Title

Protecting Against False Inferences: A Comparison Between Stability Controlled Quasi-Experiment and Difference-in-Differences Approaches

Author

Pinkelman, Colleen or

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Protecting Against False Inferences:
A Comparison Between Stability Controlled Quasi-Experiment and Difference-in-Differences Approaches

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Applied Statistics

by

Colleen Pinkelman

2020

ABSTRACT OF THE THESIS

Protecting Against False Inferences:
A Comparison Between Stability Controlled Quasi-Experiment
and Difference-in-Differences Approaches

by

Colleen Pinkelman
Master of Science in Applied Statistics
University of California, Los Angeles, 2020
Professor Chad Hazlett, Chair

When a randomized trial is not possible for evaluating the effectiveness of a new treatment, several alternatives have been proposed. Two of these methods are difference-in-differences (DID) analysis and the stability controlled quasi-experiment (SCQE). DID allows for estimation of causal effects with an assumption of "parallel trends": the trend in average non-treatment outcomes are the same between treated and comparison groups. SCQE relies on an assumption of the outcome's "baseline trend": the change between cohorts is the same in the overall non-treatment outcome. We compare these two methods under a range of baseline trend assumptions. We also evaluate the methods' reliabilities in protecting against false inferences and over-confidence. Our application is the effect of a placebo health policy, which is known to have no true effect, on 30-day mortality rate among patients treated for heart attack, heart failure, and pneumonia in U.S. hospitals.

The thesis of Colleen Pinkelman is approved.

Vivian Lew

Erin Hartman

Chad Hazlett, Committee Chair

University of California, Los Angeles

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to first acknowledge the pivotal guidance and feedback that my advisor, Chad Hazlett, provided throughout this process. He was always available to chat when I needed advice and taught me so much about both this field and the research process in general.

Thank you to Vivian Lew and Erin Hartman for serving on my thesis committee. Their commentary and suggestions were invaluable.

Lastly, I would like to thank my parents for inspiring in me a love of learning and for their unconditional support and Grant, for his unfailing encouragement at the moments I needed it most.

CHAPTER 1

Introduction

Although randomized experiments are the ideal setup for determining a cause and effect relationship, they sometimes pose feasibility or ethical concerns. In medical applications and post-hoc analyses in particular, randomized controlled trials (RCTs) are not always possible. One of the prominent concerns about RCTs is the ethical concern of denying a patient access to a potentially life-saving medical treatment. If alternatives to randomization exist, there may be more ethical ways to distribute treatment.

A few designs that can accommodate partial self-selection are comprehensive cohort studies [8] and patient preference trials [4]. These aim to address the problem that individuals who self-select into treatment are likely different from those who do not. However, they do not avoid the need for some degree of randomization. There are a few alternative causal inference methods that do not require randomization and therefore can better address this ethical concern. Here, we examine two causal inference methods as applied to patients treated for heart attack, heart failure, and pneumonia in United States hospitals.

1.1 Stability Controlled Quasi-Experiment

A new method called the stability controlled quasi-experiment (SCQE) was proposed in 2018 by Hazlett [5] for estimating the causal effect of a treatment on the outcomes of treated units. This method does not require randomization. It accommodates any process of treatment selection, including self-selection. It can be used for both designing an experiment or

analyzing observational data.

Rather than requiring an assumption of randomization, SCQE provides an estimated effect estimate subject to an assumption the user is willing to make regarding the change in average outcomes over time. In particular, it requires positing a quantity for which the average outcome would have changed from one cohort to the next, had there been no treatment exposure. This quantity is called δ. For notation, we use the the potential outcomes framework [6], where Y is the outcome of interest, $D = \{0, 1\}$ is a treatment indication, $Y(D)$ is the potential outcome under treatment assignment D. We use T to denote treatment status. $T = 0$ represents low or no use of the treatment, and the second cohort $T = 1$ represents the treated group. We express δ as

$$\delta = E[Y(0)|T = 1] - E[Y(0)|T = 0] \tag{1.1}$$

This stability assumption identifies the Average Treatment effect on the Treated (ATT) and is the key assumption for SCQE. In most cases, we rely on both subject matter experts and prior average changes in outcome to determine our δ choice. However, this means that the estimates of SCQE are accurate subject to the assumption of δ. It is then up to the investigator to judge the plausibility of the chosen δ values in order to defend, fail to defend, or reject the credibility of the corresponding effect estimates.

In cases where no change in average outcomes over time is expected, we use a δ of 0. We can also postulate a range of delta values for which an ATT is to be determined. This approach is useful in situations where: 1) we have multiple source of information about δ , 2) we know the true values lies within a plausible range, or 3) we would like to represent an observed past trend.

