Topics in Econometrics and Statistics Project
Doubly-Robust Difference in Differences Estimators

Omer Erhan Erbis
Matrikelnummer: 3466411
Universität Bonn

August 2022

Abstract:

Treatment effects measurement is an important task that serves many purposes for researchers and policy makers. For this purposes various Difference in Differences methods are in wide use for identification of the treatment effects. The emergence of the idea of Doubly-Robustness lit a spark for identification of the treatment effects in Difference in Differences setup and more robust models have been proposed throughout the existence of the idea. In this project, following Sant'Anna and Zhao's(2020) Semiparametric Doubly-Robust for inference procedure, the estimation Average Treatment effect on the Treated via Doubly-Robust Difference in Differences methodology will be conducted and the performance will be analyzed via Monte Carlo simulations. The procedure indeed is Doubly-Robust both for consistence and inference, it also improves upon the efficiency of the traditional Doubly-Robust estimator.

Keywords: treatment effect, difference in differences, doubly-robust, inference, semiparametric models

JEL Classification Codes: C01 C13 C14 C15 C31

Contents

1 Introduction

In the life, the causal relationship is one of the most sought out interaction between events or material objects by the scientists. This relationship defines the link between actions and objects and thus helps to understand the "way of the world". As "economics" tries to establish the most rational response in micro and macro world per neoclassical viewpoint dictates, the strong interest of economics on this link is undeniable. For other perspectives, this establishment is again beneficial for criticism and argument validity control mechanism. In addition, as a field of economics, "[e]conometrics uses economic theory, mathematics, and statistical inference to quantify economic phenomena"[1]. Thus the econometrics is deeply concerned with links between these phenomena.

Most of the time, however, the causal link might not be present or can be hard to establish. These bring the science to the definition of "correlation" where it defines the statistical linear relationship of phenomena of interest. Correlation is useful due to availability between each phenomena and it can also lead to find possible underlying causal links. Especially for researchers which are interested in specific relations often resort to this statistical variable to evaluate their ideas for possible explanations of the phenomena of interest.

In decision making process, more strongly pronounced in the interest of the policy-makers, the implementation of a suggested option often requires investment. This "choosing" action is thus crucial for the responsible party, and thus the definition of relationship between the decision and desired outcomes for future is desired to be identified as correctly as possible. The impact of a choice, statistically defined as "treatment effect" thus carries high importance in this process.

However, measuring the treatment effect poses quite a challenge. First of all, the treatment effect varies within observations, even if other aspects are controlled for. Thus, most of the time the essence of the impact is searched for in the Average Treatment Effect (ATE). Secondly, the treatment effect can be observed after the introduction of treatment: there is a unique distinction between pre-treatment and post-treatment observations. To measure the treatment effect, one has to eliminate all other observation state group effects (e.g. outcome evolution in time for panel data). Thirdly, there is an Observed Outcome Puzzle, which is also referred as "fundamental problem of causal inference" by Holland (1986) which originates from the law of the universe: Only one outcome can be observed for each subject, either treated or untreated. Unfortunately, even the quantum physics' "Quantum Superposition Principle" does not apply in econometrics since "Copenhagen Interpretation" clearly dictates that the quantum state is preserved until interaction with, or observation by the external world and econometrics experiments require observation.[2] This prevents the tests and any interpretation of the treatment, but by the Differences in Differences (DID) methodology with some (conditional) assumptions invoked, these issues can be overcome.

Difference in Differences setup is thus very useful in determining the effect of a policy application and is widely used in many applications. As in all of the econometrics methodologies, the identification is a crucial aspect to explore for the models. The Difference in Differences methodology was utilized under different approaches that identified the treatment effect with different models for different cases, but the consistency has remained a problem of correct specification of underlying model by researcher. The Doubly-Robust estimators that combined different approaches under one and ensured consistency provided either combined model being correctly specified has

[1] https://www.imf.org/external/pubs/ft/fandd/2011/12/basics.htm
[2] https://en.wikipedia.org/wiki/Schrödinger%27s_cat

proved to be a strong yet partial remedy for this particular issue. And after the introduction of doubly-robust for consistence Difference in Differences estimators, the correct identification for models became more likely and thus these estimators have been widely used in more contemporary research. However, further research revealed that the interest in the stabilization of variance with different model estimations in doubly-robust for consistence estimators could be quelled by doubly-robust for inference estimators with specific choices.
Thus, Doubly-Robust for Inference Difference in Differences estimators are of high value and the determination of these estimators is interesting. Throughout this project, the approach of Doubly-Robust for Inference Difference in Differences estimators will be based on the research of Sant'Anna and Zhao(2020).

2 Literature Review

The establishment of the causality link for researchers is crucial in understanding the dynamics of the world. In most experiments, observation of the causality link is flawed by additional introduction that affects the outcome e.g. selection bias. Neyman's "On the Application of Probability Theory to Agricultural Experiments" (1923) Fisher's revolutional "Statistical Methods for Research Workers" (1925) papers, the introduction of Randomized Control Trials(RCT) in statistics and econometrics literature provided the the basic setup for clean treatment effect observation. With the randomization, the selection bias was nullified and the only remaining difference between groups were reduced to the causal treatment introduction effect. The downside was that the framework did not fit data structure or sometimes was inapplicable for experiments, therefore the idea was not a complete solution. This pivotal work was utilized and is still being utilized in many modern applications.
The literature developed and starting with Rubin's master thesis, and later works in 70s, now widely used "potential outcome", which refers to "pairs of outcomes defined for the same unit given different levels of exposure to the treatment"[3], was introduced and formalized by Holland(1986) as Rubin Causal Model. Labour economics and development economics mostly benefited from these developments in the literature due to natural fit of objectives.
Ashenfelter(1978) and Ashenfelter and Card(1985) established the solid ground of treatment effect identification via Difference in Differences method. In their empirical post-schooling research, they found the unpredicted earnings decline prior to training program, thus they adopted the Difference in Differences method and proved the efficacy of the approach. Many later followed their footsteps, such as Card(1990), Heckman(1990), Card and Krueger(1993) and conducted impactful research. The Outcome Regression model by Heckman et al(1997) and Inverse Probability Weighting approach by Abadie(2005) alongside many others were proposed as different Difference in Differences models for treatment effect estimation that proved to be influential in the literature. Non-parametric and parametric identification methods were proposed in Difference in Differences setting, and the literature highly developed for semiparametric identification due to ease of interpretability while providing enough flexibility. With the emergence of Doubly-Robustness, the concept was applied to Difference in Differences settings e.g. Robins(2000) and enabled estimation of more consistent estimators of treatment effect.

