To Control or Not to Control? Bias of Simple Matching vs Difference-In-Difference Matching in a Dynamic Framework*

Sylvain Chabé-Ferret
Cemagref, UMR Métafort, Aubière, France

This version: July 15, 2010.

Abstract
In this paper, I examine whether the claim “the more control variables, the better” holds when matching in a dynamic context. I exhibit three situations where it is better to difference with respect to past outcomes rather than to control for them: DID-matching is unbiased whereas matching on past outcomes is biased. I also study the special case of evaluating a job training program, borrowing a credible selection rule from [Heckman, LaLonde, and Smith (1999)] and relying on the parameters of the wage process estimated by [MaCurdy (1982)]. I derive closed forms for the bias terms of the two estimators when the error terms are normally distributed. I show that DID matching is unbiased when applied symmetrically around the period of enrollment, as implemented by [Heckman, Ichimura, Smith, and Todd (1998)]. DID-matching is more robust than matching to misspecification both of the amount of information individuals have when deciding to enter the program and of the control period. I finally point to previously unnoticed experimental results that confirm the claim that DID-matching is less biased than matching.

Keywords: Matching - Difference in Difference Matching - Evaluation of Job training Programs.

JEL codes: C21, C23.

Correspondence to:
Sylvain Chabé-Ferret
Toulouse School of Economics, LERNA
21 Allée de Brienne
31000 Toulouse, France.
Email: sylvain.chabe-ferret@tse-fr.eu
Tel: +33 (0) 5 61 12 88 28.
Fax:+33 (0) 5 61 12 85 20.

*This paper has been written as I was visiting Cowles Foundation for the Research in Economics at Yale University. I thank Don Andrews and Philip Haile for their invitation and Cemagref for its financial support. I thank Joe Altonji, Sidd Chib, Jim Heckman, Stefan Hoderlein, Fabian Lange, Yoonseok Lee, Julie Subervie, Petra Todd, Neese Yildiz, Ed Vytlacil and seminar participants at Yale and UPenn for their valuable comments and suggestions on this research. All remaining errors are my own.
1 Introduction

Matching is an econometric estimator widely used to estimate treatment effects. Its validity rests on the assumption that all selection bias is due to variables that are observed by the analyst.\(^1\) When considering the practical implementation of matching methods, it is widely believed that the more pre-treatment variables one can control for, the lower the bias of the estimated treatment effect.\(^2\) In a theoretical analysis, Heckman and Navarro-Lozano (2004) show that this statement does not hold in general: controlling for more variables may increase or decrease bias, depending on the correlation between the added control variable and the remaining unobserved covariates. But in a practical implementation, this statement has been proved to be true by Dehejia and Wahba (1999, 2002) in the context of the evaluation of job training programs. In their re-analysis of LaLonde (1986)’s critique, these authors show that controlling for past outcomes goes a long way in reducing the bias of propensity score matching with respect to an experimental benchmark.\(^3\) Heckman, Ichimura, Smith, and Todd (1998) also show that adding past wages greatly decreases the bias of propensity score matching estimators. These results are coherent with the widely known fact that participants into a job training program experience a decrease in their earnings prior to entering the program, an empirical regularity that has been termed “Ashenfelter’s dip” after Ashenfelter (1978). If Ashenfelter’s dip is the consequence of persistent shocks leading people to enter the program, not controlling for pre-treatment outcomes may lead to underestimating the effect of the program, because earnings after entering the program would have been lower for the treated than for the untreated.

The aim of this paper is to investigate both the theoretical and practical validity of the statement “the more control variables, the better” in the case of DID-matching, with special emphasis on the role of pre-treatment outcomes when evaluating the effect of job training programs on earnings. Difference-In-Difference (DID) matching has been proposed by Heckman, Ichimura, Smith, and Todd (1998); Heckman, Ichimura, and Todd (1997) as an extension of matching robust to selection on unobservables. Taking the first difference of outcomes before

\(^1\)This assumption has been called selection on observables (Heckman and Robb, 1985), conditional exogeneity (Imbens, 2004) or ignorability (Rubin, 1978).


\(^3\)Smith and Todd (2005) argue that this is the special structure of Dehejia and Wahba (1999, 2002)’s sample that makes the most part in the bias reduction.
and after treatment gets rid of unobserved fixed time and individual effects that may be correlated to both treatment and outcomes. This estimator has been found to be the closest to the experimental benchmark by Heckman, Ichimura, and Todd (1997) and Heckman, Ichimura, Smith, and Todd (1998) in their analysis of JTPA and by Smith and Todd (2005) in their re-re-analysis of the NSW experiment studied by LaLonde (1986) and Dehejia and Wahba (1999, 2002). As Abadie (2005) points out, DID-matching can also be viewed as a nonparametric extension to simple DID: the parallel trend (or independent increments) assumption justifying DID can be replaced by a weaker version conditional on observed covariates. With DID matching, individuals with different observed characteristics may experience different increments. Abadie (2005) cites past outcomes as a case in point: because of Ashenfelter’s dip, simple DID may confound the fading out of past shocks with the effect of the treatment, leading to an upward bias on the estimated treatment effect. Applying DID conditional on past outcomes seems to be able to solve the bias due to Ashenfelter’s dip by including past outcomes in the set of control variables and at the same time to be able to get rid of the remaining selection on unobservables by differencing post-treatment outcomes with respect to pre-treatment outcomes. Note that there is a previously unnoticed contradiction here: controlling on past outcomes, both matching and DID-matching converge to the same quantity in large samples:

\[
DID_{k,k'}(x,y) = \mathbb{E}[Y_k - Y_{k'} | D_k = 1, X = x, Y_{k'} = y] \\
- \mathbb{E}[Y_k - Y_{k'} | D_k = 0, X = x, Y_{k'} = y] \\
= \mathbb{E}[Y_k | D_k = 1, X = x, Y_{k'} = y] - \mathbb{E}[Y_k | D_k = 0, X = x, Y_{k'} = y] \\
- \left( \mathbb{E}[Y_{k'} | D_k = 1, X = x, Y_{k'} = y] - \mathbb{E}[Y_{k'} | D_k = 0, X = x, Y_{k'} = y] \right) \\
= y - y = 0 \\
= M_{k,k'}(x,y),
\]

where \( k \) (resp. \( k' \)) is a post- (resp. pre-) treatment period, \( Y_t \) is the observed outcome of interest in period \( t \), \( X \) is a set of control variables and \( D_k \) is equal to one when treated in period \( k \) and zero otherwise and no-one is treated in period \( k' \). In view of this result, we can restate the aim of this paper: when observing pre-treatment outcomes, is it better to control for them or to difference them out?

The main result of this paper is that in three credible cases, making treated and untreated
individuals completely similar in all observed dimensions results in a biased estimate of the treatment effect whereas not controlling (but differencing) leads to an unbiased inference. The three cases exhibited in this paper have in common the assumption justifying DID, i.e. additive separability of individual fixed effects. In the first case, there is no dip (i.e. selection does not depend on past outcomes), but there is strong persistence in the outcome process. If the autoregressive process is linear, innovations in outcomes are independent of participation. In the second case, Ashenfelter’s dip is supposed to be transient: innovations to outcomes are not persistent, so that differencing with respect to a period prior to selection recovers an unbiased estimate of selection bias due to unobserved fixed effects. Finally, in the third case, I allow for both a persistent Ashenfelter’s dip and selection on past outcomes. Under additive linearity of the selection and outcome equations and normality of the error terms, I prove that DID applied symmetrically around the date of treatment is unbiased, generalizing a result of Ashenfelter and Card (1985) and Heckman and Robb (1985) to selection rules including fixed effects.

The intuition for these results is that in the presence of unobserved individual characteristics fixed through time and correlated with entry into the treatment, treated and untreated individuals with the same pre-treatment outcomes would have different post-treatment outcomes in the absence of the treatment, because dynamic behavior depends on unobserved heterogeneity. Another way of understanding this result is that first-differencing is not guaranteed to eliminate all the effects of unobserved individual fixed effects, depending on the covariates that are included as control variables. If some control variables are correlated to the individual fixed effects, simple intuitions about how DID “differences out” these fixed effects are not valid.

In this paper, I also provide closed form formulas for bias for DID-matching controlling and not controlling on pre-treatment outcomes in a realistic toy-model of participation into a job training program. I borrow the description and parameter values for the earnings dynamics from MaCurdy (1982) and a model of participation into a job training program from Heckman and Robb (1985) and Heckman, LaLonde, and Smith (1999). Equipped with this toy model, I study how bias of matching and DID-matching varies with the amount of information the individual has on her wage process and the number of periods between observation of pre-treatment outcomes and treatment. Some interesting results appear: even when controlling for outcomes just before entry, matching is more biased than DID-matching. Symmetric DID-
matching is unbiased, but this property rests on the assumption that the agent knows the value of the shock to her earnings at the date of treatment. If this property is not fulfilled, matching on the last period before treatment is superior to DID-matching with respect to that same period but not to symmetric DID-matching. Symmetric DID-matching thus outperforms the best matching estimate, although MaCurdy (1982)'s estimates imply that the variance of the unobserved fixed effects is zero.

Finally, I point to previously unnoticed results in Heckman, Ichimura, Smith, and Todd (1998) and Smith and Todd (2005) showing that the theoretical point made in this paper (namely that it is sometimes better to differentiate rather than to control for past outcomes) is validated in empirical applications. In these papers, among all DID-matching estimators (which are already the closest to the experimental benchmark), the ones without past outcomes as control variables are closer to the experimental benchmark, thereby confirming the main point I make in this paper.

The results in this paper have consequences for the practical utilization of matching as an empirical tool. First, when panel data on outcomes before the program exist, researchers may want to compute estimates of treatment effects with both matching and DID-matching. If these two estimates do differ, the source of discrepancy must be investigated. A second important result in this paper is that if unobserved heterogeneity is believed to be an important component of both outcomes and selection, DID-matching, under the maintained additive separability assumption, performs better than simple matching on past outcomes. Third, if the outcome process is characterized by low autocorrelation of the idiosyncratic time-varying shocks, DID-matching performs better. Fourth, when transient shocks are autocorrelated and selection is based on the outcomes at one particular period (the selection period), applying DID-matching symmetrically around the date of treatment is a way to minimize bias. The importance of individual fixed effects and of the auto-correlation of time-varying shocks may vary depending on the context, but a useful exercise for applied researchers could be to frame the evaluation problem with a credible selection equation and to evaluate the way each of the candidate estimators would behave with a credible dynamic process for potential outcomes.

This paper is organized as follows: in section 2 I study the conditions under which DID-matching with time-varying covariates that are not past outcomes is unbiased; in section 3
I present three general cases in which controlling for past outcomes leads to biased estimates whereas not doing so preserves unbiasedness; in section 4 I illustrate these findings by studying bias of matching and DID-matching using MaCurdy (1982)’s estimates of the dynamics of the earnings process to calibrate a toy model of earnings and entry into a job training program; in section 5 I provide a discussion of these results and direction for further research.