Given our assumed δ, we calculate the ATT with the unbiased estimate for the non-treatment outcome among the whole group in the pre-treatment period. In other words, this is the mean observed non-treatment outcome in the pre-treatment period shifted by δ.

The group average is then a weighted combination of the average non-treatment outcome among the untreated and the average non-treatment outcome among the treated (calculated by applying the law of iterated expectations):

$$\begin{aligned} E[Y(0)|T=0] &= E[Y(0)|T=1] - \delta \\ &= E[Y(0)|D=1,T=1]\pi_1 \\ &+ E[Y(0)|D=1,T=1](1-\pi_1) - \delta \end{aligned} \tag{1.2}$$

As a final step, we compute the ATT as the difference between the treatment and non-treatment potential outcomes, taken among the treated. Since the average outcome for the treated had they not directly taken the treatment is not directly observable, we use the strategy mentioned above of inferring this quantity using the mean observed non-treatment outcome in the pre-treatment period shifted by δ. This is represented as

$$\begin{aligned} ATT &= E[Y(1)|D=1,T=1] - E[Y(0)|D=1,T=1] \\ &= E[Y|D=1,T=1] - \frac{E[Y|T=0] - E[Y|D=0,T=1](1-\pi_1) + \delta}{\pi_1} \end{aligned} \tag{1.3}$$

1.2 Difference in Differences Analysis

Initially, the difference-in-differences (DID) and SCQE approaches may appear similar. DID requires observed outcomes of units who were and were not exposed to treatment at two time periods. Researchers rely on the "parallel trend" assumption, which dictates that in the pre-intervention period, some set of observed and unobserved covariates influence the levels and trends in the outcomes for both the treatment and comparison groups [3]. The method uses the trend in outcomes in the control group to impute the trend in non-treatment outcomes in the treated group. Notation for DID includes $D = 0, 1$ for treatment indication, $T = 0, 1$ for pre- and post-treatment time periods, respectively, and $Y(T)$ for the observed outcome at time T. The parallel trends assumption is expressed by

$$\begin{aligned} &E[Y(0)|T=1, D=1] - E[Y(0)|T=0, D=1] \\ &= E[Y(0)|T=1, D=0] - E[Y(0)|T=0, D=0] \end{aligned} \tag{1.4}$$

If the parallel trends assumptions holds, DID provides the estimated ATT. For the analysis that follows, we explore an application of both methods on the same dataset to compare their results.

1.3 The Relationship between Parallel Trends and δ

In comparing the two methods, we also find it important to clarify the relationship between parallel trends and δ. DID can be considered a "special case" of SCQE in that we determine δ from the change over time in units that are not treated, and we assume that the change in non-treatment outcomes is the same for both treated and non-treated individuals.

In SCQE, δ could be set equal to the change that was observed over time in the untreated group. This mathematically implies that the trend in non-treatment outcomes for the treated units equals that in the untreated units, which are both equivalent to δ. DID learns δ through the change of units not eligible for treatment over time and assumes that the change in non-treatment outcomes is the same for both groups. The differences between the two methods are 1) SCQE has a built-in sensitivity analysis, and 2) the assumptions that each method require differ in important ways. In many cases, a practitioner may have more information about the plausible range of baseline trends than the degree to which the trend differs between the treatment and control groups. This is primarily because the latter introduces additional complexity from the method of selection into treatment. For instance, one could conduct a sensitivity analysis with DID by allowing the researcher to posit a different trend in the control group and the treated group. In this scenario, the practitioner would be inferring more about the selection method than about the trends themselves.

1.4 An Example Implementation

To clarify the difference in implementation between the two methods, we outline a simple example. Suppose a researcher would like to evaluate the effect of a medical treatment for some condition on patient recovery rates. In DID, the researcher must first assess the plausibility of the parallel trends assumption in this case. They can then proceed with estimation of the DID effect. Once the effect is obtained, they can use it to inform their understanding of the treatment.

SCQE requires slightly more of the practitioner during the setup phase. First, the researcher must have some degree of knowledge or a reasonable assumption about the trend in outcome with no treatment. Frequently, this is obtained via analysis of the pre-intervention trends or by consulting a subject matter expert. Given this information, they then posit a δ value or range of δ values that they deem plausible for this scenario. After this is determined, they proceed with SCQE calculations. The returned effect estimate or estimates are conditioned on the assumed δ value. When interpreting the results, the researcher must factor this assumption into their interpretation of the treatment's effect on recovery rates.

CHAPTER 2

Methods

2.1 Data Preparation

We sourced our data from Hospital Compare, published by Medicare. The data's primary purpose is to enable comparison of the quality of care at hospitals around the country. We included data from the years 2009-2013 [2]. The raw data includes 5,856 hospitals. To ensure data quality, we first excluded hospitals with missing, incomplete, or inconsistent data in each year. The resulting dataset contains 5,548 hospitals. For this study, we imagine a treatment intervention that took place between 2011 and 2012. Since the treatment is imagined, we know it has a true effect of 0. We introduce this placebo treatment in the interest of evaluating each method's ability to uncover a zero effect.

