Why Does Difference In Difference Matching Work?*

Sylvain Chabé-Ferret†
Toulouse School of Economics and Inra, Lerna

This version: August 26, 2014.

Very preliminary – comments welcome

Abstract

Difference In Difference (DID) Matching is one of the nonexperimental methods of causal inference that reproduces the results of Randomized Controlled Trials (RCTs) the best. One intuitive explanation for this success is that DID & Matching combine their strengths: DID differences out the permanent confounders while Matching on pre-treatment outcomes captures transitory shocks. I show that this intuitive explanation is incorrect: it is both inconsistent theoretically and does not perform well in simulations of a model of earnings dynamics and selection into a Job Training Program (JTP). I show that DID Matching performs well when it is implemented symmetrically around the treatment date and does not condition on pre-treatment outcomes. I explain why this is the case and bring evidence from earlier experiments comparing DID Matching with RCTs that strongly support this result. These results have powerful consequences for the way DID Matching is implemented in practice, its application to programs over that JTPs and for the neverending quest for a reliable nonexperimental method of causal inference.

Keywords: Difference in Difference Matching - Selection Model - Treatment Effects.

JEL codes: C21, C23.

*I thank Koen Jochmans for very constructive comments on a preliminary version of this work presented at the French Econometrics Conference. I also thank conference participants for their useful feedback and suggestions. All remaining errors are my own.

†Correspondence to: Sylvain Chabé-Ferret, Toulouse School of Economics – Lerna, 21 Allée de Brienne, 31015 Toulouse Cedex 6, France. Email: sylvain.chabe-ferret@tse-fr.eu. Tel: +33 (0)5 61 12 88 28. Fax:+33 (0)5 61 12 85 20.

1 Introduction

Developing reliable methods for causal inference is crucial for testing scientific theories and evaluating the effect of public or private interventions. The main difficulty with causal inference is to tell causation from correlation, the causal effect from selection bias. Selection bias occurs because there exist confounding factors that simultaneously determine the outcomes of interest and who receives the intervention. Randomization corrects for selection bias by allocating the intervention independently of the confounders. Unfortunately, randomization is not without its problems (ethical and political limits to feasibility, randomization bias). As early as Yule (1899), economists and statisticians seek to develop methods disentangling the effect of an intervention from that of confounding factors without resorting to randomization. Since LaLonde (1986), these methods have been subjected to tests comparing their results with an experimental benchmark. The most robust method to survive these tests is the combination of Difference In Difference with conditioning on observed covariates (Heckman, Ichimura, Smith, and Todd, 1998; Smith and Todd, 2005).[1]

Most of the evidence that we have on the performance of DID Matching is limited to one application though: the effect of Job Training Programs (JTPs) on earnings. We do not really know why DID Matching performs so well in these experiments, what is the precise way to implement it that yields the best results and whether such nice behavior can be expected in other applications as well. Intuitively, one would expect that DID and Matching combine their strengths: DID gets rid of permanent confounders while Matching capture selection due to transitory shocks. This is especially relevant in the case of JTPs: it has been widely documented that participants in a Job Training Program (JTP) have permanently lower earnings but also experience a transitory decrease in earnings before entering the program (see Heckman, LaLonde, and Smith (1999) for a survey).

In this paper, I show that this intuitive explanation for the good performance of DID Matching is inconsistent theoretically. In a model where selection bias is a combination

[1]When conditioning is performed nonparametrically, this method has been coined DID Matching (Blundell and Costa Dias, 2000) or semiparametric DID (Abadie, 2005). When conditioning is performed using a parametric model, this approach has been coined quasi-DID (LaLonde, 1986).

of a permanent fixed effect with transitory shocks, combining DID with conditioning on pre-treatment outcomes is either irrelevant or inconsistent. It is irrelevant because it is only consistent when simple Matching is also consistent. Worse, combining DID with conditioning on pre-treatment outcomes generates bias when DID is consistent. An important case in point is when selection bias forms and dissipates at the same pace and applying DID symmetrically around the treatment date is consistent (Chabé-Ferret, 2014). Conditioning on pre-treatment outcomes breaks the symmetry of the dip and thus generates bias.

Results of Monte Carlo simulations of a model of earnings dynamics and self-selection in a JTP parameterized with realistic parameter values show that the bias generated by conditioning on pre-treatment outcomes can be sizable and that Symmetric DID still dominates DID Matching even when both are inconsistent. Taking a closer look at the papers comparing DID Matching with an experimental benchmark, I find that their positive results stem from the application of DID symmetrically around the treatment date rather than from the combination of DID with conditioning on pre-treatment outcomes. In these papers, conditioning on pre-treatment outcomes actually worsens the performances of Symmetric DID.

Together with those of Chabé-Ferret (2014), these results push toward the use of Symmetric DID instead of combining DID with conditioning on pre-treatment outcomes in cases where we suspect selection on transitory shocks and on a permanent fixed effect. These results are not a rebuttal of DID Matching per se, but only of the combination of DID with Matching on pre-treatment outcomes. Actually, it is the combination of Symmetric DID with Matching on covariates that are constant over time that performs best at reproducing experimental results. Indeed, nonparametrically conditioning on covariates allows to capture differential time trends among participants and non participants. This paper shows that this intuition suffers from a very important exception: it is not possible to capture differential trends due to pre-treatment outcomes and at the same time to difference out unobserved fixed effects. In the course of the paper, I derive necessary and sufficient conditions for Matching and DID to be consistent. These results are of separate interest. Matching is consistent if and only if selection is due to transitory shocks.

Matching on pre-treatment outcomes thus does not correct for selection on a permanent unobserved fixed effect. DID is consistent if and only if selection is due to a fixed effect.

Related literature

The approach of using a model of outcome dynamics and selection in a program to study the properties of nonexperimental estimators is rooted in an ancient literature. Ashenfelter (1978) formalizes the evaluation problem as a combination of selection on a fixed effect and on transitory shocks. Heckman (1978), Heckman and Robb (1985) and Ashenfelter and Card (1985) combine the selection equation with the outcome dynamics equation and introduces Symmetric Differencing. Bassi (1984) acknowledges that combining differencing with conditioning on pre-treatment changes in outcomes suffers from Nickell (1981)'s problem: pre-treatment earnings are correlated with transitory shocks. LaLonde (1986) introduces the quasi-differencing approach combining differencing with conditioning on pre-treatment outcomes. DID Matching was introduced by Heckman, Ichimura, and Todd (1997) and Heckman, Ichimura, Smith, and Todd (1998). Chabé-Ferret (2014) compares Matching, DID and Symmetric DID when only one observation of pre-treatment outcomes is available.

This paper is structured as follows: Section 2 formally introduces the setting, the estimators and their bias; Section 3 presents the sufficient conditions for the consistency of Matching, DID and DID Matching; Section 4 presents the results of simulations of a model of earnings dynamics and selection in a JTP calibrated with realistic parameter values; Section 5 presents evidence from comparison of JTPs with an experimental benchmark confirming that combining DID with conditioning on pre-treatment outcomes does less well than Symmetric DID at reproducing the results of RCTs.