[3] Imbens and Woolridge(2009)

3 Data Structure

The data for the Difference in Differences method can be narrowed down under two main categories: panel data and repeated cross-section data. The panel data naturally tracks the previous observations and obtains two outcomes for both before and after the treatment while repeated cross-sectional data does not. Most of the time, the property of the panel data makes it more attractive and more reliable in finding the treatment effect correctly, but in some instances the collection of such data may be too costly or not viable. For these instances the repeated cross-section data is collected and used in assessing the treatment effect. However, since the repeated cross-section data is less restricted and does not track unique observations as panel data does it allows wider range of sampling schemes, thus it requires further adjustment in order to find the correct treatment effect.

4 Theoretical Approach

4.1 Notation

The notation for the ATT calculation can become quickly confusing, thus early introduction of some basic arguments can be helpful.

- i represents individual, t represents time.
- Y_{it} refers to outcome, $Y_{it}(0)$, $Y_{it}(1)$ refers to potential outcome.
- Observe that $D_{i0} = 0$ *for* $\forall i$. Thus once again reducing the notation, $D_{i1} = D_i$ refers to receiving treatment at $t = 1$.
- X_i refers to constant and pre-treatment covariates.
- T_i is a binary variable that takes $T = 1$ if the observation is in post-treatment period, otherwise $T = 0$
- $\lambda = P(T = 1) \in (0, 1)$ represents the probability of observation belonging to post-treatment period.
- The realized outcome in the panel data setup is:

$$Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$$

- The observed outcome in the repeated cross-sectional data setup is:

$$Y_i = T_iY_{i1} + (1 - T_i)Y_{i0}$$

Note: "i"s will be dropped in the rest to reduce notation load.

Additional notation will be introduced gradually in the related context.

4.2 Assumptions

In the Difference in Differences setup, there are certain assumptions that provides the necessary conditions for identification to be established. These assumptions are standard assumptions in the Difference in Differences setups, and are considerably mild assumptions.

With the data being either panel data and repeated cross-section, the data assumptions differ slightly.

Assumption 1 - ***A(1)*** *The Data Structure:*

$$\begin{aligned}
&\boldsymbol{A(1.a)} - \textit{If the data is panel data: } \{Y_0, Y_1, D, X\}_{i=1}^{n} \ \ \textit{i.i.d.} \\
&\boldsymbol{A(1.b)} - \textit{If the data is repeated pooled cross-section data: } \{Y, D, X, T\}_{i=1}^{n} \ \ \textit{is} \\
&\qquad \textit{from i.i.d. draws from the mixture distribution} \\
&\qquad P(Y \leq y, D = d, X \leq x, T = t) = t * \lambda * P(Y \leq y, D = d, X \leq x | T = 1) + \\
&\qquad\qquad\qquad\qquad (1-t) * (1-\lambda) * P(Y \leq y, D = d, X \leq x | T = 0) \\
&\qquad \textit{where } \ (y, d, x, t) \in \mathbb{R} \times \{0,1\} \times \mathbb{R}^k \times \{0,1\} \ \textit{and} \\
&\qquad (D, X) \ \textit{joint distribution invariant to } T.
\end{aligned}$$

It can be seen that the repeated cross-sectional data allows multiple sampling possibilities, however it restricts the the joint distribution of (D, X) changes as it can introduce bias and break down the estimation as argued in Hong(2013).

One of the key assumptions in Difference in Differences methods is the Parallel Trends Assumption (PTA). Since the counterfactual outcomes after-treatment assignment period could not be observed, the treatment effect is unidentifiable as it is, and Parallel Trends Assumption provides the flexibility to identify the treatment effect without strongly straining the inference link. Even though the PTA is plausible, compositional differences under characteristic sub-samples that make up treated and untreated groups can be criticised, and researchers came up with a milder alternative that is Conditional Parallel Trends Assumption (CPTA) that yields enough flexibility for identification. CPTA simply provides parallel trends conditioned on some characteristics.

Assumption 2 - ***A(2)*** - *Conditional Parallel Trends Assumption (CPTA):*

$$\mathbb{E}[Y_1(0) - Y_0(0) | D = 1, X] = \mathbb{E}[Y_1(0) - Y_0(0) | D = 0, X] \quad a.s.$$

One other necessary assumption is the Overlap Condition Assumption, which basically ensures the existence of observations for ATT calculation and provides necessary existence probability of observations under sub-populations. This is again a very common assumption in conditional DID setups.

Assumption 3 - ***A(3)*** - *Overlap Condition Assumption:*

$$P(D = 1) > \epsilon \textit{ and } P(D = 1 | X) \leq 1 - \epsilon \ , \quad \epsilon > 0$$

As mentioned, there are two DID approaches that are going to be considered for the calculation of the ATT. The first one that is going to be considered is the Heckman's Outcome Regression (OR model) and the second one is the Inverse Probability Weighting Regression model.

4.3 Outcome Regression

The most basic and intuitive approach in estimation of the ATT would be to model the outcome evolution. Although it was in the matching setting and more concerned in selection bias, the

outcome regression model by Heckman et al (1997) suggests that that the ATT definition with the potential outcomes first requires the basic assumptions for identification.
First of all, as n_{treat} refers to sample size of the treatment group, the ATT is defined as follows:

$$\tau = \mathbb{E}[Y_1(1) - Y_1(0)|D_i = 1] = \mathbb{E}[Y_1|D_i = 1] - \mathbb{E}[Y_1(0)|D_i = 1]$$

which is clearly unidentified as $\mathbb{E}[Y_1(0)|D_i = 1]$ term is a counterfactual value that is not available as previously discussed. This problem is suggestively solved with the use of the basic assumptions (consult Appendix 1.a.), and thus the ATT can be expressed as:

$$\tau^{reg} = \mathbb{E}[Y_1|D_i = 1] - \left[\mathbb{E}[Y_0|D_i = 1] + \frac{1}{n_{treat}} \sum_{i|D_i=1} (m_{0,1}(X_i) - m_{0,0}(X_i))\right]$$

which is identified where $m_{d,t}(x) = E[Y_t|D = d, X = x]$. $m_{0,0}(X_i)$ and $m_{0,1}(X_i)$ are population parameters, therefore can estimated in the Outcome Regression by $\hat{\mu}_{0,0}(X_i)$ and $\hat{\mu}_{0,1}(X_i)$ respectively, thus solving for the ATT. This yields the ATT estimation in Outcome Regression Model as such:

$$\hat{\tau}^{reg} = \frac{1}{n_{treat}} \sum_{i|D_i=1} Y_1 - \left[\frac{1}{n_{treat}} \sum_{i|D_i=1} Y_0 + \frac{1}{n_{treat}} \sum_{i|D_i=0} (\hat{\mu}_{0,1}(X_i) - \hat{\mu}_{0,0}(X_i))\right]$$

