

CAUSAL INFERENCE OF PANEL DATA MODELS WITH ENDOGENEITY

By

Minkyu Kim

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Economics—Doctor of Philosophy

2024

## **ABSTRACT**

This dissertation studies three different models and estimation strategies to conduct the causal inference conditioning on the lagged outcome. The first and second chapters develop estimation strategies under a feedback environment where both past outcome and unobserved heterogeneity affect the policy intervention, resulting in non-random heterogeneous treatment timings assignment. The third chapter focuses on the context of the imputation method.

The first chapter uses dynamic panel data linear models, which have been useful for studying the state dependence issue, to study the policy evaluation conditioning on the lagged outcome. For the estimation of dynamic linear model with endogenous/feedback treatment, the first chapter compares the two distinctive approaches based on different assumptions. The first chapter compares Arellano and Bond (1991) and Wooldridge (2000) approaches through Monte Carlo simulation, under the environment where the size of the treated group monotonically increases as time goes by. Under the correctly specified density assumption, Wooldridge (2000) approach outperforms that of Arellano and Bond (1991). Monte Carlo simulation also shows that Wooldridge (2000) approach has similar level of bias with that from Arellano and Bond (1991) while retaining relative efficiency even under some level of misspecifications.

The second chapter proposes a dynamic panel data model with non-negative outcomes in the context of the causal inference. This GMM approach provides a useful tool for economists when they want to include unobserved heterogeneity and the lagged outcome in the model without imposing a linear relationship between the outcome and its lag. The framework in the second chapter further allows past dependent variables to have feedback to the current period treatment assignment, which can be a driving source of heterogeneous intervention timing. The second chapter provides the asymptotic properties of the resulting estimator of treatment effects using Monte Carlo simulations. Lastly, the second chapter revisits the policy intervention during the opioid crisis in United States to estimate the effect of must-access Prescription Drug Monitoring Programs (PDMPs) on child maltreatment. The GMM approach with non-negative outcome models yields significantly different policy implications depending on the inclusion of lagged outcome, and the potential feedback from

the past outcome.

The third chapter conduct a Monte Carlo simulation study to propose an imputation estimator that accounts for the lagged outcome included in the potential outcome model. This chapter identified the homogeneous average treatment effects for treated group through our imputation method and our simulation study suggests that the imputation method is easy to implement and works well in the general environments. the chapter also found that the usual two way fixed effects approach suffers bias in the most of scenarios under the dynamic environment. Under the unit root process of the dependent variable, the bias from the two way fixed effects could be amplified.

Copyright by  
MINKYU KIM  
2024

This thesis is dedicated to my grandparents.

## ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Jeffrey M. Wooldridge for his persevering guidance and his indulgence. I want to deliver my sincere thanks to Kyoo il Kim for his encouragement and patience saved me during the endgame of the program. To Ajin Lee, I want to appreciate you for your kindness and directions. Also special thanks to Zhehui Luo for her valuable comments and assistance.

I do appreciate to the scholars and economists who helped me during my PhD program, including but not limited to, Soren Anderson, Jay Pil Choi, Todd Elder, Tim Vogelsang, Carolina Caetano, Nail Kashaev, Florian Gunsilius, Enrique Pinzon, Yuya Sasaki, and Jangsu Yoon. I have another special thanks to Myoung-jae Lee and Sangsoo Park for their warm welcoming for each of my summer visit to South Korea.

I must confess that I owe too much to the community of Michigan State University. I again want to express my gratitude to everyone I met in Michigan, including but not limited to, Nick Brown, Narae Park, Jaemin Ryu, Jeonghwan Choi, Inkyu Kim, Akanksha Negi, Jaeyeon Lee, Xu Dong, Jun-Tae Park, Dongming Yu, Alyssa Carlson, Jeongmin Ha and Elizabeth Kayoon Hur. I also owe my survival to my friends in cohort, Steven Wu-Chaves and Andrew Earle. I also want to appreciate to Jay Feight for his departmental support and to Su Hwan Chung for his patience. Among the others, I want to deliver my utmost gratitude to Taeyoon Hwang for every moment we had together during this long ride.

I want to give my thanks to my good old friends, including but not limited to, Tae Hun Chang, Juno Kim, Jeong Hun Lee, and Wonjun Choi. I acknowledge that I owe my sanity to Sumin Lee and Kippeum Lee.

Finally, I want to appreciate to my family. I am deeply grateful to my parents K.T. Kim and H.Y. Lee, and my brother M.J. Kim and his wife J.H.Yun. Indeed I could have survived this madness of pandemic solely thanks to my family. If allowed, I want to dedicate my dissertation to my grandparents, who helped me find the meaning of my PhD program.

## TABLE OF CONTENTS

CHAPTER 1	ESTIMATING ENDOGENOUS TREATMENT EFFECTS USING DYNAMIC PANEL DATA MODELS . . . . .	1
1.1	Introduction . . . . .	1
1.2	Framework . . . . .	4
1.3	Results . . . . .	9
1.4	Discussion . . . . .	14
CHAPTER 2	CAUSAL INFERENCE OF DYNAMIC PANEL MODEL WITH NON-NEGATIVE OUTCOME . . . . .	15
2.1	Introduction . . . . .	15
2.2	Econometric Model . . . . .	18
2.3	Identification . . . . .	22
2.4	Estimation . . . . .	28
2.5	Monte Carlo Simulation . . . . .	32
2.6	Application: Effect of Prescription Drug Monitoring Program on Child Mal- treatment . . . . .	36
2.7	Conclusion . . . . .	47
CHAPTER 3	IMPUTATION ESTIMATION OF DYNAMIC POTENTIAL OUT- COME MODELS FOR INTERVENTION ANALYSIS . . . . .	48
3.1	Introduction . . . . .	48
3.2	Model . . . . .	50
3.3	Identification . . . . .	52
3.4	Estimation . . . . .	55
3.5	Simulations . . . . .	56
3.6	Conclusions . . . . .	61
BIBLIOGRAPHY	. . . . .	62
APPENDIX A	APPENDIX FOR CHAPTER 1 . . . . .	68
APPENDIX B	PROOFS FOR CHAPTER 2 . . . . .	78
APPENDIX C	EXTENSIONS OF CHAPTER 2 . . . . .	81
APPENDIX D	PROOF FOR CHAPTER 3 . . . . .	88

## CHAPTER 1

### ESTIMATING ENDOGENOUS TREATMENT EFFECTS USING DYNAMIC PANEL DATA MODELS

#### 1.1 Introduction

Just as in many disciplines of science, modern applied economists have been widely using causal inference, or as known as treatment effect analysis (Lee 2016). Through the policy evaluation literature, economists have been interested in parametric estimation (Mullainathan and Spiess 2017), typically a linear model. A famous estimation strategy is to use panel data, as researchers can not only handles the dynamic relationship across observations, but it also helps them to remove unobserved heterogeneity (Wooldridge 2010). For example, parametric estimation of a linear regression model is implemented straightforwardly for both of static model (e.g., fixed effect) and the dynamic model (e.g., Arellano and Bond (1991) approach). In the context of the treatment effect, one of the giants in the literature is the difference-in-differences method<sup>1</sup>, which manipulates the exogeneity of the time constant treatment group indicator, and a uniform jump of the treatment indicator between the controlled period within the treated group.

In many cases, however, treatment assignment is not exogenous. For example, Collinson et al. (2024) considered the effect of eviction (treatment) to the economic outcome of tenants. Suppose a representative economic agent has been suffering from a poor economic outcome last year, say low earning and/or unemployment. Since one cannot pay rent properly for a year, he or she might get evicted (treated). Because of such treatment (eviction), the life of the agent gets even more harder, making worse economic outcome this or next year too. Similar logic can be applied to the effect of employment/household status to the individual poverty (Biewen 2009), and the effect of unemployment to the mental health (Zimmer 2021), and so on. In general we can call this environment as feedback model, where past dependent variable affects current/future treatment. Under the feedback environment, the treatment group indicator is endogenous, and time-varying in general.

---

<sup>1</sup>Imbens and Wooldridge (2009) provides a great introduction to difference-in-differences framework. See Angrist and Pischke (2009), Lee (2016) and references therein for the detail about difference-in-differences framework.



Another caveat of this feedback environment is the state dependence or persistence of the dependent variable. As we can see from the previous eviction example, it is not difficult to assume that the poor economic outcome of tomorrow can be affected by a poor outcome of today. In a linear regression model, putting a lagged dependent variable would be a good starting point for a remedy. Combining the feedback nature described in the previous paragraph and the lagged dependent variable on the right hand side, we need to estimate a dynamic panel data model with endogenous variables other than the lagged dependent variables. Note that I used the word “dynamic” to emphasize that we included the lagged dependent variable as a regressor.

There has been meaningful development in “dynamic” treatment effect literature, but not in the way we discussed above. Traditionally, treatment indicator has been treated as independent/exogenous random variable, where feedback is not allowed for the main equation of interest (e.g., Abbring and Heckman (2007)).<sup>2</sup> If we focus on the panel data, dynamic average treatment effect was proposed and studied based on Robins (1989)’s approach, but again the word “dynamic” focuses on the time-varying treatment indicator. For an example of formal framework of such dynamic treatment effects, see Lechner (2015) and the references therein.

On the other hand, “dynamic” linear model was studied mainly under the name of state-dependence literature. The literature focused on removing endogeneity problem between the lagged dependent variable and the unobserved heterogeneity. Typical response to this issue is to take a first differencing to the equation to remove the time constant heterogeneity, then use instruments or moment conditions (e.g., Ahn and Schmidt (1995); Blundell and Bond (1998)). One of the benchmark method among them in this literature is using the generalized method of moments (GMM, hereafter) under specific exogeneity assumptions. Notably, Arellano and Bond (1991) provided a framework to consistently estimate the dynamic linear model with regressors of lagged dependent variable and endogenous regressors, under the correct exogeneity assumption (e.g., sequential exogeneity). Arellano and Bond (1991) approach also allows researchers to find ready-made instrument, as it allows researchers to use past variables as the instruments. The

---

<sup>2</sup>For duration analysis, dependent variable is included in the treatment hazard function. Hazard analysis makes, however, a different story, and it is beyond the scope of this paper.

approach uses first-differencing to remove the heterogeneity, so one might doubt potential inefficiency of Arellano and Bond (1991), compared to other methods who does not subtract out the heterogeneity. Another downside of the method is related to the ready-made instruments. Since it allows various ways to construct the instruments, there are numerous arbitrary decisions to make.<sup>3</sup>

Another way to tackle the feedback environment is to take itself into account and modeling it. Instead of running a GMM estimation relying on arbitrary decisions, Wooldridge (2000) assumed the joint distribution of dependent and explanatory variables and run a maximum likelihood estimation (MLE, hereafter). Instead of first-differencing the heterogeneity, one can model the conditional density of heterogeneity given the initial values,<sup>4</sup> and integrate out the heterogeneity during the MLE process. Wooldridge (2000)'s approach consistently and efficiently estimates the dynamic structural equation of interests under correct specification. It allows the state dependence, existence of heterogeneity, and take the feedback nature of policy assignment into account, by allowing lagged dependent variable and the heterogeneity included in the treatment indicator equation. Due to the advantages listed above, there are empirical applications using this approach (e.g., Biewen (2009), Welsch and Zimmer (2015, 2016), Zimmer (2021)). The downside of Wooldridge (2000) approach is the inconsistency under misspecification of the model, as the method is based on MLE.

The paper wants to fill the gap of the literature by comparing two promising estimation methods, Arellano and Bond (1991) and Wooldridge (2000) approaches, under the feedback environment, where the treatment assignment is affected by the past response, and the size of the treated group is monotonically increasing. Since we are on the context of treatment effect literature, this paper focuses on the parametric estimation of treatment dummy coefficient, instead of the persistent parameter. As outside options of the two approaches, I also simulated the results from pooled ordinary least squares (POLS, hereafter), fixed effect (FE, hereafter), and first-differencing (FD, hereafter) estimates.

---

<sup>3</sup>We can have more instruments other than the original suggestion of Arellano and Bond (1991) following the method of Roodman (2009a,b), or we can also add extra moment conditions (e.g., Arellano and Bover (1995)). See appendix for another example of GMM estimation strategies other than Arellano and Bond (1991).

<sup>4</sup>Similar idea of using the initial value can be found in Wooldridge (2005). It assumes the strict exogeneity of the treatment, contrary to Wooldridge (2000).

The rest of the paper is organized as follows. Section 1.2 describes the model of interests, and provides a quick refresher of related estimation strategies in the context of this paper. The following Section 1.3 describes the benchmark data generating process (DGP, hereafter), and shows the simulation results. In Section 1.4, I conclude with a short discussion regarding future research.

## 1.2 Framework

### 1.2.1 Dynamic Potential Response Model

We begin from a simple switching regression model with a constant treatment effect. Following the treatment effect literature,<sup>5</sup> let  $y^{(1)}$  be a potential treated response, and  $y^{(0)}$  be a potential controlled response. Then we can write a constant treatment effect model as

$$y_{it}^{(1)} = y_{it}^{(0)} + \beta_d \quad (1.1)$$

where  $\beta_d$  is the treatment effect for all observation  $i$  at time  $t$ .

Let  $D_{it}$  as the treatment indicator, who switches from zero to one if treated. Then, the observed response  $y_{it}$  can be written as

$$y_{it} = (1 - D_{it}) \cdot y_{it}^{(0)} + D_{it} \cdot y_{it}^{(1)} \quad (1.2)$$

To parametrically estimate  $\beta_d$ , we need following assumptions.

**Assumption [Unconfoundedness].** Suppose  $(y_{it}^{(0)}, y_{it}^{(1)})$  be a pair of counterfactual outcomes, defined as in the subsection 1.2.1, and let  $w_{it} = (y_{it-1}, z_{it}, c_i)$  be the set of covariates. Then  $(y_{it}^{(0)}, y_{it}^{(1)})$  is independent of  $D_{it}$ , conditional on  $w_{it}$ :

$$(y_{it}^{(0)}, y_{it}^{(1)}) \perp\!\!\!\perp D_{it} | w_{it}$$

**Assumption [Conditional expectation].** Let  $y_{it}^{(0)}$  be the potential controlled response, and  $w_{it} = (y_{it-1}, z_{it}, c_i)$  be the set of covariates. Then conditional expectation of  $y_{it}^{(0)}$  given  $w_{it}$  is

$$E(y_{it}^{(0)} | w_{it}) = \beta_t + \beta_y y_{it-1} + \beta_z z_{it} + c_i$$

---

<sup>5</sup> Angrist and Pischke (2009), Imbens and Wooldridge (2009), Wooldridge (2010), and Lee (2016) provide introductions to the treatment effect literature.

Note that  $w_{it}$  contains the unobserved heterogeneity  $c_i$ . Conditional expectation assumption implies

$$y_{it}^{(0)} = \beta_t + \beta_y y_{it-1} + \beta_z z_{it} + c_i + u_{it} \quad (1.3)$$

where  $z_{it}$  is strictly exogenous covariates, and  $c_i$  is unobserved heterogeneity, and  $u_{it}$  is an idiosyncratic error. Combining equations from (1.1) to (1.3) and two assumptions<sup>6</sup> yields our estimating regression equation:

$$y_{it} = \beta_t + \beta_y y_{it-1} + \beta_d D_{it} + \beta_z z_{it} + c_i + u_{it} \quad (1.4)$$

where the (conditional) average treatment effect is captured by  $\beta_d$ . Note that  $c_i$  is already included in  $y_{it-1}$  by the equation (1.4) applied on the time  $t - 1$ , so running POLS on equation (1.4) would be inconsistent in general.

Moreover, even though econometricians want to assume that  $D_{it}$  is (strictly) exogenous, and use FE (or FD) strategy to remove the heterogeneity,  $c_i$ , the policy assignment might not be exogenous, but predetermined by the past outcome.

$$D_{it} = 1[\gamma_y y_{it-1} + \gamma_z z_{it} + \gamma_c c_i \leq \kappa + \varepsilon_{it}] \quad (1.5)$$

As a motivation for the structural equation (1.5) for the policy assignment, consider a textbook funding program for those who were not performing well in the (centralized) exam last period. Let  $y_{it-1}$  is the exam score of the last year, and  $\kappa$  is the cutoff of the treatment qualification/assignment, and  $\varepsilon_{it}$  is to capture any idiosyncratic error components at year  $t$ . Then the textbook funding program of this year  $D_{it}$  can be written as:

$$D_{it} = 1[y_{it-1} \leq \kappa + \varepsilon_{it}] \quad (1.6)$$

which is a short version of equation (1.5). Note that both equation (1.5) and (1.6) have the lagged dependent variable from the feedback nature of the environment. Due to the endogeneity of  $(y_{it-1}, D_{it})$ , running a POLS, FE, or FD would be inconsistent for estimating  $\beta_d$  via equation (1.4), as we can see from the main result at Section 1.3. From here, one might start searching for instruments for  $y_{it-1}$  and  $D_{it}$ , and one estimation strategy allows us to use the past variables as the instruments under the exogeneity assumptions.

---

<sup>6</sup>Following the treatment effect literature, we may assume the overlap assumption as well:  $0 < P(D_{it} = 1|w_{it}) < 1$ .

### 1.2.2 Arellano and Bond (1991) Approach <sup>7</sup>

Arellano and Bond (1991) starts from time-differencing the equation (1.4) to remove unobserved heterogeneity

$$\Delta y_{it} = \Delta \beta_t + \beta_y \Delta y_{it-1} + \beta_d \Delta D_{it} + \beta_z \Delta z_{it} + \Delta u_{it} \quad (1.7)$$

Under the sequential exogeneity assumption, it consistently estimates the parameters of equation (1.7).

**Assumption [Sequential Exogeneity].** Suppose our general model of interest is

$$y_{it} = \vec{w}_{it}\beta + c_i + u_{it}, t = 1, 2, \dots, T$$

Then we assume sequential exogeneity by:

$$E(u_{it} | \vec{w}_{it}, \vec{w}_{it-1}, \dots, \vec{w}_{i1}, c_i) = 0, t = 1, 2, \dots, T$$

Note that here  $\vec{w}$  stands for the placeholder of the sequentially exogenous regressors. Under the sequential exogeneity, first-differenced error,  $\Delta u_{it}$ , would be uncorrelated with the past history of the covariates,  $\vec{w}_{it-1}^o = (\vec{w}_{i1}, \vec{w}_{i2}, \dots, \vec{w}_{it-1})$ :

$$E(\vec{w}_{it-1}^{o'} \Delta u_{it}) = \mathbf{0}, t = 2, 3, \dots, T \quad (1.8)$$

Now, we need to pick the instrument matrices. There are several ways to pick the instruments, but we only pick two of them, for the simplicity of discussion.<sup>8</sup>

For example, let's focus on a simple case with  $T = 3$ . Since the time dummy<sup>9</sup> ( $dt$ , hereafter) and  $z_{it}$  are strictly exogenous, we put them together and define them as  $\tilde{w}_{it} = (dt, z_{it})$ . Motivated by the original work of Arellano and Bond (1991), we can construct the theoretical instrument matrix

<sup>7</sup>This subsection mainly draws on Arellano and Bond (1991), and Wooldridge (2010).

<sup>8</sup>For example, Arellano and Bover (1995) proposed a broader set of moment conditions other than sequential exogeneity as a remedy for the weak instrument problem. In this paper, however, we skip this version of instrument.

<sup>9</sup>Since various estimation strategy have different identification of time variables,  $dt$  varies depending on the context of estimation strategies. For example, FE and FD drop one time variables by their construction.

as

$$IV_i = \begin{bmatrix} y_{i0} & D_{i1} & 0 & 0 & 0 & 0 & \Delta\tilde{w}_{i2} \\ 0 & 0 & y_{i0} & D_{i1} & y_{i1} & D_{i2} & \Delta\tilde{w}_{i3} \end{bmatrix} \quad (1.9)$$

Even with a small total time period of  $T = 3$ , the dimension of instrument matrix gets huge pretty quickly, yielding unnecessarily many overidentification restrictions. One alternative is to collapse the instrument matrix and reduce the number of conditions (Roodman (2009a,b)). Then we have the following instrument matrix

$$IV_i = \begin{bmatrix} y_{i0} & D_{i1} & 0 & 0 & \Delta\tilde{w}_{i2} \\ y_{i0} & D_{i1} & y_{i1} & D_{i2} & \Delta\tilde{w}_{i3} \end{bmatrix} \quad (1.10)$$

We use the instruments (1.9) and (1.10) and run the GMM estimation on the equation (1.7). Under the sequential exogeneity assumption, we can consistently estimate  $\beta = (\Delta\beta_d, \beta_y, \beta_d, \beta_z)'$ . As a practical matter, since the treatment starts from  $t = 2$ ,  $D_{i1} = 0$  for all observation. In this paper, therefore, we drop  $D_{i1}$  from equation (1.9) and (1.10), then use them as the instruments. Subsection 1.3.1 explains the DGP of this paper in details.

### 1.2.3 Wooldridge (2000) Approach <sup>10</sup>

Let  $x_{it} \equiv (y_{it}, D_{it})$ , as both  $y$  (of past) and  $D$  is the main regressor of interests. Omit the index  $i$  for the simplicity of discussion. In Wooldridge (2000) approach, we start from assuming the conditional distribution of interests:

$$D(y_t | D_t, Z_t, X_{t-1}, c) = D(y_t | D_t, z_t, X_{t-1}, c), t = 1, 2, \dots, T \quad (1.11)$$

$$D(D_t | Z_t, X_{t-1}, c) = D(D_t | z_t, X_{t-1}, c), t = 1, 2, \dots, T \quad (1.12)$$

where  $X_{t-1} = (x_{t-1}, \dots, x_0)$ , and  $Z_t = (z_t, z_{t-1}, \dots, z_1)$  subsequently. Since we are using a panel data, and for the nature of dynamic model, we assume that we have  $x_0 = (y_0, D_0)$ . The two equalities comes from the strict exogeneity of  $z_t$ . In other word, equation (1.11) states that once present  $z_t$  is controlled, along with  $D_t$ ,  $X_{t-1}$ , and  $c$ , neither past nor the future  $z_s$  affects  $y_t$  for all  $s \neq t$ . Equation (1.12) states the similar meaning for the treatment dummy accordingly.

---

<sup>10</sup>This subsection mainly draws on Wooldridge (2000).

By assuming parametric conditional densities of  $y$  and  $D$ , the joint density of  $x_t$  conditional on  $(Z_T, X_{t-1}, c)$  can be written as

$$p_t(x_t|z_t, X_{t-1}, c; \beta_0, \gamma_0) = f_t(y_t|D_t, z_t, X_{t-1}, c; \beta_0) \cdot g_t(D_t|z_t, X_{t-1}, c; \gamma_0) \quad (1.13)$$

where  $f_t$  is the parametric conditional density of  $y_t$  with the vector of parameter  $\beta_0$ , while  $g_t$  is the corresponding density of  $D_t$  with the parameter vector  $\gamma_0$ .