2 The model, the estimators and their bias

The model

I use a simple selection model exhibiting selection both on a fixed effect and on transitory shocks. The outcomes in the absence of the treatment depend on time and individual fixed effects and on transitory shocks (Equation (1a)). Transitory shocks are persistent: they follow an AR(1) process with $|\rho| < 1$ (Equation (1b)).

$$Y_{i,t}^0 = \delta_t + \mu_i + U_{it} \tag{1a}$$

$$\text{with } U_{i,t} = \rho U_{i,t-1} + v_{i,t} \tag{1b}$$

$$D_{i,k} = \mathbb{1}[t \geq k]\mathbb{1}[\underbrace{\theta_i + \gamma Y_{i,k-1}^0}_{D_{i,k}^*} \geq 0]. \tag{1c}$$

Treatment is offered at period k. Selection into the program depends on an individual fixed effect θ_i and on outcomes at date $k-1$ (Equation (1c)). The two critical parameters for selection are γ and $\rho_{\theta,\mu}$ (the correlation of the fixed effect μ_i with the unobserved shifter of participation θ_i). When $\gamma = 0$, selection is due to the fixed effect only. When $\rho_{\theta,\mu} = 0$, selection is on the observed pre-treatment outcome $Y_{i,k-1}^0$ only. The key question that this paper answers is whether combining DID with conditioning on $Y_{i,k-1}^0$ corrects for selection bias when both γ and $\rho_{\theta,\mu}$ are different from zero.

In order to focus the burden of selection bias on these two key parameters, I make the following assumptions: $\rho \neq 0$, $\sigma^2 > 0$, $\sigma_{U_0}^2 > 0$, $\sigma_\mu^2 > 0$, $\sigma_\theta^2 > 0$. $v_{i,t}$ are *i.i.d.* mean-zero shocks with finite variance σ^2 and $U_{i,0}$ is a mean-zero shock with variance $\sigma_{U_0}^2$. $v_{i,t} \perp\!\!\!\perp (\mu_i, \theta_i), \forall t$ and $U_{i,0} \perp\!\!\!\perp (\mu_i, \theta_i, v_{i,t}), \forall t$. I assume normally distributed error terms all along.

Examples

Although admittedly very simple, the model described by equation (1) has several virtues. First, it encapsulates in the simplest possible setting the problem that DID Matching is trying to solve: selection on a fixed effect and on transitory shocks. Second, it accounts for various types of realistic selection processes: namely self-selection in a Job Training Program (JTP) and a cutoff eligibility rule.

Self selection Assuming no idiosyncratic trend, no MA terms and limited information, setting $\gamma = -\rho$ and $\theta_i = \frac{\alpha_i}{r} - c_i$, the model of entry into a JTP studied in Chabé-Ferret (2014) simplifies to the model described by equation (1).

Eligibility rules As argued in Chabé-Ferret (2014), a program allocated when a running variable falls below some eligibility threshold can also be described by equation (1c). In that case, $\gamma = -1$ and θ_i accounts for measurement error in the variable determining eligibility.

The estimators and their bias

The parameter of interest is the average effect of the treatment on the treated (ATT) τ periods after the treatment date. I consider three estimators of the ATT: Matching, DID and DID Matching (DIDM). I only study asymptotic bias, and therefore focus on population formulae.

$$B(M_{k,\tau,1}) = \mathbb{E}[\mathbb{E}[Y^0_{i,k+\tau}|D_{i,k}=1,Y^0_{i,k-1}] - \mathbb{E}[Y^0_{i,k+\tau}|D_{i,k}=0,Y^0_{i,k-1}]|D_{i,k}=1] \tag{2a}$$

$$B(DID_{k,\tau,\tau'}) = \mathbb{E}[Y^0_{i,k+\tau} - Y^0_{i,k-\tau'}|D_{i,k}=1] - \mathbb{E}[Y^0_{i,k+\tau} - Y^0_{i,k-\tau'}|D_{i,k}=0] \tag{2b}$$

$$\begin{aligned} B(DIDM_{k,\tau,1,\tau'}) = \mathbb{E}\Big[&\mathbb{E}[Y^0_{i,k+\tau} - Y^0_{i,k-\tau'}|D_{i,k}=1,Y^0_{i,k-1}] \\ &- \mathbb{E}[Y^0_{i,k+\tau} - Y^0_{i,k-\tau'}|D_{i,k}=0,Y^0_{i,k-1}]|D_{i,k}=1\Big]. \end{aligned} \tag{2c}$$

The Matching estimator compares the expected outcomes of the treated τ periods after the treatment to those of the untreated conditional on $Y^0_{i,k-1}$. $Y^0_{i,k-1}$ is the last pre-treatment outcome observed before the treatment is taken and intuitively the one containing the most relevant information for selection. Following Chabé-Ferret (2014), the bias of Matching is the expected difference in potential outcomes in the absence of the treatment between participants and non participants, conditional on $Y^0_{i,k-1}$, integrated over the distribution of $Y^0_{i,k-1}$ conditional on $D_{i,k} = 1$ (Equation (2a)).

The DID estimator compares the post-treatment difference in outcomes to the difference that existed some τ' periods before the treatment. Following Chabé-Ferret (2014), the bias of DID is equal to the change over time in the difference in potential outcomes between participants and non participants in the absence of the program (Equation (2b)).

The DIDM estimator compares the evolution in the differences in outcomes before and after the program after conditioning on $Y^0_{i,k-1}$. The bias of DIDM is equal to the change over time in the difference in potential outcomes between participants and non participants in the absence of the program conditional on $Y^0_{i,k-1}$, integrated over the distribution of $Y^0_{i,k-1}$ conditional on $D_{i,k} = 1$ (Equation (2c)).

Consistency

As I want to state general results on the model parameters for each of the estimators to be consistent, I have to define the sets of periods k, τ and τ' for which I want the biases of the various estimators to cancel. The usual practice is to use the estimators without restricting their validity to any particular subset of the possible treatment dates (k) or lag between treatment and observation of outcomes (τ). Thus, I will define consistency in this model as requiring that the estimators are valid for all $k > 0$ and for all $\tau \geq 0$. This requirement mainly serves to weed out special cases in which the estimators are consistent for a peculiar combinations of dates. Similarly, for DID and DIDM, I define consistency as the fact that the bias of the estimator is zero regardless of the pre-treatment period $k - \tau'$ used to construct the estimator. This reflects the common practice of DID, to the best of my knowledge.

