FIN 550: Big Data Analytics

Problem Set #2

Select whether this is an individual or group submission. No more than 3 members per group. Beyond the fact that all group members may submit the same answers, each submission must be separate work.

⧠ Individual Submission  
⧠ Group Submission. List group member names:\_\_\_\_\_ Yunzhe Yu, \_ Ives He, Hanbin Yan \_\_\_\_\_\_\_\_\_\_\_

**Problem set deliverables**

You should submit the following three files as part of your problem set solution:

1. A completed version of this file, containing group member names and solutions to Problem 1.
2. A file named “Case-Executive-Summary.pdf” with the executive summary report for Problem 2.
3. An R script named “Case-Code.R” for Problem 2.

# Difference-in-Differences (30 points)

The primary intent of medical malpractice law is to protect patients against professional negligence by a health care provider, which results in injury or death to the patient. However, proponents of medical liability reform argue that in fact these laws limit patient access to health care by driving doctors out of business or encouraging doctors not to use high-risk but potentially beneficial procedures. On June 11, 2003, Texas Governor Perry signed House Bill 4, a medical liability reform that greatly limited the amount of damages for which a physician could be held liable. You are provided the data table below, which indicates the number of doctors per 100,000 patients for Texas as well as for states neighboring the Lone Star State (nickname for TX).

|  |  |  |
| --- | --- | --- |
| State | Year | Doctors |
| Texas | 1998 | 152 |
| Texas | 2002 | 158 |
| Texas | 2006 | 175 |
| Neighbors | 1998 | 196 |
| Neighbors | 2002 | 189 |
| Neighbors | 2006 | 180 |

1. (16 points) Propose an estimate of the impact of Bill 4 on the number of doctors (per 100,000 patients) using only data for Texas.
   1. Provide a brief description of your method.

We employ a “before and after” estimator to analyze the impact of Act 4 by comparing the number of physicians in Texas prior to and following the legislation. Specifically, we use 2002 as the baseline (“before”) period, with 158 physicians per 100,000 residents, and 2006 as the follow-up (“after”) period, with 175 physicians per 100,000 residents.

* 1. Does this estimate suggest Bill 4 increased or decreased the number of practicing physicians? By how much?

Impact = Post-reform value - Pre-reform value

= 175 - 158

= 17 doctors per 100,000 patients

This estimate suggests Bill 4 increased the number of practicing physicians by 17 doctors per 100,000 patients, representing about a 10.8% increase from the pre-reform level.

* 1. Discuss the key assumption required for your estimate to be valid (i.e. no bias).

The key assumption underlying this analysis is that, in the absence of Bill 4, the number of physicians in Texas would have remained at the 2002 level (158 per 100,000 residents). In other words, we attribute any change observed between 2002 and 2006 to Bill 4, rather than to other factors. This is commonly referred to as the "parallel trends" assumption—the notion that the pre-intervention trend would have continued unchanged if the policy had not been implemented.

* 1. Discuss a scenario under which this assumption would be violated.

If Texas was experiencing general population and economic growth during this period, it might have naturally attracted more physicians, regardless of Bill 4. In this case, some or all of the observed increase in physician numbers could have occurred even without the reform, leading to an upward bias in our estimate. For instance, if Texas was already expanding its medical sector through the growth of hospitals or medical schools, this expansion would have independently contributed to the increase in physician numbers, unrelated to the liability reform.

This scenario underscores the risk of relying solely on before-and-after comparisons without accounting for other underlying trends or incorporating appropriate comparison groups.

1. (16 points) Propose an estimate of the impact of Bill 4 on the number of doctors using only data for 2006.
   1. Provide a brief description of your method.

We will employ a cross-sectional estimation approach to compare Texas with its neighboring states in 2006 (post-reform):Texas (treatment group): 175 physicians per 100,000 residents. Neighboring states (control group): 180 physicians per 100,000 residents

* 1. Does this estimate suggest Bill 4 increased or decreased the number of practicing physicians? By how much?

Impact = Texas value - Neighbors value

= 175 - 180

= -5 doctors per 100,000 patients

This estimate suggests Bill 4 decreased the number of practicing physicians by 5 doctors per 100,000 patients (or by about 2.8% compared to neighbors' level).

* 1. Discuss the key assumption required for your estimate to be valid (i.e. no bias).

The key assumption is that, in the absence of Bill 4, Texas would have had the same number of physicians per 100,000 residents as its neighboring states in 2006. In other words, any observed difference between Texas and its neighbors in 2006 is attributed to Bill 4 rather than other factors. This approach relies on the assumption that Texas and its neighboring states are comparable in all relevant aspects, except for the implementation of the reform.

* 1. Discuss a scenario under which this assumption would be violated.

A clear scenario in which this assumption would be violated is if Texas and its neighboring states had systematically different physician levels even before Bill 4 due to factors such as: Differences in state-level healthcare policies, Variations in population demographics, Levels of urbanization, Cost of living, Medical education capacity.

In fact, data from 1998 shows that Texas already had fewer physicians per capita compared to its neighbors (152 vs. 196), indicating pre-existing differences between the states. As a result, our cross-sectional comparison in 2006 is likely biased, as it attributes all observed differences to Bill 4, while many of these differences were present even before the reform.

1. (16 points) Instead, construct a difference-in-differences estimate of the impact of Bill 4 on the number of doctors.
   1. Does this estimate suggest Bill 4 increased or decreased the number of practicing physicians? By how much?

First difference (Texas):

2006 - 2002 = 175 - 158 = +17 doctors per 100K

Second difference (Neighbors):

2006 - 2002 = 180 - 189 = -9 doctors per 100K

DiD = (Texas change) - (Neighbors change)

= 17 - (-9) = +26 doctors per 100K

This estimate suggests Bill 4 increased the number of practicing physicians by 26 doctors per 100,000 patients. This represents about a 16.5% increase relative to Texas's pre-reform level.

* 1. Discuss the key assumption required for your estimate to be valid (i.e. no bias).

The analysis relies on the "parallel trends" assumption: in the absence of Bill 4, Texas would have experienced the same trend in physician numbers as its neighboring states. This assumes that any differences between Texas and its neighbors that could affect the growth rate of physicians are either constant over time or accounted for within the estimated reform effect.

* 1. Discuss a scenario under which this assumption would be violated.

The assumption would be violated if Texas and its neighbors were already on different trajectories prior to the reform. For instance, if Texas had been implementing other healthcare reforms or experiencing faster economic growth that would have naturally led to a faster increase in physician numbers compared to its neighbors, even without Bill 4, our estimate would be biased upward.

* 1. Set up a test to evaluate whether this key assumption appears to be plausible, using available data. Based on the results of this test, does the assumption appear to be valid?

We can test for parallel pre-trends by comparing trends before the reform (1998-2002):

Texas change 1998-2002:

158 - 152 = +6 doctors per 100K (+1.5 per year)

Neighbors change 1998-2002:

189 - 196 = -7 doctors per 100K (-1.75 per year)

The pre-reform trends indicate that Texas was gaining physicians while its neighboring states were losing them, even before Bill 4. This suggests a likely violation of the parallel trends assumption. Since the trends were diverging prior to the reform, our Difference-in-Differences (DiD) estimate may incorrectly attribute these pre-existing differences to the impact of Bill 4. As a result, the estimate of 26 more doctors per 100,000 residents is likely biased upward. We should be cautious in interpreting the DiD estimate as the causal effect of Bill 4.

# Teenage Driving and Mortality (70 points)

Complete the data case, “Teenage Driving and Mortality.” The case is available on Canvas. The case deliverables—an executive summary and R script—should be included with your problem set solutions.