Now, the joint density of  $X_T = (x_T, X_{T-1}, \dots, x_1)$  given  $(Z_T, x_0, c)$  should be look like

$$p(X_T|Z_T, x_0, c; \beta_0, \gamma_0) = \prod_{t=1}^T p_t(x_t|z_t, X_{t-1}, c; \beta_0, \gamma_0) \quad (1.14)$$

Unlike Arellano and Bond (1991) GMM approach, MLE method does not cancels out the unobserved heterogeneity. Instead, Wooldridge (2000) proposed to model the parametric density of heterogeneity given  $Z_T$  and the initial value  $x_0$ , as a practical matter. Following Wooldridge (2000), we model the conditional density of the heterogeneity as follows

$$c|Z_T, x_0 \sim N(\lambda_1 + \lambda_y y_0 + \lambda_d D_0 + \lambda_z \bar{z}, \sigma_\lambda^2) \quad (1.15)$$

$$a|Z_T, x_0 \sim N(0, \sigma_\lambda^2) \quad (1.16)$$

Note that integrating out  $c$  is equivalent to integrating out  $a$ , once we put  $\bar{z}$  and  $x_0$  as additional regressors in the equations. Using the specified conditional density of heterogeneity  $h(c|Z_T, x_0; \lambda)$ , we integrate out the heterogeneity:

$$p(X_T|Z_T, x_0; \theta_0) = \int_{R^{dim(c)}} p(X_T|Z_T, x_0, c; \beta_0, \gamma_0) \cdot h(c|Z_T, x_0; \lambda_0) \nu(dc) \quad (1.17)$$

where  $\theta_0 \equiv (\beta'_0, \gamma'_0, \lambda'_0)'$ , and  $\nu(\cdot)$  is a suitable measure for the integration. Then finally, combining: i) equation (1.4) for the main equation; ii) equation (1.5) for the policy equation; iii) equations (1.13) through (1.17) for the MLE construction yields (with abuse of notation)

$$\ell(X_{iT}^0, Z_{iT}; \theta) = \log \left[ \int_{R^{dim(c)}} \left( \prod_{t=1}^T f_{it} \cdot g_{it} \right) dF_{c_i|Z_T, x_0; \lambda} \right] \quad (1.18)$$

where  $X_{iT}^0 = (x_{iT}, \dots, x_{i1}, x_{i0})$ , and  $dF_{c_i|}$  stands for the conditional CDF of the heterogeneity. The consistency of MLE  $\hat{\theta}$  from the likelihood (1.18) comes from the consistency of the MLE (Wooldridge (2000)).

In this paper, I used the following for the likelihood function:

$$f_{it} = \frac{1}{\sigma_u} \phi \left( \frac{y_{it} - w_{it}^y \beta_w - a_i}{\sigma_u} \right) \quad (1.19)$$

$$g_{it} = [\Phi(-w_{it}^D \gamma_w - \gamma_c a_i)^{D_{it}} \cdot (1 - \Phi(-w_{it}^D \gamma_w - \gamma_c a_i))^{1-D_{it}}]^{1-D_{it-1}} \quad (1.20)$$

where  $w_{it}^y = (dt, y_{it-1}, D_{it}, z_{it}, y_{i0}, \bar{z}_i)$  and  $w_{it}^D = (1, y_{it-1}, z_{it}, y_{i0}, \bar{z}_i)$ . Conditional density function (1.19) comes from the conditional distribution of idiosyncratic shock  $u_{it}|w_{it}^y, a_i \sim N(0, \sigma_u^2)$ , the main equation (1.4), and the heterogeneity assumptions (1.15) and (1.16). Similarly, conditional density function (1.20) comes from putting a probit assumption on the policy equation (1.5), and heterogeneity assumptions (1.15) and (1.16). The last upper right power term came from the DGP of this paper described in the Section 1.3. For the practical matter, I integrated out  $a_i$  instead of  $c_i$ , which is equivalent once we controlled  $\bar{z}_i$  and  $y_{i0}$  in the regressor terms. To boost the integration process, I used the numerical integration Hermite-Gaussian quadrature.<sup>11</sup>

### 1.3 Results

The Monte Carlo experiment focuses on estimating  $\beta_d$  of equation (1.4), and compares the bias and relative efficiency between various estimation strategies. This paper is mainly interested in comparing Arellano and Bond (1991) approaches with two different instruments under the sequential exogeneity assumption, and Wooldridge (2000) approach based on the correctly specified model. For the Arellano and Bond (1991), AB1 stands for the result based on the original version of instrument matrix in the paper, while AB2 stands for the result based on the collapsed instrument suggested by Roodman (2009a,b). I also reported the results from POLS, FE, FD run on equation (1.4) or (1.7), depending on the context. For the replication, I ran each of the estimation 1,000 times, and calculated the average of bias, standard deviations, and the root mean squared error (rMSE, hereafter).

#### 1.3.1 Benchmark DGP

Dynamic feedback nature of this paper makes it difficult to explain the DGP in a short list of equations. Let me instead explain the DGP using a story about a treatment program whose

---

<sup>11</sup>For the theoretical details of numerical integration in general, see Davis and Rabinowitz (2007).



participation rates are increasing over time.

Suppose we have a data sample with  $T = 3$ . Let  $t = 1$  be a controlled period, where nobody gets treated, and  $t = 2, 3$  are treatment period, where treated sample gets bigger, monotonically, as time goes by<sup>12</sup>.

We start from period  $t = 1$ , where no observation is treated. Since  $D_{i1} = 0$  for all units, I dropped  $D_{i1}$  from the equation (1.4)

$$y_{i1} = \beta_1 + \beta_y y_{i0} + \beta_z z_{i1} + c_i + u_{i1}$$

where  $y_{i0}$  is drawn from standard normal distribution. For the simplicity of discussion,  $z_{it}$  and  $u_{it}$  are also independently drawn from standard normal distributions for all  $t = 1, 2, 3$ . Using the drawn  $z_{it}$ 's and  $y_{i0}$ , and independently drawn  $a_i$ , I constructed  $c_i$  following the equations (1.15) and (1.16):

$$c_i = \lambda_1 + \lambda_y y_{i0} + \lambda_z \bar{z}_i + a_i$$

$$a_i | Z_{iT}, x_{i0} \sim N(0, 1)$$

where  $\bar{z}_i$  stands for the time average of  $z_{it}$ . Note that  $\lambda_d D_{i0}$  was omitted from the equation (15) as the treatment comes from  $t = 2$  and  $D_{i0} = 0$  for all observations.

At  $t = 2$ , treatment is assigned based on the policy equation (1.5)

$$D_{i2} = 1[\gamma_y y_{i1} + \gamma_z z_{i2} + \gamma_c c_i \leq \kappa + \varepsilon_{i2}]$$

where  $\varepsilon_{it}$  is drawn independently from standard normal distribution for all  $t = 2, 3$ . Then I used this  $D_{i2}$  to construct  $y_{i2}$  following equation (1.4)

$$y_{i2} = \beta_2 + \beta_y y_{i1} + \beta_d D_{i2} + \beta_z z_{i2} + c_i + u_{i2}$$

At  $t = 3$ , I designed the policy assignment as

$$D_{i3} = \{1[\gamma_y y_{i2} + \gamma_z z_{i3} + \gamma_c c_i \leq \kappa + \varepsilon_{i3}]\}^{1-D_{i2}} \cdot 1^{D_{i2}} \quad (1.21)$$

---

<sup>12</sup>For this paper, I used  $T = 3$  for the simplicity of discussion. Maximal length of time can be extended in the future research.

We can interpret the equation (1.21) as follows: If the observation was not treated in the previous period ( $D_{i2} = 0$ ), then they are assigned to the treatment program following the equation (1.5), as we did in the previous period; If the observation was treated last period ( $D_{i2} = 1$ ), then they are also treated this period too ( $D_{i3} = 1$ ). We can imagine this treatment program as one without out-movers, and the size of the treated group is monotonically increasing as time goes by. Note that the  $1^{D_{i2}}$  term is redundant and we can ignore or drop them during the actual estimation process, as they are always one.<sup>13</sup>

Now we have  $y_{i3}$  based on the equation (1.4)

$$y_{i3} = \beta_3 + \beta_y y_{i2} + \beta_d D_{i3} + \beta_z z_{i3} + c_i + u_{i3}$$

The structural parameters are set as follows:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_y \\ \beta_d \\ \beta_z \\ \sigma_u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.5 \\ 2 \\ 1 \\ 1 \end{pmatrix} \quad (1.22)$$

$$\begin{pmatrix} \gamma_y \\ \gamma_z \\ \gamma_c \\ \kappa \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad (1.23)$$

---

<sup>13</sup>I put it to help readers to understand the intuition of this DGP. Estimating equation (1.21) by probit after we drop the redundant term yield the conditional density function (1.20) in the previous section.

$$\begin{pmatrix} \lambda_1 \\ \lambda_y \\ \lambda_z \\ \sigma_\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 1 \end{pmatrix} \quad (1.24)$$

### 1.3.2 Simulation Results

As discussed above, we concentrate on the parameters of equation (1.4), especially on the estimates of  $\beta_d$ . This paper also briefly shows the simulation result for the estimation of  $\beta_y$ , which has similar implication. For the full result including the coefficients of time dummies and exogenous regressor, see appendix for the detail.

Table 1.1 shows the simulation results for  $\beta_d$  depending on the estimation strategies. For the simulations, I ran 1,000 repetition for each sample size from 500 to 10,000. As we can see, POLS, FE, FD's are biased, and the biases do not go away by the increasing sample size.

Table 1.1 Treatment Effect under Feedback Environment

$\beta_d = 2$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.675	0.113	0.684	-0.209	0.126	0.244	-0.007	0.121	0.122
$N = 1,000$	-0.674	0.079	0.678	-0.210	0.092	0.229	-0.006	0.088	0.089
$N = 5,000$	-0.670	0.036	0.671	-0.206	0.043	0.211	0.000	0.040	0.040
$N = 10,000$	-0.670	0.026	0.670	-0.206	0.031	0.208	0.000	0.029	0.029
$\beta_d = 2$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.601	0.143	0.618	-0.345	1.222	1.270	-0.162	1.634	1.642
$N = 1,000$	-0.602	0.100	0.610	-0.138	0.777	0.789	-0.045	0.857	0.859
$N = 5,000$	-0.599	0.048	0.601	-0.022	0.325	0.326	0.001	0.339	0.339
$N = 10,000$	-0.599	0.034	0.600	-0.007	0.223	0.223	0.005	0.235	0.235

Compared to Arellano and Bond (1991) estimates, Wooldridge (2000) method outperforms from a small sample size ( $N = 500$ ). For each sample size, Wooldridge (2000) estimates have smaller bias on average, smaller standard deviations, and smaller rMSE than those of Arellano and Bond (1991) results.

Another interesting point to pay attention is the difference between AB1 and AB2. According to Table 1.1, using collapsed instrument (AB2) works better than the original work (AB1) of Arellano and Bond (1991) for the all sample size. For both of AB1 and AB2, bias goes down towards zero as sample size goes up. Standard deviations and rMSEs also improve as the sample size goes to 10,000. Notably, such convergence is also faster in AB2 than AB1. For example, between  $N = 500$  and  $N = 1,000$ , the method based on the collapsed instrument (AB2) had approximately 72% decrease in the average bias, while that of original instrument (AB1) had approximately 60% decrease in the average bias.

Table 1.2 shows simulation results for the estimates of  $\beta_y$ , which was the traditional concern of the state-dependence/persistence literature. As we can see, Wooldridge (2000) method still overwhelms the competitions. For each sample size, Wooldridge (2000) approach have the smallest bias, the smallest standard deviation, and the smallest rMSE. Similar to Table 1.1, POLS, FE, FD are biased, and the biases do not go away, even though the sample size goes up.

It is also interesting to see that the collapsed instrument method (AB2) still keeps some level of relative efficiency over that of original instrument (AB1). For each sample size, AB2 has smaller bias and rMSE over AB1. For the standard deviation, however, AB1 outperforms AB2.

Table 1.2 Persistence under Feedback Environment

$\beta_y = 0.5$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	0.488	0.018	0.488	-0.124	0.021	0.126	-0.002	0.020	0.022
$N = 1,000$	0.488	0.012	0.488	-0.124	0.015	0.124	0.000	0.016	0.016
$N = 5,000$	0.488	0.006	0.488	-0.123	0.007	0.123	0.000	0.007	0.007
$N = 10,000$	0.488	0.004	0.488	-0.123	0.005	0.123	0.000	0.005	0.005
$\beta_y = 0.5$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.250	0.024	0.251	-0.111	0.386	0.402	-0.055	0.563	0.565
$N = 1,000$	-0.249	0.017	0.250	-0.049	0.258	0.263	-0.014	0.306	0.306
$N = 5,000$	-0.248	0.008	0.249	-0.007	0.110	0.110	0.001	0.123	0.123
$N = 10,000$	-0.249	0.005	0.249	-0.003	0.076	0.076	0.002	0.088	0.088

As a robustness check, I have studied additional simulations for: i) misspecified idiosyncratic

error; ii) misspecified heterogeneity distribution; iii) both error and heterogeneity are misspecified. Interestingly, for all cases, Wooldridge (2000) MLE approach works well and has similar implication to the correctly specified case. For the misspecified cases, even though Arellano and Bond (1991) GMM approaches starts to have less bias than those of MLE when the sample size gets close to 10,000, Wooldridge (2000) approach works more efficient than GMM (AB1 and AB2) while having similar level of bias. See appendix for the detail of misspecified cases.

#### **1.4 Discussion**

In this paper, I compared two estimation strategies, one of Arellano and Bond (1991) and another of Wooldridge (2000), in the context of treatment effect analysis. Under the correct specification, Wooldridge (2000) approach outperforms those of Arellano and Bond (1991) approaches, regardless of which instruments they are using. For outside options, POLS, FE, FD are inconsistent overall.

In future work, I would compare them under more misspecification including asymmetric distributions or heteroskedastic probit instead of usual probit. Another interesting question would be comparing them under the limited dependent variable, typically zero-one variable case. For example, I can compare an average partial effect from probit/logit method based on Wooldridge (2000), and the coefficient from a linear probability model based on Arellano and Bond (1991). Moreover, there are plenty of rooms for empirical illustration using these methods.

## CHAPTER 2

### CAUSAL INFERENCE OF DYNAMIC PANEL MODEL WITH NON-NEGATIVE OUTCOME

#### 2.1 Introduction

Econometricians introduced non-negative outcome models<sup>1</sup> into the economics literature by relaxing distributional assumptions (Gourieroux et al. 1984a,b) and developing a framework for the panel data analysis (Hausman et al. 1984) from the very early stage. The literature further relaxes the distribution assumption and mean-variance assumption when it comes to static panel data fixed-effects models (Wooldridge 1999), allowing applied economists to utilize the rich information of the panel data. Advances in non-negative outcome models attracted many researchers to use the exponential specification including the fields as, but not limited to, recreation demand (Ozuna and Gomez 1995), international trade (Santos Silva and Tenreiro 2006, 2011), R&D investment (Guceri and Liu 2019), crime (Lindo et al. 2018), and health (Balestra et al. 2021, Horn et al. 2022). Econometric literature also developed a straightforward approach to estimate the average treatment effects on treated for the non-negative outcome models (Lee and Kobayashi 2002, Ciani and Fisher 2018, Lee and Lee 2021). After adjusting the parallel trends assumption into the ratio form, practitioners can conduct the causal inference to estimate the effects in the proportional change interpretation. None of them above, however, considers dynamic nature of economic variables that is very common in the practice.

A recent paper in the health economics forms a good example of dynamic nature of economic variables. In Evans et al. (2022b), they study the impact of the prescription drug monitoring program (PDMP) on the child maltreatment. Under the assumption that abuse of controlled substance can affect child maltreatment, they estimate the effects of PDMP on child maltreatment, relying on the linear two-way-fixed-effects (TWFE) specification. They, however, do not consider two important

---

<sup>1</sup>There is no consensus on how to call the literature. Some authors prefer Poisson model, another group of people refer it under the title of count data model. I choose to call the literature as the non-negative outcome models as: i) Poisson regression does not necessarily rely on Poisson distribution assumption; ii) as long as the dependent variable has non-negative value, it does not necessarily have to have a count nor integer form. For the general introduction and overview, Cameron and Trivedi (2013) provides a comprehensive references on non-negative models literature.

factors that can change the policy implication: i) If the child maltreatment is affected by the abuse of materials, addictive properties of such substances can create a persistent behavior of outcome variables; ii) The different timing of the policy intervention might be assigned non-randomly. For example, The timing of PDMP implementation can be affected by the last period's socioeconomic status (e.g., child maltreatment) of each state.

In this paper I propose a dynamic panel data model with non-negative outcome, which i) allows applied economists to include the lagged outcome in their exponential models; ii) allows feedback from the past outcome to affect the treatment assignment, when the policy has a staggered intervention; iii) does not impose a linear relationship between the outcome variables; Applied economists can also benefit from my framework when they are focussing on the proportional interpretation, as the exponential mean specification of my paper relies on the weaker assumptions compared to those the usual log-level linear models assume.

My paper contributes the econometrics literature by suggesting a unified framework to estimate the conditional average treatment effects with the proportional change interpretation. My contribution can be viewed from two different perspectives of econometrics literature.

First, from the perspective of treatment effects literature, I fill the gap of staggered intervention literature by proposing a way to allow arbitrary correlation between the staggered assignment and the past outcomes and the unobserved heterogeneity. Modern prevalence of panel data models brought attention of both applied practitioners and theoretical econometricians when it comes to the causal inference in economics. The popularity comes from the rich information of the panel data structure which allows researchers to remove or control the unobserved heterogeneity (or unit fixed effects). Recent explosion of TWFE literature is also motivated from the real world problems where different units can enter into the treatment status at different timing (staggered intervention) and the heterogeneous intervention might happen over time (Callaway and Sant'Anna 2021, De Chaisemartin and d'Haultfoeuille 2020, Sun and Abraham 2021, Goodman-Bacon 2021, Borusyak et al. 2024). The literature points out that there might be a negative weights problem, and econometricians have been interested in solving the problem (Dube et al. 2023). The literature,

however, does not say much about the treatment timing assignment mechanism. If the heterogeneous intervention is affected by the past outcome, ignoring such mechanism could bias the treatment effects even when the effects themselves are homogeneous. Therefore, my paper focus on the endogeneity of the staggered intervention and develop a causal inference framework that allows feedback into the consideration of the applied economists.

Furthermore, my paper motivates treatment effects literature to include the lagged outcome into their consideration. Recent developments in the literature try to include the lagged outcome into the controlled variable (Caetano et al. 2022, Dube et al. 2023), but the progress gets relatively less attention<sup>2</sup>, and the situation does not improve much when it comes to the non-negative outcome models. My study contributes to the literature by developing non-negative outcome models that allows including lagged outcome during the causal inference.

Second, from the perspective of dynamic panel model literature, I propose a dynamic non-linear model by capitalizing an exogeneity assumption that has not been viewed as a tool for causal inference. Specifically, my paper relies on assumption in the form of sequential exogeneity, which can be traced back to the dynamic linear panel data models literature (Arellano and Bond 1991, Ahn and Schmidt 1995, Arellano and Bover 1995, Blundell and Bond 1998). Combined with the sequential exogeneity assumption, the transformation technique suggested by Chamberlain (1992) and Wooldridge (1997) allows the feedback of lagged dependent variable to affect policy assignment of current period. This assumption was proposed to estimate the persistence parameter, but my paper implies that the assumption can be a useful device to estimate the conditional average treatment effects. The advantage of the sequential exogeneity assumption comes from the fact that it is weaker than the usual strict exogeneity assumption, which is commonly and implicitly assumed in the treatment effects literature.

The rest of the paper is organized as follows. Section 2.2 introduces my model and argues that the dynamic model with multiplicative form has a less restrictive data structure compared to those with

---

<sup>2</sup>Including lagged outcome as covariates while assuming parallel trends assumption can be either difficult or even impossible. Literature have been suspicious about mixing parallel trends assumption and the lagged outcome together (Angrist and Pischke 2009, Chabé-Ferret 2017, Ding and Li 2019).



the additive form. Section 2.3 discusses the identification of conditional average treatment effects (CATE) and its necessary assumptions. Section 2.4 explains the generalized method of moments (GMM, hereafter) strategies suggested by Wooldridge (1997) and their necessary assumptions, and explains the details of the adjustment when we want to use the framework in the context of causal inference. Section 2.5 presents the Monte Carlo simulation to show that the framework works well and shows asymptotic behavior. In Section 2.6, I revisit the causal inference of Prescription Drug Monitoring Program during the opioid crisis in United States (Evans et al. 2022b) to show that, under the staggered intervention where different units have different entrance timing, path dependence of outcome or feedback nature of dynamic treatment assignment might yield a significantly different policy implication. After the empirical application, I conclude the paper with a short discussion about the future research.

## 2.2 Econometric Model

### 2.2.1 Notations, Setup, and Potential Outcome Model

In this paper, I rely on the standard Neyman-Rubin causal model as the basic setup. Let  $Y_{it}$  denote the dependent variable at time  $t \in \vec{T} \equiv \{1, 2, \dots, T\}$ . Then, one can write the potential outcome model as a additive form of two potential status, treated and controlled:

$$Y_{it} = D_{it} \cdot Y_{it}^{(1)} + (1 - D_{it}) \cdot Y_{it}^{(0)} \quad (2.1)$$

where  $Y_{it}^{(1)}$  is the potential outcome of observation  $i$  in the treated state at time  $t$ , while  $Y_{it}^{(0)}$  is the potential outcome of the controlled state. Let  $D_{it}$  be the treatment indicator which is one if the observation observation  $i$  is treated at time  $t$ , and zero when controlled. Let  $X_{it}$  be a  $1 \times K_x$  vector of contemporaneous conditioning covariates which is strictly exogenous. For each  $t = 1, \dots, T$ , an econometrician observe  $(Y_{it}, X_{it}, D_{it})$  and has a complete panel of them, so that we can focus on the balanced panel case. We further assume that the lagged outcome at  $t = 1$ ,  $Y_{i0}$  is exogenously given for all  $i = 1, \dots, N$ , but we do not observe  $X_{i0}$ . Furthermore, define the ever treated indicator  $G_i = \max_{s \in \vec{T}} D_{is}$  which is one if the unit is ever treated within the period  $t = 1, 2, \dots, T$ .

This paper is interested in the environment where the policy intervention starts at  $t = 2$  so that

we have  $t = 1$  as controlled period, which should be familiar to the researchers who are interested in staggered intervention studies. Furthermore, throughout this paper, I assume the no reversibility assumption, which means that there is no exit once the unit has entered into the treated status. To be specific, without loss of generality,

$$D_{is} \geq D_{it} \quad \forall s \geq t, \quad \forall t = 1, \dots, T$$

This is a standard assumption that is commonly assumed, either explicitly or implicitly, in the staggered intervention literature. My paper also focus on the same setup where the treatment status is absorbing<sup>3</sup>.

One might notice that my model still use the classical additive form model (2.1), even though the literature of non-negative outcome models usually impose exponential mean specification. In this paper, I am interested in estimating the conditional average treatment effects with the ratio form:

$$\frac{E(Y_{it}^{(1)} - Y_{it}^{(0)} | Y_{it-1}, \dots, Y_{i1}, X_i, D_{it} = 1)}{E(Y_{it}^{(0)} | Y_{it-1}, \dots, Y_{i1}, X_i, D_{it} = 1)} = e^\tau - 1$$

where we can interpret the ratio as the proportional change of potential outcome relative to the controlled outcome, conditional on covariates. For the algebraic details, see Section 2.3.

