Analysis Plan

Project Name: Increasing Local Government Compliance with Reporting Requirements for the State and Local Fiscal Recovery Fund

Project Code: 2112D

Date Finalized: February 8, 2022

How this document is to be used:

This document is associated with the Analysis Plan Commitment gate in the OES project process. For a step-by-step guide to Analysis Plan Commitment and a summary of roles and responsibilities, see the Project Process Guide.

This document serves as a basis for distinguishing between planned confirmatory analyses and any exploratory analyses that might be conducted on project data. This is crucial to ensuring that results of statistical tests are properly interpreted and reported. For the Analysis Plan to fulfill this purpose, it is essential that it be finalized and date-stamped before we take possession of outcome data. Once this plan is finalized, a date is entered above, and the document is posted publicly on our team website.

If any analyses are described that will not be included in the OES abstract or reported to the agency partner, then explicitly identify these in order to streamline reanalysis.

Project Description

As part of the American Rescue Plan, the Department of the Treasury ("Treasury") is administering the State and Local Fiscal Recovery Fund (SLFRF), which provides $350 billion in funding for eligible state, local, territorial, and Tribal governments ("grantees") to respond to the COVID-19 emergency and bring back jobs. Grantees have substantial flexibility over how they use funds to meet local needs—including replacing lost fiscal revenue, support for households, small businesses, impacted industries, essential workers, and the communities hardest hit by the crisis. These funds can also be used to make necessary investments in water, sewer, and broadband infrastructure.

A notable feature of SLFRF is that it is the largest transfer of fiscal funds to small cities and towns – referred to as Non-Entitlement Units (NEUs) – in over forty years. NEUs are local governments typically serving a population of 50,000 or less. There are approximately 22,000 NEUs, who have been allocated a total of $19.5 billion to spend over the next five years.

This transfer represents the potential for a major transformation of local finances. However, Treasury faces serious challenges identifying and communicating with NEUs in order to track whether and how they are spending funds.

Unlike other levels of state and local government, such as counties and states, it is somewhat rare for NEUs to receive funding directly from federal agencies. Treasury has little prior experience working directly with these governments. As a result, states play a large role in mediating Treasury's relationship with NEUs. To allocate funds to NEUs, Treasury first developed a list of potentially eligible governments using Census Gazeteer data on the 2019 population count of incorporated places and county subdivisions. States received a lump sum payment for the NEUs in their jurisdiction in an amount proportional to their 2019 population counts, and are responsible for distributing payments to NEUs. In order to receive funding, NEUs need to provide information and documentation to states. However, to remain compliant with Treasury's guidelines, they must also send annual reports to Treasury about their spending.

A major question for Treasury, therefore, is how to best incentivize NEUs to accomplish program goals and comply with reporting requirements, given the sheer number of NEUs and their infrequent contact with federal agencies. Currently, only around 300 of the 22,000 NEUs have signed up to Treasury's portal for reporting. The purpose of this project is to pilot and test improvements to Treasury's communication strategy to increase portal signup.

Preregistration Details

This Analysis Plan will be posted on the OES GitHub before outcome data are analyzed. The plan will be gated while it is reviewed by Treasury.

Hypotheses

We hypothesize that adding a concise plan of action to the email body — the "concise summary treatment" — will increase the proportion of NEUs on the portal and will lead to more timely signup.

Importantly, we assume there is no effect of the concise summary treatment on the probability of opening the email.

We hypothesize that changing the subject line from the default to the "simplified subject line treatment" will increase the proportion of NEUs on the portal and will lead to more timely signup.

We hypothesize that the simplified subject line treatment will increase the probability of recipients opening the emails.

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

Data Source(s):
We expect to receive two datasets: one tracking the email communication campaign and the other the reporting portal. We expect that the communication data will reveal whether an email was

opened or not. The reporting portal data will list the date at which an NEU joined the reporting portal.

Outcomes to Be Analyzed:
Recall we measure outcomes at two points: one day after the email is sent (February 11, 2022), and then again two weeks after the email is sent (February 24, 2022). There are two primary outcomes of interest. First, the proportion of NEUs that have signed up to the portal by the measurement date. Second, the proportion of NEUs that have opened the email by the measurement date. As a secondary measure that captures the timing of signup, we also plan to measure the number of days that an NEU has been on the portal by the measurement date (0 for those who have not yet signed up).

Imported Variables:
We expect to merge three datasets into one. We will start with the randomization dataset, which will contain primary and secondary email addresses, the randomization ID used to randomize, the treatment conditions, as well as other information on the NEUs in the study (state, name, etc.). Into this dataset, we will merge two variables from the portal data: first, whether the NEU has signed up for the portal by the first and second measurement dates; second, the date at which those who have signed up did sign up. Finally, we will merge in the date at which recipients opened the email (blank if they have not opened).