2 DID-matching with time-varying covariates: when it is better to control

In this section, I first state the case for DID-matching: under additive separability of the unobserved fixed effect in the outcome equation, DID-matching can recover the average treatment effect on the treated even when selection is on the unobservable fixed effect. Second, I prove that the validity of DID-matching may break down when time-varying control variables are correlated to the unobserved fixed effect. Finally, I show that for time varying control variables that are not past outcomes, controlling for all observed periods seems the best way to curb bias. I build on Ashenfelter and Card (1985)’s simple setting. By progressively extending their simple framework, the main problems appear and a series of results can be stated.

2.1 The case for DID-matching with covariates fixed through time

As in Ashenfelter and Card (1985), write the observed outcome as: \( Y_{it} = \mu_i + \delta_t + \beta D_{it} + \epsilon_{it} \), where \( \mu_i \) and \( \delta_t \) are unobserved respectively individual and time fixed effects, \( \epsilon_{it} \) is an i.i.d. shock independent of \( \mu_i \) and \( \beta \) is a fixed parameter. Participation in the program whose average effect we would like to measure is indicated by \( D_{it} \). Participation is decided in period \( k \) according to the following rule: \( D_{ik} = 1 [\mu_i + u_{ik} \leq 0] \), where \( u_{ik} \) is an i.i.d. shock independent from all other variables in the model. Before period \( k \), no-one gets treated \( (D_{it} = 0, \ t < k) \) and after getting treated in period \( k \), people are considered treated forever: \( D_{it} = D_{ik}, \ t \geq k \). In this simple setting, selection is on the permanent unobserved component of earnings and only individuals with low lifetime income enter the program.

In this model, the simple difference-in-difference (DID) estimator comparing increments \( \tau' \) periods before and \( \tau \) periods after treatment recovers the treatment on the treated parameter:
\[ E[Y_{i,k+\tau} - Y_{i,k-\tau'}|D_{ik} = 1] - E[Y_{i,k+\tau} - Y_{i,k-\tau'}|D_{ik} = 0] = \beta = E[Y_{i,k+\tau}^1 - Y_{i,k+\tau}^0|D_{ik} = 1], \]

where \( Y_{it}^1 \) is the value of \( Y_{it} \) if individual \( i \) receives treatment in period \( t \) and \( Y_{it}^0 \) is the value of \( Y_{it} \) if individual \( i \) does not receive the treatment in period \( t \).

The previous setting is highly restrictive in that we need to assume unconditional linear separability of both time and individual fixed effects from the treatment effect and that we make no use of potential observed control variables. A semi-parametric version of the previous argument could take the following form:

\[ Y_{it} = g(D_{it}, \delta_t, X_i, \epsilon_{it}) + \mu_i \]

with \( g \) and \( h \) unrestricted functions and \( X_i \) observed variables fixed through time and potentially correlated to \( \mu_i \) but neither to \( u_{it} \) or \( \epsilon_{it} \), for all \( t \). In this much less restrictive setting, treatment on the treated is identified by DID-matching. This is because the independent increments condition is fulfilled:

\[
E[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0|D_{ik} = d, X_i] = E[g(0, \delta_{k+\tau}, X_i, \epsilon_{i,k+\tau}) - g(0, \delta_{k-\tau'}, X_i, \epsilon_{i,k-\tau'})]
\]  

(5)

does not depend on \( d \). Together with a support condition, this restriction implies that treatment on the treated is identified:

\[
E[Y_{i,k+\tau}^1 - Y_{i,k+\tau}^0|D_{ik} = 1]
= E[E[Y_{i,k+\tau} - Y_{i,k-\tau'}|D_{ik} = 1, X_i] - E[Y_{i,k+\tau} - Y_{i,k-\tau'}|D_{ik} = 0, X_i]|D_{ik} = 1].
\]  

(6)

This identification strategy is powerful: it allows for different growth rates of the outcome variables in the treated and untreated groups, due to different observed initial conditions. It is even possible to accommodate interaction between observed variables and fixed effects and to allow the treatment effects to depend on unobservables (i.e. essential heterogeneity à la [Heckman, Urzua, and Vytlacil (2006)] can be allowed for in this setting): \( Y_{it} = g^a(D_{it}, d_t, X_i, \epsilon_{it}^a, D_{it}\mu_i^1) + g^b(X_i, \epsilon_{it}^b, D_{it}\mu_i^1 + (1 - D_{it})\mu_i^0) \) and \( D_{ik} = 1 [h(\mu_i^1, \mu_i^0, X_i, u_{ik}) \leq 0] \).

The independent increment condition is also fulfilled in this model, meaning that selection on unobserved gains to the treatment does not jeopardize the identification strategy. We have thus proved the following result:

**Result 1 (Time-unevying control variables)** DID-matching identifies treatment on the treated if control variables are fixed through time, unobserved fixed time and individual effects
are additively separable and selection bias is only due to the unobserved fixed effects terms.

Crucial to this result is the fact that \( \mu_i \) and \( \delta_t \) are linearly separable in the outcome equation. This means that DID-matching is not stable to monotonic nonlinear transformations of the outcome variable (Athey and Imbens [2006]). It is for example well-known that applying DID in logs and in levels may lead to opposite signs for the estimated treatment effects (Meyer, Viscusi, and Durbin [1995]). A second restriction of the setting presented so far is that we have only considered control variables that are fixed through time. In many applied settings, there exists potential control variables that may vary between \( k' \) and \( k \). In labor economics, age, marital status, number of children are natural candidates as time-varying control variables. In the studies of firms, past levels of capital stock and number of employees are also often included as controls. Do we really know that DID-matching still identifies treatment effects in this case? In the remaining of this section, I carefully examine whether the crucial independent increments conditions is fulfilled for two types of time-varying candidate variables: fully exogenous variables and variables correlated to the fixed effects. The case of past-outcomes is addressed in section 3.

2.2 DID-matching with exogenous time-varying covariates

In this section, I study how allowing for time-varying exogenous covariates impacts on the validity of DID-matching. The number of children is an example of a candidate variable that varies through time and may explain both wages and selection into a job training program. When considering it exogenous, we make the assumption that the number of children is independent of the unobserved fixed effect, here the ability of the individual. This is an heroic assumption, that is relaxed in the following section.

Let’s write outcome and selection as a function of time-varying exogenous covariates: \( Y_{it} = g(D_{it}, \delta_t, X_{it}, \epsilon_{it}) + \mu_i \) and \( D_{ik} = 1[ h(\mu_i, X_{ik}, u_{ik}) \leq 0] \). I write \( X_{i,k+\tau} = l(X_{ik}, \eta_{ik,k+\tau}) \): covariates in period \( k+\tau \) depend on values of the covariates in period \( k \) and on an i.i.d. shock independent of all other variables in the model. This creates scope for selection on observables: if covariates where not autocorrelated, selection depending on \( X_{ik} \) would not matter for the comparison of outcomes \( \tau \) periods ahead. Symmetrically, I also pose that there is a relation between controls at periods \( k \) and \( k - \tau' \): \( X_{ik} = m(X_{i,k-\tau'}, \eta_{ik}) \). \( X_{i,k-\tau'} \) is also assumed to be independent
from the unobserved fixed effect and all time varying shocks occurring at \( t \geq k - \tau' \). Note that treatment has no effect on covariates in either periods: this case is more akin to the past outcomes case studied in the next section.

The most natural control variable in this setting is \( X_{ik} \). It is easy to show in that case that the independent increments condition is satisfied conditional on \( X_{ik} \) and DID-matching controlling for time-varying exogenous covariates at the time of selection is unbiased:

\[
\begin{align*}
\mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 | D_{ik} = 1, X_{ik} = x] &= \mathbb{E}[g(0, \delta_{k+\tau}, X_{i,k+\tau}, \epsilon_{i,k+\tau}) - g(0, \delta_{k-\tau'}, X_{i,k-\tau'}, \epsilon_{i,k-\tau'}) | h(\mu_i, x, u_{ik}) < 0, X_{ik} = x] \\
&= \mathbb{E}[g(0, \delta_{k+\tau}, l(x, \eta_{i,k+\tau}), \epsilon_{ik}) - \mathbb{E}[g(0, \delta_{k-\tau'}, X_{i,k-\tau'}, \epsilon_{i,k-\tau'}) | m(X_{i,k-\tau'}, \eta_{ik}) = x] \\
&= \mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 | D_{ik} = 0, X_{ik} = x].
\end{align*}
\]

The first equality follows from the additive separability of the fixed effect. The last two equalities follow from the joint independence of \((\mu_i, u_{ik}, \eta_{ik}, X_{i,k-\tau'})\) from \((\eta_{i,k+\tau}, \epsilon_{ik}, \epsilon_{i,k-\tau'})\).

However, controlling only for \( X_{i,k+\tau} \) or \( X_{i,k-\tau'} \) leads to a biased estimation because the independent increments condition is not fulfilled. For example, controlling for \( X_{i,k+\tau} \):

\[
\begin{align*}
\mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 | D_{ik} = 1, X_{i,k+\tau} = x] &= \mathbb{E}[g(0, \delta_{k+\tau}, X_{i,k+\tau}, \epsilon_{i,k+\tau})] \\
&\quad - \mathbb{E}[g(0, \delta_{k-\tau'}, X_{i,k-\tau'}, \epsilon_{i,k-\tau'}) | h(\mu_i, m(X_{i,k-\tau'}, \eta_{ik}), u_{ik}) < 0, l(m(X_{i,k-\tau'}, \eta_{ik}), \eta_{i,k+\tau}) = x] \\
&\neq \mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 | D_{ik} = 0, X_{i,k+\tau} = x].
\end{align*}
\]

Conditional on \( X_{i,k+\tau} \), the independent increment condition fails because both outcomes and participation depend on \( X_{i,k-\tau'} \), which can still vary. The intuition for this result is the following: among people with the same value for \( X_{i,k+\tau} \), the treated and untreated have a different distribution of \( X_{ik} \). This is because selection depends on \( X_{ik} \) and this variable is stochastically related to \( X_{i,k+\tau} \), but the former variable is not degenerate conditional on the latter. One way around this problem would be to control for both \( X_{i,k-\tau'} \) and \( X_{i,k+\tau} \). We thus have proved the following result:

**Result 2 (Exogenous time-varying control variables)** When time-varying control vari-
ables are independent of the individual fixed effects, DID-matching identifies treatment on the
treated if the set of control variables includes the value of the covariates at the period when
treatment is decided or if it includes the values of the covariates at the pre- and post-treatment
periods when outcomes are observed.