It also has to be noted that in case of sample imbalances (more probable in repeated cross-section data) n_{treat} values differ for different periods, thus should be calculated for each $t = 0, 1$ and used accordingly in sample analogues. This simple modelling is a clean solution for the ATT estimation, but as it is clear, the correct estimators for $\hat{\mu}_{0,0}(X_i)$ and $\hat{\mu}_{0,1}(X_i)$ can sometimes be hard to justify since the form of the paramater is unknown.

4.4 Inverse Probability Weighting Regression

The other approach that is going to be considered in ATT estimation would be the Inverse Probability Weighting method. Instead of focusing on correct modeling of the outcome evolution, IPW Regression approach focuses on the correct modeling of the Propensity Score as the "balancing score", that is the probability of a treatment assignment given the characteristics. By Rosenbaum and Rubin (1983) and Heckman et al (1997), it was established that with the correct weighting scheme, i.e. with "balancing score", one could impose the same distribution onto the temporal(realized) outcomes that the weighted difference(between $t = 0, 1$) of the temporal outcomes would yield the ATT. The ingenuity of this approach is that under some circumstances, correct modeling the outcome evolution can be more complicated than the modeling of the propensity score, thus IPW Regression approach being more desirable for some research than OR approach.
For panel data setup, the ATT under this proposal can be defined as:

$$\tau^{ipw,p} = \frac{1}{\mathbb{E}[D]} \mathbb{E}\left[\frac{D - p(X)}{1 - p(X)}(Y_1 - Y_0)\right]$$

and for repeated cross-section data setup, ATT can be defined as:

$$\tau^{ipw,rc} = \frac{1}{\mathbb{E}[D]} \mathbb{E}\left[\frac{D - p(X)}{1 - p(X)} \frac{T - \lambda}{\lambda(1 - \lambda)} Y\right]$$

where $p(X) = P(D = 1|X)$ is the propensity score, a population parameter (consult Appendix 1.b. for calculation).
With the IPW approach idea, Sant'Anna and Zhao(2020) propose the following IPW estimator for panel data setup for the ATT, suggested by Abadie(2005):

$$\hat{\tau}^{ipw,p} = \frac{1}{\mathbb{E}_n[D]} \mathbb{E}_n\left[\frac{D - \hat{\pi}(X)}{1 - \hat{\pi}(X)} (Y_1 - Y_0)\right]$$

and the following IPW estimator for repeated cross-sectional data setup for the ATT:

$$\hat{\tau}^{ipw,rc} = \frac{1}{\mathbb{E}_n[D]} \mathbb{E}_n\left[\frac{D - \hat{\pi}(X)}{1 - \hat{\pi}(X)} \frac{T - \lambda}{\lambda(1 - \lambda)} Y\right]$$

where $\mathbb{E}_n[.]$ implies the sample average of the variable and $\hat{\pi}(X)$ is an estimator of true propensity score. This estimation clearly differs from OR approach by focusing on $\hat{\pi}(X)$ estimator rather than $\hat{\mu}_{0,0}(X_i)$ and $\hat{\mu}_{0,1}(X_i)$ estimators, thus enabling researcher to have alternative method.

4.5 Doubly-Robust DID Regression

The aforementioned existing DID approaches model different objectives, thus depending on the circumstances, each may present an applicable use for calculation of ATT. However, instead of using each approach separately, integration of such models may be more optimal in finding the correct ATT value. Such models are called "Doubly-Robust" models, where approaches with different objectives are combined under one model that incorporates both, in a way that preserves consistency under the assumption of either one of the approaches is specified correctly (not necessarily both). This application eases the requirement for correct specification of underlying model by researcher's ability, that increases the probability of correct identification for the ATT by preserving consistency under correct specification of either model, thus gaining the double-robustness. With the previous OR and IPWR approaches, Sant'Anna and Zhao(2020) propose the following industry-standard DRDID ATT estimands in panel-data ($\tau^{dr,p}$) and repeated cross-section ($\tau_1^{dr,rc}$ and $\tau_2^{dr,rc}$) cases as:

$$\begin{aligned}
\tau^{dr,p} &= \mathbb{E}\left[\left(w_1^p(D) - w_0^p(D, X; p(X))\right)\left(\Delta Y - m_{0,\Delta}^p(X)\right)\right] \\
\tau_1^{dr,rc} &= \mathbb{E}\left[\left(w_1^{rc}(D, T) - w_0^{rc}(D, T, X; p(X))\right)\left(Y - m_{0,Y}^{rc}(T, X)\right)\right] \\
\tau_2^{dr,rc} &= \tau_1^{dr,rc} + \left(\mathbb{E}[m_{1,1}^{rc}(X) - m_{0,1}^{rc}(X)|D = 1] - \mathbb{E}[m_{1,1}^{rc}(X) - m_{0,1}^{rc}(X)|D = 1, T = 1]\right) \\
&\quad - \left(\mathbb{E}[m_{1,0}^{rc}(X) - m_{0,0}^{rc}(X)|D = 1] - \mathbb{E}[m_{1,0}^{rc}(X) - m_{0,0}^{rc}(X)|D = 1, T = 0]\right)
\end{aligned}$$

where:

$$
\begin{aligned}
&w_1^p(D) = \frac{D}{E[D]} \quad \text{and} \quad w_0^p(D,X;p) = \frac{\dfrac{p(X)(1-D)}{1-p(X)}}{E\left[\dfrac{p(X)(1-D)}{1-p(X)}\right]} \\
&w_1^{rc}(D,T) = w_{1,1}^{rc}(D,T) - w_{1,0}^{rc}(D,T) \quad \text{and} \quad w_1^{rc}(D,T) = w_{1,1}^{rc}(D,T) - w_{1,0}^{rc}(D,T) \\
&w_1^{rc}(D,T) = \frac{D \cdot \mathbb{1}\{T=t\}}{E[D \cdot \mathbb{1}\{T=t\}]} \quad \text{and} \quad w_0^{rc}(D,T,X;p) = \frac{\dfrac{p(X)(1-D)\cdot \mathbb{1}\{T=t\}}{1-p(X)}}{E\left[\dfrac{p(X)(1-D)\cdot \mathbb{1}\{T=t\}}{1-p(X)}\right]} \\
&\Delta Y = Y_1 - Y_0 \\
&m_{d,\Delta}^p(X) = m_{d,1}^p(X) - m_{d,0}^p(X) \\
&m_{d,\Delta}^{rc}(X) = \mu_{d,1}^p(X) - m_{d,0}^{rc}(X), m_{d,Y}^{rc}(T,X) = T \cdot m_{d,1}^{rc}(X) - (1-T)\cdot m_{d,0}^{rc}(X)
\end{aligned}
$$

and propose the following Doubly-Robust Differences in Differences Estimators:

$$
\begin{aligned}
\tau^{dr,p} =& \mathbb{E}\left[(\hat{w}_1^p(D) - \hat{w}_0^p(D,X;\hat{\pi}))\left(\Delta Y - \hat{\mu}_{0,\Delta}^p(X)\right)\right] \\
\tau_1^{dr,rc} =& \mathbb{E}\left[(\hat{w}_1^{rc}(D,T) - \hat{w}_0^{rc}(D,T,X;\hat{\pi}(X)))\left(Y - \hat{\mu}_{0,Y}^{rc}(T,X)\right)\right] \\
\tau_2^{dr,rc} =& \tau_1^{dr,rc} + (\mathbb{E}[\hat{\mu}_{1,1}^{rc}(X) - \hat{\mu}_{0,1}^{rc}(X)|D=1] - \mathbb{E}[\hat{\mu}_{1,1}^{rc}(X) - \hat{\mu}_{0,1}^{rc}(X)|D=1,T=1]) \\
&- (\mathbb{E}[\hat{\mu}_{1,0}^{rc}(X) - \hat{\mu}_{0,0}^{rc}(X)|D=1] - \mathbb{E}[\hat{\mu}_{1,0}^{rc}(X) - \hat{\mu}_{0,0}^{rc}(X)|D=1,T=0])
\end{aligned}
$$

where:

$$
\begin{aligned}
&\hat{\mu}_{d,t}^p(X) \text{ is estimator for true outcome regression } m_{d,t}^p(x) = E[Y_t|D=d, X=x] \\
&\hat{\mu}_{d,t}^{rc}(X) \text{ is estimator for true outcome regression } m_{d,t}^{rc}(x) = E[Y_t|D=d, T=t, X=x] \\
&\hat{\pi}(X) \text{ is an an estimator for true } p(X) \\
&\hat{w}_1^p(.), \hat{w}_0^p(.), \hat{w}_1^{rc}(.), \hat{w}_0^{rc}(.) \text{ are estimators for true (weights) } w_1^p(.), w_0^p(.), w_1^{rc}(.), w_0^{rc}(.)
\end{aligned}
$$

Theorem 1: Under the Assumptions A(1-3):

a.) $\tau^{dr,p} = \tau$ *if either* $\pi(X) = p(X)$ *or* $\mu_{0,\Delta}^p(X) = m_{0,\Delta}^p(X)$ *a.s.*

b.) $\tau_1^{dr,rc} = \tau_2^{dr,rc} = \tau$ *if either* $\pi(X) = p(X)$ *or* $\mu_{0,\Delta}^p(X) = m_{0,\Delta}^p(X)$ *a.s.*

With Theorem 1, the proposed DRDID estimators are consistent as long as one of the underlying nuisance models are correctly specified, which is the desired strength of the Doubly-Robust estimators. One should also note that $\tau_1^{dr,rc}$ does not rely on OR model, thus the robustness is expected to be weaker.

4.6 Improved Doubly-Robust DID Regression

The usefulness of Doubly-Robustness property is extensive and many current day applications, including Machine Learning examples,[4] make use of this handy property to improve results. However, most of the time, Doubly-Robustness property is used in consistency improvements and theoretically the variance requires further investigation for correction due to either model being misspecified changing the estimator variance. Under semiparametric models family, Sant'Anna and Zhao(2020) thus propose a estimation method , namely Improved Doubly-Robust Difference in Differences estimator, by extending Vermeulen and Vansteelandt (2015) idea. The proposed method guides through first-stage estimation phase of the process among many alternatives while aiming for Doubly-Robust for inference estimation.

Since proposed DRDID estimator requires estimation procedures for $\pi(X)$ and $\mu_{0,\Delta}$, the proposal of Sant'Anna and Zhao(2020) under semiparametric family for these are linear regression models for the outcome of interest, and logistic model for the propensity score due to commonality of use. They also propose inverse probability tilting estimator[5] and the weighted least squares estimator for the first-stage estimators to obtain insensitivity to estimation effect from the first stage that proves necessary for Doubly-Robust for Inference property. For the asymptotic properties(asymptotic normality) of the estimator and influence function calculations that satisfy Doubly-Robust for Inference property, one can consult to Sant'anna and Zhao(2020).

