Matthew Draper
12/14/18

The Effect of Anticorruption Interventions on Per-Capita Corruption Convictions

Anticorruption training programs for government officials have become more frequent and intensive in the past two decades (OECD 2013). Review of the literature reveals that these programs are not evenly distributed, but are targeted where corruption is perceived to be relatively high. Theorists have speculated that anticorruption interventions would change behavior, and that they would be most effective when they changed the incentives faced by officials, either by increasing the possibility of detection or increasing the severity of punishment. By contrast, such interventions are said to be least effective when they focus on generalized “values education” (Klitgaard 1988, Rose-Ackerman, 1977). Meta-analyses of anticorruption interventions support this theory, finding a large short-term impact for monitoring and incentives-based education and a smaller but nevertheless significant impact for systemic rule changes.

This paper proposes to study the relationship between anticorruption interventions programs and corruption outcomes in the United States. I hypothesize that public officials who have experienced anticorruption interventions with an incentives focus will exhibit fewer corruption convictions per capita than officials who have not been subjected to any anticorruption interventions. We will evaluate this hypothesis via an experiment with both a large n and small n component.

Difference-in-Differences Design (Large n Component)

To undertake this study, we will employ a difference-in-differences design blending elements of a cross-section design. Difference-in-differences designs aim to isolate differential effects caused by the treatment, but depend on the assumption of parallel trends. (Manly 1992, Cook and Campbell 1979). This experiment’s design can be diagrammed as follows:

$$\begin{array}{lll} N\ O_1\ O_2\ O_3\ O_4\ O_5 & X & O_6\ O_7\ O_8\ O_9\ O_{10} \\ N\ O_1\ O_2\ O_3\ O_4\ O_5 & & O_6\ O_7\ O_8\ O_9\ O_{10} \end{array}$$

Assignment to treatment is nonrandom. Anticorruption programs tend to be implemented in the wake of particularly severe corruption scandals, which are likely to occur more frequently in high-corruption states. As interventions in state governments are typically conducted at the state level, this quasi-experiment will compare pairs of U.S. states with similar rates of public corruption convictions per capita in which one state implemented an anticorruption program and the other did not. This will allow us to create post-hoc independence between treatment and outcomes by conditioning on observables.[1] We will accomplish this by matching treatment and control units on the dependent variable (the per capita corruption conviction score).

The sampling frame will be all US states over a 20-year timeframe (1998 - 2018). We will examine the states on two-year intervals, which yields a sample size of 50 states x 10 2-year periods, for a total sample size of 500. Sampling will be purposive, which will cause difficulties in establishing causation, but as we are not able to randomly assign U.S. states to anticorruption programs, purposive sampling will allow us to choose experimentally relevant characteristics such as the presence or absence of anticorruption programs.

This quasi-experiment will first establish maturational trends with a series of observations prior to the intervention. These observations will take the form of public corruption convictions per capita, a widely available measure (Cordis and Milyo 2016, Public Integrity 2015). Our primary unit of observation will be public officials in U.S. state governments, and it is expected that a number of individuals will pass into and out of these groups over the course of the study. After establishing baseline levels, the treatment group will receive an incentives

[1] $E(Y_i^0|X, D=1) = E(Y_i^0|X, D=0)$

intervention, and the control group will receive no intervention. By hypothesis, we expect that post-treatment measures of the treatment group will show lower per-capita corruption conviction scores than controls.

In order to make a valid comparison between the treatment and control groups, it is vital that treatment assignment and outcome are independent. This entails that the groups ought to be as similar as possible on all dimensions relating to the outcome. To avoid selection threats, we will attempt to mimic random assignment via a matching strategy. In brief, we will estimate the effect of receiving treatment by accounting for all the covariates that predict receiving the treatment. The effect of treatment will therefore be the average difference in outcomes between all matched pairs. We will want to match states from similar regions, with similar public cultures and social norms.[2] To accomplish this research goal, we will employ a matching strategy on observable characteristics based on a propensity score algorithm, dropping unmatched units without common support (assuming conditional independence – see footnote #1). Given the similarity of our proposed matched pairs on both observable and unobservable characteristics, we expect a broad region of common support, which will improve external validity.