## 2.2.2 Parametric Model

Suppose that the potential outcome model satisfies

$$E(Y_{it}^{(d)} | c_i, \vec{Y}_{it-1}, X_i; \theta) = c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_{it} \beta_x + \tau d) \quad (2.2)$$

for treated ( $d = 1$ ) and controlled ( $d = 0$ ) groups, where  $\theta = (\beta_t, \beta_y, \beta_z, \tau, \beta'_x)'$  is  $K \times 1$  vector of parameters<sup>4</sup>,  $\vec{Y}_{it-1} = \{Y_{it-1}, Y_{it-2}, \dots, Y_{i1}\}$  is the history of outcome variables until  $t - 1$ ,  $c_i$  is the  $1 \times 1$  unobserved heterogeneity (or unit fixed effects),  $\beta_t$  are common time-specific effects,

---

<sup>3</sup>For a generalized framework with dynamic treatment effects where, for example, treated units can exit after receiving treatment, one can consult the structural nested mean models (SNMM) literature (Robins 1986, 1994, 1997). This generalized framework, however, reaches beyond the scope of this paper.

<sup>4</sup>This paper assumes the common parameters between the treated and controlled outcome, and I use  $\theta$  instead of  $\theta_0$  to indicate the true parameter of the model. The only exception is the section 2.4.1., where I denote  $\theta_0$  as the true parameter. I am abusing notation in 2.4.1 as I don't have to list parameters with  $\beta_{t0}, \beta_{y0}, \beta_{z0}$ , which look like separate parameter for controlled potential outcome.

$Z_{it-1} = 1[Y_{it-1} = 0]$  is the indicator function whether the past outcome is zero, and  $X_{it}$  is the contemporaneous covariates which is strictly exogenous<sup>5</sup>. The model is motivated by Crépon and Duguet (1997) who suggested a multiplicative function approach, but the specific functional form can be also found in an alternative suggestion of Windmeijer (2006). Again, after some algebraic details and assumptions described in Section 2.3, the model (2.2) becomes

$$E(Y_{it}|c_i, \vec{Y}_{it-1}, X_i, D_{it}; \theta) = c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + \tau D_{it} + X_{it}\beta_x) \quad (2.3)$$

The model further simplifies depending on the value of past outcomes:

If  $Y_{it-1} = 0$ , then  $Z_{it-1} = 1$  and the function becomes

$$E(Y_{it}|c_i, Y_{it-1} = 0, \vec{Y}_{it-2}, X_i, D_{it}; \theta) = c_i \cdot \exp(\beta_t + \beta_z + \tau D_{it} + X_{it}\beta_x).$$

If  $Y_{it-1} > 0$ , then  $Z_{it-1} = 0$  and the function is taking the form of

$$\begin{aligned} E(Y_{it}|c_i, Y_{it-1} > 0, \vec{Y}_{it-2}, X_i, D_{it}; \theta) &= c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1}) + X_{it}\beta_x + \tau D_{it}) \\ &= c_i \cdot Y_{it-1}^{\beta_y} \cdot \exp(\beta_t + X_{it}\beta_x + \tau D_{it}). \end{aligned}$$

The model has exactly the same form as the linear feedback model (LFM) suggested by Blundell et al. (1995, 2002) when  $Y_{it-1} = 0$ . When the lagged outcome has a positive value ( $Y_{it-1} > 0$ ), my model differs from the LFM as the LFM separates the lagged outcome from the exponential function in an additive form:

$$E(Y_{it}|c_i, \vec{Y}_{it-1}, X_i, D_{it}; \theta) = \beta_y Y_{it-1} + c_i \cdot \exp(\beta_t + X_{it}\beta_x + \tau D_{it}).$$

Notice that we need to pull out the lagged outcome term from exponential term either in multiplicative or additive form. When we leave the persistence term inside the exponential function, the entire outcome dynamics tends to explode in a short time (Cameron and Trivedi 2005, Windmeijer 2006).

---

<sup>5</sup>I put  $X_i$ , the entire history of  $X_{it}$  for  $t = 1, \dots, T$ , in the conditioning set on the left hand side to emphasize that the covariate vector  $X_{it}$  is considered strictly exogenous. Since I am assuming that they are strictly exogenous, just conditioning on  $X_{it}$  has the same model specification.

Compared to the conventional LFM, there are two advantages to use my multiplicative form for the dynamic panel models with non-negative outcomes. First, it allows researchers to include lagged outcome without imposing restrictive assumptions on dynamics of dependent variables. For this linear separation, suppose we have the unit root (i.e.,  $\beta_y = 1$ ). Then one can observe that LFM is imposing a monotonically increasing series for the dynamics of outcome variables. As a result, LFM implicitly assumes that once the dependent variable took off from zero ( $Y_{it-1} > 0$ ), then it is almost impossible to have zero values in the future

$$E(\Delta Y_{it} | c_i, Y_{it-1} > 0, \vec{Y}_{it-2}, X_i, D_{it}; \theta) = c_i \cdot \exp(\beta_t + X_{it}\beta_x + \tau D_{it}) \geq 0.$$

Even if we relax the unit root assumption and allow relatively small values for the persistence parameter  $\beta_y < 1$ , the additive term,  $\beta_y Y_{it-1}$ , still generates a lower bound for the future values. As a result, the data cannot have zero values in the observation in the future, once the unit experienced positive outcome. Such monotonicity could be a restrictive assumption for broad practitioners. By the nature of multiplicative form, however, my model yields a different result. By rewriting equation (2.3), we have a ratio between the two non-negative variables,

$$E \left[ \frac{Y_{it}}{Y_{it-1}^{\beta_y}} | c_i, Y_{it-1} > 0, \vec{Y}_{it-2}, X_i, D_{it}; \theta \right] = c_i \cdot \exp(\beta_t + X_{it}\beta_x + \tau D_{it}) \geq 0$$

where its non-negativeness makes sense as ratios of non-negative values are always non-negative. Unlike LFM, my framework with the multiplicative specification does not impose monotonic outcome dynamics while it still allows to include lagged outcome as covariates.

Second, my model (2.3) also allows the researchers to impose feedback of the lagged outcome to affect today's treatment assignment. Even if the researchers decide not to include the lagged outcome as covariates, they might be still interested in allowing yesterday's outcome to affect today's policy intervention (feedback). I also want to emphasize that the model also allows to have an arbitrary correlation between the treatment indicator and unobserved heterogeneity, regardless of allowing the feedback. In section 2.4, I explain that estimation of model (2.3) requires no structural equation

for treatment assignment mechanism, which means that my model allows arbitrary relationship between the treatment assignment and both of lagged outcome and unobserved heterogeneity.

## 2.3 Identification

### 2.3.1 Motivation of Conditional Average Treatment Effects

Static panel data models which exclude lagged outcome as covariates have standardized framework to apply the causal inference with a panel data, not only for linear models but also for non-linear models as difference-in-differences with non-negative outcomes. For example, proportional effect on the treated group at the post period can be identified as a ratio of difference of potential outcomes and the base of potential controlled outcome (Lee and Kobayashi 2002, Lee and Lee 2021), which is characterized by

$$\frac{E \left[ Y_{it}^{(1)} - Y_{it}^{(0)} | G_i = 1 \right]}{E \left[ Y_{it}^{(0)} | G_i = 1 \right]} \quad (2.4)$$

This might be enough when econometricians are getting their data from randomized trial where systematic gap between the observations are controlled by experiment design. In the real world, however, economists have been obtaining observational data to study the causal effects of policy interventions, where controlling covariates could get crucial issues. Therefore, conditional average treatment effects (CATE) has been a major parameter of interest in the ratio-based treatment effects literature (Lee and Kobayashi 2002, Lee and Lee 2021, Yadlowsky et al. 2021), where the effects of interest can be written as

$$\frac{E \left[ Y_{it}^{(1)} - Y_{it}^{(0)} | W = w, G_i = 1 \right]}{E \left[ Y_{it}^{(0)} | W = w, G_i = 1 \right]}. \quad (2.5)$$

There are three advantages in estimating the treatment effects in the form of (2.5). By taking ratios, we can cancel out the unobserved heterogeneity where simple subtraction between the exponential terms will not yield such result. Second, the ratio form is useful when an econometrician estimates the percentage change. Notice that the ratio of potential outcomes themselves,  $(Y_{it}^{(1)} - Y_{it}^{(0)})/Y_{it}^{(0)}$ , will not be well defined when  $Y_{it}^{(0)} = 0$ . By relying on the exponential specification, however, an econometrician can bypass the problem and the ratio (2.5) would be interpreted as

the proportional change. Specifically, (2.5) shows the proportional change relative to the potential outcome as the base level, conditional on the covariates  $W$  at  $w$ . Finally, under the exponential mean model, conventional difference estimator for average treatment effects could introduce the unnecessary heterogeneity (Yadlowsky et al. 2021, Lee and Lee 2021). To be specific, suppose the ratio (2.5) is constant  $\tau$  and we assume that the denominator is the function of covariate  $W$ . Then the numerator is equal to  $\tau$  times the denominator, whose heterogeneity might not be our special interest. For these reasons, static models with non-negative outcome structure have been interested in estimating ratio of conditional expectations when they want to conduct the causal inference.

For the dynamic panel data models, however, relatively less is known in the context of the causal inference with the lagged outcome. For example, an integration of parallel trends and sequential ignorability has been producing less progress until now (Roth et al. 2023). Structural nested mean literature in the statistics and biostatistics (Robins 1986, 1994, 1997) relies on a specific device called the blip function, whose application does not necessarily coincide with the convention of staggered intervention<sup>6</sup>. Therefore, this paper motivated its identification of treatment effect with lagged outcome from a work of Ding and Li (2019) who compared the difference-in-differences approach with lagged-dependent-variable adjustment in the context of nonparametric identification.

Suppose that difference-in-differences approach is interested in identification and estimation of the parameter  $\tau_{DiD}$ , which is equivalent to the average treatment effects for treated under the parallel trends assumption. In the original language of Ding and Li (2019), they define their difference-in-differences estimator as

$$\hat{\tau}_{DiD} = (\bar{Y}_{1,t+1} - \bar{Y}_{1,t}) - (\bar{Y}_{0,t+1} - \bar{Y}_{0,t})$$

which can be written as, after translating notations into those described in Section 2.2,

$$\tau_{DiD} = E[Y_{it} - Y_{it-1} | G_i = 1] - E[Y_{it} - Y_{it-1} | G_i = 0]$$

Rewriting above equation yields

$$\tau_{DiD} = (E[Y_{it} | G_i = 1] - E[Y_{it-1} | G_i = 1]) - (E[Y_{it} | G_i = 0] - E[Y_{it-1} | G_i = 0])$$

---

<sup>6</sup>For the recent development of combination of structural nested mean model and the parallel trends assumption, one might consult the recent paper of Shahn et al. (2022)

$$= (E[Y_{it}|G_i = 1] - E[Y_{it}|G_i = 0]) - (E[Y_{it-1}|G_i = 1] - E[Y_{it-1}|G_i = 0])$$

which is analogous to the expression of Ding and Li (2019) after changing places of terms

$$\hat{\tau}_{DiD} = (\bar{Y}_{1,t+1} - \bar{Y}_{0,t+1}) - (\bar{Y}_{1,t} - \bar{Y}_{0,t})$$

For the average treatment effects with lagged outcome, Ding and Li (2019) defined their estimator for the conditional average treatment effect on treated:

$$\hat{\tau}_{LDV} = (\bar{Y}_{1,t+1} - \bar{Y}_{0,t+1}) - \hat{\beta}_y(\bar{Y}_{1,t} - \bar{Y}_{0,t}) = (\bar{Y}_{1,t+1} - \hat{\beta}_y\bar{Y}_{1,t}) - (\bar{Y}_{0,t+1} - \hat{\beta}_y\bar{Y}_{0,t})$$

where the AR(1) persistence parameter  $\beta_y$  comes from the conditional mean functional form assumption of Ding and Li (2019)

$$E(Y_{it}|Y_{it-1}, G_i) = \beta_0 + \beta_y Y_{it-1} + \tau G_i.$$

Note that unobserved heterogeneity  $c_i$  does not show up on Ding and Li (2019), and the conditional mean function is using the ever-treated group indicator  $G_i$  instead of treatment effects indicator  $D_{it}$ . Their equation for conditional mean specification works in their  $2 \times 2$  setup. When  $T \geq 3$ , conditioning on  $G_i$  could make a restriction that does not fit well to the feedback environment. Recall the definition of  $G_i$  which is a non-linear function of entire history of treatment assignment. Conditioning on  $G_i$  is similar to conditioning on entire history of  $D_{it}$ , which is equivalent to imposing a strict exogeneity of treatment indicator<sup>7</sup>. In this paper, I suggest the exponential analog of conditional average treatment effects inspired by the definition of Ding and Li (2019). The rest of this section introduces nonparametric and parametric identification of conditional average treatment effects along with necessary assumptions.

### 2.3.2 Identification of Average Treatment Effects with Ratio Form

I begin the identification procedure from suggesting necessary assumptions. The first identification assumption is the homogeneous treatment effect assumption in the ratio form.

---

<sup>7</sup>For example, suppose  $T = 3$  and the unit  $j$  is treated at the last period 3. If we use the specification with  $G_i$  as suggested in Ding and Li (2019), conditional mean of  $Y_{j2}$  is already conditioning on the future treatment assignment,  $D_{j3}$  as the unit  $j$  is in the ever-treated group  $G_j = \max D_{js} = D_{j3} = 1$ . It would be impossible for  $Y_{j2}$  to affect  $D_{j3}$  as  $D_{j3}$  is already fixed by conditioning on  $G_j$ . Similar argument (persistence cannot affect assignment) could be found from the literature (Ghanem et al. 2022) in the context of the necessary and sufficient condition of parallel trends assumption.

**Assumption 2.1. Homogeneous Treatment Effects.** For each  $t = 2, 3, \dots, T$ , conditional average treatment effects is homogeneous. That is,

$$\frac{E \left[ Y_{it}^{(1)} - Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]} = \frac{E \left[ Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]} - 1 = \exp(\tau) - 1,$$

where  $\tau \in \mathbb{R}$  is the non-stochastic constant unknown number.

This assumption is analogous to the linear model where econometricians assume

$$E \left[ Y_{it}^{(1)} - Y_{it}^{(0)} | D_{it} = 1 \right] = \tau, \quad t = 1, \dots, T$$

Note that  $\exp(\tau) = \xi$  and  $\tau = \ln(\xi)$  for any constant  $\xi \in \mathbb{R}_{++}$ , so the right hand side can be any positive value. One can relax this assumption by allowing time-varying treatment effects relatively easily, but I want to focus on the case when we have a homogeneous treatment effects and see the impact of introducing lagged outcome or feedback into our consideration. By assuming homogeneous effects, we can also bypass the negative weights problem in the staggered intervention literature.

The second assumption is an analog of Ignorability assumption in Ding and Li (2019).

**Assumption 2.2. Sequential Ignorability.**

$$Y_{it}^{(d)} \perp\!\!\!\perp D_{it} | (c_i, \vec{Y}_{it-1}, X_i), \quad d = 0, 1.$$

The assumption differs from the assumption 2 of Ding and Li (2019) as: i) I explicitly include the unobserved heterogeneity  $c_i$  into consideration while Ding and Li (2019) ignores it; ii) I conditioned on the history of lagged outcomes instead of one lag,  $Y_{it-1}$ , as I want to allow more than two period in my framework ( $T \geq 2$ ); iii) I used time-varying  $D_{it}$  while Ding and Li (2019) used time-constant  $G_i$ . This does not creates much problem when  $T = 2$ , but as the time length gets bigger ( $T > 2$ ), using  $G_i$  works as if the researcher imposes strict exogeneity of  $D_{it}$ , which rules out the feedback environment.

Under these assumptions, we can have a nonparametric identification of the conditional average treatment effects in the ratio form.



**Proposition 2.1. Identification of CATE with Ratio Form.** Under the Neyman-Rubin causal model (2.1), homogeneous treatment effects (Assumption 2.1), and sequential ignorability (Assumption 2.2), conditional average treatment effects could be identified by

$$\frac{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0 \right]} - 1 = \frac{E \left[ Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]} - 1 = \exp(\tau) - 1$$

Furthermore, it is also true that

$$\frac{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0 \right]} - 1 = \frac{E \left[ Y_{it}^{(1)} | \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]} - 1 = \exp(\tau) - 1 \quad \square$$

For the proof, see appendix B.1. Under the homogeneous treatment effects (Assumption 2.1), econometricians can further identify unconditional average treatment effects.

**Corollary 2.1. Identification of Average Treatment Effects with Ratio Form.** Under the same assumptions as in Proposition 2.1, average treatment effects would be identified by

$$\frac{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0 \right]} = \frac{E \left[ Y_{it}^{(1)} | D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | D_{it} = 1 \right]} = \exp(\tau).$$

The proof of the corollary use the same iteration of expectation using the conditional distribution of  $D(\vec{Y}_{it-1}, X_i | D_{it} = 1)$ , as we did in the Proposition 2.1.

For the parametric identification of average treatment effects with the ratio form, I need additional assumptions. Let me begin with the parametric potential outcome model by repeating the equation (2.2):

$$E(Y_{it}^{(d)} | c_i, \vec{Y}_{it-1}, X_i; \theta) = c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_{it} \beta_x + \tau d), \quad d = 0, 1.$$

Observe that the potential outcome model (2.2) is a parametric version of homogeneous treatment effects as the only difference between the controlled and treated outcome comes from the gap between  $\exp(\tau)$ . Next, we assume that the conditional mean function of outcome model.

$$E(Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it}; \theta) = c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_{it} \beta_x + \tau D_{it}) \quad (2.6)$$

Combining the potential outcome model (2.2) and the conditional mean of outcome (2.6) yields an interesting result, which is already provided in Lee and Kobayashi (2002). In this paper, I extended the conditioning set to show that we can achieve the parametric identification even after conditioning on the lagged outcomes.

**Proposition 2.2. Identification without Selection Bias.** Suppose we assume the standard Neyman-Rubin causal model (2.1) and the parametric potential outcome model (2.2). If we further assume the conditional mean of outcome (2.6), then we have

$$E(Y_{it}^{(d)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = d; \theta) = E(Y_{it}^{(d)} | c_i, \vec{Y}_{it-1}, X_i; \theta) \quad (2.7)$$

If we assume (2.1), (2.2), and (2.7) instead, then we have the outcome model (2.6) as a result.

For the proof, see appendix B.2. As discussed in Lee and Kobayashi (2002), (2.7) means that there is no selection bias, or we can just regard (2.7) as the mean-independence assumption. The Proposition 2.2 basically says that under the mean-independence assumption (2.7) and potential outcome model with homogeneous treatment effects (2.2), conventional Neyman-Rubin causal model (2.1) provides the parametric identification of the outcome model (2.6) to estimate the conditional average effects in the ratio form. One can also regard the conditional mean specification (2.6) makes a necessary and sufficient condition for the mean-independence assumption (2.7), given (2.1) and (2.2).

The following proposition presents my main identification result:

**Proposition 2.3. Parametric Identification of CATE with Ratio Form.** Suppose the standard Neyman-Rubin causal model (2.1), parametric potential outcome model with homogeneous treatment effects (2.2), and the mean-independence assumption (2.7) holds. Then the conditional average treatment effect is identified by the form of ratio

$$\frac{E(Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1)}{E(Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0)} - 1 = \frac{E[Y_{it}^{(1)} - Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]}{E[Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]} = \exp(\tau) - 1. \quad (2.8)$$

For the proof, see appendix B.3. Again, identified parameter  $\tau$  would be interpreted as the proportional change relative to the potentially controlled outcome as the base level.

## 2.4 Estimation

In this section, I explain the Wooldridge (1997) GMM framework to estimate the parametric model (2.3). This approach has three advantages in the study of causal inference with the lagged outcome. First, the framework allows us to include lagged outcome as covariates, which was allowed in the static fixed effects Poisson models. Second, it also allows the treatment indicator  $D_{it}$  to be arbitrarily correlated with both of unobserved heterogeneity  $c_i$  and the lagged outcome (feedback). Here, I want to emphasize that we do not need to impose any specific structural models to capture the relationship between the treatment indicator and confounding heterogeneity or feedback. Finally, statistical inference is straightforward as the asymptotic behavior of the resulting estimator, including consistency and asymptotic normality, is followed by the conventional theory of GMM.

### 2.4.1 Wooldridge (1997) GMM Approach

Wooldridge (1997) estimates the multiplicative model in the form of

$$y_{it} = c_i \cdot \mu(W_{it}; \theta_0) \cdot u_{it} \quad (2.9)$$

where  $c_i$  is the unobserved heterogeneity, and  $u_{it}$  is the unobserved time-varying error term.  $W_{it}$  stands for the vector of covariates at  $t$ . Note that the lagged outcome  $Y_{it-1}$  and the function of lagged outcome could be included in the  $W_{it}$ .  $\theta_0$  is a  $K \times 1$  real column vector of unknown true parameters<sup>8</sup>. The key assumption to estimate the model above comes from the dynamic panel data model literature (Chamberlain 1992, Wooldridge 1997).

**Assumption 2.3. Sequential Exogeneity.** For each  $t = 1, 2, \dots, T$ , suppose the model follows the functional form of (2.9), and let  $u_{it}$  and  $\{W_{i1}, W_{i2}, \dots, W_{it}\}$  is defined as described above. We assume

$$E(u_{it} | c_i, W_{it}, W_{it-1}, \dots, W_{i1}) = 0, \quad \forall t = 1, 2, \dots, T \quad \square$$

Note that the Assumption 2.3 is weaker than the usual assumption used in the panel fixed-effects literature. For example, strict exogeneity assumption in Wooldridge (1999) assumes that

---

<sup>8</sup>As mentioned in section 2.2, I am denoting  $\theta_0$  as the true parameter instead of  $\theta$ . I abuse notation in this section as I need the notation  $\theta$  during the GMM procedure (13).

entire history of covariates should be considered in the conditioning set

$$E(u_{it}|c_i, W_{iT}, W_{it-1}, \dots, W_{i1}) = 1 \quad \forall t = 1, 2, \dots, T.$$

To construct an orthogonality condition, define  $r_{it}$  following the transformation suggested by the literature (Chamberlain 1992, Wooldridge 1997).

$$r_{it} \equiv y_{it} - y_{it+1} \cdot \frac{\mu(W_{it}; \theta_0)}{\mu(W_{it+1}; \theta_0)}, \quad t = 1, 2, \dots, T-1 \quad (2.10)$$

Observe that the second part of the right hand side is trying to mimic  $y_{it}$  using the model specification, and leave the term  $u_{it+1}$  only. To be specific, combined with the model (2.9),  $r_{it}$  would be

$$r_{it} = c_i \cdot \mu_{it} \cdot u_{it} - (c_i \cdot \mu_{it+1} \cdot u_{it+1}) \cdot \frac{\mu_{it}}{\mu_{it+1}} = c_i \cdot \mu_{it} \cdot (u_{it} - u_{it+1})$$

where  $\mu_{it} = \mu(W_{it}; \theta_0)$ . Then we have the moment conditions to use.

**Lemma 2.4.1** *Lemma 2.1. of Wooldridge (1997) Under the multiplicative model (2.9) and Assumption 2.3. (Sequential Exogeneity),*

$$E(r_{it}|c_i, W_{it}, W_{it-1}, \dots, W_{i1}) = 0 \quad \square$$

The proof is already provided by Wooldridge (1997). The direct result of this lemma is the following moment condition

$$E[W_i' r_i(\theta_0)] = E[\psi_i(\theta_0)] = 0 \quad (2.11)$$

where  $r_i(\theta_0) = (r_{i1}(\theta_0), r_{i2}(\theta_0), \dots, r_{iT-1}(\theta_0))'$  is the  $(T-1) \times 1$  vector, while the matrix  $W_i$  is a  $L \times L$  matrix constructed in the following block-diagonal form:

$$W_i \equiv \begin{pmatrix} \vec{W}_{i1} & 0 & \cdots & 0 & 0 \\ 0 & \vec{W}_{i2} & \cdots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \cdots & 0 & \vec{W}_{iT-1} \end{pmatrix} \quad (2.12)$$

where  $\vec{W}_{it}$  is the  $1 \times L_t$  vector of functions of  $W_{i1}, \dots, W_{it}$  for each  $t = 1, \dots, T-1$ , and  $L = L_1 + L_2 + \dots + L_{T-1}$ . For the details of specific form of instrument matrix (IV matrix, hereafter) and necessary adjustments, see Section 2.4.2.