2.3 DID-matching with endogenous time-varying covariates

In this section, I consider the more realistic case in which time-varying covariates are correlated
to unobserved individual fixed effects: \( X_{i,k+\tau} = l(X_{ik}, \mu_i, \eta_{i,k+\tau}) \). Contrary to the result in
the previous section, conditioning on \( X_{ik} \) is not enough to guarantee the independence of
increments:

\[
\begin{align*}
\mathbb{E}[ Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 | D_{ik} = 1, X_{ik} = x ] &= \mathbb{E}[ g(0, \delta_{k+\tau}, l(x, \mu_i, \eta_{i,k+\tau}), \epsilon_{i,k+\tau}) | h(\mu_i, x, u_{ik}) < 0, m(X_{i,k-\tau'}, \mu_i, \eta_{i,k}) = x ] \\
& - \mathbb{E}[ g(0, \delta_{k-\tau'}, X_{i,k-\tau'}, \epsilon_{i,k-\tau'}) | h(\mu_i, x, u_{ik}) < 0, m(X_{i,k-\tau'}, \mu_i, \eta_{i,k}) = x ] \\
& \neq \mathbb{E}[ Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 | D_{ik} = 0, X_{ik} = x ]
\end{align*}
\]

(12)

The independent increments condition is not fulfilled conditional on \( X_{ik} \) and thus result 2
breaks down when time varying control variables are endogenous. The intuition for this result
is subtle: when comparing treated and untreated people with the same value for \( X_{ik} \), we do not
compare the “same” average people: they differ in their mean level of \( \mu_i \): those who participate
do so because they have different values of the unobserved individual fixed effects than those
who do not. For example, among women with the same age and number of kids, only women
with a higher long term propensity to have children and with lower wage ability may enter
the program. It is only by accident that they have the same number of children as those who
do not enter the program. In the exogenous case, this is not a problem as all the effect of
different \( \mu_i \) is canceled out by first-differencing the outcome. But in the endogenous controls
case, different \( \mu_i \) mean different \( X_{i,k+\tau} \) and \( X_{i,k-\tau'} \). For example, women entering the program
will have more children in period \( k + \tau \) (and thus lower wages) than non-participant women
with the same number of children in period \( k \).

By a slight abuse of notation, I keep the same names in different sections for functions having different
arguments but the same role in the model.
On the other hand, controlling for time-varying covariates in both pre- and post-treatment periods restores the independent increments property and the unbiasedness of DID-matching:

\[
\begin{align*}
E\left[Y_{t,k+\tau}^0 - Y_{t,k-\tau'}^0 | D_{ik} = 1, X_{i,k+\tau} = x, X_{i,k-\tau'} = \tilde{x}\right] \\
= E[g(0, \delta_{k+\tau}, x, \epsilon_{i,k+\tau})] - E[g(0, \delta_{k-\tau'}, \tilde{x}, \epsilon_{i,k-\tau'})] \\
= E[Y_{t,k+\tau}^0 - Y_{t,k-\tau'}^0 | D_{ik} = 0, X_{i,k+\tau} = x, X_{i,k-\tau} = \tilde{x}].
\end{align*}
\]

We thus have proved the following result:

**Result 3 (Endogenous time-varying control variables)** With endogenous time-varying covariates, DID-matching controlling only on covariates at the time of selection is biased. DID-matching is unbiased if the set of control variables includes the value of the covariates at the time when both pre- and post-treatment outcomes are measured.

The general lesson from this analysis in this section is that “the more control variables, the better” is indeed a correct statement for DID-matching when considering control variables that are not past outcomes. This result nevertheless requires that covariates observed after the treatment has been received have not been modified by the treatment. If \(X_{i,k+\tau}\) is altered by the treatment (for example if women participating in a job training program do have less children because of their participation), the treatment effect recovered by DID-matching controlling on both pre- and post-treatment covariates is a partial treatment effect, net of the effect of the treatment on covariates. On the other hand, if post-treatment covariates are not included in the set of control variables, DID-matching is biased for both partial and complete treatment effects.

3 Past outcomes as covariates: when it is better not to control

In this section, I study whether or not it is better to control for past outcomes when applying DID-matching, i.e. whether it is better to control for past outcomes or to difference them out. In this section, I prove first that when unobserved individual fixed effects directly enter...
the selection equation along with past outcomes (because unobserved gains from or costs to
the treatment are correlated to these fixed effects), matching controlling on past outcomes is
biased. The intuition for this result is exactly the same as for the results in previous sections:
treated and untreated individuals with the same past outcomes differ in their unobserved fixed
effects, and thus have different average long run values for their outcomes, toward which they
will start converging after selection is decided. Conditional on past outcomes, ignorability
does not hold, leading to biased inference by matching. I then exhibit three special cases in
which DID-matching not controlling for past outcomes is unbiased whereas simple matching
controlling for past outcomes is biased. These results are, to my knowledge, the first examples
of bias due the inclusion of too many control variables in a dynamic setting. Furthermore, these
results have sharper predictions than the one in Heckman and Navarro-Lozano (2004). In their
paper, including more control variables may ever increase or decrease bias. In the present
paper, including more control variables always creates bias where none would exist otherwise.
I consider three distinct cases, that can be related to the nature of Ashenfelter’s dip. In the
first case, there is no Ashenfelter’s dip, i.e. selection is only on the persistent component of the
outcome variable. Some restrictions on the way past outcomes enter the equation for current
outcomes are necessary to get to the result of no bias for DID-matching. In the second case, I
consider a very transitory Ashenfelter’s dip: shocks determine entry into the program, but they
do not persist across time. Third and finally, I consider a situation in which Ashenfelter’s dip
is persistent. In that case, under linearity of the selection and outcome equations, normality of
the error terms and complete knowledge by the individuals of the shocks to their earnings up to
selection, I prove that DID-matching applied symmetrically around the treatment date yields
an unbiased estimate of the average effect of the treatment on the treated, whereas matching
controlling on past outcomes is biased.

3.1 No Ashenfelter’s dip: past outcomes only determine current
outcomes

The selection equation considered in this section does not include past outcomes: $D_{ik} =
1\left[h(\mu, u_{ik}) \geq 0\right]$, where $u_{ik}$ is an i.i.d. shock independent of all the other unobserved shocks
in the model. Selection is thus only driven by the long run component of outcomes. On the
Contrary, I allow outcomes to be positively autocorrelated, but assume linearity and separability in the different components: \( Y_{it}^0 = \rho Y_{i,t-1}^0 + \delta_t + \epsilon_{it} + \mu_i \), where \( \rho \) is a positive autocorrelation coefficient strictly inferior to one. I further assume that \( Y_{i,k-\tau'} \) is a random draw around the long run value of the process \( Y_i^* = \frac{\mu_i}{1-\rho} \): \( Y_{i,k-\tau'}^0 = Y_i^* + \epsilon_{i,k-\tau'} \). This amounts to assuming that the process is far enough from its initial conditions to have reached its stationary distribution.

In that setting, DID-matching not controlling for past outcomes offers an unbiased estimation of the average effect of the treatment on the treated, whereas matching controlling on past outcomes is biased for that parameter. The intuition for this result is that unconditional mean increments are independent from \( \mu_i \) with a linear process starting at a random point around its long run value, a point previously noted by Blundell and Bond (1998). Individuals with different \( \mu_i \) exhibit thus the same mean increments, due to common additive time shocks. As \( \mu_i \) is by assumption the only source of selection bias, DID-matching is unbiased. On the contrary, conditioning on past outcomes focuses on groups of individuals that are not at their long run equilibrium: participants and non participants having the same values of past outcomes have different long run equilibrium values to which they start converging immediately after the conditioning period. In that case, ignorability is violated, as well as conditional independence of increments.

Formally, it is useful to write the linear process for outcomes as an error correction model around \( Y^* \). For example, we know that: \( Y_{i,k-\tau'}^0 - Y_i^* = \rho (Y_{i,k-\tau'}^0 - Y_i^*) + \delta_{k-\tau'} + \epsilon_{i,k-\tau'} \). Extending this reasoning to period \( k + \tau \) gives the following result:

\[
Y_{i,k+\tau}^0 - Y_i^* = \sum_{j=0}^{\tau+\tau'-1} \rho^j \delta_{k+\tau-j} + \sum_{j=0}^{\tau+\tau'} \rho^j \epsilon_{i,k+\tau-j}.
\] (16)

When differencing between the two periods \( k - \tau' \) and \( k + \tau \), the long run value depending on the fixed effect is differentiated out, and only time shocks and idiosyncratic shocks unrelated to participation remain. The mean increment among the treated is thus equal to a weighted

---

\(^6\)Extension to an additive nonlinear markov process is non trivial because we have to check the independence of increments when an ergodic long run distribution is hit by common time shocks. This is left for further research.
average of time effects, and is equal to the mean increment between the untreated:

$$\mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0|D_{ik} = 1] = \sum_{j=0}^{\tau+\tau'-1} \rho^j \delta_{k+j}$$

$$= \mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0|D_{ik} = 0].$$

The independent increments condition is thus verified unconditionally in this model, leading to an unbiased estimation by DID-matching. This is not true when conditioning on $Y_{i,k-\tau'}$:

$$\mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0|D_{ik} = 1, Y_{i,k-\tau'} = y] - \mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0|D_{ik} = 0, Y_{i,k-\tau'} = y]$$

$$= (\rho^{\tau+\tau'} - 1) \left( \mathbb{E}[\epsilon_{i,k-\tau'}|h(\mu_i, u_{ik}) < 0, \frac{\mu_i}{1-\rho} + \epsilon_{i,k-\tau'} = y] - \mathbb{E}[\epsilon_{i,k-\tau'}|h(\mu_i, u_{ik}) \geq 0, \frac{\mu_i}{1-\rho} + \epsilon_{i,k-\tau'} = y] \right) \neq 0. \quad \text{(19)}$$

The last inequality follows because by simultaneously conditioning on the treatment status and past outcomes, we select people with distinct distributions for $\mu_i$ and at the same time the same value for past outcomes: this is only possible if we allow for different distributions of $\epsilon_{i,k-\tau'}$. Treated and untreated individuals having the same value of past outcomes differ by how they got there: untreated individuals have higher long run unobserved determinants of earnings but have experienced a negative shock while treated individuals have a lower long run component but have experienced a favorable shock. The influence of these shocks progressively dissipating, treated and untreated individuals converge to their distinct long run values for their outcomes. Another way to understand this result is by replacing $\epsilon_{i,k-\tau'}$ by $y - \frac{\mu_i}{1-\rho}$ in equation (19):

$$\mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0|D_{ik} = 1, Y_{i,k-\tau'} = y] - \mathbb{E}[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0|D_{ik} = 0, Y_{i,k-\tau'} = y]$$

$$= \frac{1 - \rho^{\tau+\tau'}}{1-\rho} \left( \mathbb{E}[\mu_i|h(\mu_i, u_{ik}) < 0, \frac{\mu_i}{1-\rho} + \epsilon_{i,k-\tau'} = y] - \mathbb{E}[\mu_i|h(\mu_i, u_{ik}) \geq 0, \frac{\mu_i}{1-\rho} + \epsilon_{i,k-\tau'} = y] \right) \neq 0. \quad \text{(20)}$$

Equation (20) makes clear that even though differencing seems to get rid of the long run fixed effect, this one reappears because we condition on past outcomes, whose value depend on $\mu_i$.