3 Theoretical results

DID conditioning on pre-treatment outcomes is irrelevant

Theorem 1 *The three following statements are equivalent:*

(i) $\forall k > 0$, $\forall \tau \geq 0$, $\forall \tau' > 0$, $B(DIDM_{k,\tau,1,\tau'}) = 0$

(ii) $\forall k > 0$, $\forall \tau \geq 0$, $B(M_{k,\tau,1}) = 0$

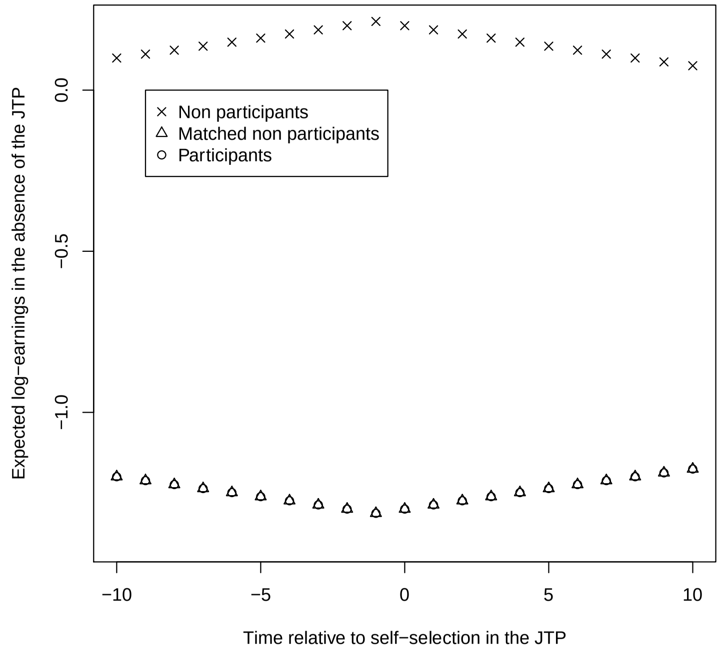
(iii) $\rho_{\theta,\mu} = 0$.

Theorem 1 shows that combining DID with Matching on pre-treatment outcomes is consistent if and only if Matching also is. Thus DID Matching does not bring any additional identifying power on top of Matching. Moreover, Theorem 1 also shows that Matching and DID Matching are consistent if and only if selection is on observed transitory shocks only. Thus, both these methods are inconsistent when selection is due to a fixed effect only, or when it is a combination of a fixed effect with a transitory shock. As a consequence, combining DID with Matching on pre-treatment outcomes is not the silver bullet we would hope it is: it does not combine the strengths of its forefathers, but only inherits the virtues of Matching.

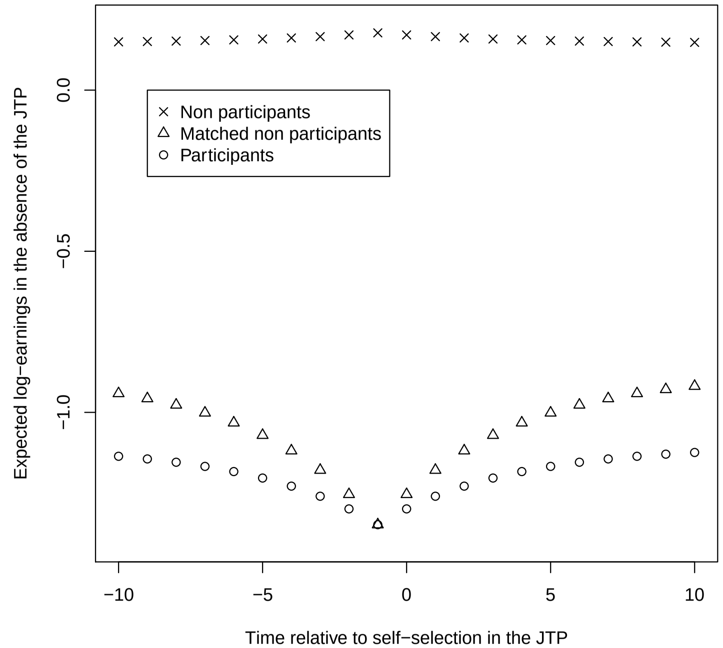
Figure 1 illustrates these results. It shows the the expected value of the outcomes in the absence of the treatment around the treatment date for the participants (circles), the non participants (crosses) and the matched non participants (triangles), *i.e.* the non participants with the same distribution of $Y^0_{i,k-1}$ as the participants. The difference between participants and non participants measures selection bias. The difference between participants and non participants measures the bias of Matching. The difference between the bias of Matching before and after the treatment date measures the bias of DID Matching.

In Figure 1(a), selection is on transitory shocks only ($\rho_{\theta,\mu} = 0$). As expected from Theorem 1, Matching is consistent since participants and matched non participants are aligned at every period after the treatment date: the matched non participants perfectly proxy the counterfactual outcomes of the participants. DID Matching does not generate

Figure 1 – DID Matching conditioning on pre-treatment outcomes is either irrelevant or inconsistent



(a) DID Matching is irrelevant



(b) DID Matching is inconsistent

Note: this figure plots the average potential outcomes in the absence of the treatment for three groups: participants (○) ($\mathbb{E}[Y^0_{i,k+\tau}|D^\iota_{i,k} = 1]$), non participants (×) ($\mathbb{E}[Y^0_{i,k+\tau}|D^\iota_{i,k} = 0]$) and matched non participants (△), *i.e.* non participants that have the same potential outcomes at period $k-1$ as the participants ($\mathbb{E}[\mathbb{E}[Y^0_{i,k+\tau}|D^\iota_{i,k} = 0, X_i, Y_{i,k-1}]|D^\iota_{i,k} = 1]$). In panel (a), selection is due to the transitory shocks only. In panel (b), selection is due to the fixed effect and transitory shocks.

bias when Matching is consistent since pre-treatment outcomes are also aligned, despite the fact that they have not been explicitly conditioned on. This is because $Y^0_{i,k-1}$ is a sufficient statistics for selection in that case.

In Figure 1(b), selection is both on the fixed effect and on transitory shocks. As expected from Theorem 1, Matching is biased, which is apparent because participants and matched non participants are not perfectly aligned after the treatment date. DID Matching is biased because the difference between participants and matched non participants varies over time: the difference between participants and matched non participants before the treatment date does not proxy for the difference that exists after.

The intuition for this result is that, conditional on $Y^0_{i,k-1}$, the participants and the matched non participants do not have the same distribution of fixed effects and transitory shocks. The matched non participants have larger fixed effects on average: this is the reason why they do not enter the treatment. The matched non participants have the same observed pre-treatment outcomes as the participants because they experience more negative transitory shocks before eligibility is decided. Matching is biased because once eligibility has been decided, the matched non participants start diverging from the participants to their higher long run mean. DID Matching is also biased because the history of transitory shocks that makes participants and non participants similar at $k-1$ in terms of pre-treatment outcomes generates differences before that date, and these differences vary over time, making it impossible to use them to proxy for the bias of Matching.

DID conditioning on pre-treatment outcomes generates bias

When simple DID is consistent

The following theorem shows that DID is consistent if and only if selection is due to the fixed effect only:

Theorem 2 *The following two statements are equivalent:*

(i) $\forall k > 0$, $\forall \tau \geq 0$, $\forall \tau' > 0$, $B(DID_{k,\tau,\tau'}) = 0$

(ii) $\gamma = 0$.