The initial data contains 5 conditions (heart attack, heart failure, pneumonia, hip/knee, and all-cause). For this analysis, we focus on the three most commonly observed conditions: heart attack, heart failure, and pneumonia. We further filter the data by including only providers that had at least one observation for each condition in each of our considered years. In each year, there is one observation for 30-day mortality rate for each condition in each hospital. We separate the datasets by condition to run three parallel analyses.

The final heart attack, heart failure, and pneumonia data sets contain 2,378, 3,517, and 3,771 measurements for mortality rate at the hospital level, respectively. Running three analyses in parallel improves our ability to generalize the result. We utilize 4 time periods (2009, 2010, 2011, 2012) for our investigation. The two final time periods (years 2011 and

2012) are used as the pre- and post-treatment introduction periods, respectively, for the SCQE and DID implementations. Data from 2009 and 2010 is used to help inform the choice of δ.

In order for our inferences from the SCQE and DID methods to be valid, we should have reason to believe that there were no other treatments that took place during the post-intervention time period. Otherwise, we might find an effect where there should be none because our treatment is a placebo. Since our data is comprised of three different medical conditions across thousands of hospitals in the US, we find it very unlikely that there would be important changes across all these dimensions in the same year. We have no way to verify this with absolute certainty, but we see no evidence to the contrary. Based on the Medicare site itself, we also find that the mortality rate metric formulation is stable throughout these years. With these assurances in mind, we proceed with the analysis.

2.2 SCQE δ Choice

A key assumption required in SCQE analysis is the proposed value or range of values for the average change in non-treatment outcomes, called δ. In other words, this is the change in mortality rate that we would expect if there had been no treatment effect. In the true experimental setting, this value is unobservable. It is recommended to use domain knowledge and prior observed rates of change to inform the choice of δ. In this application, we use the first three pre-intervention measurements to determine the expected year over year change for 30-day mortality rate in non-treated hospitals. We have no reason to believe that there are other effects influencing the outcomes during this pre-treatment period.

To inform our choice of δ we first calculate the year over year change in mortality rate from 2006 to 2007, and 2007 to 2008, respectively. We then plot these values and examine the summary statistics each condition are given in Table 3.1 below. Based on these known quantities, we will include δ assumptions of -1%, -0.5%, 0%, 0.5%, 1%, and the approximate

minimum and maximum δ's for each condition. Although δ is fundamentally unknowable, we believe this to be a plausible assumption for the range of δ values based on trends seen in the pre-intervention period. We keep this estimation method in mind while interpreting the results to avoid potential overconfidence.

Table 2.1: Pre-Intervention Trends in Mortality Rate for Each Condition

Measure	Heart Attack	Heart Failure	Pneumonia
Minimum	-4.9%	-6.9%	-9.2%
25th Percentile	-0.8%	-0.7%	-0.8%
Median	0%	0%	0%
75th Percentile	0.8%	0.8%	0.8%
Maximum	8.0%	8.1%	8.1%

2.3 Simulation

We imagine the placebo treatment took place in 2012. We aim to estimate the effect of this treatment on mortality rate for each of the three conditions. We first calculate each hospital's trend for each outcome, based on times $T = 0$ and $T = 1$. We adjust the probability to be mean-centered. We also multiply by a constant, a, which was chosen to give a reasonable distribution of probabilities.

$$P(selection) = \frac{e^{c(a_1 - a_2)}}{1 + e^{c(a_1 - a_2)}}$$

In the above, c is the chosen constant (75 in this case). a_1 is given by trend for each hospital between the pre and post-treatment intervention time periods, and a_2 is the mean trend across hospitals. This selection method means that hospitals with a higher trend in mortality rate have a higher likelihood to have been selected into treatment. The below

plots give the distribution of hospitals' probability of selection into treatment (Figure 2.1) and the relationship between a hospital's trend and its probability of selection (Figure 2.2).

