Analysis Plan

Project Name: Dallas Small Business Continuity Fund Impact Evaluation
Project Code: 2008-Dallas
Date Finalized: May 12, 2021

1 Project Description

In spring 2020, the Dallas City Council used CARES Act funding from the U.S. Department of Housing and Urban Development (HUD) to set up the Small Business Continuity Fund (SBCF). The funding was made available to Dallas businesses both as grants of up to 10,000 USD and as low-interest (0-1%) loans of up to 50,000 USD.

The program was oversubscribed: the city received almost 4,000 applications requesting a total of 26M USD. In anticipation of oversubscription, the City of Dallas Office of Economic Development designed a lottery system to disburse the funds in a manner that was both equitable and targeted to the areas of the city most in need.

We are interested in the causal effect of receiving a grant or a loan on business-level outcomes such as the ability to generate revenue and remain open, as well as the differential impact of such funding for different types of businesses and business-owners. The use of random assignment to distribute funding greatly strengthens our ability to characterize these causal effects. In this document, we describe how we plan to conduct the analysis. Section 2 describes how the lottery worked and how we can interpret it using the analogy of a block-randomized experiment with uneven probabilities of assignment. Section 3 describes the outcome data we plan to obtain and how we plan to merge it with the lottery data in order to construct the panel dataset to be used in analyses. Section 4 describes the statistical models we will use to estimate effects.

2 Research Design Overview

2.1 Description of lottery implementation

In both the loan and grant lotteries, businesses' applications were first filtered to remove those who were clearly ineligible, such as applications from businesses outside of Dallas.[1]

Loan lottery implementation. For the loan lottery, the process proceeded as follows:

1. **Randomly assigning the order of invitations:** 396 eligible businesses[2] were assigned a random number and sorted according to that number. This breaks any link between features of applicants (e.g., how quickly the business submitted an application) and their priority for funding.
2. **Setting a target number of businesses to offer loans to:** The Dallas City Council first decided on a target number of businesses to offer a loan (35 at the time of writing).

1. Initial eligibility criteria were as follows: a) Businesses must be located in Dallas; b) businesses must operate out of a physical location within Dallas city limits ('brick and mortar' edifice, farmers market, and other types of physical locations, which may include a home-based business, but not a P.O. box; c) the business must be able to demonstrate that they have experienced a loss of revenue of at least 25% due to COVID-19; d) the business must have annual revenues under $1.5 million for 2019; and, e) all recipients of federal funds must be eligible to work in the United States.

2. A total of 1270 businesses applied, but many did not meet the initial eligibility criteria and therefore were not included in the lottery.

3. **Going down that randomized list, inviting businesses to submit documentation for loan funding:** program staff proceeded through the randomly-ordered queue and invited the businesses to submit documentation to remain in consideration for funding. Many businesses initially invited to submit documentation did not actually receive a loan because they were ineligible (e.g., due to cash flow issues) or withdrew from the process. Program staff continued down the queue until they reached the Council-determined minimum number of businesses to provide with loans. The invitation email for loan applicants was sent on May 22, 2020.

Table 1 provides a hypothetical example of such a system, in which eight businesses are entered into the lottery and three are to be funded. Note that, while three businesses are funded, five businesses are *invited* to submit documentation for possible funding because the second and third businesses in the queue were ineligible or withdrew.

Business ID	Random order	Invited	Funded	Status
8	1	1	1	Funded
2	2	1	0	Withdrew
5	3	1	0	Ineligible
4	4	1	1	Funded
3	5	1	1	Funded
1	6	0	0	Uninvited
7	7	0	0	Uninvited
6	8	0	0	Uninvited

Table 1: Hypothetical lottery in which the aim is to fund three businesses.

Grant lottery implementation. The grant lottery worked in a slightly different manner. A total of 1056 businesses were entered, with an initial 2.5 million dollars in funding.[3] By January 2021, the size of the available funds had been increased to 8 million USD. Moreover, just under half of businesses invited to submit documents for funding did not make it through to the final phase of funding (due to unresponsiveness, ineligibility discovered after documentation review, and withdrawal). As such, the city was eventually able to invite every business to submit documents for an eventual award. Instead of randomizing some businesses to get or not get an invitation, therefore, the grant lottery effectively randomized the timing with which businesses received an invitation (and eventual funding).