Theorem 2 shows that DID is consistent if and only if $\gamma = 0$. Theorem 1 shows that Matching and DID Matching are consistent if and only if $\rho_{\theta,\mu} = 0$. DID, Matching and DID Matching are all consistent if and only if $\gamma = \rho_{\theta,\mu} = 0$, *i.e.* in the trivial case when there is no selection bias. When there is selection bias, both Matching and DID Matching are inconsistent when DID is consistent (*i.e.* when $\gamma = 0$ and $\rho_{\theta,\mu} \neq 0$). Conditioning on pre-treatment outcomes generates bias when simple DID is consistent.[2]

Figure 2 illustrates this result. Figure 2(a) shows that DID is consistent when selection is on the fixed effect only ($\gamma = 0$). Indeed, selection bias is constant over time in that case and the difference between participants and non participants at any pre-treatment date is a consistent proxy for post-treatment selection bias. Figure 2(a) also shows that Matching is biased because conditioning on $Y^0_{i,k-1}$ fails to account for selection on the fixed effect. Indeed, $Y^0_{i,k-1}$ suffers from measurement error due to the transitory shocks. Since these shocks are persistent, the bias of Matching varies over time, before and after treatment, which makes it impossible for DID Matching to capture it.

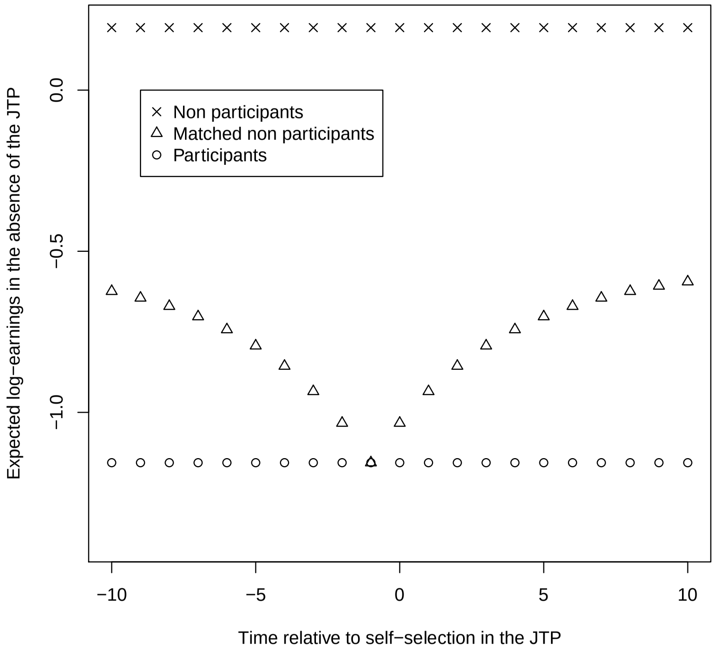
When Symmetric DID is consistent

In the model described by equation (1), DID applied symmetrically around $k-1$ is consistent when the outcome process is stationary (see e.g. on Figure 1(b)). This is because the dip due the transitory shock forms and dissipates at the same pace around $k-1$ (see Chabé-Ferret (2014)). Because the same property holds conditional on $Y^0_{i,k-1}$, DID Matching is consistent, but irrelevant, in that case.

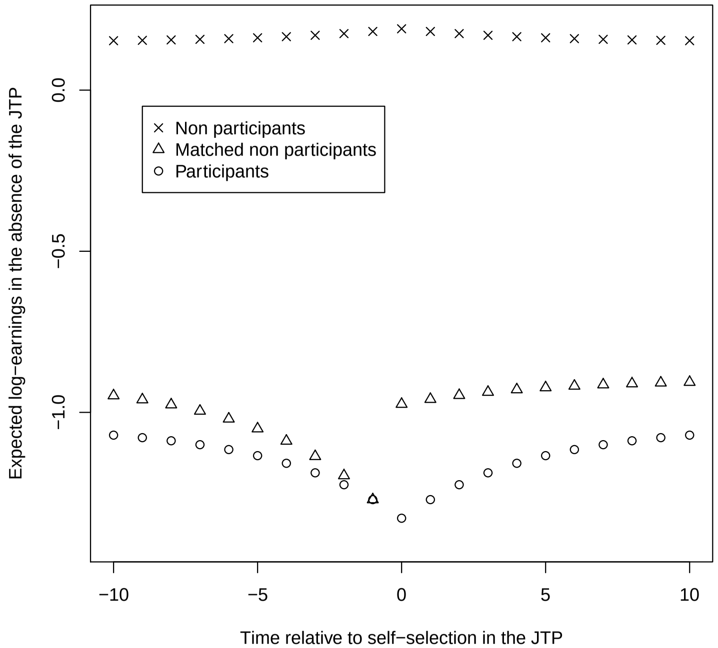
In the the self-selection model under full information studied by Chabé-Ferret (2014), DID applied symmetrically around the treatment date is consistent while Symmetric DID

[2]When $\rho = 0$, measurement error is not autocorrelated, the bias remaining after Matching is constant over time and DID Matching is consistent. But DID is also consistent making DID Matching irrelevant. Matching is biased when $\rho = 0$ since pre-treatment outcomes are an imperfect proxy for the unobserved fixed effect.

Figure 2 – DID Matching generates bias when DID and Symmetric DID are consistent



(a) DID is consistent



(b) Symmetric DID is consistent

Note: this figure plots the average potential outcomes in the absence of the treatment for three groups: participants (○) ($\mathbb{E}[Y^0_{i,k+\tau}|D^\iota_{i,k}=1]$), non participants (×) ($\mathbb{E}[Y^0_{i,k+\tau}|D^\iota_{i,k}=0]$) and matched non participants (△), *i.e.* non participants that have the same potential outcomes at period $k-1$ as the participants ($\mathbb{E}[\mathbb{E}[Y^0_{i,k+\tau}|D^\iota_{i,k}=0, X_i, Y_{i,k-1}]|D^\iota_{i,k}=1]$). In panel (a), selection is due to the fixed effect only. In panel (b), selection is due to the fixed effect and transitory shock with full information.

Matching is not. In this model, the selection equation becomes:

$$D_{i,k}^f = \mathbb{1}[t \geq k]\mathbb{1}[\underbrace{\theta_i^f + \gamma^f Y_{i,k}^0}_{D_{i,k}^{*f}} \geq 0]$$

Theorem 3 states the result:

Theorem 3 *In the self-selection model under full information, the following two statements hold when* $\sigma_{U_0}^2 = \frac{\sigma^2}{1-\rho^2}$*:*

(i) $\forall k > 0$, $\forall \tau > 0$, $B(DID_{k,\tau,\tau}) = 0$,

(ii) $\forall k > 0$, $\forall \tau > 0$, $B(DIDM_{k,\tau,1,\tau}) \neq 0$ *except if* $a\sigma_\mu = -\gamma^f \rho(\sigma_\mu^2 + \frac{\sigma^2}{1-\rho^2})$.

Theorem 3 shows that when the outcome process is stationary, Symmetric DID is consistent. Figure 2(b) illustrates this case: selection bias forms and dissipates at the same pace, it is symmetric around the treatment date k. Conditioning on $Y_{i,k-1}^0$ breaks the symmetry of the dip, and renders Symmetric DID Matching inconsistent. On Figure 2(b), the bias of Matching decreases as it gets closer to $k-1$, increases sharply at date k because the last shock before selection, and decreases thereafter. Only when the asymmetry of the dip perfectly compensates the bias due to the last shock, the bias of Symmetric DID Matching is zero, but this is not very attractive condition. Symmetric DID Matching would be consistent when conditioning on $Y_{i,k}^0$, but it is infeasible since the potential outcomes of the participants are unobserved.