Sant'Anna and Zhao(2020), combining all these, proposed the following three-step, two-stage, Improved Doubly-Robust DID Estimators for panel-data ($\hat{\tau}^{dr,p}$) and repeated cross-section ($\hat{\tau}_1^{dr,rc}$ and $\hat{\tau}_2^{dr,rc}$) cases:

$$
\begin{aligned}
\hat{\tau}^{dr,p}_{imp} =& \mathbb{E}_n \left[\left(\hat{w}_1^p(D) - \hat{w}_0^p(D, X; \hat{\gamma}^{ipt})\right) \left(\Delta Y - \hat{\mu}^{lin,p}_{0,\Delta}(X; \hat{\beta}^{wls,p}_{0,\Delta})\right)\right] \\
\hat{\tau}^{dr,rc}_{1,imp} =& \mathbb{E}_n \left[\left(\hat{w}_1^{rc}(D,T) - \hat{w}_0^{rc}(D, T, X; \hat{\gamma}^{ipt})\right) \left(Y - \hat{\mu}^{lin,rc}_{0,Y}(X; \hat{\beta}^{wls,rc}_{0,1}, \hat{\beta}^{wls,rc}_{0,0})\right)\right] \\
\hat{\tau}^{dr,rc}_{2,imp} =& \hat{\tau}^{dr,rc}_{1,imp} + \left(\mathbb{E}_n \left[\left(\frac{D}{\mathbb{E}_n[D]} - \hat{w}_1^{rc}(D,T) \right) \left(\hat{\mu}^{rc}_{1,1}(X; \hat{\beta}^{ols,rc}_{1,1}) - \hat{\mu}^{rc}_{0,1}(X; \hat{\beta}^{ols,rc}_{0,1}) \right) \right] \right) \\
& - \left(\mathbb{E}_n \left[\left(\frac{D}{\mathbb{E}_n[D]} - \hat{w}_1^{rc}(D,T) \right) \left(\hat{\mu}^{rc}_{1,0}(X; \hat{\beta}^{ols,rc}_{1,0}) - \hat{\mu}^{rc}_{0,0}(X; \hat{\beta}^{ols,rc}_{0,0}) \right) \right] \right)
\end{aligned}
$$

where Γ, B being respective parameter spaces for the semiparametric model,

$$
\begin{aligned}
&\hat{\gamma}^{ipt} = \underset{\gamma \in \Gamma}{\arg\max} \quad \mathbb{E}_n[DX'\gamma - (1-D)e^{X'\gamma}] \\
&\pi(X, \gamma) = \Lambda(X'\gamma) = \frac{e^{X'\gamma}}{1 + e^{X'\gamma}} \ \textit{is the propensity score model proposal} \\
&\hat{\beta}^{wls,rc}_{0,t} = \underset{b \in B}{\arg\min} \quad \mathbb{E}_n[\frac{\Lambda(X'\hat{\gamma}^{ipt})}{1 - \Lambda(X'\hat{\gamma}^{ipt})}(Y - X'b)^2 | D = 0, T = t] \\
&\hat{\beta}^{ols,rc}_{1,t} = \underset{b \in B}{\arg\min} \quad \mathbb{E}_n[(Y - X'b)^2 | D = 1, T = t] \\
&\mu^{rc}_{d,t}(X; \beta^{rc}_{d,t}) = \mu^{lin,rc}_{d,t}(X; \beta^{rc}_{d,t}) = X'\beta^{rc}_{d,t} \ \textit{is the outcome evolution model proposal}
\end{aligned}
$$

[4] Chernozhukov et al (2017)

[5] Graham et al. (2012)

5 Simulation Methodology

The theoretical performance of the suggested estimators require application for experimental proof, and thus the most commonly used verification method of Monte Carlo simulations are going to be constructed to verify the properties of the estimators.

5.1 Data creation

The randomized data will be based on Kang and Schafer (2007)'s Data Generation Process (DGP) as it is in Sant'Anna and Zhao(2020). This approach is chosen due to the underlying data process satisfies the suggested linearity in variables for Outcome Regression suggestion and logistic probability distribution for propensity score IPW regression.
The observed data in the simulation will be $\{Y_0, Y_1, D_i, Z_i\}$ for $\forall i$. We will create the characteristic covariates of $Z, X \in \mathbb{R}^4$ as:

$$\begin{aligned}
X &\sim N(0_{4\times 1}, I_4) \\
\tilde{Z}_{4\times 1} &= (e^{0.5X_1}, 10 + \frac{X_2}{1+e^{X_1}}, (0.6 + \frac{X_1X_3}{25})^{25}, (20 + X_2 + X_4)^2)^T \\
Z_{4\times 1} &= \frac{\tilde{Z} - \mathbb{E}[\tilde{Z}]}{Var[\tilde{Z}]^{1/2}}
\end{aligned}$$

Standardized Z and X are going to be utilized under DGPs to construct the specification cases to assess the performance of the suggested estimator. DGPs will produce in a way that OR model will hold only for DGP1 and DGP2 while IPW model will hold only for DGP1 and DGP3. This holds intuitively since the observed data and model input match will satisfy specification, otherwise not holding true. This approach is to simulate correct specification and mis-specification cases of models by the researcher.
The data generation processes, where $\epsilon_0, \epsilon_1(d) \sim N(0,1)$, $U \sim U(0,1)$ and for a generic $W \in \mathbb{R}^4$ $\nu(W, D) \sim N(D \cdot f_{reg}(W), 1)$, will follow as such:

- DGP1:

$$\begin{aligned}
Y_0(0) &= f_{reg}(Z) + \nu(Z, D) + \epsilon_0 & Y_1(d) &= 2 \cdot f_{reg}(Z) + \nu(Z, D) + \epsilon_1(d) \\
p(Z) &= \frac{e^{f_{ps}(Z)}}{1 + e^{f_{ps}(Z)}} & D &= \mathbb{1}\{p(Z) \leq U\}
\end{aligned}$$

- DGP2:

$$\begin{aligned}
Y_0(0) &= f_{reg}(Z) + \nu(Z, D) + \epsilon_0 & Y_1(d) &= 2 \cdot f_{reg}(Z) + \nu(Z, D) + \epsilon_1(d) \\
p(X) &= \frac{e^{f_{ps}(X)}}{1 + e^{f_{ps}(X)}} & D &= \mathbb{1}\{p(X) \leq U\}
\end{aligned}$$

- DGP3:

$$\begin{aligned}
Y_0(0) &= f_{reg}(X) + \nu(X, D) + \epsilon_0 & Y_1(d) &= 2 \cdot f_{reg}(X) + \nu(X, D) + \epsilon_1(d) \\
p(Z) &= \frac{e^{f_{ps}(Z)}}{1 + e^{f_{ps}(Z)}} & D &= \mathbb{1}\{p(Z) \leq U\}
\end{aligned}$$