Under the moment condition (2.11), the GMM estimator  $\hat{\theta}$  is a solution of the following minimization problem

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left[ \sum_{i=1}^N W_i' r_i(\theta) \right]' \tilde{\Omega}^{-1} \left[ \sum_{i=1}^N W_i' r_i(\theta) \right] \quad (2.13)$$

where  $r_{it}(\theta)$  is defined as in (2.10) which is the function of  $\theta$ , and  $\tilde{\Omega}$  is a GMM weighting matrix. The optimal weighting matrix with the moment condition (2.11) would be a consistent estimator of  $\Omega_0 \equiv E(\psi_i(\theta_0)\psi_i'(\theta_0))$ , which could be written as

$$\tilde{\Omega} \equiv \frac{1}{N} \sum_{i=1}^N W_i' r_i(\tilde{\theta}) r_i(\tilde{\theta})' W_i \quad (2.14)$$

where  $\tilde{\theta}$  is an arbitrary  $K \times 1$  vector. Recall that  $\tilde{\theta}$  does not necessarily have to be a consistent estimator of  $\theta_0$ . For example, one can use initially inconsistent estimator including fixed effects Poisson estimator to construct the weighting matrix. In fact,  $\tilde{\Omega}$  can be an arbitrary  $L \times L$  vector. Therefore, one can try the following two step GMM approach to estimate  $\hat{\theta}$ : In the first step, plug a  $L \times L$  identity matrix,  $\tilde{\Omega} = I_{L \times L}$  and get  $\tilde{\theta}$  which is could be an non-optimal estimate of  $\theta_0$ . Note that this  $\tilde{\theta}$  is still a consistent estimator of  $\theta_0$ . Then, use  $\tilde{\theta}$  and repeat the procedure (2.13), but this time, we use the estimator of optimal weight defined in (2.14). This procedure does not require the arbitrary choice of initial  $\tilde{\theta}$  by replacing  $\tilde{\Omega}$  with the identity matrix. One can choose arbitrary  $\tilde{\theta}$  instead. The asymptotic normality is a straightforward result of the conventional theory of GMM.

**Theorem 1. Asymptotic Normality of GMM.** Under the regularity conditions (Hansen (1982)) and the moment condition (2.11), the solution  $\hat{\theta}$  to  $\sum_{i=1}^N \psi_i(\theta) = 0$  achieves

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightarrow^d N(0, \Lambda_\psi)$$

where  $\Lambda_\psi \equiv (R_0' \Omega_0^{-1} R_0)^{-1}$ ,  $R_0 \equiv E(W_i' \nabla_\theta r_i(\theta_0))$ , and  $\Omega_0 \equiv E(\psi_i(\theta_0)\psi_i'(\theta_0))$   $\square$

### 2.4.2 Construction of IV Matrices

If an econometrician wants to utilize the Wooldridge (1997) GMM approach in the context of staggered intervention, one needs to deviate from the original work of Wooldridge (1997). In this section, I briefly discuss necessary adjustment applied to Wooldridge (1997) to fit my model (2.3) and the staggered intervention environment. I can write the model as

$$\mu(W_{it}; \theta) = \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + \tau D_{it} + X_{it} \beta_x)$$

while  $\theta' = (\beta_t, \beta_y, \beta_z, \tau, \beta_x')'$  is the  $K \times 1$  vector of the parameters of interest. There are two variables that I want to assume beyond the sequential exogeneity. First, I want to assume that time dummies are non-random. Second, I want to include the strictly exogenous time-varying covariates  $X_{it}$ . Therefore, I need to change the moment condition as

$$E(u_{it}|c_i, W_{it}, W_{it-1}, \dots, W_{i1}, X_i) = 0, \quad \forall t = 1, 2, \dots, T$$

while  $X_i$  includes the time dummies and the entire history of strictly exogenous covariates,  $X_{i1}, X_{i2}, \dots, X_{iT}$ . Observe that  $Y_{it-1}$  and its functions,  $Z_{it-1}, D_{it}$  (if we allow feedback), are still inside the vector of sequentially exogenous covariates<sup>9</sup>. In accordance with the adjustment on the baseline assumption, we change the IV instrument matrix  $W_i$ . By adding  $X_i$  to the IV matrix (2.12), we have

$$W_i \equiv \begin{pmatrix} W_{i1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & X_{i1} \\ 0 & W_{i1} & W_{i2} & \cdots & 0 & 0 & \ddots & 0 & X_{i2} \\ & & & \ddots & & 0 & & 0 & \vdots \\ 0 & 0 & 0 & \cdots & 0 & W_{i1} & \cdots & W_{iT-1} & X_{iT-1} \end{pmatrix} \quad (2.15)$$

where we can find the motivation of such matrix from the dynamic panel linear model literature (Arellano and Bond 1991). One theoretical caveat of Wooldridge (1997) framework comes from the fact that the choice of IV is arbitrary<sup>10</sup>. As a natural alternative of Arellano and Bond (1991)

<sup>9</sup>If a researcher faithfully believes that the treatment assignment is strictly exogenous, and uncorrelated with the feedback from past outcome, one can put  $D_{it}$  into  $X_{it}$  instead. In this case, we still allow  $D_{it}$  to be correlated with unobserved heterogeneity.

<sup>10</sup>The original proposal of Wooldridge (1997) was not necessarily efficient. One might try to achieve the optimal IV following the guidance of Chamberlain (1992). This paper does not cover the derivation of such optimal IV as it was too difficult to obtain in the environment of this paper

type of IV matrix, I also tried another form of IV suggested by Roodman (2009a,b), who collapsed the block-diagonal matrix into the triangular matrix

$$W_i = \begin{pmatrix} W_{i1} & 0 & \cdots & 0 & 0 & X_{i1} \\ W_{i1} & W_{i2} & \cdots & 0 & 0 & X_{i2} \\ & & \ddots & & & \vdots \\ W_{i1} & W_{i2} & \cdots & W_{iT-2} & W_{iT-1} & X_{iT-1} \end{pmatrix}$$

Finally, recall that the first period is the controlled period which makes  $D_{i1} = 0$  for all  $i = 1, \dots, N$ . To avoid any zero vectors, I dropped  $D_{i1}$  from the  $W_{i1}$ .

## 2.5 Monte Carlo Simulation

### 2.5.1 Benchmark DGP

In this section, I present a Monte Carlo simulation study to check the asymptotic behavior of GMM approaches suggested in the previous section. The true DGP of this simulation study has the form of

$$Y_{it} = c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + \tau D_{it} + X_{it} \beta_x) \cdot \varepsilon_{it}$$

$$c_i = \exp(\lambda_1 + \lambda_x \bar{X}_i + \lambda_y Y_{i0} + \lambda_a a_i)$$

where  $a_i \sim N(0, 1)$ ,  $\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it}$ , and  $X_{it}$  is also following the standard normal and  $Y_{i0}$  is exogenously given. The benchmark DGP has the parameter value of  $\beta_y = 0.3$ ,  $\beta_z = 0$ ,  $\beta_x = 0.2$ , and  $\tau = 0.3$ . To allow correlations between  $c_i$  and covariates, I also put  $\lambda_1 = 0$ ,  $\lambda_x = 1$ ,  $\lambda_y = 1$ , and  $\lambda_a = 1$  in the benchmark DGP. I draw the idiosyncratic error  $\varepsilon_{it}$  from the Poisson distribution, so that we can have the expected mean as  $E(Y_{it} | \dots) = c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_{it} \beta_x + \tau D_{it})$ .

Broadly speaking, my paper consists of four types of cases based on the two categories: i) We can either use the block-diagonal IV matrices following the approach of Arellano and Bond (1991), or we can rely on the matrix collapsing technique from Roodman (2009a,b) to have triangular IV matrices; ii) We can assume that treatment assignment is either affected by the feedback from the past outcome (sequential exogeneity of treatment assignment), or we can assume that there is no feedback from the past outcome as the conventional two-way fixed-effects literature implicitly do

(strict exogeneity of treatment assignment). When the paper states that the treatment assignment is sequentially exogenous, I am assuming the following mechanism of feedback,

$$D_{it} = 1[\gamma_y Y_{it-1} \leq \kappa + c_i + N(0, 1)]$$

where the constant  $\kappa$  could be working as a cutoff for the treatment qualification. When the paper says the treatment assignment is strictly exogenous, it means that there is no feedback of lagged dependent variable, and I made the true DGP of assignment as simple as possible by putting  $\lambda_1 = \lambda_x = \lambda_y = 0$ . I will still allow, however, that the unobserved heterogeneity to be correlated with the treatment assignment by putting  $\lambda_a = 1$  as before and defining the treatment assignment to have different entrance timing

$$D_{it} = \sum_{j=2}^T Q_i \times f_j$$

while  $f_j$  is the cohort indicator function which is one if the unit is treated at time  $j$ , and  $Q_i = 1[a_i \leq \kappa_q]$  is the ever treated indicator which is correlated with the unobserved heterogeneity. Note that the time index  $j$  starts from 2, so that period 1 would be remained controlled throughout the simulation study. For the DGP with the strictly exogenous assignment, I assigned the cohorts using the uniform distribution, so that neither feedback nor heterogeneity could affect the cohort assignment mechanism.

Finally, as mentioned in Section 2.2, I designed the simulation based on the no exit condition ( $D_{is} \geq D_{it}, \forall s \geq t$ ), following the convention of the staggered intervention literature.

### 2.5.2 Results

Table 2.1 shows the Monte Carlo simulation result when the model is correctly assuming that the treatment assignment is sequentially exogenous, being affected by the feedback from the past outcome. As one can see, usual two-way fixed-effects models estimated by fixed effects Poisson method are generally biased, regardless of including lagged outcome (AR1FE) or not (TWFE). When the true parameter was 0.3, TWFE yields the negative estimates on average, while AR1FE overestimates the true parameter. For both types of IVs (IV1 for Arellano and Bond (1991) style,



and IV2 for Roodman (2009a,b) style), GMM methods are generally consistent, and their standard deviations decrease as the sample size goes up.

Table 2.1 Average of estimated coefficients and their standard deviations when the treatment is sequentially exogenous (T=4, 1000 Repetitions)

$\tau = 0.3$	$\ln(Y_{t-1} + Z_{t-1})$ included			
	TWFE	AR1FE	IV1	IV2
N=500	-0.196 (0.269)	0.565 (0.035)	0.316 (0.168)	0.313 (0.201)
N=1000	-0.246 (0.252)	0.562 (0.024)	0.307 (0.117)	0.306 (0.139)
N=2000	-0.276 (0.219)	0.562 (0.017)	0.301 (0.083)	0.302 (0.093)
N=4000	-0.330 (0.232)	0.561 (0.013)	0.302 (0.063)	0.302 (0.072)

*Note.* Standard deviations in the parentheses. TWFE used the fixed effects Poisson with strict exogeneity. AR1FE also used the same estimation strategies but with function of lagged outcome term included. IV1 used the Arellano and Bond (1991) types of instrument matrices, while IV2 used the suggestion of Roodman (2009a,b).

When the treatment assignment is indeed strictly exogenous, we can include the treatment effects indicator into the last column of IVs along with strictly exogenous covariates, and estimate the coefficient. By putting  $D_{it}$  into  $X_i$ , we can only consider the sequential exogeneity of lagged outcome and its function  $Z_{it-1}$ . Note that during the simulation, we still allowed the unobserved heterogeneity,  $c_i$ , could be arbitrarily correlated with the treatment assignment indicator. Table 2.2 shows the result of Monte Carlo simulation under the strict exogeneity assumption on the treatment indicator. As one can see, both of the usual two-way fixed-effects models are still generally biased again, even for the case when the treatment assignment is strictly exogenous. The result comes

from the design of true DGP in this simulation, which included the lagged outcome. TWFE explicitly ignores the lagged outcome term, which creates the bias. AR1FE tries to include the lagged outcome, but the fixed effects Poisson method relies on the strict exogeneity assumption, which does not allow researchers to include the lagged outcome. Since conditioning on the lagged outcome violates the strict exogeneity assumption of the fixed effects Poisson method, AR1FE could not completely remove the bias.

Table 2.2 Average of estimated coefficients and their standard deviations when the treatment is strictly exogenous (T=4, 1000 Repetitions)

		$\ln(Y_{t-1} + Z_{t-1})$ included		
$\tau = 0.3$	TWFE	AR1FE	IV1	IV2
N=500	-0.086 (0.214)	0.271 (0.036)	0.315 (0.163)	-0.231 (17.179)
N=1000	-0.048 (0.193)	0.277 (0.026)	0.300 (0.093)	11.87323 (365.938)
N=2000	-0.041 (0.179)	0.279 (0.018)	0.299 (0.065)	0.304 (0.125)
N=4000	-0.047 (0.189)	0.282 (0.013)	0.302 (0.049)	0.302 (0.094)

*Note.* Standard deviations in the parentheses. TWFE used the fixed effects Poisson with strict exogeneity. AR1FE also used the same estimation strategies but with function of lagged outcome term included. IV1 used the Arellano and Bond (1991) types of instrument matrices, while IV2 used the suggestion of Roodman (2009a,b).

Both of GMMs are converging to the true parameter as we expected. For the triangular specification of IV2 (Roodman 2009a,b), however, estimates works poorly in the small sample. In general, simulation studies suggest that the block-diagonal suggestion from Arellano and Bond (1991) works well while collapsing triangular IV reports a less preferable result compared to the

triangular IV. Appendix C.1 presents the detailed simulations including the un-optimal GMM who use the identity matrices for the initial weights. One can find a tendency that the block-diagonal IVs are generally working well and efficient compared to the triangular IVs for exponential mean specifications like in this paper.

## **2.6 Application: Effect of Prescription Drug Monitoring Program on Child Maltreatment**

This section revisits Evans et al. (2022b) to show that the persistence of outcome variable and feedback structure of assignment could yield completely different policy implication. The original work of Evans et al. (2022b) evaluates the impact of implementation of must-access Prescription Drug Monitoring Programs (PDMP, hereafter) on the child maltreatment, and concludes that the monitoring program had either no effects or even a positive impact on the child maltreatment. They emphasize the importance of policy designs that consider the substitution pattern seriously, as such unintended consequence could be coming from the strong dependence on the controlled substances. For instance, they argue that inconsiderate policy design can aggravate physical and psychological distress. For the details of their analysis and opioid crisis in the United States, see Evans et al. (2022b) and the references therein.

Before we proceed further, notice that the PDMP makes a good example of staggered intervention. As stated in the online appendix of Evans et al. (2022b), drug monitoring programs were initially enacted and implemented from the early 2000s in several states. During the 2000s, however, the policy itself was generally optional, and did not have any specific instructions/actions enforced to the supply side of prescriptions. In 2007, Nevada adopted the first substance monitoring program with mandatory measures. This must-access prescription drug monitoring program now includes: i) the report duties of all prescriptions; ii) for the prescriptions of controlled substance, providers need to consult PDMP with patient's history beforehand. After Nevada, Oklahoma (2010) and Ohio (2011) adopted their must-access provisions, and numerous states followed, expecting to reduce the abuse of controlled substance. Table 2.3 reproduces the Table B1 of the online appendix of Evans et al. (2022b) and it summarizes the staggered nature of the PDMP assignment.

Table 2.3 Entrance Years of Must-access PDMP

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
		NV			OK	OH	DE	MA	IN	CT	NH
							KY	NY	LA	NJ	RI
							NM	TN		VA	
							WV	VT			

*Note.* Must-access PDMP implementation years were taken from the online appendix of Evans et al. (2022b). Details of treatment entrance can be found in their online appendix and the references therein.

To measure the effects of must-access PDMP implementation on child maltreatment, Evans et al. (2022b) use the following Difference-in-differences specification<sup>11</sup> with the linear model:

$$Y_{ist} = c_i + \beta_t + \tau \cdot \text{PDMP}_{st} + X'_{ist}\beta_x + \varepsilon_{ist}$$

where  $i$  is the county-level index,  $s$  is the state index,  $t$  is the year index,  $c_i$  is county-wise fixed effects,  $\beta_t$  captures the time fixed effects,  $\tau$  is the parameter of interest, and  $X_{ist}$  is the vector of strictly exogenous covariates which is not affected by treatment status. They use two outcome variables to estimate the impact of PDMP to child maltreatment. The first outcome variable is the count of allegation of physical abuse and neglect per 1,000 children within the county  $i$  for the year  $t$ . The second outcome variable is the count substantiated cases of physical abuse and neglect per 1,000 children for each county  $i$  and year  $t$ .

Using the above level-level equation, Evans et al. (2022b) interpret the meaning of coefficients in the following order: i) Suppose the estimate of  $\tau$  is 2.984, which is approximately 3. This means that the PDMP implementation raise the count of alleged case of physical abuse and neglect by three per 1,000 children per year; ii) Evans et al. (2022b) then collects the sample size used during the estimation. For example, the sample mean of outcome variable used in the first estimation was 38 allegations per 1,000 children; iii) since  $2.984/38 \approx 0.0785$  is approximately 0.08, they

<sup>11</sup>In this paper, I focus on their two-way-fixed-effects (TWFE) specification. They also examined the event study model along with the simple TWFE to estimate the effects of must-access PDMPs on child maltreatment. For the details of their specifications and results, see Evans et al. (2022b).

claim that the size of the PDMP effect is about 8 percent. Interpreting the result in the percentage change is natural in this study as the outcome variable is already defined as the size of cases per 1,000 children per year. For example, it might not be intuitive to determine whether the coefficient is big or small enough, as the estimated number is always interpreted as per thousand children. I had three concerns, however, regarding their linear model, estimation strategies, and the model specifications.

First, I want to include the lagged outcome into the model as the behavior of child maltreatment could have persistence issue. For example, if a county has 40 allegations of child maltreatment last year, one would expect the county might have a similar number of allegations this year. If the abuse of controlled substance could be a source of child maltreatment, the addictive nature of controlled substance might result in the persistent pattern of child maltreatment. However, their model does not contain any lagged outcomes.

Second, I also want to estimate the model using the log-level equation, instead of constructing ex-post ratios to re-calculate the percentage change. Considering the advantage of exponential specification (Santos Silva and Tenreyro 2006, 2011), my framework could provide a reasonable estimates to study the impact of PDMP on the child maltreatment with percentage change interpretation. Note that their data set is an ideal case for the exponential model as their outcome variable is a count response. Specifically, I use the following model:

$$Y_{ist} = c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{ist-1}) + \tau \text{PDMP}_{st} + X_{ist}\beta_x) \cdot \varepsilon_{ist}$$

where the econometricians can impose  $\beta_y = 0$  depending on whether they want to include the lagged outcome into the covariates set or not. In this paper, I estimate the models both with and without lagged outcome.

Finally, I want to take the feedback environment into our consideration during the analysis. If a policy has a heterogeneous entrance timing (staggered intervention), it might have been determined by latent structure. For example, Figure 2.1 presents the heterogeneity of treatment entrance of must-access PDMP. As we can see, early takers are concentrated in the Midwest area and regions between California and Texas.

Figure 2.1 Implementation Year of must-access PDMP

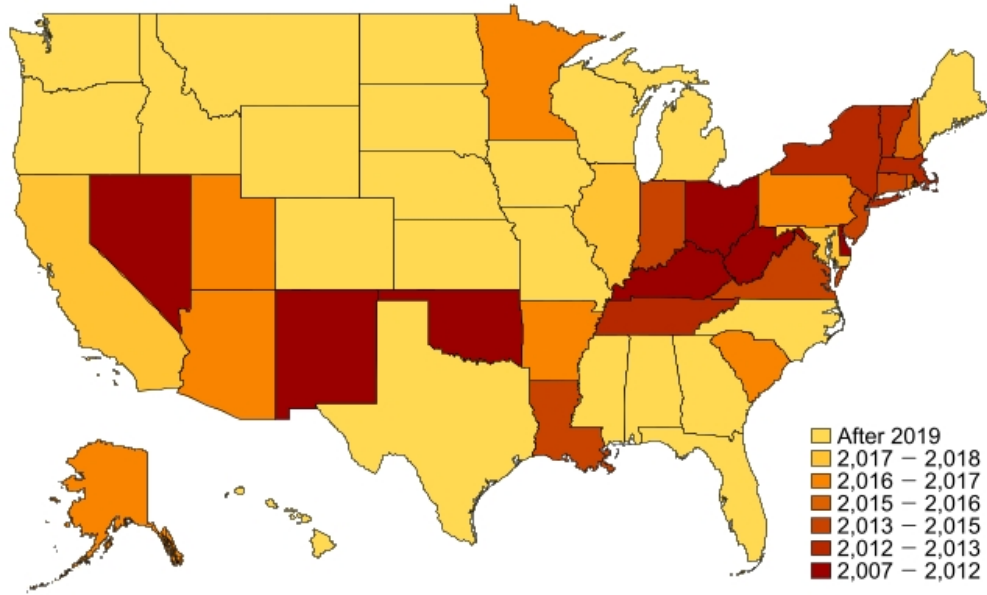
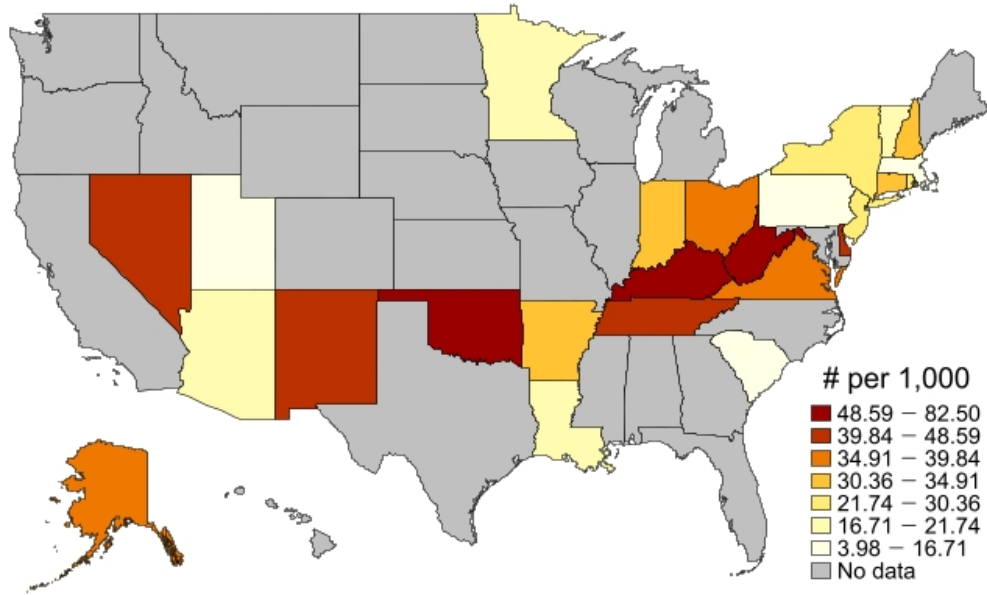


Figure 2.2 shows the regional difference of allegation count one year before the adoption of must-access PDMP. Since many states adopted must-access PDMP after 2019, those states are having no data as the data set only contains observations up to the year 2016<sup>12</sup>. However, we can find the similar pattern between Figure 2.1 (entrance timing) and Figure 2.2 (outcome right before the entrance). For example, early takers in Nevada and Oklahoma reports relatively high levels of allegation count. States in Midwest also reports relatively high level of allegations, corresponding with the pattern shown in Figure 2.1. Even though the graphical illustration does not prove that there is a structural relationship between the lagged outcome and the treatment timing decision, overlapping geographical pattern suggests that it would be worthwhile for researchers to consider about the feedback environment as a potential source of staggered intervention. It is also useful to recall that my framework does not require specific structural relationship as it allows an arbitrary correlation between the lagged outcome and the treatment assignment. To see the impact of introducing feedback environment, I estimated the exponential models with and without feedback assumption.