We thus have proved the following result:
Result 4 (No Ashenfelter’s dip) If past outcomes directly determine current outcomes, but do not directly determine selection into the treatment, in the special case of a linear autoregressive process starting at a random point around its long run equilibrium, DID-matching not controlling on past outcomes preserves unbiasedness while matching controlling on past outcomes is biased.

An illustration of result 4 is given by figure 1. This figure has been generated by the model presented in section 4. Figure 1 shows that selection bias is constant through time when there is no Ashenfelter’s dip and as a consequence DID is unbiased. On the contrary, matching, here on period $k$ outcomes, makes participants and non-participants identical at that period, but do not achieve unbiasedness in successive periods. This is because observationally identical participants and non participants have different distributions of the individual fixed effects $\mu_i$. In order to make them observationally equivalent, the matching procedure selects non participants with higher values of $\mu_i$ than those of the participants but experiencing a series of negative shocks before entry into the treatment is decided. Matched non-participants thus do not enter the treatment because they have higher values of $\mu_i$, but look identical to participants because they experience of a series of negative shocks just before participation in the program is decided.

3.2 Transient Ashenfelter’s dip: past outcomes only determine selection

In this section, I assume that the outcome process is not autocorrelated, but that selection depends on outcomes at period $k$. I write $D_{ik} = 1[h(Y_{ik}^0, \mu_i, u_{ik}) \geq 0]$ and $Y_{it} = g(D_{it}, \delta_t, \epsilon_{it}) + \mu_i$. In that general setting, unconditional DID-matching is unbiased while matching on past outcomes is biased. The intuition for unbiasedness of unconditional DID-matching is that because past outcomes do not determine current outcomes, the fact that selection depends on outcomes in period $k$ has no consequence for the distribution of outcomes in the treated and untreated groups in periods $k + \tau$ and $k - \tau'$. On the contrary, matching conditioning on period $k - \tau'$ outcomes is biased, because the distribution of shocks and fixed effects is different in the

---

7The time shocks $\delta_t$ have been all set to zero, which is equivalent to expressing outcomes as differences with respect to the average outcome in each period.
treated and untreated groups. Finally, DID-matching conditioning on $Y_{ik}$ yields an unbiased estimate of the treatment effect, because time varying shocks are not autocorrelated.

DID-matching not conditioning on past outcomes is unbiased:

$$
\mathbb{E}\left[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 | D_{ik} = 1\right]
= \mathbb{E}\left[g(0, \delta_k + \tau, \epsilon_{i,k+\tau}) - g(0, \delta_k - \tau', \epsilon_{i,k-\tau'}) | h(g(0, \delta_k, \epsilon_{ik}) + \mu_i, \mu_k, u_{ik}) < 0\right]
\leq \mathbb{E}\left[Y_{i,k+\tau}^0 - Y_{i,k-\tau'}^0 \right] | D_{ik} = 0,
$$

the second equality coming from the fact that the idiosyncratic time-varying shocks in the outcome process are uncorrelated to time shocks, individual fixed effects and across periods.

DID-matching conditioning on past outcomes, which is equivalent to simple matching, is...
biased:

\[
\begin{align*}
\mathbb{E}[Y^0_{i,k+\tau} - Y^0_{i,k-\tau'} | D_{ik} = 1, Y_{i,k-\tau'} = y] &- \mathbb{E}[Y^0_{i,k+\tau} - Y^0_{i,k-\tau'} | D_{ik} = 0, Y_{i,k-\tau'} = y] \\
&= \mathbb{E}[g(0, \delta_{k+\tau}, \epsilon_{i,k+\tau}) - g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) | h(y, \mu_i, u_{ik}) < 0, g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) + \mu_i = y] \\
&\quad - \mathbb{E}[g(0, \delta_{k+\tau}, \epsilon_{i,k+\tau}) - g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) | h(y, \mu_i, u_{ik}) \geq 0, g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) + \mu_i = y] \\
&= -\mathbb{E}[g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) | h(y, \mu_i, u_{ik}) < 0, g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) + \mu_i = y] \\
&\quad + \mathbb{E}[g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) | h(y, \mu_i, u_{ik}) \geq 0, g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) + \mu_i = y] \\
&= \mathbb{E}[\mu_i | h(y, \mu_i, u_{ik}) < 0, g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) + \mu_i = y] \\
&\quad - \mathbb{E}[\mu_i | h(y, \mu_i, u_{ik}) \geq 0, g(0, \delta_{k-\tau'}, \epsilon_{i,k-\tau'}) + \mu_i = y] \neq 0.
\end{align*}
\]

The last equality can also be obtained when trying to check the conditional ignorability condition at the heart of matching, which shows that matching is also biased in this case. In fact, simple matching controlling for past outcomes in whatever period is biased, because it cannot capture differences in fixed effects. This last equality (26) also shows that first differencing does not eliminate the individual fixed effects when conditioning on past outcomes: treated and untreated individuals differ in their distribution of \( \mu_i \). Those having the same values of past outcomes thus have different long run distributions of their outcomes, to which they converge after the conditioning period. In this model, making people completely alike thus kills the identification strategy at the heart of DID-matching. On the other hand, DID-matching conditioning only on the outcomes at the moment of selection (or at any period that is not the period with respect to which differencing is done) is unbiased (because time varying shocks to outcomes are uncorrelated) whereas there is no matching estimator that is unbiased.

We thus have proved the following result:

**Result 5 (Transitory Ashenfelter’s dip)** If past outcomes do not directly determine current outcomes but determine selection into the treatment and transitory shocks are not autocorrelated, matching conditioning on past outcomes is biased, while DID-matching not controlling on past outcomes is unbiased.

Figure 2 illustrates result 5: when Ashenfelter’s dip exist only for one period, DID-matching with respect to that period is biased, but remains valid when earlier periods as used. On the contrary, matching on outcomes observed at the date of selection does not solve the prob-
Figure 2 – Average potential outcomes when Ashenfelter’s dip is transient

Note: the outcomes are simulated from the model presented in section 4. $\tau$ indicates the number of period after assignment to treatment. $\tau'$ measures the period at which matching is performed. $\rho$ is the autocorrelation parameter.

3.3 Persistent Ashenfelter’s dip: past outcomes determine both selection and current outcomes

In this section, I study the more realistic case where both selection and outcomes depend on past outcomes. In that case, there is a persistent Ashenfelter’s dip, because past and future distance of the outcomes from their long run equilibrium are correlated to entry into the treatment: there is time varying selection bias. In this setting, I can derive another instance of bias from too much control, in a specialized example: when both selection and outcome equations are linear and error terms are normally distributed.\footnote{Extension to non normal error terms could use the Edgeworth expansion approach to general selection models in Lee (1982). This is left for further research on the generality of symmetric differencing.} In that case, it is possible to prove that DID-matching not controlling on past outcomes and applied symmetrically around
the date of enrollment is unbiased, generalizing a result by Ashenfelter and Card (1985) and Heckman and Robb (1985) to selection rules including fixed effects.

Selection into the treatment follows a linear threshold rule: \( D_{ik} = 1 \) \( [D_{ik}^* \leq 0] \), with \( D_{ik}^* = \theta \mu_i + \beta Y_{ik} + u_{ik} \). Selection thus depends on outcomes in period \( k \), as a measure of foregone earnings or because the program is means tested. Selection also depends on fixed effects: individuals having higher values of the fixed effect may for example have lower the administrative costs of participation or higher unobserved gains to the program. In this section, outcomes also follow a linear process: \( Y_{it}^0 = \rho Y_{it-1}^0 + \delta_{it} + \mu_i + \epsilon_{it} \). I assume that this process has been going on since an infinite amount of time, so that initial conditions do not matter and that the process is stationary (\( \rho \) is positive and strictly inferior to one, \( \text{Var}(Y_{it}^0) = \text{Var}(Y_{it}^0) \) and \( \text{Cov}(\mu_i, Y_{it}^0) = \text{Cov}(\mu_i, Y_{it}^0) \), \( \forall t, t' \)). Finally, I assume that all unobserved variables in the model are normally distributed and uncorrelated to each other.

In this model, there is at the same time selection on the fixed effects and a persistent Ashenfelter’s dip, because outcomes are autocorrelated. Matching on past outcomes does not solve selection bias, because treated and untreated individuals with the same past outcomes do not have the same distribution of the fixed effects, and thus do not have the same post-treatment potential outcomes. Symmetric unconditional DID-matching is on the contrary unbiased in this model. The intuition for this result is due to the fact that selection bias weakens symmetrically around \( t = k \). Selection bias is due to the correlation between \( Y_{it} \) and \( D_{ik}^* \), but this correlation is maximal at \( t = k \) and progressively weakens as time passes because \( \rho \) is inferior to one. The correlation also weakens as we go back in time, because past shocks matter less and less for selection, as the persistence of shocks is not infinite. Because this correlation weakens at the same rate whether we go in the past or in the future, symmetric DID-matching is unbiased. This intuition is embodied in the following lemma, which states that the covariances between \( D_{ik}^* \) and \( Y_{i,k+t}^0 \) and \( D_{ik}^* \) and \( Y_{i,k-t}^0 \) are equal when \( \tau = \tau' \).

**Lemma 1** In the model laid out in this section, we have, for all positive \( \tau \):

\[
\text{Cov}(Y_{i,k+t}^0, D_{ik}^*) = \text{Cov}(Y_{i,k-t}^0, D_{ik}^*) = \left( \theta + \beta \frac{1 - \rho^\tau}{1 - \rho} \right) \text{Cov}(\mu_i, Y_{ik}^0) + \beta \rho^\tau \text{Var}(Y_{ik}).
\]

**PROOF :** See in appendix A.1.  

\(^9\)See the derivation of such an equation from a rational decision rule in the next section.
Because all variables are assumed to be normally distributed in this section, the expectation of $Y_{i,k+\tau}$ conditional on participating into the program is the usual formula for the expectation of a bivariate censored random variable (Heckman, 1979). This expectation is the sum of the unconditional expectation of the outcome and of an inverted Mills ratio term that depends only on $\text{Cov}(Y_{i,k+\tau}, D_{ik}^*)$. As the expectation of $Y_{i,k+\tau}$ depends on the same covariance, the assumption of independent increments holds in this model. This is stated in the following lemma:

**Lemma 2** In the model laid out in this section, we have, for all positive $\tau$:

$$\mathbb{E}[Y_{i,k+\tau} - Y_{i,k-\tau} | D_{ik} = 1] = \mathbb{E}[Y_{i,k+\tau} - Y_{i,k-\tau}] = \mathbb{E}[Y_{i,k+\tau} - Y_{i,k-\tau} | D_{ik} = 0].$$

**Proof**: See in appendix A.2.

We thus have proved the in the model laid out in this section, symmetric DID identifies treatment on the treated. On the contrary, matching controlling on past outcomes is biased. For example, matching on the outcomes when selection occurs is biased because controlling for $Y_{ik}$ does not eliminate the bias due to the unobserved fixed effects. To prove this result, first note that, using equation (39) in section A.1 in the appendix:

$$\mathbb{E}[Y_{i,k+\tau} | D_{ik} = 1, Y_{ik} = y] = \frac{1 - \rho^\tau}{1 - \rho} \mathbb{E}[\mu_i | D_{ik} = 1, Y_{ik} = y] + \sum_{j=0}^{\tau-1} \rho^j \delta_{k+\tau-j} + \rho^\tau y. \quad (27)$$

Matching does not identify the average treatment effect on the treated because the ignorability condition is not fulfilled in this model:

$$\mathbb{E}[Y_{i,k+\tau} | D_{ik} = 1, Y_{ik} = y] - \mathbb{E}[Y_{i,k+\tau} | D_{ik} = 0, Y_{ik} = y] = \frac{1 - \rho^\tau}{1 - \rho} (\mathbb{E}[\mu_i | D_{ik} = 1, Y_{ik} = y] - \mathbb{E}[\mu_i | D_{ik} = 0, Y_{ik} = y]) \neq 0. \quad (28)$$

We thus have proved the following result:

**Result 6** When past outcomes directly determine both current outcomes and selection into the treatment, if both outcomes and selection are linear equations, error terms are normal and the process is around its stationary long run equilibrium, DID-matching not controlling on past outcomes applied symmetrically around treatment date is unbiased while matching controlling...
on past outcomes is biased.