Transformations of Variables:
From the raw data above, we will construct two versions of the following variables, one for each of the measurement dates:

1. **signed_up** - a binary variable indicating whether the NEU has signed up for the portal by the measurement date.
2. **opened_email** - a binary variable indicating whether either of the two email recipients opened the email by the measurement date.
3. **days_on_portal** - a variable counting the number of days the NEU has been signed up to the portal, including 0 for those who have not signed up, by the measurement date.

Transformations of Data Structure:
We will create a panel version of the data for graphing purposes. Specifically, we will track rates of signup and email opening by treatment status.

Data Exclusion:
NEUs for whom no valid primary or secondary email address was listed were dropped.

Treatment of Missing Data:
We do not anticipate any missing data for portal signup outcomes. It is possible that email opening data may be missing for some units. While it is possible that such data may be correlated with other outcomes, we do not expect that it will be affected by the treatment and so do not anticipate any issues of differential attrition.

Descriptive Statistics, Tables, & Graphs

We will create a time series plot tracking rates of signup and email opening by treatment status.

Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

Statistical Models:
In generic terms, the main specification will be a linear regression of the outcome on one binary treatment indicator, including fixed effects for blocks. If a pre-treatment measure of the outcome is available, we will include this in the regression as well. We will estimate coefficient standard errors using a cluster-robust, CR2, estimator clustered at the randomization ID level. We plan to use the lm_robust function from the estimatr package for R.

Confirmatory Analyses:
We expect that not everyone who we seek to treat will be treated. First, not everyone will receive emails due to the addresses malfunctioning / bounceback. Second, not everyone will be treated with the content of the email because not everyone will open the emails. With this in mind, our strategy seeks to estimate complier average causal effects (CACEs), rather than intent-to-treat effects (ITTs).

We assume the treatment does not affect bounceback. Provided we do not find a statistically significant effect of the treatment on bounceback, we plan to subset our analyses to NEUs where at least one email address did not bounce back. If we do find a statistically significant effect on bounceback, we will resort to ITT analyses and will not subset.

We define any NEU where at least one email address does not bounce back as a complier with the subject line treatment. We cannot measure spam filtering or attention by recipients to their inboxes so simply interpret such NEUs to have received the email and been treated with the subject line.

We assume that the body of the email does not affect the probability that it is opened. We test this assumption by estimating the effect of the email body treatment on the probability the email is opened. If we find a statistically significant effect, we will conduct ITT analyses without subsetting. If, instead, we find no statistically significant effect, we will subset the analysis to email openers in order to estimate the effect of the content on openers. This is where the 2x2 factorial comes in handy. We can estimate the effect of content among those the control subject line induces to open the email separately from the effect of the content among those whom the treatment subject line induces to open the email. The average of these estimates, obtained through the regression, tells us the average effect of the content on openers.

With these decision rules in place, we will conduct 6 analyses in total using the regression estimator described above:

- The CACE (or ITT) of the email content treatment on the proportion of NEUs who sign up for the portal by the two measurement dates.
- The CACE (or ITT) of the subject line treatment on the proportion of NEUs who have opened the email by the two measurement dates.
- The CACE (or ITT) of the subject line treatment on the proportion of NEUs who sign up for the portal by the two measurement dates.

Exploratory Analysis:

We will use the same regressions specification as above to conduct the following exploratory analyses:

- The CACE (or ITT) of the email content treatment on the average number of days NEUs have had an account on the portal by the measurement date.
- Heterogeneous effects analysis on all main outcomes, looking at the interaction between the treatment indicator and whether the cluster is “big” (larger than 1) or “small” (of size 1).
- Heterogeneous effects analysis on all main outcomes, looking at the interaction between the treatment indicator and whether the state has disbursed funds to NEUs by the measurement date.
- Interrupted time series analysis of the effect of the email campaign as a whole, comparing the week before the campaign to the week after. We will use the CausalImpact package for R, with default parameters.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:

As a robustness check, we will estimate ITTs for any analysis in which the main analysis sought a CACE. For example, if we fail to reject the null hypothesis of no effect for any unit with respect to the effect of the email content on the probability of opening an email, then we will estimate the CACE of the email content on the probability of having signed onto the portal by subsetting to email openers. As a robustness check, we would also estimate the ITT of the email content on the probability of having signed onto the portal by avoiding subsetting altogether.

We will calculate two tailed p-values using randomization inference with 10,000 draws from the sampling distribution of the estimators under the sharp null of no effect for any unit.

We will follow the OES SOP for multiple comparisons, and report the testwise alpha that would need to be applied in order to achieve a family-wise error rate of 5% under the global sharp null. The family of tests will include all confirmatory analyses.