Pre-analysis plan:

Combatting capture in local politics: a field experiment

October 12, 2021

Project description

That individuals with a direct monetary stake in a political decision are more likely to participate in politics is at the heart of the collective action problem. This phenomenon impacts decision making through channels from lobbying of national legislatures to participation in local government. Recent work establishes that homeowners are far more likely to participate in local government, often to oppose development and support housing prices. However, renters are also directly impacted monetarily by anti-development decisions through higher rents caused by reduced housing supply. I partner with an NGO involved in abundant housing advocacy (Abundant Housing LA) to conduct a field experiment investigating whether direct outreach highlighting the real monetary costs of non-participation in local government can encourage individuals to attend city council meetings.

This document is a pre-analysis plan of the experimental design and analytical procedures. For any issues unaddressed in this pre-analysis plan, I commit to the defaults described in Version 1.05 of the Lin, Green, and Coppock (2016) standard operating procedures, available at: http://alexandercoppock.com/Green-Lab-SOP/Green_Lab_SOP.pdf.

Background

Design

The experiment will take the following steps:

1. Renters in Los Angeles County were identified by geo-matching addresses in the voter file with Los Angeles County Department of City Planning records of multi-unit housing developments using the *FastLink* probabilistic record linkage merging package in R (Enamorado, Fifield and Imai 2019).

2. AHLA is monitoring the dates of city council public comment periods during which time feedback on housing elements (required analyses of a community's housing needs for all income levels and strategies to respond to provide for those housing needs) can be submitted.

3. Renters in the voter file will be randomly assigned to an email treatment asking them to turn out to support housing elements that expand the supply of housing in their city, or a placebo control. Multiple messages will be randomly assigned in the treatment group(s). The assignment will be block randomized by city and cluster randomized by address.

4. As participation in a public hearing is a matter of public record, I will match the names of those in the treatment group(s) with attendance in local meetings or letters of public support for increased housing supply. Attendance or sending a letter of support will represent the primary outcome measures in the experiment.

5. Results will be analyzed (details below). The primary treatment effect of interest will be defined as the covariate-adjusted treatment effect of opening an email on participating in a city council meeting by submitting a written or spoken public comment. Results will also be analyzed for spoken and written comments separately.

6. As multiple rounds of the experiment will be conducted across multiple municipalities in LA county, precision-weighted fixed effects meta-analysis will be performed to estimate an aggregate effect across municipalities.

7. An additional outcome of interest is whether individuals who attend one meeting attend future meetings. I will therefore also check for repeat attendance in future meetings.

8. Heterogenous effects analysis will be performed for density, area income, and political party affiliation.

Timeline

This pre-analysis plan was filed after treatment assignment and treatment implementation targeting three council meetings on October 12, 2021, but before treatment implementation of all remaining council meetings, and before data collection or analysis. *No outcome data has yet been gathered at the time of writing.*

Three pilot studies were conducted (Santa Monica, August 24, 2021 (N = 500), Long Beach, September 7, 2021 (N = 2000), and Long Beach September 14, 2021 N = 5000)) that inform the final research design. These pilots are identical to this pre-registration document, with the exception of: the absence of the third treatment group; the source of the consensus listed in the second treatment group (economic cost message below) randomized amongst economists, NGOs, and community leaders; and lack of cluster randomization at the address level.

All additional council meetings immediately prior to housing element deadlines between October 12, 2021 and January 2022 will be selected for treatment. In large cities (e.g., the city of Los Angeles), only up to a maximum of 15,000 individuals will be treated per council meeting for ethical purposes.

Hypotheses and treatments

Four treatment groups (including placebo control) using different messages will test the following hypotheses:

1. *Hypothesis 1:* Increased information about how to attend a meeting will increase attendance.

2. *Hypothesis 2:* Information that lack of housing supply increases rents will increase attendance more than attendance information only.