Figure 3 illustrates result 6: selection bias varies at every period, making it impossible to use a general version of DID-matching. But as selection bias is symmetrical around the selection period, applying DID-matching symmetrically around that period is unbiased. Matched non-participants are observationally identical to participants, but differ in unobserved dimensions: matched non-participants have higher values of $\mu_i$ leading them to not participate, and have experienced more negative transient shocks. The progressive fading out of these shocks biases matching.

Figure 3 – Average potential outcomes when Ashenfelter’s dip is persistent

Note: the outcomes are simulated from the model presented in section 4. $\tau$ indicates the number of period after assignment to treatment. $\tau'$ measures the period at which matching is performed. $\rho$ is the autocorrelation parameter.

4 Sensitivity of matching and DID-matching to misspecifications

In this section, I extend the analysis of the previous sections in two directions. First, I quantify the extent of the bias of matching, in order to determine whether the theoretical differences
make empirical sense. Second, I study the sensitivity of matching and DID-matching to two
types of specification problems likely to be encountered in real applications: the choice of
the control period and the information that individuals have when they decide to enter the
program. To this aim, I borrow a realistic estimates of the wage process from MaCurdy (1982)
and use a selection rule based on comparisons of expected discounted value due to Heckman
and Robb (1985) and Heckman, LaLonde, and Smith (1999). I then derive closed forms for
the bias terms of matching and DID-matching in this parametric model.

4.1 A toy model of the wage process and of entry into a job training
program

To model outcomes, I use MaCurdy (1982)’s estimates of the wage process:

\[ Y_{it}^0 = g_0(X_i, \delta_t) + \mu_i + U_{it}, \]

\[ U_{it} = \rho U_{it-1} + m_1 v_{it-1} + m_2 v_{it-2} + v_{it}, \]

where the exact formulation for the residual process comes from MaCurdy (1982)’s preferred
specification for the dynamics of wage earnings (ARMA(1,2)). I abstract from the problem of
time varying covariates other than past outcomes in this analysis. Note that MaCurdy (1982)
estimates that the wage process almost follows a random walk (with \( \rho = .99 \)) and that the
variance of the fixed effect is null (\( \sigma^2_{\mu} = 0 \)). It is useful to write current wages as a function
of past wages using equation (29) and the fact that \( U_{it} = Y_{it}^0 - g_0(X_i, d_i) - \mu_i \):

\[ Y_{ik+\tau}^0 = g_0(X_i, d_{k+\tau}) + \rho^{\tau+\tau'} Y_{ik-\tau'} + \rho^{\tau+\tau'} g_0(X_i, d_{k-\tau'}) + \mu_i(1 - \rho^{\tau+\tau'}) + \sum_{j=0}^{\tau+\tau'-1} \rho^j \nu_{ik+\tau-j}. \]

I borrow a credible selection rule from Heckman and Robb (1985) and Heckman, LaLonde,
and Smith (1999). Assuming perfect credit markets, individuals compare the expected dis-
counted value of entering the program to its costs in terms of foregone wages during the period

---

10 Baker (1997) and Guvenen (2007), among others, argue that including a random trend better captures
the dynamics of the wage process. I discuss the implications of this literature for the results in this paper in
the concluding section.
on training (period $k$), net of compensating transfers $c_i$:

$$D_{ik} = 1 \left[ \mathbb{E} \left[ \sum_{j=1}^{\infty} \frac{Y_{i,k+j}^1 - Y_{i,k+j}^0}{(1 + r)^j} - Y_{ik}^0 - c_i | \mathcal{I}_{ik} \right] \geq 0 \right], \quad (31)$$

$$= 1 \left[ \frac{\alpha_i}{r} \geq c_i + \mathbb{E}[Y_{ik}^0 | \mathcal{I}_{ik}] \right], \quad (32)$$

where $\alpha_i$ is the individual level wage gain from the program, that, for convenience, I assume constant through time and perfectly known at the time of entry, $r$ is the interest rate and $Y_{ik}^0$ are earnings in period $k$ outside of the program (they measure the opportunity cost of entering the program). Earnings in period $k$ are not observed by the econometrician for those individuals entering the program. When deciding or not to enter the program, individuals are assumed to have some information $\mathcal{I}_{ik}$ on their foregone wages. I consider two different informational contents: (i) individuals perfectly know their foregone earnings in period $k$ ($\mathcal{I}_{ik} = \mathcal{I}_{ik}^f = \{Y_{ik}^0\}$) and (ii) individuals have a limited knowledge of their foregone earnings ($\mathcal{I}_{ik} = \mathcal{I}_{ik}^l = \{Y_{i,k-1}^0, \delta_k, \delta_{k-1}, \mu_i, v_{i,k-1}, v_{i,k-2}\}$). Limited information can arise because individuals decide whether or not to enter the program in period $k$ at the end of period $k - 1$, not knowing the last innovation to their wages.

From equations (29) and (32), we can write selection into the program as a threshold crossing model: $D_{ik}^\iota = 1 \left[ D_{ik}^\iota \leq 0 \right]$, where $\iota \in \mathcal{I} = \{f, l\}$ and:

$$D_{ik}^{\iota f} = Y_{ik}^0 + c_i - \frac{\alpha_i}{r}, \quad (33)$$

$$D_{ik}^{\iota l} = g^0(X_i, \delta_k) - \rho g^0(X_i, \delta_{k-1}) + \rho Y_{i,k-1}^0 + (1 - \rho) \mu_i + m_1 v_{i,k-1} + m_2 v_{i,k-2} + c_i - \frac{\alpha_i}{r}. \quad (34)$$

I assume that all error terms are normally distributed, which implies that $D_{ik}^{\iota l}$ and $Y_{ik}^0$ are also normally distributed. The analyst wishes to estimate the effect of the program on period $k + \tau$ mean outcome and she considers controlling for past outcomes observed in period $k - \tau'$. Controlling for the actual determinants of entry into the program ($Y_{ik}^0$) is impossible, wages in period $k$ being unobserved because the individual participate in the program. I compute values of the bias for both approaches when using this infeasible estimator as a benchmark. Feasible estimators can only use observed past earnings: $Y_{i,k-\tau'}, \tau' > 0$. 

22
4.2 Bias from pure DID-matching (not controlling for past outcomes)

Expressions for the bias from DID-matching only controlling on $X_i$ can be derived from the fact that outcomes and $D_{ik}^*$ are jointly normally distributed, and that they have linear expectations. Following [Ashenfelter and Card (1985)], we know that:

$$\text{Cov}(Y_0, D^*_i) = \mu_i + \rho \sigma_i^* \sigma_Y + \rho \mu_c \sigma_{D^*} - \rho \sigma_{\mu,c} + \rho |\tau| \sigma^2.$$

From equation (36), we can see that when there is full information on the foregone wages, the covariance of wages and participation is symmetric around $k$, and thus symmetric DID-matching (with $\tau = \tau'$) is unbiased.

A more intuitive explanation on the nature of the bias of DID-matching can be attained by using equation (30) (for $\tau > 1$).\[^{13}\]

$$B^{\text{did}}(\tau, \tau', t) = \frac{\text{Cov}(Y_0, D^*_i)}{\text{Var}(D^*_i)} \left( \text{E}[D^*_i | D^*_i \leq 0] - \text{E}[D^*_i | D^*_i > 0] \right).$$

In appendix [B], I derive a closed form expression for the covariance term (for $|\tau| > 1$ and $t = f$).\[^{12}\]

\[^{11}\]I keep the conditioning on $X_i$ implicit in the remaining of the section to save on the notation.

\[^{12}\]The general formula is in the appendix.

\[^{13}\]If $\tau = 1$, in the last term, the term $\left( \rho^2 + \rho m_1 + m_2 \right)$ is deleted and replaced inside the sum by $\left( \rho^2 + \rho m_1 + 1 \right) \sum_{j=1}^{\tau-1} \rho^{j}$.
The first part of the bias term (37a) is due to initial differences in the history of shocks up to period $k - \tau'$ between treated and untreated. These initial differences persist until period $k$, leading to different participation decisions, but progressively fade away as long as we get away from the enrollment period. This progressive return to the mean wage is confounded with a causal effect of the treatment by the DID estimator because it fails to take into account that some of the initial differences in earnings are due to transient shocks. This leads to an upward bias in the estimation of the treatment on the treated parameter. The dependence of this bias term on $\rho$, the persistence of shocks, is difficult to determine: on the one hand, the higher $\rho$, the lower the discrepancy between the two periods, because the initial differences tend to persist. On the other hand, a higher $\rho$ leads to higher variance of the error terms, and thus may increase the size of the bias term.

The second part of the bias (37b) reflects the fact that some of the negative shocks to wages will get corrected in the immediate aftermath of period $k - \tau$, through the action of the MA terms. The estimated MA terms are generally negative: the negative shocks leading people to participate do not fully persist until period $k + \tau$. This bias term will also be positive because it leads to an overestimate of the initial mean wage differences between participants and non-participants.

The last part of the bias term (37c) is the consequence of shocks that have happened after the pre-treatment control period $(k - \tau')$. This bias term will generally be negative because the AR term dominates the MA terms (with the estimates in MaCurdy (1982), we have $\rho^2 + pm_1 + m_2 \approx 0.5$) and because mean shocks are negatively correlated to participation through the participation equation (the lower the wage, the higher the probability of participation). From equation (35), we know that this term offsets the two previous terms when $\tau = \tau'$. In appendix B, I derive closed form expressions for these bias terms under the assumption that the i.i.d MA terms are normal with variance $\sigma^2$. Values of these bias terms are plotted in figure 4 for MaCurdy (1982)'s estimates of the wage process.