Figure 2.1: Distribution of selection probabilities for hospitals

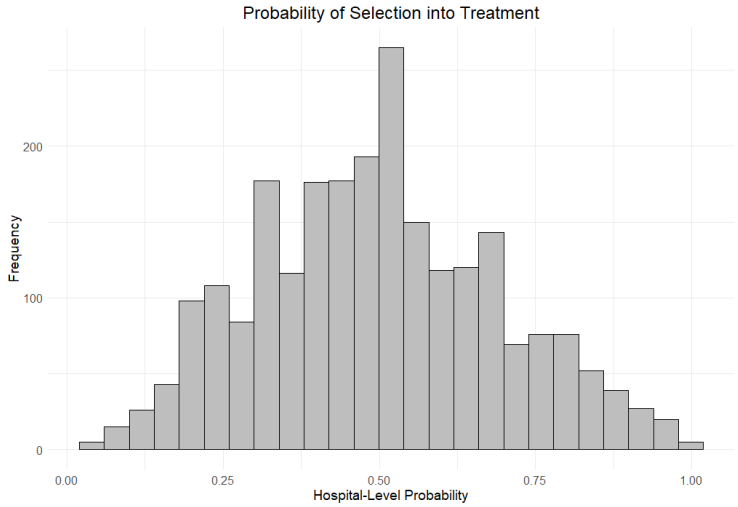
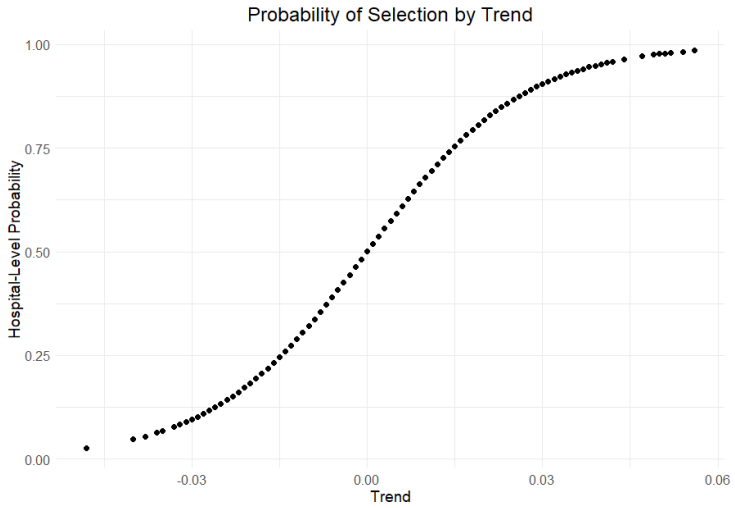


Figure 2.2: Selection probability for hospitals by trend in mortality rate between times T=0 and T=1



The placebo treatment has an effect of zero. We can then evaluate our methods against this known effect [1]. The goal of the method is to estimate the zero effect by including it within the given confidence interval of the estimate. Specifically, we compare the DID estimator to the SCQE method using the range of δ values chosen above. We consider the

"coverage" of a method to be 1 when the CI includes zero, and 0 when it does not. We simulate 500 times for each of the conditions and compare the coverage of each. In each simulation, we select a random subset of 100 hospitals for evaluation. Our unit of analysis is the mortality rate for one of the three conditions at an individual hospital. Given this simulation method, the expected bias is towards positive results. The placebo treatment is assigned more often to units whose outcomes are increasing over time, which means that DID will appear to show positive (harmful) effects.

2.4 SCQE Package

To implement the SCQE analysis, we use the SCQE package for R. The package was developed in 2020 and is available for download from GitHub. It allows a user to study both the one and two cohort cases, using either summary statistics from the population or the full data. In our case, we have a one cohort case with the full data for each simulation available. It also allows the user to plot the SCQE object in order to visualize the results. This paper is the first documented application of the package.

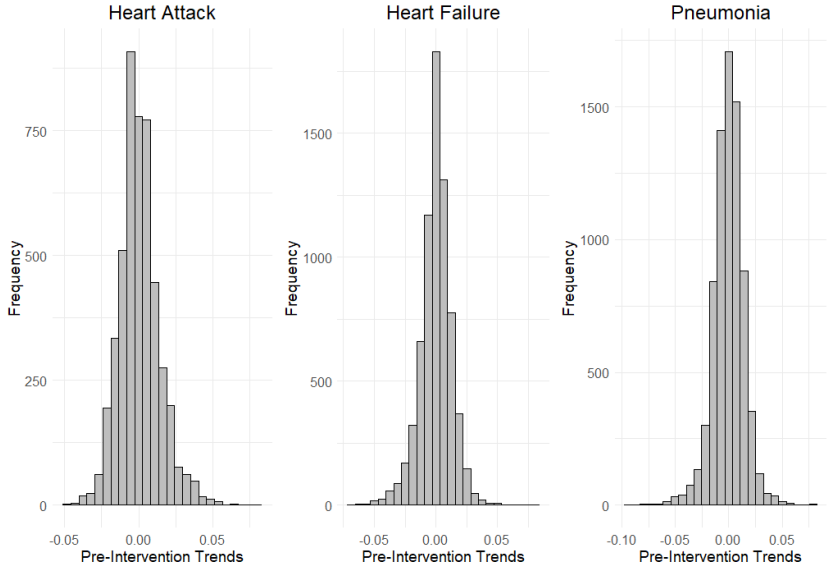
CHAPTER 3

Results

3.1 SCQE Results

We consider our plausible range of δ's to be between the 25th and 75th quantiles of observed pre-intervention trends. The summary statistics are provided in Table 2.1 in the above chapter, and the full distribution of pre-intervention trends for each of the three conditions is given in Figure 3.1 below. Given the summary statistics and distribution of the pre-trends, we find this to be a reasonable and conservative estimation of the plausible δ values. We choose to err on the conservative side in terms of δ range to reduce the chance that we capture a zero effect as a result of having a wide, permissive δ assumption.