4 Simulation results

In this section, I use simulations in order to gauge the size of the bias of DID conditioning on pre-treatment outcomes and how it compares with DID, especially when applied symmetrically around the treatment date and with simple Matching. I use a model of earnings dynamics and selection into a JTP calibrated with realistic parameter values (see Chabé-Ferret (2014) for details on the parameterizations and simulations).

Figure 3 presents the results of simulations of a model with no fixed effect in the outcome process. Matching does not perform well in this model because the outcome process combines an AR(1) with MA terms (see Chabé-Ferret (2014)). Figure 3(a) shows that Symmetric DID is consistent when agents have full information, as expected from Theorem 3. Figure 3(a) also shows that conditioning on pre-treatment outcomes generates severe bias (of the order of 100% of the treatment effect). This is because conditioning breaks the symmetry of the dip, as illustrated in Figure 2(b). Figure 3(a) also shows that combining DID with Matching is less bias than simple Matching. Figure 2(b) helps to understand why: the pre-treatment difference in outcomes is of the same sign as the post-treatment difference and is smaller in absolute value. Differencing after Matching goes some way in reducing selection bias.

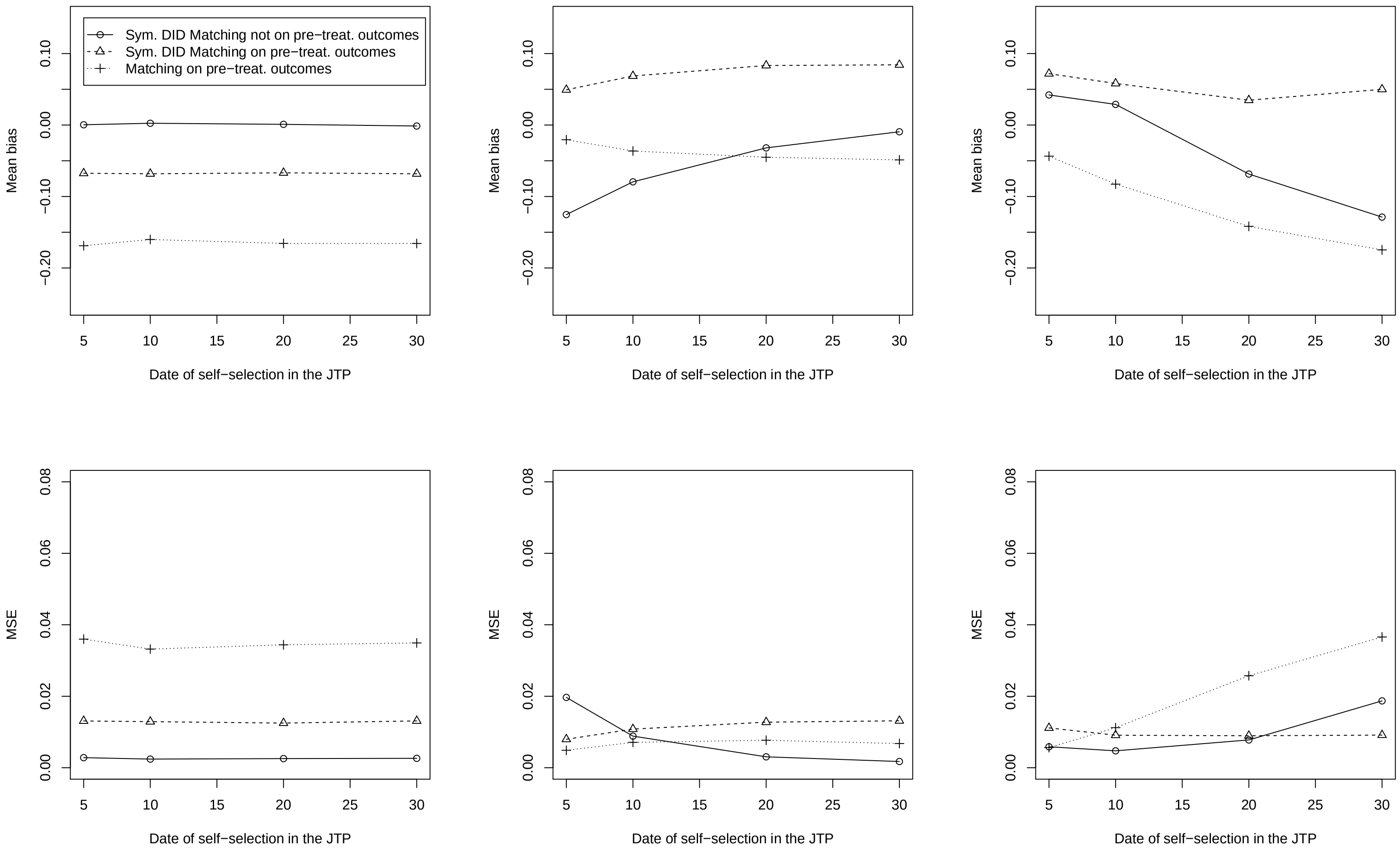
When the dip is no longer symmetric, Symmetric DID is inconsistent, whether conditioning on pre-treatment outcomes or not. This is the case for example when the outcome process is not stationary because the variance of the initial shock is smaller than the long run variance (Figure 3(b)) or when the income process exhibits an idiosyncratic trend ((Figure 3(c)). In the latter case, not conditioning fares better in the beginning and the middle of the life cycle, but performs less well when the trend dominates the bias term later in life. In the former case, not conditioning fares better in the middle and the end of the life cycle but performs less well at the beginning of the agents' career, when the dip is strongly asymmetric.

5 Revisiting experimental estimates

The results of studies comparing Symmetric DID with and without conditioning on pre-treatment outcomes to an experimental benchmark confirm the theoretical predictions that conditioning on pre-treatment outcomes increases bias.

Heckman, Ichimura, Smith, and Todd (1998) compare nonexperimental estimates of the JTPA program obtained with Matching and DID Matching to the experimental benchmark, making use of the random allocation of the program. They implement DID symmetrically

Figure 3 – Simulations of the bias and MSE of Symmetric DID and Matching in a realistic selection model