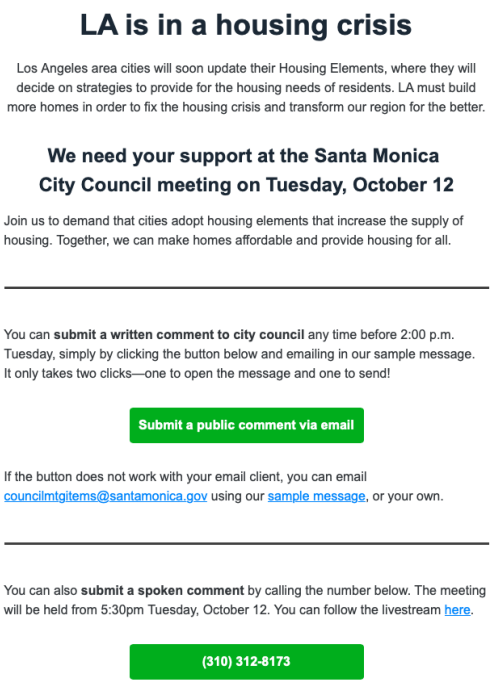
3. *Hypothesis 3:* Information that renters tend not to participate and that this lack of participation is costly (i.e., costly abstention theory) will increase attendance more than attendance information or rental price increase information only.

4. Funding permitting, future rounds of the experiment may add a nominal direct monetary incentive treatment arm. This pre-analysis plan will be updated in this event.

As these are competing hypotheses, each treatment group will contain a message designed to test each hypothesis. All three treatment groups will be analyzed in the aggregate compared to placebo, in addition to each treatment group separately to assess which has the largest effect.

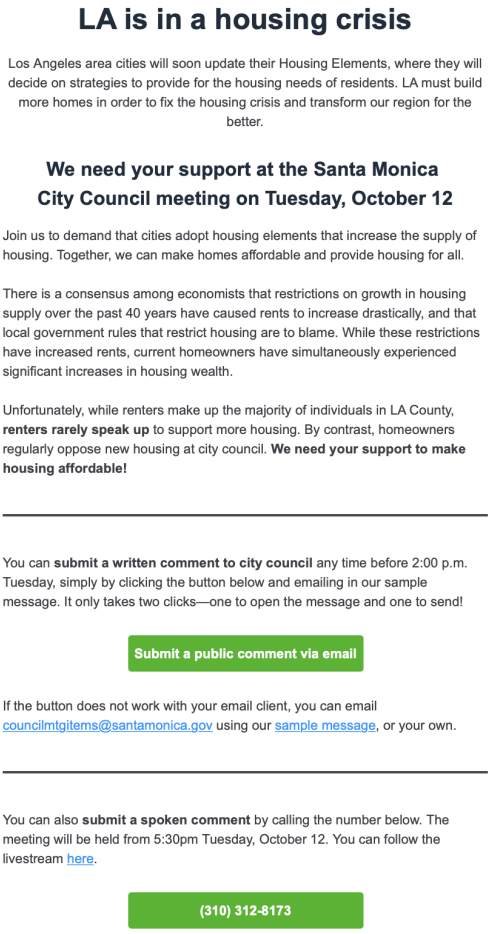
Examples of treatment messages for one council meeting can be found in Figure 1. Treatments for other council meetings will change the location, date, and attendance information only.

(a) Placebo treatment message



(b) Instructions only treatment message

(c) Economic costs message



(d) Cost of abstention message

Figure 1: Treatment groups

Treatment assignment

Likely renters in the voter file were randomly assigned to an email treatment asking them to attend a meeting, or a placebo control.

Individuals were block randomly assigned by city and cluster randomly assigned by address, with 10% probability of assignment to a placebo message with no information on how to attend a meeting, 30% probability of assignment to the attendance instructions only treatment (T1), 30% probability of assignment to the informational treatment regarding how lack of housing supply impacts rents (T2), and 30% probability of assignment to a treatment identical to T2 but which also highlights lack of renter participation (T3).

Outcomes

The primary treatment outcomes of interest are a binary outcome indicating whether an individual:

(1) Attended (virtually) a city council meeting, or
(2) Wrote a message to city council.

This data will be collected by matching the names of individuals in each treatment group from the voter file with public records documenting both attendance and letters of support (both of which require a registration by name).

Treatment effect heterogeneity

I will examine the following heterogenous treatment effects:

1. The density of the building (i.e., number of units) in which an individual lives.
2. The median income of the area in which an individual lives.
3. Turnout in local elections.

I may also conduct analysis using machine learning methods designed to automate the detection of heterogeneous treatment effects (e.g., Bayesian Additive Regression Trees). However, this analysis will be exploratory in nature, in contrast to the pre-registered variables denoted above.

Estimation procedures

My primary estimand is the CACE, and the estimator will include covariate adjustment. I will therefore estimate treatment effects using the Lin estimator (Lin 2013) and include the following pre-treatment covariates in the regression specification: *city, units, gender, age, building age, primary language spoken, vote history, and party affiliation.* Missing covariates will be mean imputed. As units were cluster randomly assigned by address, standard errors will be clustered at the address level.