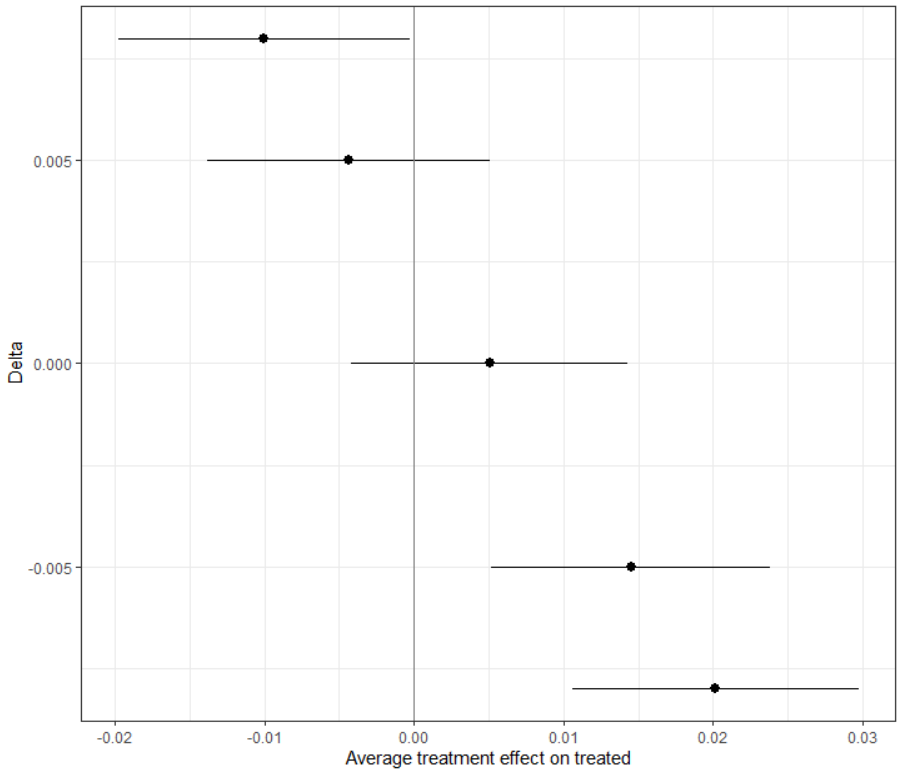
Figure 3.1: Distribution of Pre-Intervention Trends in Mortality Rate for Each Condition



Using this specified range, we run simulations for each of the three conditions. The SCQE method produces a range of estimates based on the provided choices of δ. Figure 3.2 is an example plot for one simulation, to illustrate the outcome of an SCQE analysis for a single hospital. Each horizontal line estimate represents the effect estimate obtained from the δ value on the y-axis. The point gives the exact effect estimate obtained, while the whiskers represent the respective size of the 95% confidence interval.

For example, we see that assuming a δ value of 0 gives an estimated effect of 0.005 with lower and upper bounds of approximately -0.004 and 0.014. Here, the practitioner would have to assume a δ value of at least 0.008 to make a conclusions that the ATT was negative. Conversely, the practitioner would have to assume a δ of less than -0.0028 to conclude there was a positive ATT.