<sup>12</sup>One can also use the initial outcome level at year 2006 for the baseline instead. In this case, we can still conclude that early takers between California and Texas, and the Midwest states are showing a similar pattern. For the details, see the appendix C.3.

Figure 2.2 Allegations of Child Maltreatment, 1 Year Before PDMP



### 2.6.1 Data

For the empirical illustration in this section, I used the data set provided by Evans et al. (2022b), whose replication package (Evans et al. 2022a) is publicly available in OPEN-ICPSR website. For my GMM approach, I need a specific environments in the panel data set. First, I lose the first period observation (2006 in the data set) as I use them for the equation of the second period (2007) when I include the lagged outcome ( $y_{2006}$ ) into the model. Second, I lose the last period observation (2016 in the data set) to construct the moment condition which use the future period outcome for the ratio. Therefore, my framework will lose at least two period, initial one and the last one. Therefore, I could not use Nevada sample within my application as the data set has no observation in year 2005<sup>13</sup>. To make a fair comparison with TWFE specification with my dynamic models, I cut the sample before 2007, and used the data starting from 2008. This allowed me to use year 2009 as the controlled period while losing samples in 2008, in exchange for including the lagged outcome<sup>14</sup>.

<sup>13</sup>The data starts from the year 2006. If I include the Nevada units, year 2007 is the first treatment period while year 2006 become the controlled period. Since the data has no 2005 observation, I cannot include the lagged outcome in the controlled period. If I decided to lose 2006 observations in exchange for including lagged outcome, my framework does not have controlled period as Nevada starts the first PDMP from 2007. In this case, I cannot make a fair comparison between my framework and TWFE/DiD specification as they have at least one controlled period.

<sup>14</sup>Although it is not reported in the manuscript, regressions with/without Nevada sample (3 observations) does not create a significant difference, which is mentioned in Evans et al. (2022b) footnote 35.

### 2.6.2 Estimation Results

Table 2.4 shows the estimation result for the alleged child maltreatment per 1,000 children. To make a fair comparison between the percentage change interpretation, I calculated the approximate percentage change following the original work of Evans et al. (2022b), where they divided the coefficients of linear models by sample average of dependent used during the estimation. The ex-post percentage interpretation is also reported in the Panel A. As one can see, the usual TWFE strategy with the linear model has a significantly different implication when we compared it to those with exponential specification reported in Panel B.

First, it is noticeable that all of coefficients estimated by GMM in Panel B reports negative estimates. When we allow both of persistence and the feedback environment, GMM1 with un-optimal identity weight reports that PDMP decreases the size of child maltreatment allegation by 7.3% with 1% significance level. GMM2 with optimal weighting matrix reports that PDMP still decreases the amount of child maltreatment allegations around 2.9% at the 10% significance level. When we dropped the persistence, GMM2 estimates reports 12% reduce in child maltreatment with 5% significance level. Panel B suggests a weak but existing evidence that PDMP might have reduced the amount of child maltreatment allegations. It is a noticeable result as the original paper of Evans et al. (2022b) was reporting a positive but insignificant effects.

If the practitioners changed the model specification from the linear one to the exponential one, they would still obtain a different implication. Fixed effects Poisson method, which does not assume persistence nor feedback structure, reports that the implementation of PDMP raised the allegations of child maltreatment by 7.4%. The magnitude of the effect is similar to the result from TWFE (approximately 10.6%), while the exponential specification (FePoi) provides statistically more significant result ( $F = 13.69$ ) compared to the original linear model from Evans et al. (2022b).



Table 2.4 Effects of PDMP for Child Maltreatment Allegations

Panel A.	Coefficients			% Interpretation			
Linear Models	TWFE	LyStD	LySqD	TWFE	LyStD	LySqD	
Alleged Cases							
PDMP	4.150	-0.031	0.177	0.106	-0.000	0.005	
	(2.769)	(0.289)	(0.219)				
Allow Feedback	NO	NO	YES	NO	NO	YES	
Lagged Y controlled	NO	YES	YES	NO	YES	YES	
Panel B.							
Exponential Models	FePoi	GMM1	GMM1	GMM1	GMM2	GMM2	GMM2
Alleged Cases							
PDMP	0.074***	-0.101*	-0.096	-0.073***	-0.120**	-0.032	-0.029*
	(0.020)	(0.060)	(0.023)	(0.020)	(0.054)	(0.018)	(0.017)
Allow Feedback	NO	YES	NO	YES	YES	NO	YES
Lagged Y controlled	NO	NO	YES	YES	NO	YES	YES

*Note.* Standard deviations in the parentheses. Significance levels are \*\*\*(1%), \*\*(5%), and \*(10%) respectively. LyStD and LySqD were estimated using the GMM approach following the Arellano and Bond (1991) instrument while reporting only the second stage estimates. FePoi is the analogy of TWFE into the exponential mean specification. GMM1 is the first stage GMM estimates while GMM2 stands for the second stage GMM estimates. % Interpretation in the Panel B were calculated by dividing the average size of outcome used during the estimation, following the approach in Evans et al. (2022b). All GMM results including LyStD and LySqD are based on the IV matrix suggested by Arellano and Bond (1991).

If practitioners want to keep the linear model, but also want to include the lagged outcome into their model, they can use the GMM framework of dynamic panel linear model literature mentioned in Section 2.1. LyStD and LySqD in Panel A report the results estimated by Arellano and Bond (1991) approach. Table 2.4 still suggests that choice between the linear model and the exponential model can create a remarkable difference during the policy evaluation. For example, linear model considering the feedback and the persistence (LySqD) argues that there the PDMP has a relatively weak effects around 0.5% on the child maltreatment. When we consult the exponential model, it

has an opposite policy implication: PDMP reduced the size of child maltreatment by 2.9%, or even a 7.3%.

Table 2.5 Effects of PDMP for Child Maltreatment Substantiations

Panel A	Coefficients			% Interpretation			
Linear Models	TWFE	LyStD	LySqD	TWFE	LyStD	LySqD	
Substantiated Cases							
PDMP	1.006	0.254**	0.294***	0.117	0.030	0.034	
	(0.661)	(0.127)	(0.078)				
Allow Feedback	NO	NO	YES	NO	NO	YES	
Lagged Y controlled	NO	YES	YES	NO	YES	YES	
Panel B.							
Exponential Models	FePoi	GMM1	GMM1	GMM1	GMM2	GMM2	GMM2
Substantiated Cases							
PDMP	0.097***	0.024	0.059	0.050*	-0.135	0.028	0.018
	(0.025)	(0.094)	(0.033)	(0.029)	(0.088)	(0.032)	(0.029)
Allow Feedback	NO	YES	NO	YES	YES	NO	YES
Lagged Y controlled	NO	NO	YES	YES	NO	YES	YES

*Note.* Standard deviations in the parentheses. Significance levels are \*\*\*(1%), \*\*(5%), and \*(10%) respectively. LyStD and LySqD were estimated using the GMM approach following the Arellano and Bond (1991) instrument while reporting only the second stage estimates. FePoi is the analogy of TWFE into the exponential mean specification. GMM1 is the first stage GMM estimates while GMM2 stands for the second stage GMM estimates. % Interpretation in the Panel B were calculated by dividing the average size of outcome used during the estimation, following the approach in Evans et al. (2022b). All GMM results including LyStD and LySqD are based on the IV matrix suggested by Arellano and Bond (1991).

Table 2.5 presents the estimation result for the substantiated cases of physical child abuse and neglect, instead of alleged cases. Similar to the previous Table 2.4, the linear model (TWFE) and the exponential model (FePoi) share similar implication (around 10% increase of substantiations). By consulting the exponential model over linear model, however, researchers can find additional significance at 1% level ( $F = 15.05$ ). Linear models with persistence (LyStD and LySqD) also

reports that there is a significant increase in the substantiation (around 3%) due to the implementation of PDMP. If the researchers relied on conventional methods, they would end up with the conclusion that PDMP significantly raise the size of child maltreatment substantiations, contrary to the expectation of the policy makers. If they consulted my GMM framework, however, they would have different conclusion: From Table 2.5, there is no evidence that the PDMP raised the substantiations of child maltreatment.

The discrepancy between the estimation strategies come from several source. First, assuming persistence (including lagged outcome) can change the coefficient. Since we conditioned on the lagged outcome, variations of outcome variable explained by the persistence would be fixed, resulting in a different estimates for the treatment effects. Second, assuming feedback environment can change the coefficient. If the staggered intervention is indeed affected by the lagged outcome, assuming feedback environment could solve the potential source of endogeneity of treatment assignment, resulting in different estimates.

One can further argue that we are getting the different estimates because we are comparing the level equation to the log equation. To conduct a fair comparison, an economist can also think about the following log-level linear equation, where we put the natural log to the outcome variable to interpret the coefficient as a proportional change.

$$\ln(Y_{ist}) = c_i + \beta_t + \tau \cdot \text{PDMP}_{st} + X'_{ist}\beta_x + \varepsilon_{ist}$$

Table 2.6 compares the log-level linear equations and the exponential forms when we focus on alleged cases. As we can see, model specification can lead to different conclusions. For example, economists can consider persistence and feedback environment seriously and estimate the log-level equation using LySqD specification. Table 6 reports that LySqD estimated the effect of 2% increase in the allegations with significance level at 1% ( $F = 16$ ). Meanwhile, two exponential specification with persistence and feedback reported negative effects (-7.3% and -2.9% respectively). Even though they have different significance levels, log linear models and exponential models presents completely opposite policy implication in this data set, which was already anticipated in Santos Silva and Tenreiro (2006, 2011). Briefly speaking, the log-linear

model has completely different assumption on the error terms compared to the errors in the usual exponential mean specification. This gap comes from the benefit of exponential specification which allows  $Y_{it} = 0$ , while log-linear model does not allow it.

Table 2.6 Effects of PDMP for Child Maltreatment Allegations

Panel A.							
Log-Linear Models	TWFE	LyStD	LySqD				
Alleged Cases							
PDMP	0.055	0.013*	0.020***				
	(0.055)	(0.006)	(0.005)				
Allow Feedback	NO	NO	YES				
Lagged Y controlled	NO	YES	YES				
Panel B.							
Exponential Models	FePoi	GMM1	GMM1	GMM1	GMM2	GMM2	GMM2
Alleged Cases							
PDMP	0.074***	-0.101*	-0.096	-0.073***	-0.120**	-0.032	-0.029*
	(0.020)	(0.060)	(0.023)	(0.020)	(0.054)	(0.018)	(0.017)
Allow Feedback	NO	YES	NO	YES	YES	NO	YES
Lagged Y controlled	NO	NO	YES	YES	NO	YES	YES

*Note.* Standard deviations in the parentheses. Significance levels are \*\*\*(1%), \*\*(5%), and \*(10%) respectively. LyStD and LySqD were estimated using the GMM approach following the Arellano and Bond (1991) instrument while reporting only the second stage estimates. FePoi is the analogy of TWFE into the exponential mean specification. GMM1 is the first stage GMM estimates while GMM2 stands for the second stage GMM estimates. All GMM results including LyStD and LySqD are based on the IV matrix suggested by Arellano and Bond (1991).

When we focus on the substantiated cases (Table 2.7), log-level linear equation still reports the positive effects, while exponential mean specification (GMM1 and GMM2) does not yield significant effects of PDMP on the child maltreatment. General implication stays the same as we had in Table 2.5: it would be an impetuous conclusion if an economist argues that PDMP raised the substantiated cases of child maltreatment.

Table 2.7 Effects of PDMP for Child Maltreatment Substantiations

Panel A.							
Log-Linear Models	TWFE	LyStD	LySqD				
Substantiated Cases							
PDMP	0.073	0.056***	0.075***				
	(0.060)	(0.014)	(0.012)				
Allow Feedback	NO	NO	YES				
Lagged Y controlled	NO	YES	YES				
Panel B.							
Exponential Models	FePoi	GMM1	GMM1	GMM1	GMM2	GMM2	GMM2
Substantiated Cases							
PDMP	0.097***	0.024	0.059	0.050*	-0.135	0.028	0.018
	(0.025)	(0.094)	(0.033)	(0.029)	(0.088)	(0.032)	(0.029)
Allow Feedback	NO	YES	NO	YES	YES	NO	YES
Lagged Y controlled	NO	NO	YES	YES	NO	YES	YES

*Note.* Standard deviations in the parentheses. Significance levels are \*\*\*(1%), \*\*(5%), and \*(10%) respectively. LyStD and LySqD were estimated using the GMM approach following the Arellano and Bond (1991) instrument while reporting only the second stage estimates. FePoi is the analogy of TWFE into the exponential mean specification. GMM1 is the first stage GMM estimates while GMM2 stands for the second stage GMM estimates. All GMM results including LyStD and LySqD are based on the IV matrix suggested by Arellano and Bond (1991).

## 2.7 Conclusion

This paper studies the identification strategy of proportional treatment effects conditional on the lagged outcome, and provide a GMM estimation strategy which allows the treatment effects to be correlated with unobserved heterogeneity and past outcome. Monte Carlo simulation study suggests that the IV matrices following the block-diagonal tradition of Arellano and Bond (1991) generally work well, while triangular specifications from Roodman (2009a,b) work relatively poorly. The empirical application revisiting the causal inference of PDMP on child maltreatment suggests that considering persistence or feedback environment can create a significantly different policy implications depending on which assumptions we apply.

This paper focuses on the homogeneous treatment effects, as I wanted to see the impact of introducing i) persistence of outcome variable or ii) feedback environment. A natural extension for the future research would be allowing conditional treatment effects to be time-variant. This will further require ex-post reconstruction of time constant effects, which will reach far beyond the scope of this paper.

## CHAPTER 3

### IMPUTATION ESTIMATION OF DYNAMIC POTENTIAL OUTCOME MODELS FOR INTERVENTION ANALYSIS

#### 3.1 Introduction

The rich information of panel data helps economists to identify models with the unobserved (time constant) heterogeneity, which is not allowed in the cross-sectional or time-series data. There are at least two approaches to handle this unobserved heterogeneity using the (linear) panel data models. First, econometrician can cancel out the time constant heterogeneity by time differencing, then instrumentize the differenced variable with the exogenous information under the sequential exogeneity assumption (Arellano and Bond 1991). Panel data allows us to do “time-differencing” as we can obtain multiple observations across time for the same unit. This approach is particularly useful when we want to control the lagged outcome as we are instrumentizing them using the exclusion restrictions anyway. The caveat of this approach comes from the fact that differencing will also remove variations in the time-differenced variables and the instruments too. Therefore, the econometrics literature has been focusing on finding more moment conditions to wrestle with potential weak IV problem (e.g., Ahn and Schmidt (1995), Arellano and Bover (1995), Blundell and Bond (1998), and Roodman (2009a,b, 2020) most recently). The literature extensively reviews even for the case with nonlinear models (Arellano and Honoré 2001, Arellano and Bonhomme 2012), but there has been relatively less attention from the literature in the context of causal inference. Except an extensive development in biostatistics literature based on Robins (1986, 1994, 1997) which usually imposes sequential ignorability assumptions for the entire path of covariates and lagged/realized outcomes, there has been some recent developments who tried to i) incorporate sequential exogeneity into the covariate balancing (Viviano and Bradic 2021); ii) integrating autoregressive integrated moving average (ARIMA) model with the Rubin Causal Model (Menchetti et al. 2023); iii) taking design-based perspective to avoid super-population assumptions (Bojinov et al. 2021). None of them, however, tried to integrate sequential exogeneity in the context of causal inference relying on parallel trends (Roth et al. 2023).

Such hesitations coincide with the recent development of econometrics literature regarding parallel trends. Since Angrist and Pischke (2009) described that there is a “bracketing relationship” between the Difference-in-Differences (DiD, hereafter) and the dynamic linear models, literature has been accumulating warning signs when people want to mix parallel trends and dynamic model. Chabé-Ferret (2015, 2017) investigates the merits of DiD and matching on pre-treatment outcomes, and pointed out that mixing DiD and lagged outcome might not work well, recommending to control initial outcome instead. Ding and Li (2019) generalized the result of Angrist and Pischke (2009) into a nonparametric analogue, arguing DiD and dynamic model always have a monotonic relationship. As a result, mixing DiD (parallel trends) and sequential ignorability (regressor of lagged outcome) has been discussed as “an exciting new literature” or “an interesting area for future research” (Roth et al. 2023).

Another literature that was hesitant to adopt dynamic model was regression imputation/adjustment approach. In our best knowledge, it is believed that it was the seminal work of Heckman et al. (1997) who brought attention of econometricians into the method of imputation. It is also notable that the main concern of Heckman et al. (1997) was the estimation of the average treatment effects (ATE) by DiD in the context of matching methods. Recent development of this literature still works in the context of DiD (Gobillon and Magnac 2016, Gardner 2021, Wooldridge 2021, Caetano et al. 2022, Borusyak et al. 2024), and most of them, if not all, are not interested in incorporating the lagged outcome into their conditional covariates set during the DiD.

In this paper, therefore, we conduct a simulation study to mix dynamic model with imputation approach. The ultimate goal of this paper is to estimate the average treatment effects for treated groups based on the regression imputation approach while allowing lagged outcome included in the potential outcome models. We designed our imputation estimator mainly motivated by Borusyak et al. (2024), but details can be traced back to various works in the regression imputation/adjustment literature (Gardner 2021, Wooldridge 2021). We suggest easy steps to construct the estimator, and demonstrated that our imputation estimator works well by Monte Carlo simulation.

We believe our work can contribute two distinct literature that were mentioned above. First,



we showed that sequential exogeneity assumption used in dynamic panel linear models (Arellano and Bond 1991) can be also utilized in the context of causal inference to construct imputation estimator. Next, our work can provide evidence that regression imputation/adjustment estimators can be extended into the dynamic models too.

Our methods are similar in principle to Menchetti and Bojinov (2022) and Menchetti et al. (2023) who predict counterfactual outcomes using classical time series models. One benefit of their approach is that they do not require a control unit to fit the time series model. While beneficial to applications lacking untreated units, their methodology will not allow for the capture of time-varying secular trends as in a classical DiD analysis. Further, we derive our estimator in a fixed- $T$  setting, but show how the main idea can extend to large samples as well.

The rest of the paper is structured as follows. Section 3.2 describes our dynamic potential outcome model, parameter of our interests, and necessary assumption in this paper. Section 3.3 discuss the identification of our treatment effects. Section 3.4 summarize our estimation strategy of imputation estimator in three simple steps, which was inspired by Borusyak et al. (2024). Section 3.5 represents the our monte carlo simulation results, and we close our paper in section 3.6.

### 3.2 Model

We assume a balanced panel of  $N$  individuals over  $T$  time periods. We first consider the case of a common treatment timing where some individuals are treated immediately after period  $T_0 > 1$  and the others remain untreated. Let  $D_{it}$  is an indicator function to be 1 if the unit  $i$  at time  $t$  is treated and 0 otherwise. Then We define  $D_i$  as an ever-treated indicator variable, which is 1 if unit  $i$  is treated through the entire history and 0 otherwise. Since the treatment assignment  $D_{it}$  can either have 1 or 0, we can also write  $D_i$  as a nonlinear function of entire history of treatment assignment. For example, we can write the ever-treated indicator as  $D_i \equiv \max(D_{it})_{t=1}^{t=T}$ .

We adopt a potential outcome notation:  $y_{i,t}(1)$  is the outcome of unit  $i$  at time  $t$  if they were exposed to treatment at  $T_0 + 1$ . The counterfactual untreated outcome is  $y_{i,t}(0)$ . We are primarily interested in dynamic Average Treatment Effects on the Treated (ATT), defined as

$$\tau_t \equiv \mathbb{E} [y_{i,t}(1) - y_{i,t}(0) \mid D_i = 1] \quad (3.1)$$

We similarly define  $\tau_{i,t} = y_{i,t}(1) - y_{i,t}(0)$  as unit  $i$ 's effect of treatment at time  $t$ . We put no restrictions on the objects  $\tau_{i,t}$  other than the necessary regularity conditions for the appropriate sample averages to converge. The parameter  $\tau_t$  is similar to the object of interest in Callaway and Sant'Anna (2021) when there is only a single treated group. These parameters can be further aggregated to identify different effects of the treatment. For example, we could average  $t = T_0 + 1, \dots, T$  for an overall average treatment effect.

We consider the following model for untreated potential outcomes:

$$y_{i,t}(0) = \alpha y_{i,t-1}(0) + c_i + \theta_t + u_{i,t} \quad (3.2)$$

where  $c_i$  is a unit-specific fixed effect,  $\theta_t$  is a constant time effect, and  $u_{i,t}$  is a mean-zero idiosyncratic shock. The fixed effect  $c_i$  captures permanent discrepancies between the treated and never-treated units and the time effect  $\theta_t$  captures secular changes in the macroeconomy<sup>1</sup>. The non-standard component is the lagged untreated potential outcome  $y_{i,t-1}(0)$ . The AR(1) structure captures transitory shocks that can affect treatment decisions, like a worker's temporary layoff before a work training program (Ashenfelter 1978). It is also responsible for the failure of unconditional parallel trends (PT). The usual PT assumption states that variation in the untreated potential outcomes is equal across treatment status. It is formally written as

$$\mathbb{E} [\Delta y_{i,t}(0) \mid D_i = 1] = \mathbb{E} [\Delta y_{i,t}(0) \mid D_i = 0] \quad (3.3)$$

where  $\Delta$  be the first-differencing transformation so that  $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$ . Let  $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,T})'$  and  $\boldsymbol{\tau}_i = (\tau_{i,T_0+1}, \dots, \tau_{i,T})'$ . We make the following assumptions on the data:

**Assumption 3.1. Sampling.** The random components  $(c_i, \mathbf{u}_i, y_{i,0}, \boldsymbol{\tau}_i, D_i)$  are iid with finite fourth moments.

**Assumption 3.2. Dynamic Completeness.**  $\mathbb{E} [y_{i,t}(0) \mid y_{i,t-1}(0), \dots, y_{i,1}(0), y_{i,0}, c_i] = \alpha y_{i,t-1} + \theta_t + c_i$  for all  $t$ .

---

<sup>1</sup>The interactive fixed effects model, or factor model, nests the two-way error model as a special case. We explicitly cover the general factor model later in the section.

**Assumption 3.3. No Anticipation.**  $y_{i,t} = y_{i,t}(0)$  if unit  $i$  is not subject to treatment at time  $t$ .

**Assumption 3.4. Exogeneity.**  $\mathbb{E}[u_{i,t} \mid c_i, y_{i,t-1}, \dots, y_{i,1}, D_i] = 0$

Assumption 3.1 is a standard random sampling assumption as in Callaway and Sant’Anna (2021) and can be relaxed at the expense of notational complexity. Assumption 3.2 implies that the model in equation (3.2) is dynamically complete in the sense that all correlation between units are captured by the lagged outcome and fixed effect. We could extend the dynamic specification to a general autoregressive distributed lag model without changing the fundamental identifying argument, meaning we can allow for additional lags along with lagged independent variables. We would only need additional time periods. We point out that Assumptions 3.1 and 3.2 imply almost no distributional assumptions on the treated potential outcomes  $y_{i,t}(1)$ . Thus we can allow for almost arbitrary treatment effect heterogeneity, so long as the aggregate parameters can be consistently estimated. Assumption 3.3 is also standard in the literature and rules out anticipatory effects. We can test the assumption by redefining the treatment date to before  $T_0 + 1$  and estimate a treatment effect at  $T_0$ .