Businesses were randomly assigned to a rank and this rank determined the date at which invitations to submit documents for funding were sent. The random ranking was weighted to ensure that the first 125 businesses would be from "high-poverty or low-income areas." Specifically, businesses entered into the grant lottery were first sorted into two groups, depending on whether they answered "Yes" or "No" to the following question: "Is your business located in a high poverty or low income area as indicated by the shaded areas on this map?" The map on Figure 1 was provided.

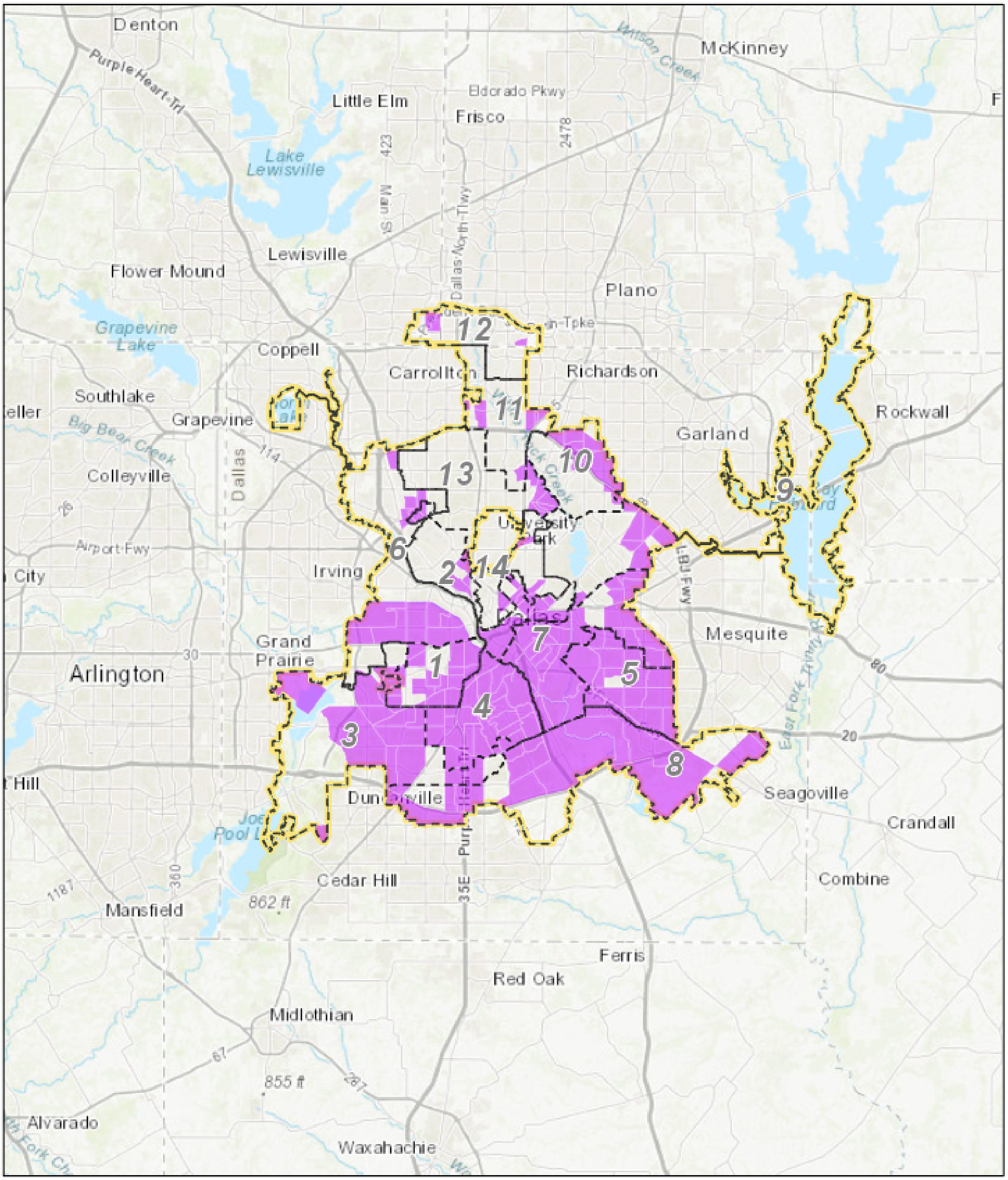
All businesses who answered "Yes" were given a randomly-assigned rank. Those randomly assigned to ranks 1 through 125 were set aside. Those businesses who ranked above 125 were recombined with the businesses who answered "No" to the question above. Then, the pooled businesses were once again randomly sorted, and ranked 126 through to 1,056. Table 2 lists the dates at which emails were sent to invite the randomly-ranked businesses to submit documentation for grants.