Figure 4 confirms the sign of the different components of bias and the unbiasedness of symmetric-DID. It also shows that DID-matching is sensitive to the information the agents have when deciding to enter the program: when they do not know the last innovation to their wages, symmetric-DID is no longer unbiased. This is because the correlation between treatment
and shocks occurring after period $k - \tau'$ is less negative.

Figure 4 – Bias from DID-matching (not controlling on past outcomes) for $\tau = 4$

Note: $a$, $b$ and $c$ stand respectively for the components (37a), (37b) and (37c) of the bias term of DID matching derived in the appendix (see equations (58a), (58b) and (58c)). (shock) corresponds to the value of the corresponding terms when $v_{ik}$ is not observed when the individual decides to participate, i.e. to limited information: $\iota = l$. All values have been computed according to MacCurd’s estimates of the parameters.

4.3 Bias from pure matching (controlling for past outcomes)

From equation (30), we can derive the bias that arises from matching on $Y_{ik-\tau'}$ (for $\tau > 1$):

$$B^m(y, \tau, \tau', l) = \left(1 - \rho^{\tau+\tau'}\right) (\mathbb{E}[\mu_i | Y_{ik-\tau'} = y, D_{ik} = 1] - \mathbb{E}[\mu_i | Y_{ik-\tau'} = y, D_{ik} = 0])$$

$$+ \rho^{\tau+\tau'-2} \left( (\rho m_1 + m_2) (\mathbb{E}[v_{ik-\tau'} | Y_{ik-\tau'} = y, D_{ik} = 1] - \mathbb{E}[v_{ik-\tau'} | Y_{ik-\tau'} = y, D_{ik} = 0]) 
+ \rho m_2 (\mathbb{E}[v_{ik-\tau'-1} | Y_{ik-\tau'} = y, D_{ik} = 1] - \mathbb{E}[v_{ik-\tau'-1} | Y_{ik-\tau'} = y, D_{ik} = 0]) \right)$$

$$+ \rho^{\tau-2} \left( \rho^2 + \rho m_1 + m_2 \right) \sum_{j=1}^{\tau'-1} \rho^j (\mathbb{E}[v_{ik-j} | Y_{ik-\tau'} = y, D_{ik} = 1] - \mathbb{E}[v_{ik-j} | Y_{ik-\tau'} = y, D_{ik} = 0]).$$

The first part of the bias term is due to differences in unobservables fixed through time.
between treated and untreated. This term arises because participation into the program is partly determined by differences in $\mu_i$ because of the opportunity cost of participating in the program in terms of foregone wages.

The second part of the bias term arises because of the moving average components of the wage process. In terms of the participation equation, this term means that among people with the same pre-treatment wage, only those with a recent negative shock to their wages have decided not to enroll into the program. This shock being transitory (40% of the shock will have disappeared after one year), these people tend to have higher wages in period $k$ after a wage decrease in period $k-1$. With higher potential wages in period $k$, they thus decide not to enroll into the program, whose opportunity cost has increased. After the adjustment in period $k$ due to the MA(1) term, the rest of the positive wage shock tend to persist through time, due to the large AR(1) term. Wages for participants thus persist, but at a higher level than those of untreated people who had the same wages in period $k-1$. We thus tend to underestimate the impact of the program by confounding it with the fading out of a transient shock that lead people to participate in the first place.

The last component of the bias is due to shocks posterior to period $k-\tau'$. We could think that this term does not depend on $y$, because it is due to i.i.d. shocks posterior to the period at which we control. But this intuition is false. Entry into the program depends on all shocks until period $k$. Thus, conditionally on treatment status, shocks posterior to $k-\tau'$ are correlated with $Y_{i,k-\tau'}$. Among individuals with the same value of $Y_{i,k-\tau'}$, participants are the one who experience the more negative shocks, lowering the opportunity cost of participation. This shock tending to persist, this bias term is also negative: using matching, we interpret these difference in persistent shocks leading to selection as a lack of effect of the program.

Figure 5 illustrates these results by plotting the conditional expectation of wages in the absence of the treatment for participants, non participants and matched non participants. Even when controlling on the last observed outcome before treatment (i.e. $Y_{i,k-1}$), we cannot drive bias to zero by using matching. Individual with the same wages at period $k-1$ differ both in their previous shocks and their last shock. When prospective participants have full information about $Y_{ik}$ (figure 5(a)), matched non-participants differ because they have more negative shocks before the control period and experience a large positive shock in the treatment period, which
Figure 5 – Average potential outcomes with MaCurdy’s estimates of the wage process

Note: the outcomes are simulated from the model presented in section 4. τ indicates the number of period after assignment to treatment. τ' measures the period at which matching is performed. ρ is the autocorrelation parameter. (shock) corresponds to the value of the corresponding terms when vi,k is not observed when the individual decides to participate, i.e. to limited information: i = l.
leads them not to participate. When this last shock is unknown to the prospective participants when they decide to enroll (figure 5(b)), participants and matched non-participants still differ in their composition of old and recent shocks. Non-participants have experienced higher shocks in the past, and suffer from two negative shocks just before period \( k - 1 \). As these two shocks are corrected by the MA terms, non-participants have higher values of \( D_{ik}^* \) because of higher anticipated wages in period \( k \)\(^{14}\).

Figure 6 – Comparison of bias when controlling (simple matching) and not controlling (DID matching) on past outcomes for \( \tau = 4 \)

Note: the value of the bias for DID (resp. simple) matching are the absolute values of the terms in equation (37) (resp. 38) computed according to MaCurdy (1982)'s estimates of the parameters for normal error terms (see the derivation in appendix B.1. (shock) corresponds to the value of the corresponding terms when \( v_{ik} \) is not observed when the individual decides to participate, i.e. to limited information: \( \iota = l \).

Figure 6 presents the absolute values of the bias terms of DID matching (equation 37) and simple matching (equation 38) with the values of the parameters estimated by MaCurdy (1982). A series of results emerge from this figure:

1. Not controlling on past outcomes has generally lower bias than controlling on past outcomes, apart from the case of the infeasible estimator using unobserved period \( k \) outcomes.

\(^{14}\)This suggests that controlling for the last three periods before enrollment would make matching unbiased.

\(^{15}\)Derivation of these bias terms in the normal case can be found in appendix B.
as controls.

2. Controlling on past outcomes is the least biased when controlling for the period closest to enrollment (i.e. $\tau' = 1$). In that case, matching controlling on past outcome is nevertheless more biased than DID-matching with respect to the same period with the parameter values estimated by [MaCurdy](1982).

3. Symmetric DID-matching is the least biased estimator under full information: it is less biased than matching controlling on the period closest to enrollment.

4. DID-matching is less sensitive than matching to the misspecification of the control period: when shifting from period $k$ (infeasible and unbiased) to period $k - 1$, the bias of matching increases by 200 % of the treatment effect, while shifting from period $k - 4$ to period $k - 5$ increases bias of DID-matching by 15 % of the treatment effect.

5. When individuals do not know the last shock to their earnings prior to entry, matching on the last period before entry is less biased than DID-matching with respect to the same period. Bias still represents 90 % of the treatment effect. Symmetric DID is also biased, but less so (60 % of the treatment effect).

## 5 Discussion

In this paper, I study whether it is always better to control for more covariates in a dynamic context. I focus on the comparison of the relative biases of matching on past covariates vs Difference in Difference matching. I first show that intuitions about how first differencing removes additive fixed effects are not valid while simultaneously conditioning on time-varying control variables, even if these ones are exogenous. When these variables are not past outcomes, I show that controlling for values of these variables at both the pre- and post- treatment periods is necessary to restore unbiasedness of DID-matching. In this case, we nevertheless face a risk of overcontrolling.

With past outcomes, or generally when the control variables are modified by the treatment, it is impossible to control on post-treatment values. I show that in three credible cases, matching on past outcomes is biased whereas DID-matching, or a symmetric version of it, is unbiased.
This is to my knowledge the first instances of bias from too much control in a dynamic context.

In order to quantify the extent of the bias generated by matching and to study the robustness of both estimators to the control period and the information that the agent have when deciding to enter the program, I study the special case of the evaluation of a job training program. Borrowing a credible selection rule from [Heckman, LaLonde, and Smith (1999)] and relying on the parameters of the wage process estimated by [MacCurdy (1982)], I derive closed forms for the bias terms of the two estimators when the error terms are normally distributed.

I show that DID matching performs better when used symmetrically around the period of enrollment, as implemented by [Heckman, Ichimura, Smith, and Todd (1998)]. Matching is more biased than DID-matching even when using the first pre-enrollment period. Symmetric DID-matching is less sensitive to misspecifications of the control period and very slightly less sensitive to the misspecification of the information set of the agents at the date of enrollment. These results are to my knowledge the first one that prove in a practical application that not controlling for past outcomes may be better than controlling for them.

These results point to some previously unnoticed results in [Heckman, Ichimura, Smith, and Todd (1998)] and [Smith and Todd (2005)]. [Heckman, Ichimura, Smith, and Todd (1998)] compares the relative ability of different set of control variables to reproduce the experimental results of the evaluation of Job Training Partnership Act (JTPA) thanks to matching and DID matching. When using a crude control set not including wages at the date of enrollment, the average bias of the estimator is of 73 % of the treatment effect, lower than that obtained thanks to the set of variables with the higher predicting power including wages at enrollment (120 % of the treatment effect) (see their table XII p.1062). [Smith and Todd (2005)] use two set of control variables when estimating the bias of propensity score matching and DID propensity score matching in the National Support for Work experimental study: the first set (they name it the Lalonde set) does not contain past income while the second set (the Dehejia and Wahba (DW) set) does contain past income. When they apply DID matching with the first set of controls, the bias is of respectively -2 %, 22 % and -16 % of the treatment effect when using the most efficient matching estimators (respectively nearest neighbour matching with one neighbour restricted to the common support, local linear matching with a small bandwidth (1.0) and local linear regression adjusted matching with the same bandwidth). When using past outcomes as control
variables, the bias is larger: respectively -105 %, -137 % and -137 % with the same estimators (see their table 6, p.340).

For applied researchers, results in this paper clearly plead for a great care when choosing the appropriate set of control variables for matching. Applying symmetric DID-matching excluding past outcomes of the control set would be a useful robustness test. The most useful advice that can be given to applied researchers from the perspective of this paper is to specify a model of outcomes and of selection into treatment, and to simulate, in the particular case at hand, the sensitivity of matching and DID-matching to different misspecifications. For example, in the case of the effect of job training programs on wages, the results in this paper clearly plead for the use of symmetric DID-matching as the least biased estimator.