Finally, we require Assumption 3.4 to estimate the potential outcome model written as in equation (3.2). By conditioning on the ever-treated indicator function  $D_i$ , we are also allowing a general correlation between the treatment indicator  $D_{it}$  and the unobserved heterogeneity  $c_i$  for all time. Therefore, one can say that the lagged outcomes are sequentially exogenous up until each time period  $t$ , while the treatment indicator  $D_{it}$  or ever-treated indicator  $D_i = \max(D_{it})_{t=1}^{t=T}$  is strictly exogenous. For the detail of various versions of exogeneity levels in the panel data literature, see Wooldridge (2010) and the references therein.

### 3.3 Identification

We now describe identification of the dynamic ATTs. We follow a similar strategy to the popular “imputation” approach in the panel data literature (Gardner 2021, Wooldridge 2021, Borusyak et al. 2024, Brown and Butts 2023). We start by writing the model for untreated potential outcomes (3.2):

$$y_{i,t}(0) = \alpha y_{i,t-1}(0) + \mu_i + \theta_t + u_{i,t}$$

In this paper, we use the untreated outcomes to estimate the model's parameters, then subtract the estimated untreated potential outcomes from the post-treatment treated observations to estimate the ATT. Our model for treatment effects is

$$\tau_{i,t} = y_{i,t}(1) - y_{i,t}(0)$$

which places no restrictions on the treated counterfactual and allows arbitrary dynamics in the treatment effect parameters. Since this paper is interested in the impact of dynamic model on the two way fixed effects approach, and suggest imputation approach as an alternative, we further assume that the treatment effect is homogeneous ( $\tau_{i,t} = \tau_t = \tau$ ). By narrowing down our interests to this specific environment, we can make a fair comparison between the imputation and the simple two way fixed effects, as the imputation approach will allow us to estimate various parameters for different timings, while the simple two way fixed effects will give us one parameter as a basis.

The original identifying result depends on Arellano and Bond (1991) estimation of the autoregressive parameter  $\alpha$ . First, note that taking first differences removes the unit effect  $c_i$ . We have

$$\Delta y_{i,t}(0) = \alpha \Delta y_{i,t-1}(0) + \Delta \theta_t + \Delta u_{i,t} \quad (3.4)$$

One issue to consider is the “exogeneity” of the treatment effect parameters. Consider the Arellano and Bond (1991) moments modified to use only control outcomes:

$$\mathbb{E} \left[ (1 - D_i) y_{i,s} (\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \Delta \theta_t) \right] = 0 \quad (3.5)$$

where  $(1 - D_i)$  appears to use only the controlled observations. This moment condition holds under Assumption 3.4 as one can see from the dynamic panel literature (Arellano and Bond 1991). Once we control for past outcomes and unobserveables, the path of the untreated outcomes is independent of treatment and differ in levels only up to the time effects. We assume common treatment timing with no sample switching.

Turning back to the primary moment conditions, we assume  $s < t - 2$ . Then observe that

$$\begin{aligned}\mathbb{E} \left[ (1 - D_i) y_{i,s} (\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \Delta \theta_t) \right] &= \mathbb{E} \left[ (1 - D_i) y_{i,s} \Delta u_{i,t} \right] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ y_{i,s}(0) \Delta u_{i,t} \mid D_i = 0 \right] \right].\end{aligned}$$

Applying iterated expectations to the expectation conditional on  $D_i = 0$  gives zero moment condition. We summarize this conclusion in a lemma:

**Lemma 3.3.1** *Under Assumptions 3.1, 3.2, and 3.4,  $\alpha$  and  $\theta_t$ s are identified by*

$$\mathbb{E} \left[ y_{i,s} (\Delta y_{i,t} - \alpha \Delta y_{i,t-1} - \Delta \theta_t) \mid D_i = 0 \right] = 0 \quad (3.6)$$

where  $s < t - 2$ .

Once we have  $\alpha$  and  $\theta_t$ , we have

$$y_{i,t}(0) - \alpha y_{i,t-1}(0) - \theta_t = c_i + u_{i,t} \quad (3.7)$$

so  $\mathbb{E} [c_i \mid D_i = 0]$ , and  $\mathbb{E} [c_i \mid D_i = 1]$  are identified and estimable ex-postly. In this paper, we use the pre-treatment observations to impute  $\mathbb{E} [c_i \mid D_i = 1]$ , whose approach can be traced back to Gardner (2021), Wooldridge (2021), and Borusyak et al. (2024). For the detailed form of imputed heterogeneity, see the next section and Appendix C.1.

Identification of the ATTs follows an iterative procedure. First, consider  $t = T_0 + 1$ , the first treated period. In this stage, we can impute  $\mathbb{E} [y_{i,T_0+1}(0) \mid D_i = 1]$  to get our ATT estimate:

$$\begin{aligned}\mathbb{E} [y_{i,T_0+1}(0) \mid D_i = 1] &= \mathbb{E} [\alpha y_{i,T_0}(0) + c_i + \theta_{T_0+1} + u_{i,T_0+1} \mid D_i = 1] \\ &= \alpha \mathbb{E} [y_{i,T_0} \mid D_i = 1] + \mathbb{E} [c_i \mid D_i = 1] + \theta_{T_0+1}\end{aligned}$$

where the second equality comes from the construction that  $T_0$  is a controlled period. Therefore  $\tau_{T_0+1}$  is identified by taking the difference between the observed outcome and the imputed untreated potential outcome over the treated sample.

**Theorem 3.3.1** *Under Assumption 1, and given  $(\alpha, \mathbb{E} [c_i \mid D_i = 1], \theta_2, \dots, \theta_T)$ ,  $\tau_{T_0+1}$  is identified by*

$$\mathbb{E} [y_{i,T_0+1} - \alpha y_{i,T_0} - c_i - \theta_{T_0+1} \mid D_i = 1] = 0 \quad \square$$

We can continue inductively to identify the treatment effects past  $T_0 + 1$ . Suppose  $\tau_t = \mathbb{E}[y_{i,t}(1) - y_{i,t}(0) \mid D_i = 1]$  is identified. Then  $\tau_{t+1}$  can be achieved recursively:

$$\begin{aligned}\tau_{t+1} &= \mathbb{E}[y_{i,t+1}(1) - y_{i,t+1}(0) \mid D_i = 1] \\ &= \mathbb{E}[y_{i,t+1} - \alpha y_{i,t}(0) - c_i - \theta_{t+1} \mid D_i = 1] \\ &= \mathbb{E}[y_{i,t+1} - \alpha(y_{i,t}(1) - \tau_t) - c_i - \theta_{t+1} \mid D_i = 1] \\ &= \mathbb{E}[y_{i,t+1} - \alpha y_{i,t} - c_i - \tau_{t+1} \mid D_i = 1] - \alpha \tau_t\end{aligned}$$

The following theorem summarize our identification result.

**Theorem 3.3.2** *Under Assumption 1, and given  $(\tau_t, \alpha, \mathbb{E}[c_i \mid D_i = 1], \theta_2, \dots, \theta_{t+1})$  for  $t > T_0 + 1$ ,  $\tau_{t+1}$  is identified by*

$$\tau_{t+1} = \alpha \tau_t + \mathbb{E}[y_{i,t+1} - \alpha y_{i,t} - c_i - \theta_{t+1} \mid D_i = 1] \quad \square$$

An interesting component of this AR(1) dynamic specification is that the treatment effects also follow an AR(1) dynamic process.

### 3.4 Estimation

In this section, we summarize the estimation procedure. Estimation of average treatment effects on treated groups can be conducted by regression imputation approach in the following three steps.

1. Using controlled observations only, regress the potential outcome equation (3.2) and estimate  $\alpha$  and  $\theta_t$  by Arellano and Bond (1991)
2. Impute  $c_i$  by

$$\hat{c}_i = \frac{1}{T_0} \sum_{t=1}^{T_0} (y_{i,t} - \hat{\alpha} y_{i,t-1} - \hat{\theta}_t)$$

3. Estimate  $\tau_{T_0+1}$  as

$$\hat{\tau}_{T_0+1} = \frac{1}{N_1} \sum_{i=1}^N D_i \cdot (y_{i,T_0+1} - \hat{\alpha} y_{i,T_0} - \hat{\theta}_{T_0+1} - \hat{c}_i) \quad (3.8)$$

Then for  $t > T_0 + 1$ , estimate

$$\hat{\tau}_t = \hat{\alpha} \hat{\tau}_{t-1} + \frac{1}{N_1} \sum_{i=1}^N D_i \cdot (y_{i,t} - \hat{\alpha} y_{i,t-1} - \hat{\theta}_t - \hat{c}_i) \quad (3.9)$$

where  $N_1 = \sum_{i=1}^N D_i$  and  $N_0 = \sum_{i=1}^N (1 - D_i)$  be the number of treated and untreated units respectively.

The general construction of this estimation procedure is motivated by the approach suggested in

Borusyak et al. (2024), where the original work was focusing on the static models without lagged outcomes in their model. In this paper, we extend the linear panel data models from static versions to dynamic version, and check whether regression imputation approach still works.

In this step 1, we exchange the ordinary least square into the generalized method of moments method to estimate the dynamic panel model. We chose Arellano and Bond (1991) for the simplicity of use, but different researcher can also apply various estimation strategies depending on their context. After step 1, we still do not have the unobserved heterogeneity part as the step 1 estimates the model in the form of differenced equation. Again, inspired by regression imputation literature (Gardner 2021, Wooldridge 2021, Borusyak et al. 2024), we imputed  $E[c_i|D_i = 1]$  by using the pre-treatment periods ( $t \leq T_0$ ), whose identification can be found in Appendix C.1. In step 3, we estimate the treatment effects for treated group. The effects for the first treatment period is captured by the gap between the data ( $y_{i,t} = y_{i,t}(1)$ ) and the fitted/imputed counterfactual ( $\widehat{y}_{i,t}(0) = \widehat{\alpha}y_{i,T_0} + \widehat{\theta}_{T_0+1} + \widehat{c}_i$ ). It is interesting to point out that, after the first period, the treatment effect parameters ( $\tau_t$ ) also follows AR(1) process, and their estimation can be achieved accordingly.

If we are willing to assume further that the lagged outcomes are non-stochastic, we might be able to try inference guided by the results in Borusyak et al. (2024). Since it is not reasonable to assume that the past dependent variables are non-stochastic, we narrow our focus to the simulation study in this paper.

### 3.5 Simulations

We present the results of Monte Carlo simulations to compare our imputation approach with the usual TWFE approach under the dynamic potential outcome model. Even though we are assuming a homogeneous treatment effects in this paper, we estimated the treatment effect for the two different treated periods ( $\tau_4$  for the first treatment period and  $\tau_5$  for the final period). For the two way fixed effects, we captured the homogeneous treatment effects by regressing the standard two way fixed effects equation with a large set of dummy variables

$$y_{i,t} = c_i + \theta_t + \tau \cdot D_{it} + u_{i,t}$$

### 3.5.1 Benchmark DGP

In this paper, we designed the last period  $T = 5$  and the last controlled period  $T_0 = 3$ , so that the intervention happens at  $t = 4, 5$ . The untreated potential outcome model follows the equation (3.2)

$$y_{i,t}(0) = \alpha y_{i,t-1}(0) + c_i + \theta_t + u_{i,t}$$

where  $\alpha \in \{0.5, 1.0\}$  is the persistence parameter that captures path dependence in the dynamics of outcome variables,  $c_i$  is the unobserved heterogeneity, and  $\theta_t$  is the time fixed effects. We draw  $c_i$  from the standard normal distribution, and we put time trends are all zero for the simplicity of discussion. We defined our strictly exogenous treatment assignment as  $D_i = c_i + N(0, 1)$  so that the assignment is correlated with the unobserved heterogeneity. For the uncorrelated assignment, we drew the assignment from the standard normal distribution.

### 3.5.2 Results

Table 3.1 shows the simulation results with the treatment assignment correlated with the unobserved heterogeneity under the presence of persistence in the outcome variable dynamics. One can find the pattern that various parameters report the same results in Table 3.1. Under the dynamic potential outcome models, imputation approach has the smaller bias than the two way fixed effects approach. We can also observe the shrinking bias from the imputation approach as the sample size increases, and the standard deviation also decrease as the sample size goes up.



The biases from the two way fixed effects approach, however, does not go away. Under the dynamic potential outcome model, the two way fixed effects estimator converged to a biased target from the small sample size ( $N = 300$ ), and the estimate does not change much even with much bigger size of sample ( $N = 1,000$ ). The two way fixed effects approach also reports the biggest root mean squared error (rMSE), dominated by big size of bias.

Table 3.1 Monte Carlo Simulation under Strictly Exogenous Assignment ( $\alpha = 0.5$ )

Regression Imputation								TWFE			
$\tau_4 = 1$				$\tau_5 = 1$				$\tau = 1$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.039	0.164	0.168	300	0.036	0.233	0.235	300	0.543	0.133	0.560
500	0.027	0.120	0.123	500	0.026	0.178	0.180	500	0.553	0.105	0.563
1000	0.012	0.088	0.089	1000	0.012	0.133	0.133	1000	0.552	0.074	0.557

Regression Imputation								TWFE			
$\tau_4 = 2$				$\tau_5 = 2$				$\tau = 2$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.039	0.164	0.168	300	0.036	0.233	0.235	300	0.543	0.133	0.560
500	0.027	0.120	0.123	500	0.026	0.178	0.180	500	0.553	0.105	0.563
1000	0.012	0.088	0.089	1000	0.012	0.133	0.133	1000	0.552	0.074	0.557

Regression Imputation								TWFE			
$\tau_4 = -1$				$\tau_5 = -1$				$\tau = -1$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.039	0.164	0.168	300	0.036	0.233	0.235	300	0.543	0.133	0.560
500	0.027	0.120	0.123	500	0.026	0.178	0.180	500	0.553	0.105	0.563
1000	0.012	0.088	0.089	1000	0.012	0.133	0.133	1000	0.552	0.074	0.557

*Note.* Repetition size is 1,000 for all simulations.  $N$  is the size of unit observations in the panel. The treatment assignment is correlated with the unobserved heterogeneity.

Table 3.2 presents the result under the unit root process of outcome variable. Regression Imputation method is still retain considerably less bias compared to the simulated estimates from the two way fixed effects method. Imputation method, even under the unit root process, reported shrinking standard deviations along with reducing bias as observation size goes up. If we compare table 3.1 and 3.2, we can see that the biases in the two way fixed effects method magnify as the persistence parameter gets bigger. By increasing the magnitude of persistence from 0.5 to 1.0,

biases of TWFE estimates now have five times bigger magnitude than those from the previous table. The standard deviations of TWFE estimators are also having twice as bigger size in Table 3.2 compared to those in Table 3.1.

Table 3.2 Monte Carlo Simulation under Strictly Exogenous Assignment ( $\alpha = 0.5$ )

Regression Imputation								TWFE			
$\tau_4 = 1$				$\tau_5 = 1$				$\tau = 1$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.017	0.153	0.154	300	0.029	0.318	0.320	300	2.805	0.284	2.820
500	0.016	0.119	0.120	500	0.031	0.247	0.248	500	2.810	0.222	2.819
1000	0.004	0.086	0.086	1000	0.007	0.176	0.176	1000	2.818	0.162	2.823

Regression Imputation								TWFE			
$\tau_4 = 2$				$\tau_5 = 2$				$\tau = 2$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.017	0.153	0.154	300	0.029	0.318	0.320	300	2.805	0.284	2.820
500	0.016	0.119	0.120	500	0.031	0.247	0.248	500	2.810	0.222	2.819
1000	0.004	0.086	0.086	1000	0.007	0.176	0.176	1000	2.818	0.162	2.823

Regression Imputation								TWFE			
$\tau_4 = -1$				$\tau_5 = -1$				$\tau = -1$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.017	0.153	0.154	300	0.029	0.318	0.320	300	2.805	0.284	2.820
500	0.016	0.119	0.120	500	0.031	0.247	0.248	500	2.810	0.222	2.819
1000	0.004	0.086	0.086	1000	0.007	0.176	0.176	1000	2.818	0.162	2.823

*Note.* Repetition size is 1,000 for all simulations.  $N$  is the size of unit observations in the panel. The treatment assignment is correlated with the unobserved heterogeneity.

Sometimes, economists want to estimate the causal effect of natural experiment or lab experiment, where treatment assignment is completely random. Table 3.3 shows the simulation result when treatment assignment is not even correlated with the unobserved heterogeneity. As mentioned in the section 3.5.1, we have independently drawn the assignment as a standard normal so that it is not correlated with the unobserved heterogeneity. The pattern described in the previous tables continues to hold in Table 3.3. Imputation estimators retained shrinking bias and standard deviations with respect to increasing sample size, and working well compared to the two way fixed effects method in root mean squared error.

Table 3.3 Monte Carlo Simulation under Strictly Exogenous Assignment ( $\alpha = 0.5$ )

Regression Imputation								TWFE			
$\tau_4 = 1$				$\tau_5 = 1$				$\tau = 1$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.003	0.102	0.102	300	0.005	0.126	0.126	300	0.001	0.134	0.134
500	-0.002	0.080	0.080	500	-0.003	0.100	0.100	500	-0.001	0.106	0.106
1000	-0.001	0.054	0.054	1000	-0.001	0.072	0.072	1000	0.001	0.077	0.077

Regression Imputation								TWFE			
$\tau_4 = 2$				$\tau_5 = 2$				$\tau = 2$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.003	0.102	0.102	300	0.005	0.126	0.126	300	0.001	0.134	0.134
500	-0.002	0.080	0.080	500	-0.003	0.100	0.100	500	-0.001	0.106	0.106
1000	-0.001	0.054	0.054	1000	-0.001	0.072	0.072	1000	0.001	0.077	0.077

Regression Imputation								TWFE			
$\tau_4 = -1$				$\tau_5 = -1$				$\tau = -1$			
N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE	N	Bias	S.D.	rMSE
300	0.003	0.102	0.102	300	0.005	0.126	0.126	300	0.001	0.134	0.134
500	-0.002	0.080	0.080	500	-0.003	0.100	0.100	500	-0.001	0.106	0.106
1000	-0.001	0.054	0.054	1000	-0.001	0.072	0.072	1000	0.001	0.077	0.077

*Note.* Repetition size is 1,000 for all simulations.  $N$  is the size of unit observations in the panel. The treatment assignment is correlated with the unobserved heterogeneity.

Unlike the previous tables above, The two way fixed effects methods are working competitive in Table 3.3. Recall that the lagged outcome variable is still in the dynamic potential outcome models in Table 3.3. In case where treatment assignment is uncorrelated with unobserved heterogeneity, the two way fixed effects method has the smallest bias when the sample size is small ( $N = 300$ ) and the root mean squared error also decreases as the sample size grows. Combining Table 3.1 and Table 3.3, one can also mention that the two way fixed effects estimator could bear bias when we doubt zero correlation between the assignment and the unit fixed effects in the data, while the data also could have path dependence or persistence in the outcome variable sequence.

### 3.6 Conclusions

In this paper, we expanded the regression imputation method to study the causal inference under potential outcome models with lagged outcome. We designed our imputation estimator inspired by Arellano and Bond (1991) and Borusyak et al. (2024), and the Monte Carlo study suggests that the imputation estimator might work even for the potential outcome model with dynamic form. The usual two way fixed effects estimator, however, suffered severe bias from the existence of path dependence, and such bias could be amplified when there is a unit root process in the dependent variable sequence. The simulation study also suggests that ignoring persistence might require heavy restriction on the treatment assignment (selection) mechanism.

The result from this paper can be extended to the various directions. First, we want to allow staggered intervention beyond common intervention. In this paper, we focused on the common intervention as we were interested in comparing imputation approach with the two way fixed effects approach under the dynamic setting. Furthermore, combined with staggered intervention, we can also allow heterogeneous effects for different timings.

## BIBLIOGRAPHY

- Abbring, Jaap H, and James J Heckman.** 2007. "Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation." *Handbook of econometrics* 6 5145–5303.
- Ahn, Seung C, and Peter Schmidt.** 1995. "Efficient estimation of models for dynamic panel data." *Journal of Econometrics* 68 (1): 5–27.
- Angrist, Joshua D, and Jörn-Steffen Pischke.** 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Arellano, Manuel, and Stephen Bond.** 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The Review of Economic Studies* 58 (2): 277–297.
- Arellano, Manuel, and Stéphane Bonhomme.** 2012. "Identifying distributional characteristics in random coefficients panel data models." *The Review of Economic Studies* 79 (3): 987–1020.
- Arellano, Manuel, and Olympia Bover.** 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of econometrics* 68 (1): 29–51.
- Arellano, Manuel, and Bo Honoré.** 2001. "Panel data models: some recent developments." In *Handbook of econometrics*, Volume 5. 3229–3296, Elsevier.
- Ashenfelter, Orley.** 1978. "Estimating the effect of training programs on earnings." *The Review of Economics and Statistics* 47–57.
- Balestra, Simone, Helge Liebert, Nicole Maestas, and Tisamarie B Sherry.** 2021. "Behavioral Responses to Supply-Side Drug Policy During the Opioid Epidemic." Working Paper 29596, National Bureau of Economic Research.
- Biewen, Martin.** 2009. "Measuring state dependence in individual poverty histories when there is feedback to employment status and household composition." *Journal of Applied Econometrics* 24 (7): 1095–1116.
- Blundell, R, R Griffith, and F Windmeijer.** 1995. "Dynamics and correlated responses in longitudinal count data models." In *Statistical Modelling: Proceedings of the 10th International Workshop on Statistical Modelling Innsbruck, Austria, 10–14 July, 1995*, 35–42, Springer.
- Blundell, Richard, and Stephen Bond.** 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of econometrics* 87 (1): 115–143.
- Blundell, Richard, Rachel Griffith, and Frank Windmeijer.** 2002. "Individual effects and dynamics in count data models." *Journal of econometrics* 108 (1): 113–131.

- Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard.** 2021. “Panel experiments and dynamic causal effects: A finite population perspective.” *Quantitative Economics* 12 (4): 1171–1196.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess.** 2024. “Revisiting event-study designs: robust and efficient estimation.” *Review of Economic Studies* rdae007.
- Brown, Nicholas, and Kyle Butts.** 2023. “Dynamic Treatment Effect Estimation with Interactive Fixed Effects and Short Panels.” Technical report.
- Caetano, Carolina, Brantly Callaway, Stroud Payne, and Hugo Sant’Anna Rodrigues.** 2022. “Difference in differences with time-varying covariates.” *arXiv preprint arXiv:2202.02903*.
- Callaway, Brantly, and Pedro HC Sant’Anna.** 2021. “Difference-in-differences with multiple time periods.” *Journal of econometrics* 225 (2): 200–230.
- Cameron, A Colin, and Pravin Trivedi.** 2005. *Microeconometrics: methods and applications*. Cambridge University Press.
- Cameron, A Colin, and Pravin K Trivedi.** 2013. *Regression analysis of count data*. Volume 53. Cambridge university press.
- Chabé-Ferret, Sylvain.** 2015. “Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes.” *Journal of Econometrics* 185 (1): 110–123.
- Chabé-Ferret, Sylvain.** 2017. “Should we combine difference in differences with conditioning on pre-treatment outcomes?”
- Chamberlain, Gary.** 1992. “Comment: Sequential moment restrictions in panel data.” *Journal of Business & Economic Statistics* 10 (1): 20–26.
- Ciani, Emanuele, and Paul Fisher.** 2018. “Dif-in-dif estimators of multiplicative treatment effects.” *Journal of Econometric Methods* 8 (1): 20160011.
- Collinson, Robert, John Eric Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie Van Dijk.** 2024. “Eviction and poverty in American cities.” *The Quarterly Journal of Economics* 139 (1): 57–120.
- Crépon, Bruno, and Emmanuel Duguet.** 1997. “Estimating the innovation function from patent numbers: GMM on count panel data.” *Journal of Applied Econometrics* 12 (3): 243–263.
- Davis, Philip J, and Philip Rabinowitz.** 2007. *Methods of numerical integration*. Courier Corporation.
- De Chaisemartin, Clément, and Xavier d’Haultfoeuille.** 2020. “Two-way fixed effects estimators with heterogeneous treatment effects.” *American Economic Review* 110 (9): 2964–2996.