3. A total of 2633 businesses applied, but many did not meet the initial eligibility criteria and therefore were not included in the lottery.

Invitation email date (2020)	Random lottery rank
05/22	1-250
8/25	251-350
9/16 and 9/18	351-550
10/19	551-600
10/27	601-700
11/2	701-800
11/18	801-850
12/2	851-950
12/14	951-1056

Table 2: Dates at which invitation emails were sent to grant lottery entrants.

Targeted SBCF Areas



December 9, 2020

1:514,295

0 3.75 7.5 15 mi

0 5 10 20 km

Sources: Esri, HERE, Garmin, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), (c) OpenStreetMap contributors, and the GIS User Community, City of Dallas GIS Services

Figure 1: Map of low-income areas targeted in the SBCF grant lottery.

2.2 Analysis strategy

We leverage the fact that the timing of the invitations to submit documentation for funding was randomized in order to estimate the effect of funding receipt on business survival. Our datasets on business outcomes, described in greater detail below, measure businesses over many periods (daily, weekly, monthly, etc.). Thus, our unit of analysis is a unit-period. Unit-periods are randomly assigned to invitations for funding in different waves (with some businesses never invited, as in the case of the loan lottery). In this respect, the design is similar to a stepped wedge randomized controlled trial.

The randomized timing of invitations avoids confounding that could otherwise occur between the timing of an invitation and other temporal shocks affecting business outcomes. For instance, in early June, Dallas allowed restaurants to expand their indoor dining capacity to up to 75%, then going back to 50% capacity in late June.[4] Many businesses, regardless of their treatment or control status, might have experienced improved outcomes in the two weeks of expanded capacity. By using "not yet invited" businesses as a control for invited businesses during periods like those, we can obtain estimates of the average effect of funding that are not biased by time-varying confounders.

The key feature of stepped wedge designs is that the probability of a given unit-period being assigned to treatment varies over time: periods are like experimental blocks with different probabilities of assignment to treatment. To obtain consistent estimates of treatment effects, it is important to calculate these probabilities and account for them in the analysis. In this study, we will assume that the rank-specific schedule of invitation emails is fixed. In other words, businesses ranked 701-800 in the grant lottery would always have received invitation emails on November 2 and those ranked 801-850 would have received emails on November 18, irrespective of the particular businesses that happened to be assigned to the different ranks.[5]

Under this assumption, we can simulate the lottery invitation process thousands of times under different random assignments, and thereby obtain each business-period's probability of having been invited by that period. Denoting a binary indicator for business i having been invited by period t as $Z_{it} \in \{0, 1\}$, and the simulated probability of assignment $\widehat{Pr(Z_{it} = 1)} = p_{it}$, we can construct inverse propensity weights: $\frac{1}{Z_{it}p_{it}(1-Z_{it})(1-p_{it})}$. We describe the regression analyses these weights are employed in below.

While the timing of invitations is randomized, actual receipt of funds is endogenous to unobserved characteristics of the business. Before receiving any funds: businesses must accept the invitation and return the requested supplementary documents, these must be reviewed by program staff at the City, additional reviews and follow-up may be conducted, then the request for funds needs to be processed through the financial department. As noted above, some invited businesses did not receive funds because they were found ineligible, withdrew, or became unresponsive to follow-up requests from the city. In turn, unobserved attributes of the business—for instance, whether an on-staff accountant is available to respond to document requests—are likely correlated with whether a business invited to complete these steps actually completes them.