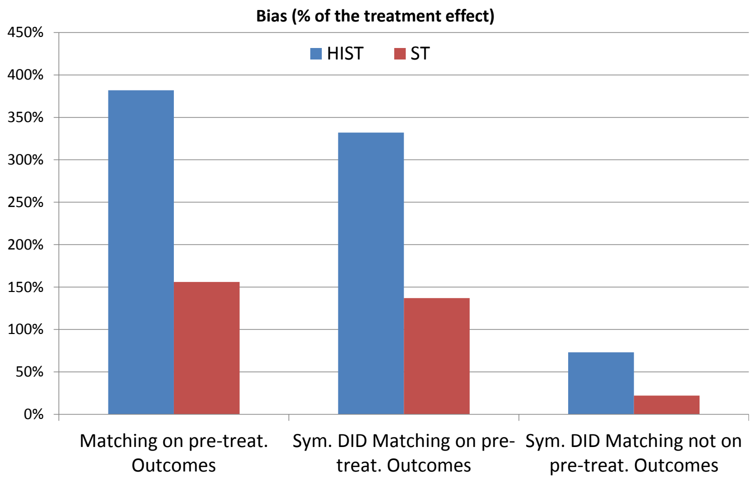


(a) Symmetric dip (b) Asymmetric dip – no trend (c) Asymmetric dip – trend

Note: the bias of both estimators is estimated $\tau = 4$ periods after the JTP. Matching conditions on $(E_i, \left\{Y_{i,k-j}\right\}_{j=1}^{d})$, for $d \in \{1, 2, 3\}$ and $k \in \{5, 10, 20, 30\}$. Symmetric DID conditions on E_i. The mean bias and the mean squared error (MSE) are calculated thanks to 500 Monte-Carlo replications. Each sample contains 1000 individuals with roughly 100 to 200 participants. The bias is estimated using LLR Matching on the propensity score with a biweight kernel. The bandwidth is set to .15 and the trimming level is set to .4. The parameterization of the wage process is presented in Table **??**. The full parameterization of the model is presented in Table **??** in appendix **??**. For the "long run" (resp. "short run") simulations, the initial conditions of the earnings process are drawn from the ergodic limit distribution (resp. from the distribution of the idiosyncratic shock $v_{i,t}$).

around the treatment date. They vary the set of control variables when assessing the performances of DID Matching. With a coarse set of predictors (only variables that are constant over time like age schooling and marital status), the bias of Symmetric DID (resp. Matching) is equal to 73% (resp. 670%) of the experimental treatment effect. When including pre-treatment earnings (model PII), the bias of Symmetric DID (resp. Matching) worsens (resp. improves) and equals 332% (resp. 382%) of the treatment effect. So, conditioning on pre-treatment earnings increases the bias of Symmetric DID Matching.

Figure 4 – Symmetric DID Matching not conditioning on pre-treatment outcomes reproduces the results of RCTs the best



Note: the figure presents the bias of various estimators estimated using randomly allocated JTPs. HIST stand for Heckman, Ichimura, Smith, and Todd (1998) and ST for Smith and Todd (2005). The results of the bias of Matching and DID Matching from HIST are from their Table XIII on p.1062. The coarse set of predictors does not condition on pre-treatment earnings while the set PII does. The results of the bias of Matching from ST are from their Table 5 p.336 and the bias of DID Matching is from their Table 6 p.340. The LaLonde set of predictors does not contain pre-treatment earnings while the DW set does. The sampe is the full LaLonde sample. The Matching estimator used for the comparisons is the local linear Matching with a small bandwidth (1.0).

Smith and Todd (2005) compare the ability of Matching and DID Matching to reproduce the results of the famous NSW experiment already analyzed by LaLonde (1986). They apply DID roughly symmetrically around the treatment date, since outcomes are measured in 1975 and 1978, and treatment allocated between 1976 and 1977. They vary the set of control variables when assessing the performances of Matching and DID Matching. With a coarse set of controls, the bias of DID Matching (resp. Matching) is of -2%, 22% and -16%

(resp. -405%, -402%, -388%) of the experimental treatment effect, with the smaller (and most efficient) bandwidth. On the same sample but with a larger set of controls, including pre-treatment outcomes, the bias of DID Matching (resp. Matching) is of -105%, -137% and -137% (resp. -95%, -156%, -159%). Again, conditioning on pre-treatment earnings improves the Matching estimates but increases the bias of DID Matching.

6 Conclusion

This paper shows that the hope of correcting for selection bias due to a combination of a fixed effect and transitory shocks by combining DID with conditioning on pre-treatment outcomes is unfulfilled. Theoretically, combining DID with conditioning on pre-treatment outcomes is consistent only when differencing is irrelevant. Worse, it generates bias when DID is consistent. Performing DID symmetrically around the treatment date allows to solve for selection bias combining a fixed effect with transitory shocks when the outcome process is stationary. Simulations using earnings dynamics and selection into a JTP as an example show that Symmetric DID outperforms DID conditioning on pre-treatment outcomes even when the outcome process is not stationary and both estimators are inconsistent. Empirically, Symmetric DID performs better at reproducing experimental results when not conditioning on pre-treatment outcomes.

Combined with those of Chabé-Ferret (2014), that shows that Symmetric DID dominates Matching on pre-treatment outcomes in a similar setting, these results push for the use of Symmetric DID for solving selection bias that is a combination of a fixed effect and transitory shocks. This estimator performs well both theoretically, in simulations and experimentally in situations in which a dips forms and dissipates around the treatment date.

Directions for further research include a more general investigation of the dgps under which Symmetric DID is consistent (or works well) and the identification of estimation procedures valid when the outcome process is non stationary.

References

Abadie, A. (2005): "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72(1), 1–19.

Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, 60(1), 47–57.

Ashenfelter, O., and D. Card (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *The Review of Economic Statistics*, 67(4), 648–660.

Bassi, L. J. (1984): "Estimating the Effect of Training Programs with Non-Random Selection," *The Review of Economics and Statistics*, 66(1), 36–43.

Blundell, R., and M. Costa Dias (2000): "Evaluation Methods for Non-Experimental Data," *Fiscal Studies*, 21(4), 427–468.

Chabé-Ferret, S. (2014): "Analysis of the Bias of Matching and Difference-In-Difference under Alternative Earnings and Selection Processes," Mimeo.

Heckman, J. J. (1978): "Longitudinal Studies in Labor Economics: A Methodological Review," University of Chicago.

Heckman, J. J., H. Ichimura, J. A. Smith, and P. E. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1099.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64(4, Special Issue: Evaluation of Training and Other Social Programmes), 605–654.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, ed.

by O. C. Ashenfelter, and D. Card, vol. 3, chap. 31, pp. 1865–2097. Elsevier, North Holland.

Heckman, J. J., and R. Robb (1985): “Alternative Methods for Evaluating the Impact of Interventions,” in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman, and B. Singer, pp. 156–245. Cambridge University Press, New-York.

LaLonde, R. J. (1986): “Evaluating the Econometric Evaluation of Training Programs with Experimental Data,” *American Economic Review*, 76, 604–620.

Nickell, S. (1981): “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, 49(6), 1417–1426.

Smith, J. A., and P. E. Todd (2005): “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?,” *Journal of Econometrics*, 125(1-2), 305–353.

Yule, G. U. (1899): “An Investigation into the Causes of Changes in Pauperism in England, Chiefly During the Last Two Intercensal Decades (Part I.),” *Journal of the Royal Statistical Society*, 62(2), 249–295.

A Proofs

Proof of Theorem 1

The following lemma proves that Matching is consistent if and only if selection is due to the transitory shock only.