- Ding, Peng, and Fan Li.** 2019. “A bracketing relationship between difference-in-differences and lagged-dependent-variable adjustment.” *Political Analysis* 27 (4): 605–615.
- Dube, Arindrajit, Daniele Girardi, Oscar Jorda, and Alan M Taylor.** 2023. “A local projections approach to difference-in-differences event studies.” Technical report, National Bureau of Economic Research.
- Evans, Mary F, Matthew C Harris, and Lawrence M Kessler.** 2022a. “Data and Code for: The Hazards of Unwinding the Prescription Opioid Epidemic: Implications for Child Maltreatment.” Nashville, TN: American Economic Association [publisher]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-06-22, <https://doi.org/10.3886/E146542V1>.
- Evans, Mary F, Matthew C Harris, and Lawrence M Kessler.** 2022b. “The hazards of unwinding the prescription opioid epidemic: Implications for child maltreatment.” *American Economic Journal: Economic Policy* 14 (4): 192–231.
- Gardner, John.** 2021. “Two-Stage Difference-in-Differences.” Technical report.
- Ghanem, Dalia, Pedro HC Sant’Anna, and Kaspar Wüthrich.** 2022. “Selection and parallel trends.” *arXiv preprint arXiv:2203.09001*.
- Gobillon, Laurent, and Thierry Magnac.** 2016. “Regional policy evaluation: Interactive fixed effects and synthetic controls.” *Review of Economics and Statistics* 98 (3): 535–551.
- Goodman-Bacon, Andrew.** 2021. “Difference-in-differences with variation in treatment timing.” *Journal of Econometrics* 225 (2): 254–277.
- Gourieroux, C, A Monfort, and A Trognon.** 1984a. “Pseudo maximum likelihood methods: Theory.” *Econometrica* 52 (3): 681–700.
- Gourieroux, C, A Monfort, and A Trognon.** 1984b. “Pseudo maximum likelihood methods: Applications to Poisson models.” *Econometrica* 52 (3): 701–720.
- Guceri, Irem, and Li Liu.** 2019. “Effectiveness of fiscal incentives for RD: quasi-experimental evidence.” *American Economic Journal: Economic Policy* 11 (1): 266–291.
- Hansen, Lars Peter.** 1982. “Large sample properties of generalized method of moments estimators.” *Econometrica: Journal of the econometric society* 1029–1054.
- Hausman, J A, B H Hall, and Z Griliches.** 1984. “Econometric models for count data with an application to the patents-RD relationship.” *Econometrica* 52 (4): 909–938.
- Heckman, James J, Hidehiko Ichimura, and Petra E Todd.** 1997. “Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme.” *The review of*

*economic studies* 64 (4): 605–654.

**Horn, Danea, Adam Sacarny, and Annetta Zhou.** 2022. “Technology adoption and market allocation: The case of robotic surgery.” *Journal of Health Economics* 86 102672.

**Imbens, Guido W, and Jeffrey M Wooldridge.** 2009. “Recent developments in the econometrics of program evaluation.” *Journal of economic literature* 47 (1): 5–86.

**Lechner, Michael.** 2015. “Treatment effects and panel data.”

**Lee, Myoung-jae.** 2016. *Matching, regression discontinuity, difference in differences, and beyond.* Oxford University Press.

**Lee, Myoung-Jae, and Satoru Kobayashi.** 2002. “Proportional treatment effects for count response panel data: effects of binary exercise on health care demand.” *Econometric Analysis of Health Data* 117–132.

**Lee, Myoung-jae, and Sanghyeok Lee.** 2021. “Difference in differences and ratio in ratios for limited dependent variables.” *arXiv preprint arXiv:2111.12948*.

**Lindo, Jason M, Peter Siminski, and Issac D Swensen.** 2018. “College party culture and sexual assault.” *American Economic Journal: Applied Economics* 10 (1): 236–265.

**Menchetti, Fiammetta, and Iavor Bojinov.** 2022. “Estimating the effectiveness of permanent price reductions for competing products using multivariate Bayesian structural time series models.” *The Annals of Applied Statistics* 16 (1): 414–435.

**Menchetti, Fiammetta, Fabrizio Cipollini, and Fabrizia Mealli.** 2023. “Combining counterfactual outcomes and ARIMA models for policy evaluation.” *The Econometrics Journal* 26 (1): 1–24.

**Mullainathan, Sendhil, and Jann Spiess.** 2017. “Machine learning: an applied econometric approach.” *Journal of Economic Perspectives* 31 (2): 87–106.

**Ozuna, Teofilo, and Irma A Gomez.** 1995. “Specification and testing of count data recreation demand functions.” *Empirical Economics* 20 543–550.

**Robins, James.** 1986. “A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect.” *Mathematical modelling* 7 (9-12): 1393–1512.

**Robins, James M.** 1989. “The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies.” *Health service research methodology: a focus on AIDS* 113–159.



- Robins, James M.** 1994. “Correcting for non-compliance in randomized trials using structural nested mean models.” *Communications in Statistics-Theory and methods* 23 (8): 2379–2412.
- Robins, James M.** 1997. “Causal inference from complex longitudinal data.” In *Latent variable modeling and applications to causality*, 69–117, Springer.
- Roodman, David.** 2009a. “How to do xtabond2: An introduction to difference and system GMM in Stata.” *The stata journal* 9 (1): 86–136.
- Roodman, David.** 2009b. “A note on the theme of too many instruments.” *Oxford Bulletin of Economics and statistics* 71 (1): 135–158.
- Roodman, David.** 2020. “xtabond2: Stata module to extend xtabond dynamic panel data estimator.”
- Roth, Jonathan, Pedro HC Sant’Anna, Alyssa Bilinski, and John Poe.** 2023. “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature.” *Journal of Econometrics*.
- Santos Silva, J M, and Silvana Tenreyro.** 2006. “The log of gravity.” *The Review of Economics and Statistics* 88 (4): 641–658.
- Santos Silva, J M, and Silvana Tenreyro.** 2011. “Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator.” *Economics Letters* 112 (2): 220–222.
- Shahn, Zach, Oliver Dukes, David Richardson, Eric Tchetgen Tchetgen, and James Robins.** 2022. “Structural nested mean models under parallel trends assumptions.” *arXiv preprint arXiv:2204.10291*.
- Sun, Liyang, and Sarah Abraham.** 2021. “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.” *Journal of Econometrics* 225 (2): 175–199.
- Viviano, Davide, and Jelena Bradic.** 2021. “Dynamic covariate balancing: estimating treatment effects over time.” *arXiv preprint arXiv:2103.01280*.
- Welsch, David M, and David M Zimmer.** 2015. “The relationship between student transfers and district academic performance: Accounting for feedback effects.” *Education Finance and Policy* 10 (3): 399–422.
- Welsch, David M, and David M Zimmer.** 2016. “The dynamic relationship between school size and academic performance: An investigation of elementary schools in Wisconsin.” *Research in Economics* 70 (1): 158–169.
- Windmeijer, Frank.** 2006. “GMM for panel count data models.” *University of Bristol Economics Working Paper* (06-591): .

- Wooldridge, Jeffrey M.** 1997. “Multiplicative panel data models without the strict exogeneity assumption.” *Econometric Theory* 13 (5): 667–678.
- Wooldridge, Jeffrey M.** 1999. “Distribution-free estimation of some nonlinear panel data models.” *Journal of Econometrics* 90 (1): 77–97.
- Wooldridge, Jeffrey M.** 2000. “A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables.” *Economics Letters* 68 (3): 245–250.
- Wooldridge, Jeffrey M.** 2005. “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity.” *Journal of applied econometrics* 20 (1): 39–54.
- Wooldridge, Jeffrey M.** 2010. *Econometric analysis of cross section and panel data*. MIT press.
- Wooldridge, Jeffrey M.** 2021. “Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators.” Technical report.
- Yadlowsky, Steve, Fabio Pellegrini, Federica Lionetto, Stefan Braune, and Lu Tian.** 2021. “Estimation and validation of ratio-based conditional average treatment effects using observational data.” *Journal of the American Statistical Association* 116 (533): 335–352.
- Zimmer, David M.** 2021. “The effect of job displacement on mental health, when mental health feeds back to future job displacement.” *The Quarterly Review of Economics and Finance* 79 360–366.

## APPENDIX A      APPENDIX FOR CHAPTER 1

### A.1. Full result of Table 1.1 and Table 1.2

For the fair comparison, I report the estimates for the parameters of equation (1.4), for each sample size.

Table A.1 Estimation Results under Feedback Environment (N = 500)

Reps=1,000	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.339	0.123	0.361	0.107	0.090	0.139	0.006	0.090	0.090
$\beta_3 - \beta_1$	-0.112	0.136	0.176	0.241	0.107	0.264	0.006	0.104	0.105
$\beta_1$	0.001	0.094	0.094				0.001	0.064	0.068
$\beta_y$	0.488	0.018	0.488	-0.124	0.021	0.126	-0.002	0.020	0.022
$\beta_d$	-0.675	0.113	0.684	-0.209	0.126	0.244	-0.007	0.121	0.122
$\beta_z$	0.575	0.044	0.576	-0.044	0.032	0.054	-0.001	0.032	0.032
Reps=1,000	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.303	0.099	0.319	0.175	0.609	0.633	0.084	0.813	0.818
$\beta_3 - \beta_2$	0.282	0.067	0.290	0.128	0.463	0.481	0.062	0.661	0.664
$\beta_y$	-0.250	0.024	0.251	-0.111	0.386	0.402	-0.055	0.563	0.565
$\beta_d$	-0.601	0.143	0.618	-0.345	1.222	1.270	-0.162	1.634	1.642
$\beta_z$	-0.109	0.034	0.114	-0.048	0.170	0.177	-0.025	0.245	0.246

Table A.2 Estimation Results under Feedback Environment (N = 1,000)

Reps=1,000	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.335	0.091	0.347	0.104	0.062	0.121	0.002	0.063	0.063
$\beta_3 - \beta_1$	-0.115	0.101	0.153	0.238	0.077	0.251	0.002	0.076	0.076
$\beta_1$	0.001	0.069	0.069				0.001	0.046	0.046
$\beta_y$	0.488	0.012	0.488	-0.124	0.015	0.124	0.000	0.016	0.016
$\beta_d$	-0.674	0.079	0.678	-0.210	0.092	0.229	-0.006	0.088	0.089
$\beta_z$	0.573	0.031	0.573	-0.044	0.022	0.050	-0.002	0.023	0.023
Reps=1,000	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.300	0.066	0.308	0.068	0.389	0.395	0.021	0.428	0.429
$\beta_3 - \beta_2$	0.282	0.047	0.286	0.055	0.305	0.310	0.015	0.356	0.357
$\beta_y$	-0.249	0.017	0.250	-0.049	0.258	0.263	-0.014	0.306	0.306
$\beta_d$	-0.602	0.100	0.610	-0.138	0.777	0.789	-0.045	0.857	0.859
$\beta_z$	-0.110	0.023	0.113	-0.023	0.114	0.116	-0.008	0.136	0.137

Table A.3 Estimation Results under Feedback Environment (N = 5,000)

Reps=1,000	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.335	0.040	0.337	0.103	0.028	0.107	0.001	0.028	0.028
$\beta_3 - \beta_1$	-0.116	0.043	0.123	0.238	0.035	0.241	0.000	0.033	0.033
$\beta_1$	0.001	0.030	0.030				0.001	0.020	0.020
$\beta_y$	0.488	0.006	0.488	-0.123	0.007	0.123	0.000	0.007	0.007
$\beta_d$	-0.670	0.036	0.671	-0.206	0.043	0.211	0.000	0.040	0.040
$\beta_z$	0.575	0.014	0.575	-0.043	0.010	0.044	0.000	0.010	0.010
Reps=1,000	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.300	0.031	0.302	0.011	0.164	0.165	0.000	0.171	0.171
$\beta_3 - \beta_2$	0.283	0.021	0.284	0.008	0.130	0.130	-0.002	0.143	0.143
$\beta_y$	-0.248	0.008	0.249	-0.007	0.110	0.110	0.001	0.123	0.123
$\beta_d$	-0.599	0.048	0.601	-0.022	0.325	0.326	0.001	0.339	0.339
$\beta_z$	-0.109	0.010	0.109	-0.003	0.049	0.049	0.000	0.055	0.055

Table A.4 Estimation Results under Feedback Environment (N = 10,000)

Reps=1,000	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.334	0.028	0.335	0.103	0.020	0.105	0.000	0.020	0.020
$\beta_3 - \beta_1$	-0.115	0.031	0.120	0.239	0.025	0.240	0.000	0.024	0.024
$\beta_1$	0.001	0.021	0.021				0.000	0.014	0.014
$\beta_y$	0.488	0.004	0.488	-0.123	0.005	0.123	0.000	0.005	0.005
$\beta_d$	-0.670	0.026	0.670	-0.206	0.031	0.208	0.000	0.029	0.029
$\beta_z$	0.575	0.010	0.575	-0.043	0.007	0.044	0.000	0.007	0.007
Reps=1,000	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$\beta_2 - \beta_1$	0.300	0.022	0.301	0.004	0.113	0.113	-0.002	0.119	0.119
$\beta_3 - \beta_2$	0.284	0.015	0.284	0.004	0.090	0.090	-0.002	0.102	0.102
$\beta_y$	-0.249	0.005	0.249	-0.003	0.076	0.076	0.002	0.088	0.088
$\beta_d$	-0.599	0.034	0.600	-0.007	0.223	0.223	0.005	0.235	0.235
$\beta_z$	-0.109	0.007	0.109	-0.001	0.034	0.034	0.001	0.039	0.039

## A.2. Blundell and Bond (1998) Approach

One might argue that Arellano and Bond (1991) is not the only one GMM estimation strategies we can use, and there could be efficient alternative other than Arellano and Bond (1991) approach. A direct extension of Arellano and Bond (1991) would be Blundell and Bond (1998) approach. In this section: i) I briefly introduce the logic of Blundell and Bond (1998); ii) show that the approach is not valid for our case when initial value is given exogenously; iii) show the simulation result for the correctly specified cases. For the theoretical details, consult Blundell and Bond (1998).

Recall that Arellano and Bond (1991) uses the following moment condition to find instruments:

$$E(\vec{w}_{it-1}' \Delta u_{it}) = \mathbf{0}, t = 2, 3, \dots, T$$

where  $\vec{w}$  is the placeholder of the sequentially exogenous regressors vector, and  $\vec{w}_{it-1}^o$  is the history of  $\vec{w}$  from period 1 to  $t - 1$ , as defined in the section 2.2. In addition to the Arellano and Bond (1991) moment conditions above, Blundell and Bond (1998) uses extra moment conditions:

$$E(v_{it} \cdot \Delta \vec{w}_{it}) = 0, t = 2, 3, \dots, T$$

where  $v_{it} = c_i + u_{it}$  is the composite error term.

Note that Blundell and Bond (1998) condition starts from  $t = 3$ , and  $t = 2$  is included just because they are observed and ready to use. As Blundell and Bond (1998) have mentioned in their paper, the extra moment condition at  $t = 2$  should depends on the initial condition DGP  $\vec{w}_{i1}$ , which is  $y_{i0}$  and redundant  $D_{i1}(= 0, \forall i)$  in this paper.

Even if we focus on  $t = 3$ , which is the actual starting point of Blundell and Bond (1998), the moment condition does not hold in our setup due to the initial value problem. At  $t = 3$ , we have three instruments:  $\Delta y_{i2}$ ,  $\Delta D_{i3}$ , and the exogenous regressor  $z_{i3}$ . Observe that  $\Delta y_{i2}$  can be rewritten as

$$\begin{aligned}\Delta y_{i2} &= \Delta \beta_2 + \beta_y \Delta y_{i1} + \beta_d \Delta D_{i2} + \beta_z \Delta z_{i2} + \Delta u_{i2} \\ &= \{\Delta \beta_2 + \beta_z \Delta z_{i2} + \Delta u_{i2}\} + \beta_y \Delta y_{i1} + \beta_d D_{i2}\end{aligned}$$

where the second equality comes from the fact that  $D_{i1} = 0$  for all  $i$ . Since  $y_{i0}$  is exogenously given and does not include the heterogeneity  $c_i$ ,  $\Delta y_{i1} = y_{i1} - y_{i0}$  does not remove the heterogeneity. Also note that remaining  $D_{i2}$  also includes heterogeneity term by construction. Therefore,  $\Delta y_{i2}$  will be correlated with the heterogeneity, which violates the Blundell and Bond (1998) moment condition. Similarly, the third instrument  $z_{i3}$  is correlated with the heterogeneity by construction, unless we impose a strong assumption that the heterogeneity is not correlated with  $\bar{z}_i$ , the time average of exogenous variable<sup>2</sup>.

Table A.5 shows the simulation result of  $\beta_d$  for the correctly specified case. Unlike all the other simulations based on GAUSS, I used Stata's user written commend suggested by Roodman (2009b) for the simplicity of computation. As we can see from the result for POLS, FE, and FD, DGP itself seems to similar between two packages. Wooldridge (2000) approach retains both of consistency and efficiency compared to any other estimation strategies. As we can see from BB1 (with original diagonal instrument matrix) and BB2 (with collapsed instrument matrix), adding violated moment conditions would give us much bigger bias for  $\beta_d$ .

---

<sup>2</sup> $\Delta D_{i3}$  might make a valid instrument, but the other instruments are already invalid and they will make the estimates biased anyway.

Table A.5 Treatment Effect under Feedback Environment

$\beta_d = 2$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.672	0.106	0.680	-0.201	0.129	0.239	0.001	1.250	1.250
$N = 1,000$	-0.674	0.080	0.678	-0.205	0.096	0.227	-0.000	0.120	0.120
$N = 5,000$	-0.671	0.034	0.672	-0.205	0.041	0.209	-0.000	0.038	0.038
$N = 10,000$	-0.671	0.026	0.672	-0.207	0.031	0.210	-0.002	0.030	0.030
$\beta_d = 2$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.592	0.148	0.610	-0.232	1.268	1.289	-0.641	4.673	4.717
$N = 1,000$	-0.598	0.106	0.607	-0.111	0.828	0.836	-0.588	7.645	7.668
$N = 5,000$	-0.598	0.045	0.599	-0.026	0.308	0.309	-0.031	0.901	0.902
$N = 10,000$	-0.601	0.033	0.602	0.004	0.214	0.214	0.026	0.622	0.623
$\beta_d = 2$				BB1			BB2		
				Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$				-1.062	0.119	1.068	-1.180	0.131	1.187
$N = 1,000$				-1.068	0.089	1.072	-1.184	0.094	1.188
$N = 5,000$				-1.066	0.040	1.067	-1.183	0.042	1.184
$N = 10,000$				-1.067	0.029	1.067	-1.184	0.030	1.184

Table A.6 shows the simulation result for  $\beta_y$  for the correctly specified case. The implication is the same as for the result of  $\beta_d$ . For this reason, I only used AB1 and AB2 as a candidate of GMM estimator for the misspecified cases as they are consistent at least, while Blundell and Bond (1998) is not even consistent in the setup of this study.

Table A.6 Persistence under Feedback Environment

$\beta_y = 0.5$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	0.488	0.018	0.488	-0.122	0.021	0.124	-0.008	0.188	0.188
$N = 1,000$	0.488	0.012	0.488	-0.123	0.015	0.124	-0.000	0.022	0.022
$N = 5,000$	0.488	0.006	0.488	-0.123	0.006	0.124	0.000	0.006	0.006
$N = 10,000$	0.488	0.004	0.488	-0.123	0.005	0.123	0.000	0.005	0.005
$\beta_y = 0.5$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.248	0.025	0.249	-0.084	0.402	0.411	-0.173	1.222	1.234
$N = 1,000$	-0.248	0.017	0.249	-0.042	0.275	0.278	-0.160	2.046	2.052
$N = 5,000$	-0.249	0.007	0.249	-0.009	0.107	0.107	-0.009	0.243	0.243
$N = 10,000$	-0.249	0.005	0.249	0.001	0.076	0.076	0.008	0.170	0.170
$\beta_y = 0.5$				BB1			BB2		
				Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$				0.416	0.019	0.416	0.395	0.021	0.396
$N = 1,000$				0.416	0.014	0.416	0.395	0.014	0.395
$N = 5,000$				0.416	0.006	0.416	0.395	0.007	0.395
$N = 10,000$				0.416	0.004	0.416	0.395	0.005	0.395

### A.3. Misspecified Models

Suppose the idiosyncratic error  $u_{it}$  is not following standard normal, but the uniform $[-2, 2]$  distribution instead<sup>3</sup>. Table A.7 and Table A.8 show the results from the misspecified idiosyncratic error. As we can see, for both of  $\beta_d$  and  $\beta_y$ , Wooldridge (2000) approach still outperform the others, and has almost identical implication to the correctly specified case.

Suppose the error is following the standard normal distribution, but the part of heterogeneity,  $a_i$ , is following the uniform $[-2, 2]$  distribution. Table A.9 and A.10 shows the result for the case. Under the DGP of misspecified heterogeneity, Arellano and Bond (1991) approaches seems to have less bias as the sample size gets close to 10,000. As we can see from columns of standard deviations, however, Wooldridge (2000) approach works very efficiently, while having similar level of bias compared to Arellano and Bond (1991) approaches.

<sup>3</sup>I chose uniform distribution between  $\pm 2$  as it has thicker tail than normal distribution, while having a similar range of support.



