Report 1 for Microeconometrics

Paper Review of *The State of Applied Econometrics*

Li Yuanchao

International Undergraduate Experimental Class

SOE, Xiamen University

April 17, 2022

1 Introduction

This report is based on Susan Athey and Guido Imbens' article[2] about causality and policy evaluation. I would discuss some traditional methods to do policy evaluation in this report. In addition, I have ran some basic regressions of these methods and I would also introduce more relevant researches and further methods about policy evaluation based on other 5 articles.

The remainder of this report would be organized as follows: part 2 reviews the article and discusses some traditional methods where I would do the regressions for DID and RDD and review another famous paper for peer effect. Part 3 is the further discussion about DID based on another 3 papers. In the end, I would give a short conclusion.

2 Basic Review

2.1 RDD Method

As for the RDD estimation, I personally think it is a very tricky method to identify casual inference, however its result may be very local. In other words, it has a strong internal validity but limited external validity.

This article mainly discusses 2 things about RDD. The first is a method called local linear estimation that involves estimating linear regression on the left and right of the threshold separately and then comparing the prediction at the threshold. The authors argue this method may have smaller bias compared with non-parametric method. In addition, this article also discuss many things about the choice of bandwidth.

I have done a simple RDD estimation according to the paper[1] mentioned in the course . To be specific, I want to test whether the rating in Dazhongdianping could motivate more visiting times using RDD method based on China's data. The data is acquired in the Peking University open data platform and being processed by myself. The codes and results are showing as follows.

```
bysort merchant: egen aver_rating=mean(rating) //generate average rating
keep if year == 2022 // keep only 1 year's data
bys merchant: egen numb=count(rating) //calculate total visiting times
bysort merchant: gen filter=_n
keep if filter == 1  // keep one dimension data
gen ln_numb=ln(numb)  // ln the visiting times
scatter ln_numb aver_rating
drop filter
rd numb aver_rating,z0(4.0) gr mbw(100) //RDD estimation best bandwidth
```

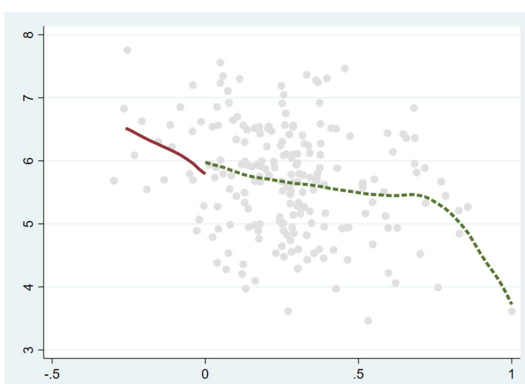

Figure 1: The Result of RDD Estimation

	(1)
Variables	ln_numb
lwald	0.19
	(0.48)
Observations	206

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: The Result of RDD Estimation

According to the above result graph and table, I find that around the 4.0 rating, there exists a slight increase for the visiting times, however the result is not significant statistically.

2.2 DID Method

I personally think that DID is the most important and popular method to do policy evaluation nowadays. In this section, I would firstly summarize the contents about DID in this article, then do a very basic DID estimation practice based on Kruger and Card's data set. In the next part of this report, I would briefly discuss more topics of interest about DDD and DID with variation in treatment timing.

This article has introduced two recent developments to the DID approach, which are the synthetic approach and the nonlinear changes-in-changes method.

As for the synthetic method, it suggests us to use a weighted average of the control sets instead of a simple control or a simple average of the sets. I personally think this would be a good approach to help us construct a better control group that is making a better guarantee for the parallel trends before shock. In addition, it could also help us do a easy selection among the control groups candidates. However, it has introduced a new question that how to settle the exact values for the weights. This paper lists some new-developed method such as minimum distance approach.

When it comes to the nonlinear changes-to-changes method, the authors say that this approach does not rely on functional form assumption for the observed entities are different, I have to admit that I do not understand it clearly, I may think it refers to the non-parametric method.