Figure 3.2: SCQE estimates for a single simulation in terms of mortality rate



We can now evaluate the coverage of the SCQE method. We consider the coverage to

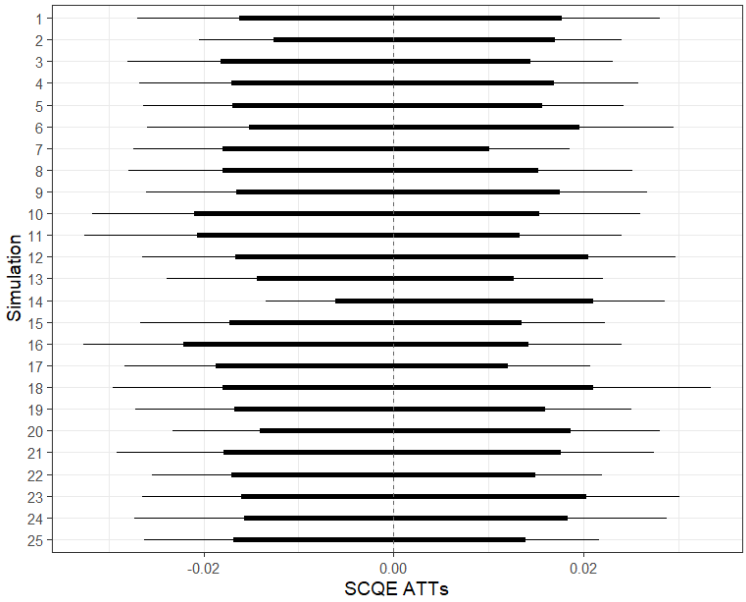
be 1 in cases where the range of effect estimates includes the zero effect case and 0 in cases that do not include the zero effect case. We find that in 100% of simulations for each of the three cases, the effect estimate interval includes the zero effect case. We also examine the mean effect estimates for the upper and lower bounds of the explored δ ranges. For both the upper and lower 95% CI bounds, the method is not off by more than about 0.15% in estimated increase or decrease in mortality rate. Based on the symmetry of these estimates, SCQE appears equally likely to slightly over- or under-estimate the true zero effect. This translates to an approximately equal chance of having a favorable or unfavorable perspective on the effect of the theoretical treatment.

Table 3.1: SCQE Coverage from Range of δ Values

Condition	Percent of Intervals Including 0	Average Lowest Effect Estimate	Average Highest Effect Estimate
Heart Attack	100%	-1.7%	1.3%
Heart Failure	100%	-1.6%	1.6%
Pneumonia	100%	-1.7%	1.5%

For a subset of the simulations, we plot the ATT estimates for each simulation result for SCQE, calculated by subtracting the treated hospitals' counterfactual non-treatment mortality rate from their observed mortality rate under the placebo treatment. The boxes represent the range of likely point estimates produced by plugging in the calculated 95% CI of δ values. The whiskers represent the 95% CI produced by the standard errors.

Figure 3.3: ATTs (in terms of mortality rate) across the first 25 simulations for SCQE



3.2 DID Results

For DID, we consider that the control and treatment group have been observed both before and after the hypothetical treatment. We obtain the DID estimate by using an interaction term in an OLS model. This method imposes a constant diff-in-diff effect across units. The estimated DID parameter gives the estimated effect of the treatment. We also consider the uncertainty of these estimates.

Across the 3 conditions (heart attack, heart failure, and pneumonia) we obtain mean effect estimates of 0.0047, 0.0049, and 0.0063 respectively. Similar to in SCQE, when the DID interval includes the zero effect case we consider the coverage to be 1, and 0 in cases where it does not. Since DID provides a point estimate rather than a range of estimates like SCQE, we expect the coverage of this method to be lower. We also measure how often 95% CI contains entirely positive or entirely negative effect values. This proportion quantifies how often DID significantly over- or under-estimates the effect.

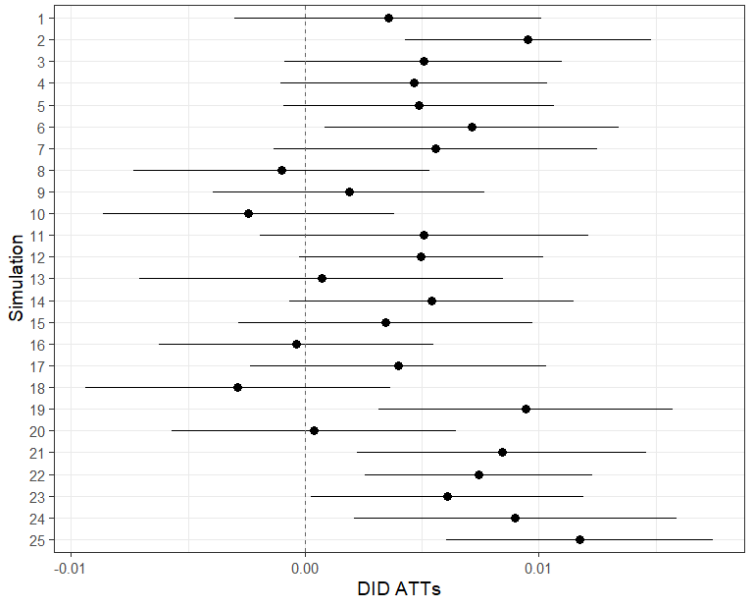
We also plot the ATT estimates obtained by DID for a subset of the simulation results

Table 3.2: DID Coverage with 95% CI

Condition	Percent of Intervals Including 0	Percent of Estimates Significantly Positive	Percent of Estimates Significantly Negative
Heart Attack	68.8%	31.0%	0.2%
Heart Failure	60.4%	39.6%	0.0%
Pneumonia	59.6%	40.4%	0.0%

in Figure 3.4. Instead of the boxes present in Figure 3.3, we observe points that correspond to the single effect estimate from DID. The whiskers represent the 95% CI produced by the standard errors.