- DGP4:

$$Y_0(0) = f_{reg}(X) + \nu(X, D) + \epsilon_0 \quad Y_1(d) = 2 \cdot f_{reg}(X) + \nu(X, D) + \epsilon_1(d)$$
$$p(X) = \frac{e^{f_{ps}(X)}}{1 + e^{f_{ps}(X)}} \qquad D = \mathbb{1}\{p(X) \leq U\}$$

where for generic $W \in \mathbb{R}^4$:

$$f_{reg}(W) = 210 + (27.4, 13.7, 13.7, 13.7)_{1\times 4} \cdot W_{4\times 1}$$
$$f_{ps}(W) = 0.75 \cdot ((-1, 0.5, -0.25, -0.1)_{1\times 4}) \cdot W_{4\times 1}$$

The correct ATT under this process is 0 as it can be clearly understood as time trend is $f_{reg}(W)$, group heterogeneity holder as $\nu(W, D)$ for both groups $D \in \{0, 1\}$. Correct modelling in this case would be to use linear regression for $m^p_{0,\Delta}(.)$ and logistic regression for $p(.)$ with correct variable (Z or X) under given DGPs. However as suggested, when there is a mismatch with observed characteristic covariates and DGPs, the error will lead to incorrect estimations as intended.
These DGPs satisfy the desired properties in the simulation process, although one can opt different sampling schemes to test, this specific example will be utilized in this project.

6 Results

The number of Monte Carlo simulations will be 550 for Panel Data and 250 for Repeated Cross-Sectional Data. The sample sizes are 1000 for panel data for each period while the size is 1000 for repeated cross-sectional data for including periods due to computational power bottleneck. The groups in repeated cross-sectional data are constructed with 0.5 distribution as Sant'Anna and Zhao(2020) also asserts that this does not largely impact the results. For benchmarking purposes, the Doubly-Robust for consistency Difference in Differences estimators will be included in the discussion where outcome evolution is estimated via Ordinary Least Squares and propensity score is estimated via Maximum Likelihood. For other estimator performances such as common two-way fixed effects estimator, which are generally not flexible enough for data structures thus remains largely biased, and for semiparametric efficiency property under correct specification of both models with influence function(nuisance function) calculations, one can look for the discussion in Sant'anna and Zhao(2020).
One working example of codes can be accessed through `https://github.com/omererhanerbis/Bonn_MS_Topics_Econometrics_2022`.

6.1 Panel Data Results

The results of the simulations confirm the findings of Sant'Anna and Zhao(2020).

6.1.1 Bias

The simulation results show that doubly-robustness property in DID setting works really well as there is little to no bias, as long as the identification of the ATT is consistent under correct

Average Bias				
	$\hat{\tau}^{dr,p}_{imp}$	$\hat{\tau}^{ipw,p}$	$\hat{\tau}^{or,p}$	$\hat{\tau}^{dr,p}$
DGP1	0.002535554	-0.017104241	0.004423463	0.002563076
DGP2	0.010968659	-1.078301544	0.009092825	0.009990636
DGP3	-0.07204745	-0.05419572	-2.99716634	-0.06184151
DGP4	-1.472536	-2.742326	-4.512268	-1.872973

specification of either model, that is under DGP1, DGP2 and DGP3. OR model fails to correctly identify ATT under DGP3 and DGP4 while IPW Regression model fails under DGP2 and DGP4. The strength of the doubly-robustness property is thus established and is very convenient to use for consistency purposes.

However, doubly-robustness property also fails if no model is correctly specified, as it was previously asserted that doubly-robust estimators only prove to be partial remedy for consistency concerns. Moreover, doubly-robustness property does not mean better performance under misspecification of both models although the results show better performance. The results shown here is due to specific samples from simulation and thus can be misleading for misspecification consistency. This is especially important, because under some circumstances, the researcher who focuses on consistency may want to adopt other estimators rather than doubly-robust ones if there is high probability of misspecification.

6.1.2 Variance

Coverage Probability				
	$\hat{\tau}^{dr,p}_{imp}$	$\hat{\tau}^{ipw,p}$	$\hat{\tau}^{or,p}$	$\hat{\tau}^{dr,p}$
DGP1	0.9472727	0.9436364	0.9563636	0.9509091
DGP2	0.9218182	0.5272727	0.9381818	0.9290909
DGP3	0.94000000	0.94363636	0.01818182	0.94363636
DGP4	0.2836364	0.3581818	0.0000000	0.3145455

Asymptotic Variance				
	$\hat{\tau}^{dr,p}_{imp}$	$\hat{\tau}^{ipw,p}$	$\hat{\tau}^{or,p}$	$\hat{\tau}^{dr,p}$
DGP1	20.66440	346.83279	19.64489	20.92164
DGP2	20.99689	356.46478	19.94511	20.93392
DGP3	327.3672	1477.9328	696.2011	647.8763
DGP4	323.8025	1442.6566	696.2256	599.8123

When it comes to variance, the first thing that shows up is the inefficiency i.e. high variance of the IPW Regression estimation for either case, even though the attractiveness of the IPW model is indeniable in many applications. On the other hand, OR model seems to perform quite efficiently for either case even though the bias should also be taken into account for practical use. Doubly-Robust estimators both perform highly efficiently under correct outcome model specification, but perform less efficiently otherwise. The Doubly-Robust for Inference estimator cannot be said to have optimal variance, although better than the regular DRDID estimator which supports the special choice of DR for inference DID estimator. The DGPs may be the reason regular doubly-robust estimator's DGP2 efficiency that is not expected, and Sant'Anna and Zhao(2020) also notes this.

6.2 Repeated Cross-Section Data Results

The results of the simulations confirm the findings of Sant'Anna and Zhao(2020) although the DGP4 consistency and DGP2 and DGP3 inconsistency of $\hat{\tau}_1^{dr,rc}$ are shocking which are not expected theoretically. Furthermore, it should also be noted that the efficiency of all repeated cross-sectional data estimators are worse than panel data case, which is natural due to information loss but the efficiency loss magnitude is noteworthy. This may result from low number of Monte Carlo simulations, and further investigation can be a legitimate idea under these findings.