In addition, I have finished a basic DID estimation based on Kruger and Card's data set, the codes and result are as follows.

```
use "http://fmwww.bc.edu/repec/bocode/c/CardKrueger1994.dta" //import data
gen did = t* treated //genereate did variable
reg fte t treated did, r //regression without control
//diff fte, t (treated) p (t) robust
reg fte t treated did bk kfc roys wendys, robust //adding control
//diff fte, t (treated) p (t) cov( kfc bk roys wendys) robust
outreg2 using intro1.xlsx, replace tstat bdec(2) tdec(2) ctitle(fte)//export result
```

	(1)
VARIABLES	FTE
DID	-2.40*
	(-1.70)
Treated	-2.32*
	(-1.85)
Diff	2.94*
	(1.90)
KFC	-9.20***
	(-10.24)
BK	0.92
	(0.98)
ROYS	-0.90
	(-0.86)
WENDYS	0.00
	(.)
Constant	21.16***
	(16.19)
Observations	801
R-squared	0.188

Robust t-statistics in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2: The Result of DID Estimation

2.3 Peer Effect

In this section, I would briefly discuss another very popular topic, that is peer effect. I suppose the objective of this field is trying to explore how people's behavior would be affected in a social network. Many relevant works have been accomplished about this effect in classrooms, labor markets and many other atmospheres.

$$Y_i = \beta_0 + \beta_{\bar{Y}} \cdot \bar{Y}_i + \beta'_X \cdot X_i + \beta'_{\bar{X}} \cdot \bar{X}_i + \beta'_Z \cdot Z_i + \varepsilon_i \tag{1}$$

This equation is the most basic model of peer effect where $\bar{Y}_i$ is the average outcome of the peer group. In addition, X_i and $\bar{X}_i$ are the individual characteristics, I think we need control them since they are confounders. Z_i is the group's characteristics. After all, the β would be a credible estimate of the peer effect.

Afterwards I have picked an article[4] about peer effect from AER . I had hoped that I could do the replication for it since it has released its codes and data. But I have to admit that since the lack of time and ability, I do not understand its codes clearly. Hence, I would only describe this article's contributions and methods instead of codes and results.

There have been many articles exploring the peer effects under ordinary workplaces, however, the articles exploring the peer effects under workplace with high financial incentive and elite players are very limited. The authors have used the data from professional golf tournaments to test for the peer effects in the professional workplaces. Afterwards, they find no evidence that the ability of partners would affect the performance of professional golfers.

$$Score_{iktr} = \alpha_1 + \beta_1 \cdot Ability_i + \gamma_1 \cdot Abi\bar{l}ity_{-i,kt} + \delta_{tc} + \varphi_1 \cdot Abi\bar{l}ity_{-i,ct} + e_{iktr} \tag{2}$$

This is the model specification of the paper where i indexes players, k indexes groups, t indexes tournaments, c indexes categories, δ_{tc} is the a set of dummy variables where I think it likes the fixed effect. Because the players are randomly assigned to different groups based on categories(the authors have tested its randomization and also drops some not random data), the authors just use OLS to do the estimation. $Ability_i$ is the individual's ability, $Abi\bar{l}ity_{-i,kt}$ is the average ability, so γ_1 could measure the effect of the average ability of playing partners on own scores. $Abi\bar{l}ity_{-i,ct}$ is the average ability of the same-category players which are the population at risk to be the individual's peers. It has been controlled for that there may exists a bias generated by the negative correlation in predetermined characteristics of peers because individuals cannot be their own peers.(The authors also use Monte Carlo simulation to show that this bias could be reasonably large). After all these works, γ_1 would be considered as a plausible estimator of the peer effect.

Based on their regression results, they find that the coefficient on own ability is strongly significant which also follows intuitions. However, the estimate of the effect of peers on own score is not statistically significant, and the point estimate is actually negative. In addition, they change other methods to measure ability to do the robustness tests and also test for two dimensions of players' skills. All the results show that there is no significant peer effect between professional golfers.

The author have given 3 potential interpretations for this phenomenon. Firstly, financial incentive may be the key factor for the players to avoid responding to social incentives; secondly, there may exists a selection that the players who are more good at dealing with peer pressure would be more likely to be the elite players; thirdly, the professional players may accept specific training not to be affected by social forces.

That is the main content for this paper and also the end of the discussion of peer effects.

2.4 Other Contents in the Paper

This paper also discuss many other aspects of casual inferences or rather I want to call them the components of a good reduced- form research. However, I would not mention them in details. I would only give some of my own thoughts.

In my opinion, to do a good reduced-form economic research, you first need an interesting topic, and then design a model to identify its casual inference clearly based on the available data. Furthermore, the required works would be the robustness test that could make your result more credible. A very popular method for robustness test is the placebo analyses, in addition you could also use any other methods to measure your variables or just change another model to do the estimation. Anyway there are many methods to do the robustness tests. Besides making your results more credible, I think another purpose for robustness tests is to verify which assumption is the most important or rather the weakest one, which may point out the direction of further researches. Then you also need to discuss the heterogeneity and the potential mechanisms of your topic which could make your research more vivid and convincing. In the end, you also need to discuss the problems of your research, maybe the external validity is not so good or any other things. I think you could find the author's opinions towards each part I mentioned above in this paper.

