Mora, R., & Reggio, I. (2017). Alternative diff-in-diffs estimators with several pretreatment periods. *Econometric Reviews*, 38 (5), pp. 465-486.

DOI: 10.1080/07474938.2017.1348683

Alternative Diff-in-Diffs Estimators with Several Pre-treatment Periods*

Ricardo Mora† Iliana Reggio‡

This version: March 2017

Abstract

This paper deals with the identification of treatment effects using difference-in-differences estimators when several pre-treatment periods are available. We define a family of identifying non-nested assumptions that lead to alternative difference-in-differences estimators. We show that the most usual difference-in-differences estimators imply equivalence conditions for the identifying non-nested assumptions. We further propose a model that can be used to test multiple equivalence conditions without imposing any of them. We conduct a Monte Carlo analysis and apply our approach to several recent papers to show its practical relevance.

Keywords: common trends, difference-in-differences, identification, treatment effect
JEL Classification: C21 C52

*We gratefully acknowledge the financial support by the Ministerio Economía y Competitividad (Spain), through grant ECO2015-65204-P. Ricardo Mora also acknowledges the financial support from the grants MDM 2014-0431, and MadEco-CM (S2015/HUM-3444). We are grateful to Julio Cáceres, Rodolfo Campos, Raquel Carrasco, Jesús Carro, Jeff Friedman, Jeff Furman, Bas van der Klaauw, and the audiences at the 25th EALE annual conference (Torino 2013), the 68th European Meeting of the Econometric Society (Toulouse 2014), the 1st annual conference of the IAAE (London 2014), and research seminars at the Institut für Volkwirtschaftslehre (University of Graz) and the Departament of Economics (UC3M).

†Corresponding author: Ricardo Mora, Universidad Carlos III de Madrid, Department of Economics, Calle Madrid 126, 28903 Getafe (Madrid), Spain. E-mail: ricmora@eco.uc3m.es

‡Department of Economics, Universidad Carlos III, Calle Madrid 126,28903 Getafe (Madrid), Spain

1 Introduction

Difference-in-differences (DID) estimators are a standard econometric tool widely used to evaluate the impact of a specific treatment on an outcome of interest (for a recent review of the literature, see Imbens and Wooldridge, 2009). In its simplest implementation, only data from two periods and two groups (treated and controls) are available. In the first period—the pre-treatment period—none of the groups is exposed to the treatment. In the second period—the post-treatment period—the treated are exposed to the treatment, whereas the controls are not. The appropriateness of the DID technique depends crucially on the non-testable Parallel Paths assumption. Parallel Paths assumes that the average change in the outcome variable for the treated in the absence of treatment is equal to the equivalent average change in the outcome variable for the controls.

We focus on applications in which several pre-treatment periods are available. In this context, most researchers apply an extension of the Parallel Paths assumption, usually referred to as the *common trends* assumption.[1] Under *common trends*, in the absence of treatment the average outcome change *from any pre-treatment period to any post-treatment period* for the treated is equal to the equivalent average outcome change for the controls. Unlike Parallel Paths, *common trends* implies the testable condition that treated and controls have the same average outcome growth in each pre-treatment period (as argued by Angrist and Krueger, 1999). In practice, researchers who detect pre-treatment trend differentials include linear or quadratic trend polynomials to account for these differentials (see, for example, Jacobson et al., 1993, Friedberg, 1998 and Wolfers, 2006). This strategy is routinely motivated either as a relaxation of the *common trends* assumption or as a robustness check exercise. However, there is never

[1] For a clear presentation of this terminology see Blundell et al., 2004, p. 578.

a strict characterization of the implicit assumptions that identify the effect for the new specifications and whether they contain testable restrictions. This is a serious gap in the DID literature whose importance goes beyond the interest of knowing which assumptions we rely on to identify the effect. In practice, the presence of testable restrictions can be exploited to minimize the risk of imposing invalid assumptions and, therefore, the risk of using inconsistent estimators of the effects.

We fill this gap and characterize the set of assumptions that identify the treatment effect for any econometric specification. To do so, we define a family of alternative Parallel assumptions. We refer to them as Parallel-(q) with q ranging from 1 to the number of pre-treatment periods. Intuitively, Parallel-(q) assumes that, in the absence of treatment, the average q-th difference in the outcome variable after treatment is equal for controls and treated. Each Parallel-(q) assumption is non-testable in isolation because it refers to the behavior of the treated in a scenario that is not observable. Hence, for alternative Parallel-(q) assumptions it is only possible to test if their identification of the treatment effect is equivalent. We show that the Parallel-(q) assumptions are non-nested in the sense that one does not imply any of the others. We also show that any of them is sufficient to obtain identification of the treatment effect.