6.2.1 Bias

Average Bias

	$\hat{\tau}_{2,imp}^{dr,rc}$	$\hat{\tau}^{ipw,rc}$	$\hat{\tau}^{or,rc}$	$\hat{\tau}_2^{dr,rc}$	$\hat{\tau}_1^{dr,rc}$
DGP1	0.009087408	0.367330811	0.186681403	0.007644166	0.273292814
DGP2	-0.003847228	-1.071766068	-0.117042671	-0.003594161	1.446197370
DGP3	-0.009806396	0.473143682	-2.664739185	0.128807889	2.595034815
DGP4	-1.5303031	-2.6847217	-4.5233528	-1.8623949	0.1412678

The results again confirm that the doubly-robustness property improves over the one approach models(excluding $\hat{\tau}_1^{dr,rc}$) while either model is correctly specified. The OR model shows significant bias when the outcome model is misspecified, as expected as in DGP3 and DGP4. IPW Regression model performs rather poorly as the propensity score model is misclassified in DGP2 and DGP4 cases. Among the doubly-robust estimators, it is easy to see that there is no significant bias(excluding $\hat{\tau}_1^{dr,rc}$) if either model is correctly specified. However, yet again the doubly-robust estimators do not necessarily perform better when no underlying model is correctly specified.

6.2.2 Variance

Coverage Probability

	$\hat{\tau}_{2,imp}^{dr,rc}$	$\hat{\tau}^{ipw,rc}$	$\hat{\tau}^{or,rc}$	$\hat{\tau}_2^{dr,rc}$	$\hat{\tau}_1^{dr,rc}$
DGP1	0.936	0.992	0.988	0.940	0.944
DGP2	0.940	0.992	0.996	0.944	0.944
DGP3	0.972	0.988	0.968	0.984	0.948
DGP4	0.960	0.976	0.964	0.968	0.944

Asymptotic Variance

	$\hat{\tau}_{2,imp}^{dr,rc}$	$\hat{\tau}^{ipw,rc}$	$\hat{\tau}^{or,rc}$	$\hat{\tau}_2^{dr,rc}$	$\hat{\tau}_1^{dr,rc}$
DGP1	40.29002	35227.23817	24941.18565	42.29931	266005.37333
DGP2	40.72289	38423.19894	29905.80799	41.15555	269053.83419
DGP3	5934.567	58584.130	37155.715	11323.453	1025321.176
DGP4	6239.585	56871.215	42314.530	9274.863	1018800.756

The first distinction when the efficiency is concerned with is that the second doubly-robust estimator that improves on the first performs much better. This confirms the insight of Sant'anna and Zhao(2020) that incorporating the outcome evolution into the doubly-robust estimation improves the estimator. Again it is not expected for the regular second DRDID estimator to perform efficiently under DGP2, but it does perform nearly as efficient as the improved DRDID estimator. But, as expected, the doubly-robust for inference estimator significantly improves on regular doubly-robust estimators under DGP3 and DGP4, which is quite impressive.

7 Conclusion

Causality is one of the most attractive relationships available between events or objects that enables the understanding of the universe. Therefore researchers and policy-makers are highly interested in finding the treatment effect that define this relationship (at least on correlation level) as accurately and efficiently as possible. Some circumstances do not allow randomized experiments, thus the Difference in Differences model is a valuable methodology with diverse applications in the literature. However, often the one dimensional approaches fall short of desired or achievable level of consistency and efficiency. With the development of Doubly-Robust estimators, application of the idea to DID setup turns up to be highly effective for both correct identification and efficient calculation purposes and the extension of the Doubly-Robust for inference models improves the idea for inference procedures with variance stabilization. To further provide meaningful, easily applicable yet flexible enough estimation for treatment effect calculations, use of semiparametric methods in the Doubly-Robust for inference models, Sant'Anna and Zhao(2020) sets cornerstone of combined these state-of-the-art estimators that utilizes all the ideas. Further research for treatment effect calculations mainly with inference concerns will highly benefit from this approach and future works may build up on the idea to improve the estimation procedures.

8 Appendix

8.1 Appendix 1. Construction of Existing ATT estimators

8.1.1 Appendix 1.a. OR Model

To start the construction, we will reiterate the ATT in potential outcomes term:

$$\tau = \mathbb{E}[Y_1(1) - Y_1(0)|D_i = 1] = \mathbb{E}[Y_1|D_i = 1] - \mathbb{E}[Y_1(0)|D_i = 1]$$

As it can be seen, $\mathbb{E}[Y_1(0)|D_i = 1]$ is not identified due to being a counterfactual outcome. However by invoking the Law of Iterated Expectations(LIE), Conditional Parallel Trends Assumption and linearity of expectations operator $\mathbb{E}[.]$, the unidentified part can be expressed as:

$$\begin{aligned}
\mathbb{E}[Y_1(0)|D = 1] &= \mathbb{E}[Y_1(0) - Y_0(0)|D = 1] + \mathbb{E}[Y_0(0)|D = 1] \\
&\overset{\text{LIE}}{=} \mathbb{E}[\mathbb{E}[Y_1(0) - Y_0(0)|D = 1, X]] + \mathbb{E}[Y_0(0)|D = 1] \\
&\overset{\text{CPTA}}{=} \mathbb{E}[\mathbb{E}[Y_1(0) - Y_0(0)|D = 0, X]] + \mathbb{E}[Y_0(0)|D = 1] \\
&= \mathbb{E}[\mathbb{E}[Y_1(0)|D = 0, X] - \mathbb{E}[Y_0(0)|D = 0, X]] + \mathbb{E}[Y_0(0)|D = 1] \\
&= \mathbb{E}[\mathbb{E}[Y_1|D_i = 0, X] - \mathbb{E}[Y_0|D = 0, X]] + \mathbb{E}[Y_0|D = 1]
\end{aligned}$$

which is identified. Carefully, it should be noted that the expectation of the LIE applied term is in terms of $D = 1$ conditioning. Thus, the ATT can be expressed as:

$$\begin{aligned}
\tau^{reg} &= \mathbb{E}[Y_1|D_i = 1, X] - \left[\mathbb{E}[Y_0|D_i = 1, X] + \frac{1}{n_{treat}} \sum_{i|D_i=1} (\mathbb{E}[Y_1|D_i = 0, X] - \mathbb{E}[Y_0|D = 0, X])\right] \\
&= \mathbb{E}[Y_1|D_i = 1, X] - \left[\mathbb{E}[Y_0|D_i = 1, X] + \frac{1}{n_{treat}} \sum_{i|D_i=1} (m_{0,1}(X_i) - m_{0,0}(X_i))\right]
\end{aligned}$$