Another important part of this paper is the discussion of machine learning and casual inference. Actually, I have to admit that I am not familiar with and not interested in machine learning methods before this course. But I think it is necessary for an economist to understand this method, because nowadays we have more and more data, it would be important for economists to deal with them and to identify some features from them. As for the machine learning itself, I think casual inference is also an important field to do more research in the future. Present machine learning maybe pay more attention to accurate prediction, but I think causality is the key factor that researches really could provide some useful guidances for the real life. I think this paper make a very good introduction towards the combination of machine learning and casual inference.

3 Additional Discussion

In the last part, I would give further discussions based on more paper about the DID method.

3.1 Is DID Really a Good Method to do Policy Evaluation

At the beginning of this part, I would list some questions for the DID method. Some scholars (like Anthony Lee Zhang of Chicago Booth) argue that the policy evaluation based on casual inference mainly referring to DID estimation is a very conservative method. That is, the research only focus on the issued policies that may not be the aggressive one and cares about the slight marginal change, however, take the more macro environment as given. As a result, these kinds of methods could not motivate us to propose entirely new policies that may change our lives totally.

In addition, there may also exist a selection process that only the significant policy evaluation researches could be published. That may lead to a phenomenon that most of researches could suggest that the existing policies are efficient and motivate the policy designers to maintain the status quo.

One alternative solution maybe is the policy experiment, but there also exists severe moral hazard about policy experiment designs. Besides the moral hazards, I may think it is impossible to implement the explicit policy experiments in countries that governments do not have strong power such as U.S. By contrast, I think it is feasible to do the policy experiments in China due to the strong intervention power of China's government.However, Shaoda Wang and David Y. Yang also point out another potential problems[6] to do policy experiments in China. The first problem is the selection bias, that is the selection of experiment places is not random. They find it is related to two factors, one is whether the local officials have the strong willingness to get promoted, the other factor is the closeness of ties between local officials and officials of specific central departments. As a result, they find that the developed cities are much likely to become the policy pilot cities, which may make the experiment sample could not be the good representatives. In addition, they also find the empirical evidence that local officials would invest more resources to ensure the success of the pilot policy due to the promotion incentives, which may lead to the inconsistent results when the policy is introduced nationwide. Both these two problems may make the result of policy experiment not so plausible.

To solve this problem to some extent and make a better evaluation towards policy experiments, I want to introduce a more advanced methodology called difference-in-difference-in-differences in the coming section.

3.2 Difference-in-Difference-in-Differences

To be honest, the first time I heard about DDD method is in a lecture of professor Fang Ying, he introduced us this method using one of his paper. So I just have found the paper[5] he mentioned in that lecture, and plan to introduce the DDD method using this article.

From my point of view, I think the situation we need to use DDD method is that there exist a confounder that would violate the parallel trend assumption and also could not be controlled using normal ways. Then we could introduce a new group that also have been affected by this confounder but indifferent with the policy shock. Then we could do the subtraction for three times to eliminate the impact of this confounder.

Taking Guo Junjie and Fang Ying's paper as an example, they want to test whether the China's pollution levy standards reform promote industrial SO2 emission reduction. Initially they have designed a DID model to do the evaluation, but they find that there exists a confounder damaging parallel trend that is the local governments' commitments to environmental protection. Unfortunately, there is no indicator for this covariable and also could not control it by two-way fixed effect. In order to solve this problem, they have selected the industrial dust as an additional variables and then do the DDD estimation. The reason for choosing industrial dust is that the industrial dust is also affected by the local governments' commitments but not affected by the SO2 restriction policy. Hence, after adding the industrial dust, the regression could give us a plausible estimation.

3.3 DID in Multiple Time Periods

Besides the problem that the parallel trend is very difficult to assure, there is also another severe problem about the DID method. That is, in the real life, the timing of some policy shocks varies from place to place, then it is impossible to set a standard DID model with $Treat_i$ and $Post_t$. To solve this kind of problem, some scholars have proposed some new designed DID models to evaluate multiple time periods policy shocks.