Cross-Case Design (Small n Component)

Prior to beginning the difference-in-differences experiment, we will investigate a selection of 15-20 case studies involving the implementation of particular public anticorruption programs. This cross-case study will attempt to shed light on a larger class of cases by focusing on the manner in which anticorruption programs are implemented, specifically the particular incentives altered by the intervention and their directional effects on different types of corrupt

[2] Some sample matched pairs with similar covariates as well as similar levels of corruption convictions per capita are Mississippi and Louisiana (2014-2015), Virginia and Delaware (2016-2017), Illinois and Illinois (2013-2014) (before and after implementation of an anticorruption program) and Oregon and Washington (2014-2015).

acts. Differentiating these effects in the quasi-experiment outlined above would be extremely challenging, but using the case study method we hope to elucidate precise causal mechanisms at work to inform our large-n experiment and provide direction for future research (Gerring 2004). This cross-case design can be diagrammed as follows:

N	**X**	**O**
N		**O**

In the cross-case component of this quasi-experiment, we will begin with a Most Similar Design, looking at variations within individual cases over time as well as variation across cases. We will sample purposively, with the goal of achieving a deeper understanding of outliers and unusual cases, including cases where anticorruption programs were implemented and no results were observed as well as cases where corruption was remediated without the intervention of an anticorruption program. We will be especially sensitive to such deviant cases for their confirmatory or disconfirmatory power, as well as influential cases such as the anticorruption programs implemented by New York and Illinois between 2010 – 2014. We will use data from the large-n component to inform the cross-case component, investigating (for instance) outliers with a high Cook's Distance (D_i) score which seem to be driving large-n conclusions.

Design Threats

The principal threats to difference-in-differences designs arise from violations of the parallel trends assumption – the assumption that in the absence of the treatment, the outcome measure (growth rate in Y) of the treatment and control groups would have moved in tandem. In other words, we are concerned that treatment and control groups may not exhibit similar trends (pre-treatment) on all the dimensions relating to research outcomes. We wish to avoid mistaking differential change for an effect of our treatment. To check validity, we will evaluate preprogram trends with a series of observations prior to administering treatment, but as past performance does not perfectly predict future performance, we cannot be certain that the parallel trends assumption has not been violated. We will also consider placebo effects, both by examining the trend in groups and outcomes not affected by the program.

As mentioned, this experiment will condition on observables by matching treatment and control units on the dependent variable (per-capita corruption convictions). It will not be necessary to condition on all possible sources of difference between the treatment and control groups. Rather, this experiment will condition on variables correlated with both treatment and outcome to preclude omitted variable bias. Likely variables on which to condition include crime rates, levels of public trust, corruption perceptions and the efficiency of criminal justice. Unobservable characteristics remain a concern. Fortunately, officials in matched U.S. states will be similar on many unobservable characteristics, and we will employ a fixed-effects strategy by comparing states from common regions and political cultures, preserving parallel trends.

Quasi-experiments are known to exhibit weaker internal validity than randomized experiments as they struggle to demonstrate causation (Trochim and Donnelly 2007). We avoid single-group threats by means of a comparison (control) group, but we must consider multiple-group threats and social threats to internal validity. The selection-history threat is salient in this context because despite their similarity U.S. states are susceptible to nonrandom events occurring in one but not both members of a matched pair, particularly the type of corruption scandals that trigger anticorruption interventions. The shorter our research timeframe, the lower the likelihood of such threats will be.

Selection-maturation threats are less relevant in our experiment, as maturation is likely to be limited across public officials as a whole, and what maturation does occur is likely to be shared across matched pairs. Selection-testing threats will be absent here because the data-collection process will not directly involve study participants, precluding differential effects. Selection-instrumentation threats are minimized by the common methodology used to collect the dependent variable (per-capita corruption convictions). Selection-mortality threats are unlikely, as the primary unit of analysis is the state level bureaucracy as a whole rather than individuals comprising it. Selection-regression threats are concerning, and to address them this experiment will not base research conclusions on statistical outliers as they will fail to find a match.