Defining an indicator for business i actually receiving funding by time t as the treatment, $T_{it} \in \{0, 1\}$, the design thus features one-sided noncompliance ("failure to treat"). In addition to estimators of the effect of invitations on business-period outcomes (intent-to-treat effects - ITTs), we also describe an instrumental variable strategy for obtaining consistent estimates of the effect of funding receipt on

4. https://www.dallasnews.com/food/restaurant-news/2021/03/17/a-pandemic-timeline-how-covid-19-turned-dallas-restaurants-u
5. This assumption amounts to a non-interference or no spillover assumption with respect to treatment assignment.

outcomes of businesses who would be funded if invited (complier average causal effects - CACEs).

3 Data and Data Structure

In this section, we describe the different datasets used in the analysis. There are three broad categories: raw data on business-level outcomes from data providers such as Yelp and PACER; internal data on grants and loans applications from the Economic Development Department of the City of Dallas; and the panel and cross-sectional datasets that result from merging and restructuring the outcome and application datasets.

3.1 Datasets on business outcomes

Self-reported opening status or pivots to remote services

Yelp data on whether the business reports either a permanent or temporary closure helps us to measure businesses' attempts to stay open and generate revenue. The Yelp data contains daily information on whether the business has permanently or temporarily closed. The data also indicates whether the business has put up a "virtual services" banner and whether the business has partnered with Grubhub to enable delivery, and (possibly) opening hours.

Bankruptcy

Data on bankruptcy are downloaded from the Federal U.S. Bankruptcy Court using PACER. The case report files contain a number of fields that describe the disposition of the case as well as identifiers for the parties. These will include Chapter 7, Chapter 11, and Chapter 13 consumer and business filings. A recent working paper shows these follow different trends (Wang et al. 2020).

Unemployment

We are working on obtaining access to the raw Quarterly Census of Employment and Wages data. For all businesses with at least one FTE, this data reports businesses' quarterly employment and wage bill. This data may be added at a later date and a separate analysis plan may be written to describe how it will be analyzed.

3.2 Internal application and lottery data

We have datasets that record the outcomes of the grant and loan lottery process, including the rank of the business (which for lottery applicants, determines the invitation email date), whether or not it received an invitation for funding, and the outcome of the invitation process (not invited, invited and funded, invited and withdrew / unresponsive, invited and ineligible).

3.3 Transformations of data structure

We will construct a dataset containing identifiers for every day in 2020 and every business in the grant or loan lotteries. Thus, the dataset will contain $N \approx (396 + 1056) \times 365 = 529,980$ rows of unit-period observations.

The dataset will include the following primary outcomes:

- `bankruptcy`: A binary indicator that is 1 if the business or business owner filed for any bankruptcy chapter, commercial or consumer, on that day or prior, and 0 otherwise. This is recorded in the PACER data.
- `virtual_services`: A binary indicator that takes the value 1 if the business offered any "virtual services" according to Yelp on that day, and 0 otherwise. This includes both enabling the virtual services banner, indicating remote services, and delivery through Grubhub.

- `closed`: An indicator that takes the value 1 if the business was "temporarily closed" or "permanently closed" according to its Yelp page on that day and 0 otherwise.

We will also include the following treatment variables and covariates:

- `invitation_wave`: Categorical variable indicating date at which that business was invited to submit funding. For those businesses never invited, date takes the value "never."
- `invited`: A binary indicator that is 1 if the business has been invited to submit documents for review by that week, 0 otherwise.
- `weeks_since_invitation`: A categorical variable used for fixed effects. For businesses ever invited to submit documents, it takes negative values corresponding to the number of weeks prior to the invitation, 0 on the week of, and positive numbers following that week. For businesses never invited, it takes the value "never," which is the reference category.
- `funded`: A binary indicator that is 1 if the business has been funded by that date and 0 otherwise.
- `week`: A categorical variable indicating the week.
- `day`: A categorical variable indicating the date.
- `weekday`: A categorical variable indicating the day of week (e.g., Monday).
- `business`: A categorical variable serving as a unique business identifier.

To these variables, we will add additional covariates and the inverse propensity weights constructed using methods described below.