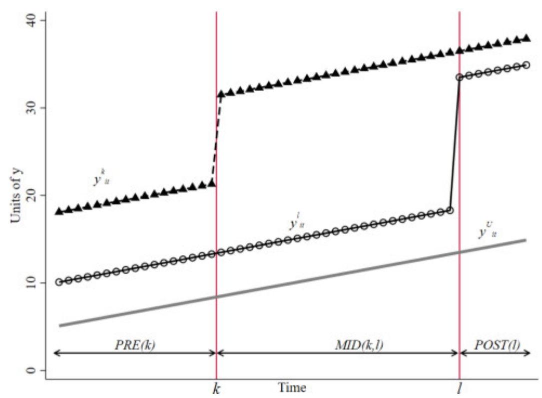

Figure 2: Three Groups Example

$$y_{it} = \alpha_i + \alpha_t + \beta^{DD} \cdot D_{it} + e_{it} \tag{3}$$

This is the most initialized model specification where the D_{it} is the dummy variable set of each entity and time period while α_i and α_t is the traditional two-way fixed effects just like the normal DID method.

However, Goodman-Bacon[3] argues that the traditional TWFE model would lead to a biased result. Afterwards, he proposes that the policy effect should be the weighted average of treament effects for each subperiod. He takes a three groups case as an example.

We could see that, as for the two policy shocks, one unaffected group and two treatment groups example, we could separate it into four 2*2 normal DID. Then we could do the normal

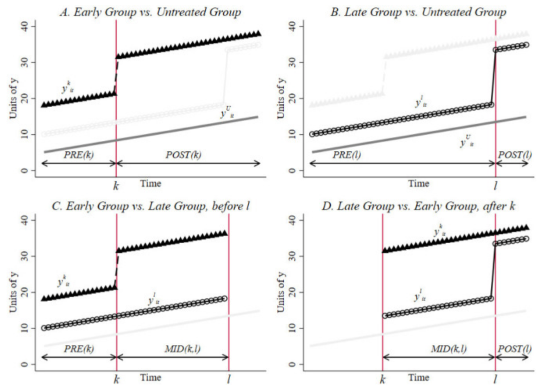

Figure 3: Four Simple 2*2 DID Estimates

DID estimation for 4 times and take the weighted average of the treatment effects to get our policy effect. As for the exact weights, Goodman-Bacon have proposed a new method to calculate them which is abstract and difficult for me to understand it clearly, so I would not list the mathematical details about that method here.

As for the disadvantages of this method, firstly it requires much stronger assumptions. Besides the parallel trend assumption, it also requires that the effect of policy should keep constant towards time. We could see the lower right DID estimation of the above image. The control group of this DID is the group which is affected by the first policy shock but unaffected by the second. I personally think only under very strong assumptions, this group could be a qualified control group. Secondly, each sub DID estimation would only use a very small part of the whole database which may make our estimation much "local".

4 Conclusion

Actually, I do not know what to write in the conclusion part for a review report, so I decided to write some not so professional thoughts in this part. To be honest, I do not think my report should be called as the review report since the majority of my report may be not so correlated with the given article although I have read all the contents of this article. I would rather call the given paper as the hook or the clew, I just try to explore many things about casual inferences of my interest.

The part in this report that really makes me impressive should be the little research in the 2.1 RDD part. Although in the report, I only display some codes, a graph and a table. I actually have spent much time on it and have experienced the feeling for doing a tiny economic research. When the teacher mentioned that original paper in the course, I was impressed by the idea. I am a heavy user of Dazhongdianping and really like to go to some restaurants based on this app. So I was very happy and keen to replicate this research using China's data. However, the data was not easy to get. I had spent much time on searching and dealing with the data. Finally, I accomplished all the works of data and run a RDD regression. Unfortunateky the regression showed me a very strange and confusing result which made me very disappointed. One day later, I suddenly realized what the problem is. After the adjustment, the result finally became much "normal" although it is not significant.

In a word, I am very glad that I have done many explorations to accomplish this report and I think I really have a much deeper understand of casual inferences.

References

[1] Michael Anderson and Jeremy Magruder. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, 122(563):957–989, 2012.

[2] Susan Athey and Guido W Imbens. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32, 2017.

[3] Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

[4] Jonathan Guryan, Kory Kroft, and Matthew J Notowidigdo. Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68, 2009.

[5] Guo Junjie, Fang Ying, and Yang Yang. Does china's pollution levy standards reform promote emissions reduction? *World Economy*, 42(1):121–144, 2019 (in Chinese).

[6] Shaoda Wang and David Y. Yang. Policy experimentation in china: The political economy of policy learning. *National Bureau of Economic Research*, 2021.