Table A.7 Misspecified Error Term

$\beta_d = 2$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.692	0.128	0.704	-0.252	0.149	0.293	-0.002	0.146	0.146
$N = 1,000$	-0.692	0.087	0.697	-0.251	0.108	0.273	0.001	0.101	0.101
$N = 5,000$	-0.688	0.040	0.689	-0.248	0.048	0.253	0.004	0.044	0.045
$N = 10,000$	-0.687	0.029	0.688	-0.246	0.034	0.249	0.006	0.032	0.033
$\beta_d = 2$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.715	0.169	0.735	-0.487	1.379	1.462	-0.396	5.432	5.446
$N = 1,000$	-0.712	0.116	0.722	-0.195	0.895	0.916	-0.092	1.110	1.114
$N = 5,000$	-0.712	0.053	0.714	-0.032	0.381	0.383	-0.002	0.399	0.399
$N = 10,000$	-0.711	0.037	0.711	-0.009	0.255	0.255	0.008	0.274	0.274

Table A.8 Misspecified Error Term

$\beta_y = 0.5$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	0.485	0.018	0.486	-0.153	0.023	0.155	0.000	0.026	0.026
$N = 1,000$	0.485	0.012	0.485	-0.153	0.016	0.154	0.000	0.017	0.017
$N = 5,000$	0.486	0.006	0.486	-0.152	0.007	0.152	0.001	0.007	0.007
$N = 10,000$	0.486	0.004	0.486	-0.152	0.005	0.152	0.001	0.005	0.005
$\beta_y = 0.5$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.300	0.026	0.301	-0.163	0.465	0.493	-0.138	1.729	1.734
$N = 1,000$	-0.299	0.018	0.300	-0.072	0.314	0.322	-0.032	0.414	0.416
$N = 5,000$	-0.298	0.008	0.298	-0.011	0.138	0.139	0.001	0.153	0.153
$N = 10,000$	-0.299	0.006	0.299	-0.003	0.093	0.093	0.004	0.106	0.106

Table A.9 Misspecified Heterogeneity

$\beta_d = 2$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.736	0.116	0.745	-0.220	0.131	0.256	-0.032	0.134	0.138
$N = 1,000$	-0.735	0.078	0.739	-0.221	0.095	0.240	-0.029	0.089	0.094
$N = 5,000$	-0.730	0.036	0.731	-0.216	0.042	0.221	-0.024	0.040	0.047
$N = 10,000$	-0.730	0.027	0.730	-0.216	0.030	0.218	-0.023	0.029	0.037
$\beta_d = 2$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.614	0.147	0.631	-0.298	1.208	1.244	-0.132	1.491	1.497
$N = 1,000$	-0.615	0.105	0.624	-0.106	0.783	0.790	-0.018	0.869	0.869
$N = 5,000$	-0.612	0.046	0.614	-0.015	0.332	0.332	0.008	0.346	0.346
$N = 10,000$	-0.611	0.034	0.612	-0.010	0.224	0.225	0.001	0.239	0.239

Table A.10 Misspecified Heterogeneity

$\beta_y = 0.5$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	0.493	0.018	0.493	-0.121	0.021	0.123	-0.002	0.026	0.026
$N = 1,000$	0.493	0.012	0.493	-0.121	0.015	0.122	-0.002	0.016	0.016
$N = 5,000$	0.494	0.006	0.494	-0.120	0.007	0.120	-0.002	0.007	0.007
$N = 10,000$	0.494	0.004	0.494	-0.120	0.004	0.120	-0.002	0.005	0.005
$\beta_y = 0.5$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.245	0.024	0.246	-0.094	0.379	0.391	-0.040	0.531	0.532
$N = 1,000$	-0.244	0.017	0.245	-0.038	0.252	0.255	-0.004	0.302	0.302
$N = 5,000$	-0.243	0.008	0.243	-0.005	0.109	0.109	0.004	0.123	0.123
$N = 10,000$	-0.244	0.005	0.244	-0.003	0.074	0.074	0.000	0.086	0.086

Finally, suppose both of the error term and heterogeneity terms are misspecified; While both of the error and the heterogeneity are following the uniform distribution, econometrician might assume that they are following the standard normal distribution, and run MLE based on that assumption. This “both wrong case” would be a realistic scenario for most of researchers. Table A.11 and Table A.12 show the result for the doubly-misspecified cases. Surprisingly, the implication is close to the previous conclusions. Wooldridge (2000) approach seems to work efficient with some bearable bias even under a realistic misspecification.

Table A.11 Misspecified Error and Heterogeneity

$\beta_d = 2$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.748	0.130	0.760	-0.266	0.151	0.306	-0.027	0.151	0.154
$N = 1,000$	-0.746	0.089	0.752	-0.267	0.106	0.287	-0.020	0.101	0.103
$N = 5,000$	-0.741	0.041	0.742	-0.261	0.050	0.265	-0.016	0.046	0.049
$N = 10,000$	-0.740	0.029	0.741	-0.259	0.034	0.262	-0.013	0.032	0.035
$\beta_d = 2$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.735	0.170	0.755	-0.444	1.422	1.489	-0.366	2.489	2.516
$N = 1,000$	-0.731	0.117	0.740	-0.195	0.996	1.015	-0.071	1.358	1.360
$N = 5,000$	-0.727	0.054	0.729	-0.035	0.413	0.415	-0.001	0.454	0.454
$N = 10,000$	-0.725	0.038	0.726	-0.022	0.275	0.276	-0.007	0.311	0.311

Table A.12 Misspecified Error and Heterogeneity

$\beta_y = 0.5$	POLS			FE			W2000		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	0.490	0.018	0.491	-0.150	0.023	0.151	-0.001	0.029	0.029
$N = 1,000$	0.491	0.013	0.491	-0.149	0.016	0.150	-0.002	0.017	0.017
$N = 5,000$	0.491	0.006	0.491	-0.148	0.007	0.149	-0.001	0.007	0.007
$N = 10,000$	0.491	0.004	0.491	-0.149	0.005	0.149	-0.001	0.005	0.005
$\beta_y = 0.5$	FD			AB1			AB2		
	Bias	S D	rMSE	Bias	S D	rMSE	Bias	S D	rMSE
$N = 500$	-0.294	0.026	0.295	-0.146	0.478	0.500	-0.130	0.895	0.904
$N = 1,000$	-0.294	0.018	0.294	-0.070	0.344	0.351	-0.025	0.504	0.505
$N = 5,000$	-0.293	0.008	0.293	-0.012	0.145	0.146	0.001	0.170	0.170
$N = 10,000$	-0.293	0.005	0.293	-0.008	0.097	0.097	-0.003	0.118	0.118

## APPENDIX B PROOFS FOR CHAPTER 2

### B.1. Proof of Proposition 2.1.

It suffices to show the result with ratios while ignoring -1. Observe that

$$\frac{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0 \right]} = \frac{E \left[ Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0 \right]} = \frac{E \left[ Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]} = \exp(\tau).$$

where the first equality holds by Neyman-Rubin model (2.1), the second equality holds by sequential ignorability (Assumption 2.2), and the third equality holds by homogeneous treatment effects (Assumption 2.1). Therefore, the ratio between the outcome models of treated and controlled group identifies the conditional average treatment effects. Furthermore, we can show that the following equality holds, indicating that under the homogeneous treatment effects, integration of heterogeneity is straightforward.

$$\frac{E \left[ Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]} = \frac{E \left[ Y_{it}^{(1)} | \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]}{E \left[ Y_{it}^{(0)} | \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]} = \exp(\tau).$$

To have this equality, take the denominator  $E \left[ Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]$  to the right hand side and apply iterated expectation to integrate out the heterogeneity  $c_i$  using the distribution  $D(c_i | \vec{Y}_{it-1}, X_i, D_{it} = 1)$ . Then put  $E \left[ Y_{it}^{(0)} | \vec{Y}_{it-1}, X_i, D_{it} = 1 \right]$  back to the denominator on the left hand side.

### B.2. Proof of Proposition 2.2. (Identification without Selection Bias)

It suffices to show that (2.6) implies (2.7) and (2.7) implies (2.6), given (2.1) and (2.2). First, observe that (omitting  $\theta$  for the simplicity)

$$\begin{aligned} E(Y_{it}^{(d)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = d) &= E(Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = d) \\ &= c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_i \beta_x + \tau d) = E(Y_{it}^{(d)} | c_i, \vec{Y}_{it-1}, X_i) \end{aligned}$$

where the first equality hold by construction, second equality holds by (2.6), and the third equality hold by (2.2). Second, notice that

$$E(Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it}) = E(D_{it} \cdot Y_{it}^{(1)} + (1 - D_{it}) \cdot Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it})$$

$$\begin{aligned}
&= D_{it} \cdot E(Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it}) + (1 - D_{it}) \cdot E(Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it}) \\
&= D_{it} \cdot E(Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i) + (1 - D_{it}) \cdot E(Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i) \\
&= D_{it} \cdot c_i \cdot \mu_{it} \cdot \exp(\tau \cdot 1) + (1 - D_{it}) \cdot \mu_{it} \cdot \exp(\tau \cdot 0) = c_i \cdot \mu_{it} \cdot \exp(\tau \cdot D_{it}) \\
&= c_i \cdot \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_i \beta_x + \tau D_{it})
\end{aligned}$$

where the first equality hold by (2.1), the second equality holds as we are conditioning on  $D_{it}$ , the third equality holds by (2.7), the fourth equality holds by (2.2) where  $\mu_{it} = \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_i \beta_x)$ , the fifth equality comes by construction of binary  $D_{it}$ .

**Remark.** One would find almost the same proof from Lee and Kobayashi (2002). What I did was extending the conditioning set to include the (sequential) history of lagged outcomes, and check whether it still works or not.

### B.3. Proof of Proposition 2.3. (Parametric Identification of CATE with Ratio Form)

Under (2.1), (2.2), and (2.7), we have (2.6) by Proposition 2.2. Then by (2.6) we have

$$\frac{E(Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1)}{E(Y_{it} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0)} = e^\tau.$$

Therefore, it suffices to show that

$$\frac{E[Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]}{E[Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]} = e^\tau.$$

Observe

$$\frac{E[Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]}{E[Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]} = \frac{E[Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i]}{E[Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i]} = \frac{c_i \cdot \mu_{it} \cdot \exp(\tau \cdot 1)}{c_i \cdot \mu_{it} \cdot \exp(\tau \cdot 0)} = e^\tau$$

where the first equality holds by (2.7), the second equality holds by (2.2) where  $\mu_{it} = \exp(\beta_t + \beta_y \ln(Y_{it-1} + Z_{it-1}) + \beta_z Z_{it-1} + X_i \beta_x)$ .

**Remark.** As mentioned in section 2.3, one might think of another identification result with a ratio-in-ratios form, which can be written as

$$\frac{E(Y_{it}/h(Y_{it-1}) | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1)}{E(Y_{it}/h(Y_{it-1}) | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 0)} = \frac{E[Y_{it}^{(1)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]}{E[Y_{it}^{(0)} | c_i, \vec{Y}_{it-1}, X_i, D_{it} = 1]}$$

where  $h(Y_{it-1}) \equiv (Y_{it-1} + Z_{it-1})^{\beta_y}$ . Considering the original definition of Ding and Li (2019), the ratio-in-ratios above would be a faithful analog.

$$\tau_{LDV} = (E[Y_{it}|G_i = 1] - E[Y_{it}|G_i = 0]) - \beta_y(E[Y_{it-1}|G_i = 1] - E[Y_{it-1}|G_i = 0])$$

Observe that including  $h(Y_{it-1})$  terms are redundant as they are all cancelled out once we condition on the lagged outcomes. For the simplicity of discussion, I removed the  $h(Y_{it-1})$  terms from my main identification result.

## APPENDIX C      EXTENSIONS OF CHAPTER 2

### C.1. Simulation Results

Table C.1 repeats the result of Table 2.1. Here IV1 and IV2 stands for the block-diagonal and triangular IV matrices respectively, and GMM1 stands for the first stage GMM who used the identity matrix as the weighting matrix of GMM. GMM2 used the optimal weighting matrix, whose result was posted on Table 2.1. As one can see, un-optimal GMM can work poorly under the small sample size, but the optimal GMM converge to the true parameter even with small sample. For both of IV specifications (IV1 and IV2), we can also find that the optimal GMM2 gets efficiency gain, as we use the optimal weighting matrix over the identity matrix. As mentioned in Section 2.5, IV2 cannot converge to the true parameter in GMM1, which makes the triangular collapsing matrix specification less competitive compared to the block-diagonal one.

Table C.1 Average of estimated coefficients and their standard deviations when the treatment is sequentially exogenous (T=4, 1000 Repetitions)

		ln( $Y_{t-1} + Z_{t-1}$ ) included				
		IV1			IV2	
$\tau = 0.3$	TWFE	AR1FE	GMM1	GMM2	GMM1	GMM2
N=500	-0.196 (0.269)	0.565 (0.035)	9770937 (3.09e+08)	0.316 (0.168)	13778.3 (378361.7)	0.313 (0.201)
N=1000	-0.246 (0.252)	0.562 (0.024)	0.355 (0.389)	0.307 (0.117)	0.754 (12.286)	0.306 (0.139)
N=2000	-0.276 (0.219)	0.562 (0.017)	0.293 (0.348)	0.301 (0.083)	2706153 (8.56e+07)	0.302 (0.093)
N=4000	-0.330 (0.232)	0.561 (0.013)	0.278 (0.321)	0.302 (0.063)	49353.76 (1458559)	0.302 (0.072)

*Note.* Standard deviations in the parentheses. IV1 used the Arellano and Bond (1991) types of instrument matrices, while IV2 used the suggestion of Roodman (2009a,b). GMM1 is the first stage GMM estimation using the identity weighting matrices, while GMM2 stands for the second stage GMM based on the optimal weighting matrices.



Table C.2 shows the extension of Table 2.2 when the treatment assignment is strictly exogenous. As in Table C.1, block-diagonal IV matrix (IV1) works well. Triangular IV2 reports less competitive performance compared to the block-diagonal specification.

Table C.2 Average of estimated coefficients and their standard deviations when the treatment is strictly exogenous (T=4, 1000 Repetitions)

		$\ln(Y_{t-1} + Z_{t-1})$ included				
		IV1			IV2	
$\tau = 0.3$	TWFE	AR1FE	GMM1	GMM2	GMM1	GMM2
N=500	-0.086	0.271	0.319	0.315	0.432	-0.231
	(0.214)	(0.036)	(0.230)	(0.163)	(3.411)	(17.179)
N=1000	-0.048	0.277	0.302	0.300	0.306	11.87323
	(0.193)	(0.026)	(0.143)	(0.093)	(0.220)	(365.938)
N=2000	-0.041	0.279	0.297	0.299	0.305	0.304
	(0.179)	(0.018)	(0.106)	(0.065)	(0.111)	(0.125)
N=4000	-0.047	0.282	0.298	0.302	0.297	0.302
	(0.189)	(0.013)	(0.106)	(0.049)	(0.171)	(0.094)

*Note.* Standard deviations in the parentheses. IV1 used the Arellano and Bond (1991) types of instrument matrices, while IV2 used the suggestion of Roodman (2009a,b). GMM1 is the first stage GMM estimation using the identity weighting matrices, while GMM2 stands for the second stage GMM based on the optimal weighting matrices.

Table C.3 presents a simulation with the case where the true DGP is designed with the strictly exogenous treatment assignment, while an econometrician decided to run the GMM assuming that the treatment is sequentially exogenous (over-misspecification of treatment assignment). The implication of this simulation study is similar to the previous table: block-diagonal IV1 works well for both of GMM1 (un-optimal) and GMM2 (optimal), while the triangular IV2 is producing poor result, including relative inefficiency even in the large sample.

Table C.3 Average of estimated coefficients and their standard deviations when the treatment is strictly exogenous (T=4, 1000 Repetitions)

		$\ln(Y_{t-1} + Z_{t-1})$ included				
		IV1			IV2	
$\tau = 0.3$	TWFE	AR1FE	GMM1	GMM2	GMM1	GMM2
N=500	-0.086 (0.214)	0.271 (0.036)	0.319 (0.231)	0.317 (0.167)	0.326 (0.349)	0.776 (14.570)
N=1000	-0.048 (0.193)	0.277 (0.026)	0.302 (0.143)	0.299 (0.106)	0.306 (0.220)	10126.74 (320226.1)
N=2000	-0.041 (0.179)	0.279 (0.018)	0.297 (0.107)	0.299 (0.066)	0.308 (0.120)	0.301 (0.156)
N=4000	-0.047 (0.189)	0.282 (0.013)	0.298 (0.106)	0.303 (0.053)	0.297 (0.172)	0.303 (0.087)

*Note.* Standard deviations in the parentheses. IV1 used the Arellano and Bond (1991) types of instrument matrices, while IV2 used the suggestion of Roodman (2009a,b). GMM1 is the first stage GMM estimation using the identity weighting matrices, while GMM2 stands for the second stage GMM based on the optimal weighting matrices.

Table C.4 shows the simulation result for the case when the true DGP is designed with the feedback (sequentially exogenous treatment assignment), while an econometrician wrongly assumes that the assignment is strictly exogenous (under-misspecification of treatment assignment). For all instrument specifications, un-optimal GMM (GMM1) is generally biased. Optimal GMM (GMM2) works relatively better, but bias can be big under the small sample size. The Pattern between the two IV specifications stays the same: Triangular IV2 is generally inefficient compared to the result from block-diagonal IV1.

Table C.4 Average of estimated coefficients and their standard deviations when the treatment is sequentially exogenous (T=4, 1000 Repetitions)

		$\ln(Y_{t-1} + Z_{t-1})$ included				
		IV1			IV2	
$\tau = 0.3$	TWFE	AR1FE	GMM1	GMM2	GMM1	GMM2
N=500	-0.196 (0.269)	0.565 (0.035)	0.542 (1.941)	0.319 (0.171)	161.765 (5065.52)	-0.062 (11.887)
N=1000	-0.246 (0.252)	0.562 (0.024)	0.377 (0.730)	0.307 (0.118)	1.636 (41.000)	0.303 (0.142)
N=2000	-0.276 (0.219)	0.562 (0.017)	0.297 (0.332)	0.302 (0.084)	197476.7 (6244753)	0.302 (0.094)
N=4000	-0.330 (0.232)	0.561 (0.013)	0.287 (0.304)	0.303 (0.066)	15916.02 (440807.8)	0.301 (0.073)

*Note.* Standard deviations in the parentheses. IV1 used the Arellano and Bond (1991) types of instrument matrices, while IV2 used the suggestion of Roodman (2009a,b). GMM1 is the first stage GMM estimation using the identity weighting matrices, while GMM2 stands for the second stage GMM based on the optimal weighting matrices.

## C.2. Empirical Application

Table C.5 and C.6 are the analog of Table 2.4 and Table 2.5, using the triangular IV collapsing method suggested by Roodman (2009a,b). They yields similar implications in general.

Table C.5 Effects of PDMP for Child Maltreatment Allegations (with Triangular IV)

Panel A	Coefficients				% Interpretation		
Linear Models	TWFE	LyStD	LySqD		TWFE	LyStD	LySqD
Alleged Cases							
PDMP	4.150	-0.031	0.177		0.106	-0.000	0.005
	(2.769)	(0.289)	(0.219)				
Allow Feedback	NO	NO	YES		NO	NO	YES
Lagged Y controlled	NO	YES	YES		NO	YES	YES
Panel B							
Exponential Models	FePoi	GMM1	GMM1	GMM1	GMM2	GMM2	GMM2
Alleged Cases							
PDMP	0.074***	-0.127	-0.183**	-0.040	-0.053	-0.045	-0.055
	(0.020)	(0.118)	(0.090)	(0.042)	(0.072)	(0.036)	(0.039)
Allow Feedback	NO	YES	NO	YES	YES	NO	YES
Lagged Y controlled	NO	NO	YES	YES	NO	YES	YES

*Note.* Standard deviations in the parentheses. Significance levels are \*\*\*(1%), \*\*(5%), and \*(10%) respectively. LyStD and LySqD were estimated using the GMM approach following the Arellano and Bond (1991) instrument while reporting only the second stage estimates. FePoi is the analogy of TWFE into the exponential mean specification. GMM1 is the first stage GMM estimates while GMM2 stands for the second stage GMM estimates. % Interpretation in the Panel B were calculated by dividing the average size of outcome used during the estimation, following the approach in Evans et al. (2022b). All GMM results including LyStD and LySqD are based on the IV matrix suggested by Roodman (2009a,b).

Table C.6 Effects of PDMP for Child Maltreatment Substantiations (with Triangular IV)

Panel A	Coefficients			% Interpretation			
Linear Models	TWFE	LyStD	LySqD	TWFE	LyStD	LySqD	
Substantiated Cases							
PDMP	1.006	0.254**	0.294***	0.117	0.030	0.034	
	(0.661)	(0.127)	(0.078)				
Allow Feedback	NO	NO	YES	NO	NO	YES	
Lagged Y controlled	NO	YES	YES	NO	YES	YES	
Panel B							
Exponential Models	FePoi	GMM1	GMM1	GMM1	GMM2	GMM2	GMM2
Substantiated Cases							
PDMP	0.097***	-0.012	0.056	0.088*	0.020	-0.052	0.053
	(0.025)	(0.094)	(0.077)	(0.048)	(0.136)	(0.066)	(0.042)
Allow Feedback	NO	YES	NO	YES	YES	NO	YES
Lagged Y controlled	NO	NO	YES	YES	NO	YES	YES

*Note.* Standard deviations in the parentheses. Significance levels are \*\*\*(1%), \*\*(5%), and \*(10%) respectively. LyStD and LySqD were estimated using the GMM approach following the Arellano and Bond (1991) instrument while reporting only the second stage estimates. FePoi is the analogy of TWFE into the exponential mean specification. GMM1 is the first stage GMM estimates while GMM2 stands for the second stage GMM estimates. % Interpretation in the Panel B were calculated by dividing the average size of outcome used during the estimation, following the approach in Evans et al. (2022b). All GMM results including LyStD and LySqD are based on the IV matrix suggested by Roodman (2009a,b).

# per 1,000

- 43.47 – 63.06
- 38.97 – 43.47
- 30.65 – 38.97
- 26.45 – 30.65
- 23.63 – 26.45
- 16.03 – 23.63
- 1.60 – 16.03
- No data

Figure 2.2 lacks the information about states where the must-access PDMP was implemented after the year 2017, as the data set only contains the outcome until the year 2016. Figure C.1 presents a graphic illustration of regional variation of alleged count using the initial observation in 2006. We can still detect an overlapping patterns in Midwest regions and states between California and Texas.

## APPENDIX D      PROOF FOR CHAPTER 3

### D.1. Imputation of Heterogeneity

To avoid confusion, let me abuse notation of covariates into  $x_{it} = (y_{i,t-1}, \theta_t)$  so that we have

$$\widehat{c}_i \equiv \frac{1}{T_0} \sum_t \left[ y_{i,t} - \widehat{\alpha} y_{i,t-1} - \widehat{\theta}_t \right] = \frac{1}{T_0} \sum_t \left[ y_{i,t} - x_{i,t} \widehat{\beta} \right].$$

where  $\widehat{\beta} \xrightarrow{p} \beta$  under the Assumption 3.4. Now, it suffices to show that

$$\frac{1}{N} \sum_i \widehat{c}_i \xrightarrow{p} \mathbb{E}[c_i] = \frac{1}{T_0} \sum_t \mathbb{E}[c_i] = \mathbb{E} \left[ \frac{1}{T_0} \sum_t c_i \right]$$

By taking summation both hand side, we have

$$\frac{1}{N} \sum_i \widehat{c}_i = \frac{1}{N} \sum_i \left[ \frac{1}{T_0} \sum_t \left[ y_{it} - x_{i,t} \widehat{\beta} \right] \right] \xrightarrow{p} \mathbb{E} \left[ \frac{1}{T_0} \sum_t \left[ y_{it} - x_{i,t} \beta \right] \right]$$

where the last convergence hold under the weak law of large numbers and Slutsky's theorem under their regularity conditions. By this convergence, we only need to show that

$$\mathbb{E}[c_i] = \mathbb{E}[y_{i,t} - x_{i,t} \beta].$$

By (3.2) and Assumption 3.3, we have  $c_i = y_{i,t} - x_{i,t} \beta - u_{i,t}$  for  $t \leq T_0$ . Under the Assumption 3.4, after iterated expectation, we have the result. To impute  $\mathbb{E}[c_i | D_i = 1]$ , one need to exchange  $N$  with  $N_1 = \sum_i^N D_i$  and multiplied  $D_i$  inside summation as one can see from equation (3.8).