_311, family = "binomial")
hypoth3_capac <- glm.nb(days ~ median_income*months_elections +
                          hall_distance +as.factor(department)
```

```
                        + as.factor(type),
              data = bos_311)
```

Discussion and Conclusion

This proposal seeks to examine whether policymakers in American cities exhibit bias—on the basis of income—in delivering services to constituents. It follows from an emerging literature which argues that legislators in the U.S. respond primarily to the preferences of high-income earners in crafting public policy (Bartels 2008; Gilens 2012). These studies, however, are plagued by significant correlations in public preferences across income groups: indeed, if, as some work has shown (Soroka and Wlezien 2008; Tausanovitch 2016), high-income and low-income preferences over public policy are both similar and move in parallel to one another, it seems unlikely that policymakers can actually differentially respond to one group over another. This proposal suggests that a more tractable way to detect bias in representation—to the extent that is exists—is by considering requests for constituent services. Doing so may allow us to reliably detect such bias, given that the propensity to lobby for government assistance is more likely to differ substantially across income groups.

The data and empirical strategy above outlines the "first cut" at examining this question. Of course, the project can and should expand beyond Boston. Numerous major cities—including New York City and San Francisco—make similar data publicly available. Second, the models outlined above do not take into account other factors—such as racial and ethnic diversity—that may also influence whether and how quickly government responds to requests for service. Another consideration may be voter turnout. Some work shows that politicians are particularly responsive to the preferences of voters (Anzia 2014). As a result, areas with higher voter turnout may be more likely to find that their city government is responsive to their concerns and requests. Future iterations of the project will incorporate these factors.

Before collecting and analyzing such data, however, an amendment to this pre-analysis plan would be filed that outlines the data source, structure, and the new specifications to be used. In the end, such revisions should go a long way in allowing the estimate of income to better reflect—within department and within task—only the unique effect of income.

In sum, the study outlined above offers a unique take on how to study representation. Studies of representation overwhelming consider either representation with respect to either particular polices and broader ideological positions. But politicians and governments more generally interact with and influence the lives of citizens in more ways than simply through policy proposal and implementation. Moreover, citizens want more from government than just policy representation. Considering these other dimensions of representation will go a long way toward expanding our understanding of the relationship between the governed and the government in the United States.

References

Anzia, Sarah F. 2014. *Timing and Turnout: How Off-Cycle Elections Favor Organized Groups.* Chicago: University of Chicago Press.

Bartels, Larry M. 2008. *Unequal Democracy: The Political Economy of the New Gilded Age.* Princeton: Princeton University Press.

Bashir, Omar S. 2015. "Testing Inferences About American Politics: A Review of the 'Oligarchy' Result." *Research & Politics* 2 (4): 1–7.

Branham, Alexander J., Stuart N. Soroka, and Christopher Wlezien. 2017. "When Do the Rich Win?" *Political Science Quarterly* 132 (1): 43–62.

Enns, Peter K. 2015. "Relative Policy Support and Coincidental Representation." *Perspectives on Politics* 13 (4): 1053–64.

Gilens, Martin. 2012. *Affluence and Influence: Economic Inequality and Political Power in America.* Princeton: Princeton University Press.

Soroka, Stuart N., and Christopher Wlezien. 2008. "On the Limits to Inequality in Representation." *PS: Political Science and Politics* 41 (2): 319–27.

Tausanovitch, Chris. 2016. "Income, Ideology, and Representation." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2 (7): 33–50.