We must now consider social threats to internal validity. Diffusion or imitation of treatment is unlikely, as are compensatory rivalry, resentful demoralization and compensatory equalization of treatment. Anticorruption training programs are costly, difficult to implement, and not subject to the control of the experimenter. It strains credulity to expect that state officials would look enviously at the anticorruption programs imposed on officials in a neighboring state, or that controls might become demoralized (or work harder) as a result of such programs conducted elsewhere. For similar reasons, Hawthorne effects are unlikely.

Other salient threats to internal validity will be selection threats and omitted variable bias. As discussed, we will employ a matching strategy to minimize selection threats and conditioning on variables correlated with treatment and outcome to minimize omitted variable bias. More generally, the credibility of all quasi-experiments hinges on our ability to persuasively rule out alternative explanations. It may be necessary to collect additional data to evaluate these threats, such as data on overall crime rates or on levels of public trust.

External validity will depend upon the representivity of our sample (Trochim and Donnelly, 2007). The non-random nature of the sample limits the scope of external validity, but replication can circumvent population differences and reassure us that the experimental findings can be validly generalized. The matching strategy will force us to drop unmatchable outliers, which will limit generalizability. The small-n component will help us improve representivity by allowing us to choose representative examples, but at the cost of generalizability.

Construct validity can be expected to be quite high in this experiment. The background concept to be examined is corruption, specifically levels of corruption in state officials. We are interested in the systematized concept of per-capita corruption convictions, which limits us to cases where corruption has been observed and brought into the legal system, which is an unavoidable mono-operation bias.[3] This operationalization gets at the heart of the background concept, but it is necessarily limited to situations where corrupt acts have been detected. We thus achieve strong translation validity by capturing the main attributes of corruption without capturing anything else.

Conclusion validity will be moderate in the large-n component of this experiment, and very low in the case study component. While our sample size is not underpowered and is large enough to insulate against Type I and Type II errors, its non-random nature reduces our confidence in “meaningful zeros”. This is acceptable, as we are not seeking to robustly demonstrate causation but rather to investigate relationships and guide future research.

Threats to the cross-case design will include low conclusion validity (high noise-to-signal ratio), no possibility of randomization, and unrepresentative results. We will sample purposively, selecting units where anticorruption programs affecting incentives were put in place and no effects were observed, as well as the most successful examples of such programs, illuminating the full range of variation on the dependent variable. By so doing, we hope to develop theory and generate additional hypotheses for future experimental testing.

[3] It might be possible to incorporate a corruption perceptions measure to correct for this mono-operations bias, but perception measures (since they are based on impressions) have significant construct validity problems of their own.

References

Cook, T.D. and Campbell, D.T. "Quasi-Experimentation: Design and Analysis for Field Settings". Rand McNally, Chicago, Illinois. (1979)

Cordis, A. and Milyo, J. "Measuring Public Corruption in the United States: Evidence From Administrative Records of Federal Prosecutions". Public Integrity Vol. 18, pp. 127-148 (2016).

Gerring, John. 2007. "Case Study Research: Principles and Practices". Cambridge: Cambridge University Press.

Huther, J. and Shah, A. "Anti-Corruption Policies and Programs". The World Bank, Policy Research Working Paper #2501 (2000)

Klitgaard, R. "Controlling Corruption". University of California Press (1988).

Manly, B. "The Design and Analysis of Research Studies". Cambridge University Press (1992).

Organization for Economic Cooperation and Development (OECD) "Ethics Training for Public Officials" (2013) https://www.oecd.org/corruption/acn/resources/EthicsTrainingforPublicOfficialsBrochureEN.pdf

Rose-Ackerman, S. "Corruption: A Study in Political Economy". Academic Press (1978).

Trochim, B. and Donnelly, J. (2007). *The Research Methods Knowledge Base*. Atomic Dog Publishing, 2007.

The Center for Public Integrity. "State Integrity Investigation" (2015) https://publicintegrity.org/topics/state-politics/state-integrity-investigation/

United Nations Development Program (UNDP) "Fighting Corruption" (2017) http://www.undp.org/content/undp/en/home/democratic-governance-and-peacebuilding/fighting-corruption.html

Wilhelm, Paul "International Validation of the Corruption Perceptions Index: Implications for Business Ethics and Entrepreneurship Education". Journal of Business Ethics, Vol. 35 No. 3 (2002) pp. 177-189.