Figure 3.4: ATTs (in terms of mortality rate) across the first 25 simulations for DID



CHAPTER 4

Conclusions

With the above results, we can conclude that the SCQE method protects against false inferences of estimating an effect when there is none much more consistently than DID does. In 100% of cases with a plausible range of δ assumptions, our SCQE effect estimate includes the zero effect case. The estimates do not have a clear bias toward over- or under-estimating the effect. The caveat to this result is that the estimates are less precise than the estimates obtained from DID. The 100% coverage is partially owed to the range of δ assumptions allowed, but we must still agree that the assumptions are reasonable for the coverage value to hold.

In contrast, DID had approximately 60%-70% coverage across the three conditions. There were very few simulations in which DID produced a significantly negative result. In cases where the effect estimate excluded 0, it was due to an overestimate. On average, DID biased toward a small, positive effect across all simulations that were done. The magnitude of this effect was similar to that of the maximum estimated SCQE effect. However, the frequent estimation above zero may translate to a rejection of further use of the placebo treatment in a real world scenario.

We also find that the two methods differ in what they require of the practitioner. DID is a slightly more straightforward in its application. Although the assumptions required are stringent, the calculation and consequent effect estimates are relatively easy to interpret. SCQE, on the other hand, demands more critical thought and decision-making during its application. Critical to the end result is the practitioner's choice of δ, and therefore more

preparation analysis is required in setting up the problem. While interpreting the result, one must also be cognizant of the assumption required in order for the estimated effect to be valid. This can have both desirable and undesired consequences. When the practitioner has a solid grasp on a reasonable δ range, inferences are robust. However, a moral hazard is created in that an individual could use the flexibility within the framework to bend the effect estimates to more closely fit a desired outcome.

A potential threat to the validity of inferences from both methods would be any systematic differences in the characteristics of those who receive treatments and those who do not. Although the methods can both accommodate non-random selection into treatment, the parallel trends and stability assumptions might be violated in cases where there is an underlying factor that affects who does and does not receive treatment. This would interfere with our interpretation of the effects. This is not a concern in this particular case since we simulate selection into treatment. There are no known underlying characteristics in this hospital dataset that could have an unintended impact on the results.

In short, we find that SCQE more reliably protects against false inferences about the effect of a placebo treatment. We recommend continuing to compare these two methods in a wider variety of scenarios to see if these findings hold. However, we believe that choice of method in situations similar to ours depends largely on the practitioner's preferences. SCQE is the best choice when parallel trends is not a reasonable assumption. Based on these results, it also seems to fare better than DID when there is a biased process of selection into treatment. The primary threat to SCQE's usefulness in this scenario is that its estimates are less precise. DID may be preferable in cases that adhere to the parallel trends requirements and have a random selection process, since it provides a single point estimate that is easier to interpret.

CHAPTER 5

Future Work

One clear weakness of this work is that there was a known absence of effect. Although this provided an advantage in being able to accurately assess the outcomes of the two methods, it is not representative of a realistic scenario. Two obvious extensions would be to also analyze: 1) a case where there is a known, non-zero effect and 2) a case where there is an unknown, posited nonzero effect. In the first, we would examine how often they reject a zero effect in favor of an effort estimate in the correct direction. These two cases might provide additional information about the methods' sensitivity in detecting an effect, instead of solely examining the methods' protection against false inferences.

In the interest of further investigating protection against false inferences, we may want to extend this analysis to a case where there is less consistency in the pre-intervention trends. For each of the conditions in our application, the change in mortality rate was consistently centered around 0. In a situation with more fluctuation in pre-intervention trends, it may be more difficult to propose the plausible δ range and therefore more difficult to obtain valid inferences. We recommend further study on SCQE's efficacy in cases with less clear and consistent pre-intervention trends.

It would also be worthwhile to compare to additional methods. Another alternative that can control for non-parallel trends is the usage of matching estimators [7]. This differs from the stability assumption required for SCQE. It instead takes a subset of the treated and non-treated groups that have a similar pre-intervention pattern, and compares within this subset. Often, researchers match not only on the pre-intervention outcomes and observables, but also

use synthetic control methods and weighting. This essentially amounts to considering the non-treated group as the counterfactual for the treated counterparts. However, it has been found in some cases to actually increase estimator bias instead of reduce it. This effect appears to be largely based on research context. It may be worthwhile to conduct a more in-depth investigation of this matching approach to the SCQE approach in the case where parallel trends is violated.

REFERENCES

[1] A Linden A Ryan E Kontopantelis. "Now trending: Coping with non-parallel trends in difference-in-differences analysis". In: *Statistical Methods in Medical Research* (2018).