8.1.2 Appendix 1.b. IPW Regression Model

For panel data case, Abadie(2005) says that under the CPTA, the following is true:

$$\mathbb{E}[Y_1(1) - Y_1(0)|D_i = 1, X] = \mathbb{E}[\frac{D - P(D = 1|X)}{P(D = 1|X)(1 - P(D = 1|X)}(Y_1 - Y_0)] = \mathbb{E}[\rho_0(Y_1 - Y_0)]$$

thus,

$$
\begin{aligned}
\mathbb{E}[Y_1(1) - Y_1(0)|D_i = 1, X] &= \int \mathbb{E}[Y_1(1) - Y_1(0)|D_i = 1, X] d(P(X|D=1)) \\
&= \int \mathbb{E}[\rho_0(Y_1 - Y_0)] d(P(X|D=1)) \\
&= \mathbb{E}[\rho_0(Y_1 - Y_0)\frac{P(D=1|X)}{P(D=1)}] \\
&= \mathbb{E}[\frac{1}{P(D=1)} \frac{D - P(D=1|X)}{(1 - P(D=1|X))}(Y_1 - Y_0)] \\
&= \frac{1}{\mathbb{E}[D]} \mathbb{E}\left[\frac{D - p(X)}{1 - p(X)}(Y_1 - Y_0)\right]
\end{aligned}
$$

since $P(D = 1) = \mathbb{E}[D]$ due to binary nature of D. This returns the proposed IPW estimator form for panel data.

Similarly, under the repeated cross-sectional data case, the Abadie(2005) argues that the following is true:

$$
\mathbb{E}[Y_1(1) - Y_1(0)|D_i = 1, X] = \mathbb{E}[\frac{T - \lambda}{\lambda(1-\lambda)} \cdot \frac{D - P(D=1|X)}{P(D=1|X)(1 - P(D=1|X)}(Y_1 - Y_0)] = \mathbb{E}[\phi_0 Y]
$$

similar calculation steps to panel data case can be applied to reach the following suggested IPW estimator for of panel data:

$$
\tau = \frac{1}{\mathbb{E}[D]} \mathbb{E}\left[\frac{D - p(X)}{1 - p(X)} \frac{T - \lambda}{\lambda(1-\lambda)} Y\right]
$$

8.2 Appendix 2. Efficient Influence Functions for Variance Calculation

The efficient influence function for the panel data case is:

$$
\begin{aligned}
\eta^{e,p}\left(Y_1, Y_0, D, X\right) = w_1^p(D)\Big(m_{1,\Delta}^p(X) - &\ m_{0,\Delta}^p(X) - \tau\Big) \\
&+ w_1^p(D)\Big(\Delta Y - m_{1,\Delta}^p(X)\Big) - w_0^p(D, X; p)\Big(\Delta Y - m_{0,\Delta}^p(X)\Big)
\end{aligned}
$$

and the efficient influence function for the repeated cross-sectional data case is:

$$
\begin{aligned}
\eta^{e,rc}(Y, D, T, X) = &\ \frac{D}{\mathbb{E}[D]}\left(m_{1,\Delta}^{rc}(X) - m_{0,\Delta}^{rc}(X) - \tau\right) \\
&+ \left(w_{1,1}^{rc}(D, T)\left(Y - m_{1,1}^{rc}(X)\right) - w_{1,0}^{rc}(D, T)\left(Y - m_{1,0}^{rc}(X)\right)\right) \\
&- \left(w_{0,1}^{rc}(D, T, X; p)\left(Y - m_{0,1}^{rc}(X)\right) - w_{0,0}^{rc}(D, T, X; p)\left(Y - m_{0,0}^{rc}(X)\right)\right)
\end{aligned}
$$

For further interest, one can refer to Sant'Anna and Zhao(2020).

9 References

- Sant'Anna, Pedro H.C. & Zhao, Jun, 2020. "Doubly robust difference-in-differences estimators," Journal of Econometrics, Elsevier, vol. 219(1), pages 101-122.
- Heckman, James J., et al. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." The Review of Economic Studies, vol. 64, no. 4, 1997, pp. 605–54. JSTOR, https://doi.org/10.2307/2971733.
- Abadie, Alberto. "Semiparametric Difference-in-Differences Estimators." The Review of Economic Studies, vol. 72, no. 1, 2005, pp. 1–19. JSTOR, http://www.jstor.org/stable/3700681.
- Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." Journal of Economic Literature, 47 (1): 5-86.
- Newey, Whitney K. "Semiparametric Efficiency Bounds." Journal of Applied Econometrics, vol. 5, no. 2, 1990, pp. 99–135. JSTOR, http://www.jstor.org/stable/2096601.
- Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings." The Review of Economics and Statistics, vol. 60, no. 1, 1978, pp. 47–57. JSTOR, https://doi.org/10.2307/1924332.
- Ashenfelter, Orley, and David Card. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs." The Review of Economics and Statistics, vol. 67, no. 4, 1985, pp. 648–60. JSTOR, https://doi.org/10.2307/1924810.
- Joseph D. Y. Kang. Joseph L. Schafer. "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data." Statist. Sci. 22 (4) 523 - 539, November 2007. https://doi.org/10.1214/07-STS227
- Splawa-Neyman, Jerzy, et al. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Statistical Science, vol. 5, no. 4, 1990, pp. 465–72. JSTOR, http://www.jstor.org/stable/2245382.
- Fisher, R.A. (1992). Statistical Methods for Research Workers. In: Kotz, S., Johnson, N.L. (eds) Breakthroughs in Statistics. Springer Series in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4380-9_6
- Holland, Paul W. "Statistics and Causal Inference." Journal of the American Statistical Association, vol. 81, no. 396, 1986, pp. 945–60. JSTOR, https://doi.org/10.2307/2289064. Accessed 2 Sep. 2022.
- Card, David. "The Impact of the Mariel Boatlift on the Miami Labor Market." ILR Review, vol. 43, no. 2, Jan. 1990, pp. 245–257, doi:10.1177/001979399004300205.
- Heckman, James. "Varieties of Selection Bias." The American Economic Review, vol. 80, no. 2, 1990, pp. 313–18. JSTOR, http://www.jstor.org/stable/2006591
- Card, D., & Krueger, A. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. American Economic Review, 84(4), 772-93.
- Robins, J. M. (2000). Robust estimation in sequentially ig- norable missing data and causal inference models. Pro- ceedings of the American Statistical Association Section on Bayesian Statistical Science, 6–10.