4 Statistical Models & Hypothesis Tests

4.1 Estimands

We are interested in two main estimands. First, the intent-to-treat effect (ITT), that is, the average difference between a state of the world in which all businesses versus no businesses were invited to be funded. Second, the complier average causal effect (CACE), that is, among businesses who would receive funding if invited, the average difference between a state of the world in which they were all funded and one in which they were not. Below we also specify subgroup-specific analyses.

4.2 Estimators

To estimate the ITT and CACE as defined above, we will run weighted linear regressions with robust standard errors clustered at the business-level. Weights are constructed using the estimated propensity scores as above. The p-values constructed from the standard errors from regression models will constitute our main test of the null hypothesis of no average effect. We will make no adjustments for multiple comparisons.

To estimate the ITT averaged over all periods, we run a regression of the following form using `estimatr` for `R`:

```
lm_robust(
  formula = outcome ~ invited,
  fixed_effects = ~ period + business,
  se_type = "stata",
  clusters = business,
```

```
    weights = ipw,
    data = df
    )
```

The coefficient on `invited` thus identifies the ITT by estimating the inverse propensity-weighted average two-period difference-in-difference for every business in the sample. Note we do not include block fixed effects as these are absorbed by the business-specific fixed effects.

To estimate the CACE, we run the following weighted instrumental variables regression:

```
iv_robust(
  formula = outcome ~ funded | invited,
  fixed_effects = ~ period + business,
  se_type = "stata",
  clusters = business,
  weights = ipw,
  data = df
  )
```

The coefficient on `funded` thus identifies the CACE by estimating a two-stage least squares regression.

4.3 Subgroup specific analyses

We will also examine conditional average treatment effects, or how the effect of the grant varies across the following strata of pre-treatment covariates:

- `women_owned` – A binary indicator for whether one or more of the business-owners identify as a woman in the intake survey administered prior to the lottery.
- `black_or_hispanic_owned` – A binary indicator for whether one or more of the business-owners identify as black or hispanic in the intake survey administered prior to the lottery.
- `lma` – A binary indicator for whether the business is located in a low or moderate income area.[6]

4.4 Robustness checks

In our main analyses, we use parametric standard error estimates in order to calculate p-values corresponding to the null hypothesis of no average effect. However, given that we are permuting the assignment vectors in order to generate estimated propensities, we will also generate the sampling distribution under the sharp null of no effect for any unit in order to construct randomization inference p-values.

6. We define these based on the business' Census tract.

Appendix

A Simulation Study

```
prop_withdrawl <- .3
N_granted <- 20
het_fx <- 1

design <-
  declare_population(
    N = 100,
    withdraws = rbinom(n = N, size = 1, prob = prop_withdrawl),
    U = rnorm(N)) +
  declare_potential_outcomes(
    D_Z_1 = 1 - withdraws,
    D_Z_0 = 0,
    Y_Z_1 = U + withdraws + D_Z_1 + het_fx * withdraws * D_Z_1,
    Y_Z_0 = U + withdraws
  ) +
  declare_estimands(
    itt = mean(Y_Z_1 - Y_Z_0),
    cace = mean(Y_Z_1[D_Z_1 == 1] - Y_Z_0[D_Z_1 == 1])
  ) +
  declare_assignment(
    rank = sample(N),
    handler = fabricate
  ) +
  declare_step(handler = arrange, rank) +
  declare_assignment(
    actual_rank = cumsum(1-withdraws),
    Z = as.numeric(actual_rank <= N_granted),
    handler = fabricate
  ) +
  reveal_outcomes(D,Z) +
  reveal_outcomes(Y,Z) +
  declare_estimator(
    formula = Y ~ Z,
    model = lm_robust,
    label = "Y~Z all",
    estimand = c("itt","cace")) +
  declare_estimator(
    formula = Y ~ D,
    model = lm_robust,
    label = "Y~D all",
    estimand = c("itt","cace")) +
  declare_estimator(
    formula = Y ~ Z,
    model = lm_robust,
```

```
    label = "Y~Z non-withdrawers",
    subset = !(Z == 1 & withdraws == 1),
    estimand = c("itt","cace")) +
  declare_estimator(
    formula = Y ~ D,
    model = lm_robust,
    label = "Y~D non-withdrawers",
    subset = !(Z == 1 & withdraws == 1),
    estimand = c("itt","cace")) +
  declare_estimator(
    formula = Y ~ D | Z,
    model = iv_robust,
    label = "IV all",
    estimand = c("itt","cace")
  )
```