[2] J Dimick A Ryan BK Nallamothu. "Medicare's public reporting initiative on hospital quality had modest or no impact on mortality from three key conditions". In: *Health Aff* (2012).

[3] JD Angrist and J Pischke. *Mostly harmless econometrics: an empiricist's companion*. Princeton: Princeton University, 2009.

[4] C Brewin and C Bradley. "Patient preferences and randomised clinical trials". In: *BMJ: British Medical Journal* (1989).

[5] C Hazlett. "Estimating causal effects of new treatments despite self-selection: The case of experimental medical treatments". In: *Journal of Causal Inference* 7 (2019).

[6] T Speed J Splawa-Neyman D Dabrowska. "On the application of probability theory to agricultural experiments. Essay on principles." In: *Statistical Science* 5.4 (1990), pp. 465–472.

[7] A Linden. "A matching framework to improve causal inference in interrupted time-series analysis". In: *Eval Clin Pract.* (2015).

[8] M Olschewski and H Scheurlen. "Comprehensive cohort study: an alternative to randomized consent design in a breast preservation trial". In: *Methods of information in medicine* (1985).

APPENDIX A

R Code

In the interest of reproducibility and as a demonstration of the SCQE package, we include the R function for our simulations below. We exclude the data cleaning and visualization steps for brevity.

```
library(did)
devtools::install_github("chadhazlett/scqe", force=TRUE)

# function to simulate SCQE/DID
    # and collect results for a given condition
simulate_methods <- function(df, deltas=deltas, n=10) {
  df$probs<-0
  df$trends<-0

  # specify selection probability variables for each provider
  for (i in 1:length(unique(df$provider))){
    trend <- (df[i,]$date4 -df[i,]$date3) # trend of T=1 - T=0
    a <- 75 # arbitrary constant
    trend_avg <- -0.0005285755 # average of all trends
    prob_selection <- exp(a*(trend-trend_avg))/
                      (1+exp(a*(trend-trend_avg)))
```

```
  df[i,]$probs <- prob_selection
  df[i,]$trends <- trend
}

# initialize variables
scqe_effects <- NULL
scqe_low_se <- NULL
scqe_high_se <- NULL
scqe_low_delta <- NULL
scqe_high_delta <- NULL
did_effects <- NULL
did_se <- NULL

for (i in 1:n){
  df_i <- NULL

  # select hospitals into treatment
  df$treated <- rbinom(nrow(df), 1, prob=df$probs)

  # random selection into simulation
  hospitals_select <- sample(1:nrow(df), 100, replace = F)
  df_i <- df[hospitals_select, ]

  # SCQE result estimate and standard errors
  post <- c(rep(0, nrow(df_i)), rep(1, nrow(df_i)))
  tx <- c(rep(0, length(df_i$treated)), (df_i$treated))
  y <- as.numeric(c(df_i[,4], df_i[,5]))
```

```
scqe_result <- scqe(treatment = tx,
                    outcome = y,
                    delta=deltas,
                    post = post)
result <- summary(scqe_result)

# Effects
scqe_effects_i <- as.numeric(result$critical.points[3])
scqe_effects <- c(scqe_effects, scqe_effects_i)

# SEs
scqe_low_se_i <- as.numeric(result$full.results[2,5])
scqe_high_se_i <- as.numeric(result$full.results[6,5])
scqe_low_se <- c(scqe_low_se, scqe_low_se_i)
scqe_high_se <- c(scqe_high_se, scqe_high_se_i)

# Upper and lower effect estimates
scqe_low_delta_i <- as.numeric(result$full.results[2,2])
scqe_high_delta_i <- as.numeric(result$full.results[6,2])
scqe_low_delta <- c(scqe_low_delta, scqe_low_delta_i)
scqe_high_delta <- c(scqe_high_delta, scqe_high_delta_i)

# DID result estimate and standard errors
did_result <- lm(y ~ tx + post + tx*post, data=df)
mod <- summary(did_result)
did_effects_i <- mod$coefficients[2]
did_effects <- c(did_effects, did_effects_i)
```

```
    did_se_i <- mod$coefficients[5]
    did_se <- c(did_se, did_se_i)

  }

  # return outputs as a list
  out <- list()
  out$scqe <- scqe_effects
  out$scqe_low_se <- scqe_low_se
  out$scqe_high_se <- scqe_high_se
  out$scqe_low_delta <- scqe_low_delta
  out$scqe_high_delta <- scqe_high_delta
  out$did <- did_effects
  out$did_se <- did_se
  out$probs <- probs
  invisible(out)

}

# run simulations
y_hf<-simulate_methods(df=hf_long, deltas=hf_deltas, n=500)
y_ha<-simulate_methods(df=ha_long, deltas=ha_deltas, n=500)
y_pn<-simulate_methods(df=pn_long, deltas=pn_deltas, n=500)
```