This work can naturally be extended to controlling for covariates that are not past outcomes, but are correlated to them, like quasi-fixed factors when evaluating the effect of a treatment on variable factors. For example, in the evaluation of investment or hiring subsidies, should we control for past values of the capital stock? A second extension to this work would include examining the relative performances of other estimators when faced with the same problem: the Change in Change (CIC) estimator of Athey and Imbens (2006), the exchangeable estimator of Altonji and Matzkin (2005).

Finally, two extensions are worth mentioning. First, it would be interesting to assess the robustness of the conclusions in this paper to alternative specifications of the wage process. Baker (1997), Guvenen (2007, 2009) and Browning, Ejrnaes, and Alvarez (2010) argue persuasively that MaCurdy (1982)’s random walk specification is not supported by the data, and that the inclusion of random coefficients (a random trend for the first two authors) captures much nicely properties of the wage process. Second, the approach in this paper suggests an alternative estimation strategy: using estimates of the wage process along the lines developed in these papers, it could be possible to simulate the counterfactual wages and then estimate treatment on the treated. Whether or not identification of the information set of participants and of the correlation between random parameters and participation is needed to perform this task is left for further research.
References


A Proofs of results in section 3

A.1 Proof of lemma 1

First, it is convenient to write $Y_{ik}$ as a function of $Y_{i,k-\tau}$ by repeatedly substituting for the value of current outcomes as a function of past outcomes:

$$Y_{ik}^0 = \mu_i \frac{1 - \rho^\tau}{1 - \rho} + \sum_{j=0}^{\tau-1} \rho^j (\delta_{k-j} + \epsilon_{i,k-j}) + \rho^\tau Y_{i,k-\tau}^0. \quad (39)$$

Then it is easy to calculate the covariance between the selection index and $Y_{i,k-\tau}$:

$$\text{Cov}(Y_{i,k-\tau}, D_{ik}^*) = \text{Cov}(Y_{i,k-\tau}, \theta \mu_i + \beta Y_{ik}^0 + u_{ik}) \quad (40)$$

$$= \left( \theta + \beta \frac{1 - \rho^\tau}{1 - \rho} \right) \text{Cov}(\mu_i, Y_{i,k-\tau}^0) + \beta \rho^\tau \text{Var}(Y_{i,k-\tau}^0). \quad (41)$$

Because the process is stationary, variances and covariances calculated at different periods are identical. This proves the first part of the lemma.

By using the same trick, it is possible to compute the forward covariance:

$$\text{Cov}(Y_{i,k+\tau}^0, D_{ik}^*) = \text{Cov}(\mu_i \frac{1 - \rho^\tau}{1 - \rho} + \sum_{j=0}^{\tau-1} \rho^j (\delta_{k+\tau-j} + \epsilon_{i,k+\tau-j}) + \rho^\tau Y_{i,k}^0, \theta \mu_i + \beta Y_{ik}^0 + u_{ik}) \quad (42)$$

$$= \frac{1 - \rho^\tau}{1 - \rho} \theta \text{Var}(\mu_i) + \theta \rho^\tau \text{Cov}(\mu_i, Y_{ik}^0) + \beta \frac{1 - \rho^\tau}{1 - \rho} \text{Cov}(\mu_i, Y_{i,k-\tau}^0) + \beta \rho^\tau \text{Var}(Y_{ik}^0). \quad (43)$$

The result can be proved by calculating the covariance between $\mu_i$ and $Y_{ik}$ and relating it to the variance of $\mu_i$. Note that, because the process is going on for an infinite amount of time, we have:

$$Y_{it}^0 = \frac{\mu_i}{1 - \rho} + \sum_{j=0}^\infty \rho^j (\delta_{t-j} + \epsilon_{i,t-j}). \quad (44)$$

Thus, $\text{Cov}(Y_{it}^0, \mu_i) = \frac{\text{Var}(\mu_i)}{1 - \rho}$. By substituting for $\text{Var}(\mu_i)$ in equation (43), we have:

$$\text{Cov}(Y_{i,k+\tau}^0, D_{ik}^*) = \left( \theta + \beta \frac{1 - \rho^\tau}{1 - \rho} \right) \text{Cov}(\mu_i, Y_{ik}^0) + \beta \rho^\tau \text{Var}(Y_{ik}^0), \quad (45)$$
which proves the result.

### A.2 Proof of lemma 2

Using the well known formulas for the expectation of a truncated bivariate normal random variable (Tallis 1961; Heckman 1979), we have, forall $\tau$ in $\mathbb{Z}$:

\[
E[ Y_{i,k+\tau}^0 | D_{ik} = 1 ] = E[ Y_{i,k+\tau}^0 | D_{ik}^* < 0 ] = E[ Y_{i,k+\tau}^0 ] - \frac{\text{Cov}(Y_{i,k+\tau}^0, D_{ik}^*)}{\sqrt{\text{Var}(D_{ik}^*)}} \frac{\phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)}{1 - \Phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)},
\]

and:

\[
E[ Y_{i,k+\tau}^0 | D_{ik} = 0 ] = E[ Y_{i,k+\tau}^0 | D_{ik}^* \geq 0 ] = E[ Y_{i,k+\tau}^0 ] + \frac{\text{Cov}(Y_{i,k+\tau}^0, D_{ik}^*)}{\sqrt{\text{Var}(D_{ik}^*)}} \frac{\phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)}{\Phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)}.
\]

We thus have:

\[
E \left[ Y_{i,k+\tau}^0 - Y_{i,k-\tau}^0 | D_{ik} = 1 \right] = E \left[ Y_{i,k+\tau}^0 \right] - E \left[ Y_{i,k-\tau}^0 \right] - \frac{\text{Cov}(Y_{i,k+\tau}^0, D_{ik}^*) - \text{Cov}(Y_{i,k-\tau}^0, D_{ik}^*)}{\sqrt{\text{Var}(D_{ik}^*)}} \frac{\phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)}{1 - \Phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)},
\]

and:

\[
E \left[ Y_{i,k+\tau}^0 - Y_{i,k-\tau}^0 | D_{ik} = 0 \right] = E \left[ Y_{i,k+\tau}^0 \right] - E \left[ Y_{i,k-\tau}^0 \right] + \frac{\text{Cov}(Y_{i,k+\tau}^0, D_{ik}^*) - \text{Cov}(Y_{i,k-\tau}^0, D_{ik}^*)}{\sqrt{\text{Var}(D_{ik}^*)}} \frac{\phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)}{\Phi \left( \frac{D_{ik}^*}{\sqrt{\text{Var}(D_{ik}^*)}} \right)}.
\]

Applying lemma 1 proves the result.
Derivation of bias terms in the labor example with normal MA terms

B.1 Not controlling for past outcomes

It is possible to derive closed form expressions for the bias terms of section 4.2 if we make the assumption that the i.i.d MA terms are normal with variance \( \sigma^2 \). I assume that the process generating the outcomes as begun sufficiently far in the past so that I can abstract from the dependence on \( t \) by considering that the MA terms are a sum of an infinite number of shocks. I moreover pose that \( \alpha_i = \alpha \) is a constant, and that \( c_i \) is a normal variable with variance \( \sigma^2_c \), independent of \( \mu_i \), whose variance is \( \sigma^2_\mu \). To obtain the biased terms, I study the joint distribution of for normal variables conditional on \( X_i = x \), under the assumption that \( X_i \) is independent from \( \mu_i \). The observed variables are:

\[
D^*_{ik} = c_i - \frac{\alpha_i}{r} + \mathbb{E}[Y^0_{ik}|I_{ik}], \quad (52)
\]

\[
Y^0_{ik-\tau'} = g(x, d_{k-\tau'}) + \mu_i + U_{ik-\tau'}, \quad (53)
\]

\[
Y^0_{ik+\tau} = g(x, d_{k+\tau'}) + \mu_i + U_{ik+\tau}. \quad (54)
\]

These variables are normally distributed, so:

\[
\mathbb{E}[Y^0_{it}|D^*_{ik} = 1] = \mathbb{E}[Y^0_{it}] + \frac{\text{Cov}(Y^0_{it}, D^*_{ik})}{\text{Var}(D^*_{ik})} \left( \mathbb{E}[D^*_{ik}|D^*_{ik} \leq 0] - \mathbb{E}[D^*_{ik}] \right) \quad \quad (55)
\]

In order to derive closed form expressions for the bias terms of DID-matching, we need to derive the joint distribution of the following error terms:

\[
U_{ik}, U_{ik-\tau'}, \{v_{ik-j}\}_{0 \leq j \leq \tau'-1}, v_{ik-\tau'}, v_{ik-\tau'-1}.
\]

\[\text{I keep the conditioning on } X_i \text{ implicit in the remaining of the section to save on the notation.}\]
This distribution is a centered normal with covariance matrix $\Sigma_1$:

$$
\Sigma_1 = \begin{pmatrix}
\sigma_U^2 & \sigma_{U_k,U_{k-\tau'}} & \sigma_U^2 \\
\sigma_{U_k,v_{i,j}} & 0 & \sigma^2 \\
\sigma_{U_i,v_{i-\tau'}} & \sigma^2 & 0 & \sigma^2 \\
\sigma_{U_i,v_{i-\tau'-1}} & \sigma^2(m_1 + \rho) & 0 & 0 & \sigma^2
\end{pmatrix},
$$

with $\sigma_U^2 = \sigma^2 \left( 1 + (m_1 + \rho)^2 + \frac{(\rho^2 + \rho m_1 + m_2)^2}{1-\rho^2} \right)$, $\sigma_{U_k,U_{k-\tau'}} = \rho^{\tau'} \sigma_U^2 + \sigma^2 \rho^{\tau'-2} \left( \rho(m_1 + m_2)(m_1 + \rho) + m_2 \right)$ and $\sigma_{U_i,v_{i,j}} = \rho^{j-2} (\mathbb{1} [j \neq 0] \mathbb{1} [j \neq 1] m_2 + \mathbb{1} [j \neq 0] \rho m_1 + \rho^2) \sigma^2$. The latter term comes from the fact that the shocks at $t$ and $t - 1$ are not fully adjusted for at period $t$: the first shock only enters directly while the second shock enters through the AR term and the MA term. All previous shocks enter in the same way, according to a weighted average of the ARMA terms. To obtain $\sigma_U^2$, note that $U_{it} = (\rho^2 + \rho m_1 + m_2) \sum_{j=0}^{\infty} \rho^j v_{it-j-2} + v_{it} + (m_1 + \rho)v_{it-1}$. The variance of $U_{it}$ is the sum of the variances of these three terms. The variance of the first term is:

$$
\text{Var}(\sum_{j=0}^{\infty} \rho^j v_{it-j-2}) = \sigma^2 \sum_{j=0}^{\infty} \rho^{2j}. \quad \text{As } \rho^2 < 1 \text{ we can write: } \sum_{j=0}^{\infty} \rho^{2j} = \frac{1}{1-\rho^2}, \text{ which gives the result.}
$$