A.1 Treatment of missing data

We distinguish between two types of missingness, each of which requires a slightly different approach. First, there is the issue of data that exists in principle but, in practice, we are unable to find a match. For example, whether or not a business had transactions occur on a given date in principle exists for all businesses, including those that closed. However, a business may not show up in our data. We call this "attrition." Second, there is the issue of data that is missing for the more fundamental reason that it is observed conditional on post-treatment outcomes. For example, the average amount of revenue on a given day will be undefined for businesses that ceased to exist. We cannot define average treatment effect estimands for such outcomes for those businesses. We call this "post-treatment missingness." Our approach to this issue is to define our outcomes such that they do not depend on post-treatment potential outcomes.

As concerns attrition, we know already that we will lack data on the outcomes of some substantial proportion of businesses. It is conceivable that some businesses fail to match in ways that are correlated with treatment. For example, if those who applied earliest put the least time into their applications and therefore introduced more misspellings or messier address data that made it more difficult to match them based on these fields. Similarly, if brick and mortar businesses were more likely to apply earlier and also more likely to have Yelp accounts where they post hours, match status might be correlated with treatment status.

Our approach to this issue will depend on a test for associations between treatment status and potential outcomes. We will conduct an F-test between two linear models, both of which will be run using the same analytic sample and inverse propensity weighting scheme as the main analyses, defined below. The first will regress an indicator for missingness on the available pre-treatment covariates used in the propensity prediction model. The second will supplement the first with a treatment term, interacted with those covariates. The F-test thus tests the null hypothesis that there is no *differential* attrition between the treatment and control group.

We will run these tests for any outcome that exhibits missingness. When a test is statistically significant, we will report the following additional robustness analyses:

1. **Poststratification**: We will run a model that predicts each unit's probability of attriting based on observed characteristics measured in the initial intake survey. Observed units are then weighted by the inverse of this propensity. This results in the upweighting of units that (1)

we *are* able to match/observe outcomes for but that (2) had characteristics that led to a high propensity of non-match/attrition.

2. **Imputation**: We will use `randomForest::rfImpute()` along with the covariates used in the propensity score analysis to impute missing values and run the main analyses on the full, imputed, dataset.

3. **Trimming bounds**: We will apply Lee 2009 trimming bounds. Suppose, for example, that there are more matches in the treatment than in the control group. In that case, we define a proportion to be trimmed, Q. Let R^1 denote the match rate in the treatment group and R^0 that in control. Then $Q = \frac{R^1 - R^0}{R^1}$. The approach requires an assumption that the treatment exerts a monotonic effect on the match rate. In this example, that means no unit would have failed to appear in the Yelp data if treated and appeared if untreated. The upper bound on the treatment effect is obtained by removing the Q% of units in the treatment with the lowest outcomes and estimating the effect as usual on this subset. Ties will be broken at random. The lower bound on the treatment effect is obtained by removing the Q% of units in the treatment with the highest outcomes and estimating the effect as usual. If the imbalance in attrition runs in the opposite direction, the opposite monotonicity assumption is imposed and the trimming is applied to the control group. Importantly, this approach does not necessarily bound the sample average treatment effect. Instead, it bounds the sample average treatment effect for always-matchers (those whose appearance in the Yelp data is unaffected by assignment to treatment).

4. **Extreme-value bounds**: This approach involves imputing missing values using the extrema of their support (Manski 1990), or using their most plausible extreme values. We will use the minimum and maximum outcomes for the extreme value bound analysis.

References

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76 (3): 1071–1102.

Manski, Charles F. 1990. "Nonparametric Bounds on Treatment Effects." *American Economic Review: Papers and Proceedings* 80:319–323.

Wang, Jialan, Jeyul Yang, Benjamin Charles Iverson, and Raymond Kluender. 2020. "Bankruptcy and the COVID-19 Crisis." *Available at SSRN 3690398.*