This will be conducted using the "lm_lin" function in the "estimatr" package in R (Blair, Cooper, Coppock and Humphreys 2019; Lin 2013). The code that will be used is as follows:

lm_lin(comment ~ treatment, covs, data = df, subset = opened == 1, clusters = address), where covs is the list of covariates above. Results will also be reported in the appendix without covariate adjustment. The code will be as follows: lm_robust(comment ~ treatment + city, data = email, subset = opened == 1, clusters = address).

Heterogenous treatment effects will be estimated by regressing the outcome variables on treatments and the interaction between the treatment and the covariate. Heterogenous treatment effects will be estimated for density, median area income, and party affiliation. This will be conducted using the "lm_robust" function in the "estimatr" package in R.

Results will be analyzed both as one large experiment with city fixed effects, as well as aggregated using precison-weighted fixed and random effects meta-analysis. The meta-analysis will both include and exclude results from the pilot studies.

CACE vs ATE

Because I expect many emails to go unread, I wish to demarcate the complier average causal effect (CACE) vs. the ATE. I will monitor if an email is opened, and look at the CACE by examining whether an email was opened as the measure of treatment receipt.

Meta-analysis

Precision-weighted[1] fixed effects meta-analysis, including the pilot studies, will be performed to estimate an aggregate effect across council meetings. For council meetings where no comments are reported in treatment or placebo, I will estimate standard errors according to the procedure described in Gelman and Hill (2006, p.17, footnote 1). Random effects meta-analysis and meta-analysis excluding the pilot studies will also be performed for robustness purposes. This will be conducted using the rma.uni function in the *metafor* package in R. The code for this analysis can be found below:

Weighted fixed effects:
rma.uni(yi = estimate, sei = std.error, weighted = TRUE, method = "FE", data = meta_cace)

Random effects:
rma.uni(yi = estimate, sei = std.error, data = meta_cace)

Tests for proper implementation

1. I will test that contact rates in treatment and placebo are the same.
2. I will test that compliers in the placebo group have similar baseline values to compliers in the treatment group. For the covariates for this test, I use the baseline covariates mentioned above for covariate adjustment.

Pilot studies

A small-scale pilot study to test mechanics was conducted in Santa Monica, CA on August 24, 2021 (N = 500), and two larger pilot studies to test outcomes were conducted in Long Beach, CA on September 7, 2021 (N = 2000) and September 14, 2021 (N = 5000).

[1] With weights equal to the inverse of the variance.

These pilot studies suggest a compliance rate of 15.5%, intent-to-treat effect of 0.0012, and complier average causal effect of 0.0077.

The power analysis below assumes this compliance rate and CACE for T2 (0.0077), a smaller effect size for T1 (0.004), and a larger effect size for T3 (0.01).

Power analysis

The LA County voter file possesses 266,057 individuals with email addresses who could be linked to a rental apartment address with posterior probability of a correct match of 99% or higher. An assumed compliance rate of 15% equals 39,909 individuals who will receive a message. However, as not all cities will undergo public comment periods for housing elements during the experiment timeline, a conservative estimate of number of individuals who will receive a message is likely around 6000.

The power analysis below assumes this compliance rate and CACE for T2 (0.0077), a smaller effect size for T1 (0.004), and a larger effect size for T3 (0.01).

Assuming a 0.1% turnout rate amongst compliers in the control group, 0.4% in the first treatment group, 0.77% in the second treatment group, and 1% in the third treatment group in a simulation reveals the statistical power between each treatment group and control in the table below. The "all treatments - control" simulation assumes a constant treatment effect of 0.77% across all treatment groups.

Estimate	Power	SE(power)
DIM (T1 - Control)	0.35	0.05
DIM (T2 - Control)	0.87	0.03
DIM (T3 - Control)	0.97	0.02
DIM (T2 - T1)	0.23	0.04
DIM (T3 - T2)	0.07	0.03
DIM (T3 - T1)	0.46	0.05
DIM (All treatments - Control)	0.99	0.01

References

Blair, Graeme, Jasper Cooper, Alexander Coppock and Macartan Humphreys. 2019. "Declaring and diagnosing research designs." *American Political Science Review* 113(3):838–859.

Enamorado, Ted, Benjamin Fifield and Kosuke Imai. 2019. "Using a probabilistic model to assist merging of large-scale administrative records." *American Political Science Review* 113(2):353–371.

Gelman, Andrew and Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Lin, Winston. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *Annals of Applied Statistics* 7(1):295–318.