Lemma 1 *$\forall k > 0$, $\forall \tau \geq 0$, $B(M_{k,\tau,1}) = 0 \Leftrightarrow \rho_{\theta,\mu} = 0$.*

Proof: First preliminary result: $\forall k > 0$, $\forall \tau \geq 0$, $B(M_{k,\tau,1}) = 0 \Leftrightarrow \sigma_{Y_{k+\tau},D^*_k}\sigma^2_{Y_{k-1}} - \sigma_{Y_{k-1},D^*_k}\sigma_{Y_{k-1},Y_{k+\tau}}$. By linearity of conditional expectations:

$$\mathbb{E}[Y^0_{i,t}|D^*_{i,k}, Y^0_{k-1}] = \mathbb{E}[Y^0_{i,t}] + \theta_{Y^0_{k+\tau},D^*_k}\left(D^*_{i,k} - \mathbb{E}[D^*_{i,k}]\right) + \theta_{Y^0_{k+\tau},Y^0_{k-1}}\left(Y^0_{i,k-1} - \mathbb{E}[Y^0_{i,k-1}]\right),$$

with $\theta_{Y^0_{k+\tau},D^*_k} = \frac{\overbrace{\sigma_{Y_{k+\tau},D^*_k}\sigma^2_{Y_{k-1}} - \sigma_{Y_{k-1},D^*_k}\sigma_{Y_{k-1},Y_{k+\tau}}}^{\mathrm{num}_{k,\tau}}}{\sigma^2_{D^*_k}\sigma^2_{Y_{k-1}} - \sigma^2_{Y_{k-1},D^*_k}}$. As a consequence,

$$B(M_{k,\tau,1}) = \theta_{Y^0_{k+\tau},D^*_k}\mathbb{E}[\mathbb{E}[D^*_{i,k}|D_{i,k} = 1, Y^0_{k-1}] - \mathbb{E}[D^*_{i,k}|D_{i,k} = 0, Y^0_{k-1}]|D_{i,k} = 1].$$

The result follows because $\mathbb{E}[D^*_{i,k}|D_{i,k}=1,Y^0_{k-1}]-\mathbb{E}[D^*_{i,k}|D_{i,k}=0,Y^0_{k-1}]>0$ and $\sigma^2_{D^*_k}\sigma^2_{Y_{k-1}}-\sigma^2_{Y_{k-1},D^*_k}>0$.

Second preliminary result: $\forall k>0$, $\forall\tau\geq 0$, $\text{num}_{k,\tau}=\sigma^2_{U_{k-1}}\sigma_\mu a(1-\rho^{\tau+1})$, with $a=\rho_{\theta,\mu}\sigma_\theta$ and $b=a+\gamma\sigma_\mu$.

$$\begin{aligned}
\text{num}_{k,\tau} &= \left[b\sigma_\mu+\gamma\rho^{\tau+1}\sigma^2_{U_{k-1}}\right]\left[\sigma^2_\mu+\sigma^2_{U_{k-1}}\right]-\left[b\sigma_\mu+\gamma\sigma^2_{U_{k-1}}\right]\left[\sigma^2_\mu+\rho^{\tau+1}\sigma^2_{U_{k-1}}\right]\\
&=\gamma\rho^{\tau+1}\sigma^4_{U_{k-1}}-\gamma\rho^{\tau+1}\sigma^4_{U_{k-1}}\\
&\quad+\sigma^2_{U_{k-1}}\left[\gamma\rho^{\tau+1}\sigma^2_\mu+b\sigma_\mu-b\sigma_\mu\rho^{\tau+1}-\gamma\sigma^2_\mu\right]\\
&\quad+\sigma^2_\mu\left[b\sigma_\mu-b\sigma_\mu\right]\\
&=\sigma^2_{U_{k-1}}\sigma_\mu(b-\gamma\sigma_\mu)(1-\rho^{\tau+1})
\end{aligned}$$

From this, we have $\rho_{\theta,\mu}=0\Rightarrow\text{num}_{k,\tau}=0$. The reciprocal follows from the fact that $\sigma_\mu>0$, $\sigma_\theta>0$, $\sigma^2_{U_{k-1}}>0$, $\forall k>0$ and $(1-\rho^{\tau+1})>0$, $\forall\tau\geq 0$. Thus $\text{num}_{k,\tau}=0\Rightarrow\rho_{\theta,\mu}=0$. ■

The following lemma proves the main result: DIDM is consistent if and only if selection is only on past outcomes.

Lemma 2 *$\forall k>0$, $\forall\tau\geq 0$, $\forall\tau'>0$, $B(DIDM_{k,\tau,1,\tau'})=0\Leftrightarrow\rho_{\theta,\mu}=0$.*

PROOF: First preliminary result: $\forall k>0$, $\forall\tau\geq 0$, $\forall\tau'>0$, $B(DIDM_{k,\tau,1,\tau'})=0\Leftrightarrow\text{num}_{k,\tau}-\text{num}_{k,-\tau'}=0$. This stems from the proof of Lemma 1.

Second preliminary result: $\text{num}_{k,-\tau'}=\sigma_\mu a(\sigma^2_{U_{k-1}}-\rho^{\tau'-1}\sigma^2_{U_{k-\tau'}})$.

$$\begin{aligned}
\text{num}_{k,-\tau'} &= \left[b\sigma_\mu+\gamma\rho^{\tau'-1}\sigma^2_{U_{k-\tau'}}\right]\left[\sigma^2_\mu+\sigma^2_{U_{k-1}}\right]-\left[b\sigma_\mu+\gamma\sigma^2_{U_{k-1}}\right]\left[\sigma^2_\mu+\rho^{\tau'-1}\sigma^2_{U_{k-\tau'}}\right]\\
&=\gamma\rho^{\tau'-1}\sigma^4_{U_{k-\tau'}}-\gamma\rho^{\tau'-1}\sigma^4_{U_{k-\tau'}}\\
&\quad+\sigma^2_{U_{k-1}}\sigma_\mu\left[b-\gamma\sigma_\mu\right]-\rho^{\tau'-1}\sigma^2_{U_{k-\tau'}}\sigma_\mu\left[b-\gamma\sigma_\mu\right]\\
&\quad+\sigma^2_\mu\left[b\sigma_\mu-b\sigma_\mu\right]\\
&=\sigma_\mu(b-\gamma\sigma_\mu)(\sigma^2_{U_{k-1}}-\rho^{\tau'-1}\sigma^2_{U_{k-\tau'}})
\end{aligned}$$

Third preliminary result: $\sigma^2_{U_t}=\frac{1-\rho^{2t}}{1-\rho^2}\sigma^2+\rho^{2t}\sigma^2_{U_0}$.

We now have that:

$$\begin{aligned}
\text{num}_{k,\tau}-\text{num}_{k,-\tau'} &= \sigma_\mu a(\rho^{\tau'-1}\sigma^2_{U_{k-\tau'}}-\rho^{\tau+1}\sigma^2_{U_{k-1}})\\
&=\sigma_\mu a\left(\underbrace{(\rho^{\tau'-1}-\rho^{\tau+1})\frac{\sigma^2}{1-\rho^2}}_{B(\tau,\tau')}+\rho^{2(k-\tau')}\underbrace{\left(\sigma^2_{U_0}-\frac{\sigma^2}{1-\rho^2}\right)(\rho^{\tau'-1}-\rho^{\tau+1}\rho^{2(\tau'-1)})}_{C(\tau,\tau')}\right)
\end{aligned}$$

From this, we have that $\rho_{\theta,\mu}=0\Rightarrow B(DIDM_{k,\tau,1,\tau'})=0$.