To obtain $\sigma_{U_k,U_{k-\tau'}}$, note that $\sigma_{U_{k+\tau'},U_{k-\tau'}} = \text{Cov}(\rho^{\tau+\tau'} U_{ik-\tau'} + \sum_{j=0}^{\tau'+\tau'-1} \rho^j U_{ik-j}, U_{i,k-\tau'})$. This leads to the following formula:

$$
\sigma_{U_{k+\tau'},U_{k-\tau'}} = \rho^{\tau+\tau'} \sigma_U^2 + \rho^{\tau+\tau'-2} \sigma^2 \left( \mathbb{1} [\tau + \tau' > 0] \rho(m_2(m_1 + \rho) + m_1) + \mathbb{1} [\tau + \tau' > 1] m_2 \right).
$$

(57)
From these expressions, we obtain the bias terms for the DID case (for $\tau, \tau' \geq 2$): \[ (58a) \]

\[
\frac{B^\text{did}}{\sigma^2 \sigma_D^2} \left( \frac{\phi(A_x)}{1 - \Phi(A_x)} + \frac{\phi(A_x)}{\Phi(A_x)} \right) = -\left( \rho^{\tau+\tau'} - 1 \right) \rho^{\tau'-2} \left( \rho^2 \frac{\sigma_U^2}{\sigma^2} + \rho m_1 + m_2 \rho^2 (m_1 + \rho) + m_2 \right)
\]

\[- \rho^{\tau+2\tau'-4} \left( \rho m_1 + m_2 (\rho^2 + 1) \right) (m_2 + \rho + m_1 + \rho^2) \] \[ (58b) \]

\[- \rho^{\tau-2} \left( \rho^2 + \rho m_1 + m_2 \right) \left( 1 - \rho^{2\tau'} \rho^2 - m_1 \rho - \frac{1}{1 - \rho^2} + m_2 \rho^2 \right) \]

\[ (58c) \]

\[
\frac{\sigma_D^2}{\sigma^2} \left( \frac{\phi(A_x)}{1 - \Phi(A_x)} + \frac{\phi(A_x)}{\Phi(A_x)} \right) = \frac{\sigma_D^2}{\sigma^2} \left( \frac{\phi(A_x)}{1 - \Phi(A_x)} + \frac{\phi(A_x)}{\Phi(A_x)} \right) = \sum_{j=0}^{\tau'-1} \rho^{j} \frac{\sigma_{v_{k,v_{k-j}}}^2}{\sigma^2} = \sum_{j=0}^{\tau'-1} \rho^{j} (1 - \mathbb{I}_{j \neq 0}) (m_2 + 1 - \mathbb{I}_{j \neq 0} \rho m_1 + \rho^2) = \sum_{j=0}^{\tau'-1} \rho^{j} (\rho^2)^j + m_1 \rho \sum_{j=0}^{\tau'-1} \rho^{j} (\rho^2)^j + m_2 \rho^2 \sum_{k=0}^{\tau'-3} (\rho^2)^k \text{ if the decision maker does not know period } k \text{ information (i.e. shock } v_{ik} \text{) when deciding to enter the program, the first term of the previous sum is changed to } \sum_{j=1}^{\tau'-1} (\rho^2)^j = \rho^2 \sum_{j=1}^{\tau'-1} (\rho^2)^j = \rho^2 \sum_{k=0}^{\tau'-2} (\rho^2)^k = \rho^2 \frac{1 - \rho^{2\tau'-2}}{1 - \rho^2}.
\]

These bias terms have the expected signs: the first term \[ (58a) \] is positive because $\rho < 1$ and $\frac{\sigma_D^2}{\sigma^2} > 1$, so that in general $\rho \frac{\sigma_D^2}{\sigma^2} + m_1 + m_2 (m_1 + \rho)$ is positive. The second term \[ (58b) \] is positive because $m_1 < m_2 < 0$. The third term \[ (58b) \] is negative because the AR term dominates the MA term, leading to positive terms inside the brackets.

---

\[ ^{17} \] When $\tau' = 1$, the second component of the bias \[ (58b) \] is equal to $-\rho^{\tau' - 1} \left( \rho m_1 + m_2 (m_1 + \rho) + m_2 (m_2 + m_1 + \rho^2) \right)$ and the third component \[ (58c) \] is equal to $-\rho^{\tau - 2} (\rho^2 + \rho m_1 + m_2 \tau > 1) m_2$.

\[ ^{18} \] If $\tau = 1$ and $\tau' > 1$, the third component \[ (58c) \] is equal to $\rho + m_1 + \rho^2 \rho m_1 + m_2 \left( \rho^2 \frac{1 - \rho^2}{1 - \rho^2} + m_1 \frac{1 - \rho^2}{1 - \rho^2} + m_2 \rho \frac{1 - \rho^2}{1 - \rho^2} \right)$. 39
B.2 Controlling for past outcomes

To derive the bias term of matching on past outcomes, I use the fact that it can be rewritten in the following way:

\[ B^m(\tau, \tau', t, y) = \mathbb{E}[Y^0_{i,k+\tau} | D^t = 1] - \mathbb{E}[\mathbb{E}[Y^0_{i,k+\tau} | D^t_{ik} = 0, Y^0_{i,k-\tau'}] | D^t = 1]. \]  

(59)

The average outcome for the treated can be obtained by results in the previous section. The main difficulty is to form the second part of the term on the right hand side of equation (59): the mean outcome of the matched non-participants. To form this quantity, first note that, because these variables are jointly normally distributed, their conditional expectation is linear:

\[ \mathbb{E}[Y^0_{i,k+\tau} | D^t_{ik}, Y^0_{i,k-\tau'}] = \mathbb{E}[Y^0_{i,k+\tau}] + \beta_{\tau, D^t} (D^t_{ik} - \mathbb{E}[D^t_{ik}]) + \beta_{\tau, \tau'} (Y^0_{i,k-\tau'} - \mathbb{E}[Y^0_{i,k-\tau'}]), \]  

(60)

with:

\[ \beta_{\tau, D^t} = \frac{\text{Cov}(Y^0_{i,k+\tau}, D^t_{ik}) \sigma_Y - \text{Cov}(Y^0_{i,k-\tau'}, D^t_{ik}) \sigma_{Y_{k-\tau', k-\tau'}}}{\sigma_{D^t}^2 - \text{Cov}(Y^0_{i,k-\tau'}, D^t_{ik})^2}, \]  

(61)

\[ \beta_{\tau, \tau'} = \frac{\sigma_{Y_{k+\tau}, Y_{k-\tau'}} \sigma_{D^t}^2 - \text{Cov}(Y^0_{i,k+\tau}, D^t_{ik}) \text{Cov}(Y^0_{i,k-\tau'}, D^t_{ik})}{\sigma_{D^t}^2 \sigma_Y^2 - \text{Cov}(Y^0_{i,k-\tau'}, D^t_{ik})^2}, \]  

(62)

\[ \sigma_{Y_{k+\tau}, Y_{k-\tau'}} = \sigma_{U_{k+\tau}, U_{k-\tau'}} + (1 - \rho_{\tau+\tau'}) \sigma_{\mu}^2. \]  

(63)

\[ \sigma_{D^t}^2 = \sigma_U^2 - \mathbf{1}[t = l] * \sigma^2 + \sigma_{\mu}^2 + \sigma_c^2 + \frac{\sigma^2}{\tau^2} + 2 * (\sigma_{\mu,c} - \frac{\sigma_{\mu,\alpha}}{\rho} - \frac{\sigma_{c,\alpha}}{\rho}), \]  

(64)

\[ \sigma_Y^2 = \sigma_U^2 + \sigma_{\mu}^2. \]  

(65)

From this, we again use the law of iterated expectation to derive the conditional expectation of non-participants’ outcomes:

\[ \mathbb{E}[Y^0_{i,k+\tau} | D^t_{ik} = 0, Y^0_{i,k-\tau'}] = \mathbb{E}[\mathbb{E}[Y^0_{i,k+\tau} | D^t_{ik}, Y^0_{i,k-\tau'}] | D^t_{ik} > 0, Y^0_{i,k-\tau'}] \]  

(66)

\[ = \mathbb{E}[Y^0_{i,k+\tau}] + \gamma_{\tau, \tau'} (Y^0_{i,k-\tau'} - \mathbb{E}[Y^0_{i,k-\tau'}]) + \gamma_{\tau, D^t} \phi(A_{xy}) / \Phi(A_{xy}), \]  

(67)
with:

\[
\gamma_{\tau, \tau'} = \beta_{\tau, D^*} \frac{\text{Cov}(Y^0_{i,k-\tau'}, D^*_ik)}{\sigma^2_Y} + \beta_{\tau', \tau}
\]

\[
\gamma_{\tau, D^*} = \beta_{\tau, D^*} \sqrt{\frac{\sigma^2_{D^*}}{\sigma^2_Y} - \frac{\text{Cov}(Y^0_{i,k-\tau'}, D^*_ik)^2}{\sigma^4_Y}},
\]

\[
A_{xy} = \bar{c} - \frac{\alpha}{\tau} + g(x, d_k) + (y - g(x, d_{k-\tau'})) \frac{\text{Cov}(Y^0_{i,k-\tau'}, D^*_ik)}{\sigma^2_Y}. \]

(68)

(69)

(70)

In order to obtain bias terms that are comparable to those calculated for DID matching, we have to integrate \(B_{m1}^{xy}\) and \(B_{m2}^{xy}\) with respect to the distribution \(F_{Y_{i,k-\tau'}|D_{ik}=1}(y)\). This distribution has the following density [Arnold, Beaver, Groeneveld, and Meeker 1993]:

\[
f_{Y_{i,k-\tau'}|D_{ik}=1}(y) = \frac{1}{\sigma_Y} \phi \left( \frac{y - g(x, d_{k-\tau'})}{\sigma_Y} \right) \frac{1 - \Phi(A_{xy})}{1 - \Phi(A_x)}. \]

(71)

We can derive expressions for the unconditional bias term after integrating out \(Y_{i,k-\tau'}|D_{ik} = 1\):

\[
\mathbb{E}[\mathbb{E}[Y^0_{i,k+\tau}|D^*_ik = 0, Y^0_{i,k-\tau}]|D^* = 1] = \mathbb{E}[Y^0_{i,k+\tau}] - \gamma_{\tau, \tau'} \frac{\text{Cov}(Y^0_{i,k-\tau'}, D^*_ik)}{\sigma_{D^*}} \frac{\phi(A_x)}{1 - \Phi(A_x)}
\]

\[
+ \gamma_{\tau, D^*} \int_{-\infty}^{+\infty} \frac{1}{\sigma_Y} \phi(A_{xy}) \frac{1 - \Phi(A_{xy})}{\sigma_Y} (y - g(x, d_{k-\tau'})) \phi \left( \frac{y - g(x, d_{k-\tau'})}{\sigma_Y} \right) dy. \]

(72)

There is no closed form expression for the last integral. I use 32-point Gauss-Hermite quadrature to compute this integral numerically.