Since $\sigma^2_\mu>0$ and $\sigma^2_\theta>0$, $\forall k>0$, $\forall\tau\geq 0$, $\forall\tau'>0$, $B(DIDM_{k,\tau,1,\tau'})=0\Rightarrow$ either $\rho_{\theta,\mu}=0$ or $\forall k>0$, $\forall\tau\geq 0$, $\forall\tau'>0$, $A(k,\tau,\tau')=B(\tau,\tau')+\rho^{2(k-\tau')}C(\tau,\tau')=0$. To prove the final result, it remains to be shown that the second condition on $A(k,\tau,\tau')$ is not fulfilled in the model. Let's assume that the result holds and show that this yields to a contradiction. Fix τ and τ' such that $\tau + 2 \neq \tau'$. $A(k, \tau, \tau')$ as a function of k has at most one real root as long as $B(\tau, \tau') \neq 0$ or $C(\tau, \tau') \neq 0$. So $A(k, \tau, \tau') = 0$, $\forall k > 0$ $\Rightarrow B(\tau, \tau') = 0$ and $C(\tau, \tau') = 0$. But $B(\tau, \tau') = 0 \Rightarrow \rho = 0$ or $\tau + 2 = \tau'$ or $\sigma^2 = 0$, a contradiction. This proves the result. ■

Proof of Theorem 2

PROOF: First preliminary result: $\forall k > 0$, $\forall \tau \geq 0$, $\forall \tau' > 0$, $B(DID_{k,\tau,\tau'}) = 0 \Leftrightarrow \text{Cov}(Y^0_{i,k+\tau}, D^*_{i,k}) - \text{Cov}(Y^0_{i,k-\tau'}, D^*_{i,k}) = 0$.

$$\mathbb{E}[Y^0_{i,t}|D^*_{i,k}] = \mathbb{E}[Y^0_{i,t}] + \frac{\text{Cov}(Y^0_{i,k+\tau}, D^*_{i,k})}{\text{Var}(D^*_{i,k})}\left(D^*_{i,k} - \mathbb{E}[D^*_{i,k}]\right)$$
$$\mathbb{E}[Y^0_{i,t}|D_{i,k} = 1] = \mathbb{E}[Y^0_{i,t}] + \frac{\text{Cov}(Y^0_{i,k+\tau}, D^*_{i,k})}{\text{Var}(D^*_{i,k})}\left(\mathbb{E}[D^*_{i,k}|D^*_{i,k} \geq 0] - \mathbb{E}[D^*_{i,k}]\right)$$
$$\mathbb{CD}[Y^0_{i,t}] = \frac{\text{Cov}(Y^0_{i,k+\tau}, D^*_{i,k})}{\text{Var}(D^*_{i,k})}\left(\mathbb{E}[D^*_{i,k}|D^*_{i,k} \geq 0] - \mathbb{E}[D^*_{i,k}|D^*_{i,k} < 0]\right)$$

The result follows because $\mathbb{E}[D^*_{i,k}|D^*_{i,k} \geq 0] - \mathbb{E}[D^*_{i,k}|D^*_{i,k} < 0] > 0$.

Second preliminary result: $\text{Cov}(Y^0_{i,k+\tau}, D^*_{i,k}) - \text{Cov}(Y^0_{i,k-\tau'}, D^*_{i,k}) = \gamma(\rho^{\tau+1}\sigma^2_{U_{k-1}} - \rho^{\tau'-1}\sigma^2_{U_{k-\tau'}})$.

$$\begin{aligned}
\text{Cov}(Y^0_{i,t}, D^*_{i,k}) &= \text{Cov}(\mu_i + U_{i,t}, \theta_i + \gamma\mu_i + \gamma U_{i,k-1}) \\
&= b\sigma_\mu + \gamma\text{Cov}(U_{i,t}, U_{i,k-1}) \\
&= b\sigma_\mu + \gamma\rho^{|t-k+1|}\sigma^2_{U_{\min\{t,k-1\}}}
\end{aligned}$$

Third preliminary result:

$$\text{Cov}(Y^0_{i,k+\tau}, D^*_{i,k}) - \text{Cov}(Y^0_{i,k-\tau'}, D^*_{i,k}) = -\gamma A(k, \tau, \tau')$$

From this, we have $\gamma = 0 \Rightarrow B(DID_{k,\tau,\tau'}) = 0$. Moreover, $\forall k > 0$, $\forall \tau \geq 0$, $\forall \tau' > 0$, $B(DID_{k,\tau,\tau'}) = 0 \Rightarrow \gamma = 0$ or $\forall k > 0$, $\forall \tau \geq 0$, $\forall \tau' > 0$, $A(k, \tau, \tau') = 0$. The same reasoning as in the proof of Lemma 2 shows that the second condition on $A(k, \tau, \tau')$ is not fulfilled in the model. This proves the result. ■

Proof of Theorem 3

PROOF: Using the same line of reasoning as the proof of Theorem 2, but modifying it accordingly, yields the following result: $\forall k > 0$, $\forall \tau \geq 0$, $\forall \tau' > 0$, $B(DID_{k,\tau,\tau'}) = 0 \Rightarrow -\gamma^f A^f(k, \tau, \tau')$, with:

$$A^f(k, \tau, \tau') = \left((\rho^{\tau'} - \rho^\tau)\frac{\sigma^2}{1 - \rho^2} + \rho^{2(k-\tau')}\left(\sigma^2_{U_0} - \frac{\sigma^2}{1 - \rho^2}\right)(\rho^{\tau'} - \rho^\tau\rho^{2\tau'})\right).$$

This proves the consistency of Symmetric DID when $\sigma^2_{U_0} = \frac{\sigma^2}{1-\rho^2}$.

In order to derive the bias of DIDM, it is useful to rewrite $D^{*f}_{i,k}$ as a function of $D^{*}_{i,k}$:

$$\begin{aligned} D^{*f}_{i,k} &= \theta^f_i + \gamma^f Y^0_{i,k} \\ &= \underbrace{\theta^f_i + \gamma^f \mu_i(1-\rho)}_{\theta_i} + \underbrace{\gamma^f \rho}_{\gamma} Y^0_{i,k-1} + \gamma^f v_{i,k} \\ &= D^*_{i,k} + \gamma^f v_{i,k}. \end{aligned}$$

Following the line of the proof of Theorem 1, we have that $B(DID_{k,\tau,\tau'}) = 0 \Leftrightarrow a\sigma_\mu A(k,\tau,\tau') + \gamma^f \rho^\tau \sigma^2(\sigma^2_\mu + \sigma^2_{U_{k-1}})$. When $\sigma^2_{U_0} = \frac{\sigma^2}{1-\rho^2}$, we have $B(DID_{k,\tau,\tau}) = 0 \Leftrightarrow \rho^{\tau-1}\sigma^2 \left(a\sigma_\mu + \gamma^f \rho(\sigma^2_\mu + \frac{\sigma^2}{1-\rho^2})\right)$. This proves the result. ■