We further establish equivalence conditions for any two consecutive Parallel-(q) assumptions. Using these results, we discuss the most usual econometric specifications and show that they always impose the equivalence of several Parallel-(q) assumptions. Thus, the most usual econometric specifications impose restrictions that result in over-identification of the treatment effect. In addition, we propose a model with fully flexible group–specific dynamics. In contrast to the most usual econometric specifications, this model can be used to test multiple equivalence conditions.

The rest of the paper is structured as follows. We define the family of alternative Parallel-

(q) assumptions in Section 2. We discuss several econometric specifications and present the fully flexible model in Section 3. In Section 4, we analyze, by means of a Monte Carlo simulation, the relative performance of four alternative models. In Section 5, we review current practice and explore the practical relevance of our proposal. Section 6 concludes. Proofs are gathered in Appendix A and the details of the review of current practice are available in Appendix B.

2 Alternative Parallel-(q) assumptions

In the simplest empirical DID application, we have two periods: one before and one after the treatment. Generally, we have $T \geq 2$ periods, treatment starts sometime after the last pre-treatment period, t^*, and finishes before the first post-treatment period, $t^* + 1$. There are at least two periods before treatment and at least one period after treatment. We first focus on the case when there is only one period after treatment and later we extend the analysis to the general case of several post-treatment periods, i.e. when $T - t^* > 1$.

Following conventional notation we define Y_{it} as the observed outcome variable for individual i at period t. Let Y_{it}^0 denote the outcome in period t when the individual receives no treatment, and Y_{it}^1 the outcome in period t when the individual receives treatment. Before treatment (i.e., $t \leq t^*$), $Y_{it}^0 = Y_{it}$, while for the post-treatment period (i.e., for $t = t^* + 1$) either Y_{it}^0 or Y_{it}^1 is observed. Let $D_i = 1$ if the individual receives treatment and $D_i = 0$ otherwise. Potential and observed outcomes are related to D_i by $Y_{it} = Y_{it}^1 D_i + Y_{it}^0 (1 - D_i)$ for $t = t^* + 1$. Finally, for notational simplicity we do not consider any covariates, but results are similar if the analysis is made conditional on a vector of covariates.

We study identification conditions for the average treatment effect on the treated,

$$\alpha = E\left[Y^1_{i,t^*+1} - Y^0_{i,t^*+1} \,|D_i = 1\right]. \tag{1}$$

The estimation of α is problematic because Y^0_{i,t^*+1} is not observable for the treated. To estimate the average counterfactual, $E\left[Y^0_{i,t^*+1} \,|D_i = 1\right]$, one needs an assumption regarding the trend behavior of the treated if untreated. For example, when $T = 2$ the DID estimator stems from the so-called Parallel Paths assumption.

The Parallel Paths assumption

For illustration of the DID identification strategy, consider the simplest case with only one period before and only one period after treatment: $T = 2$ and $t^* = 1$. Let L be the lag operator so that for any time series x_t and integer p, $L^p x_t \equiv x_{t-p}$. Let Δ denote the first difference operator so that $\Delta \equiv (1 - L)$. The Parallel Paths assumption can be expressed as follows.

Definition 1. *Parallel Paths*

$$E\left[\Delta Y^0_{i2} \,|D_i = 1\right] = E\left[\Delta Y^0_{i2} \,|D_i = 0\right]. \tag{2}$$

Parallel Paths means that, in the absence of treatment, average changes in outcome among treated are equal to the average changes among comparable controls.

Under Parallel Paths the counterfactual $E\left[Y^0_{i2} \,|D_i = 1\right]$ is built by adding the observed

average increase in the controls to the pre-treatment level of the treated:

$$E\left[Y_{i2}^{0}\,|D_i=1\right]=E\left[Y_{i1}\,|D_i=1\right] + E\left[\Delta Y_{i2}\,|D_i=0\right] \tag{3}$$

Using this counterfactual, the treatment effect, α, may be expressed as the difference in observed outcome changes among treated and controls, usually referred to as the difference-in-differences operator:

$$\alpha=E\left[\Delta Y_{i2}\,|D_i=1\right]-E\left[\Delta Y_{i2}\,|D_i=0\right] \tag{4}$$

Non-nested Parallel-(q) assumptions

Returning to the case when the number of pre-treatment periods can be larger than one, we define Parallel-(q) as follows.

Definition 2. *Parallel-(q)*

For a given q such that $q \in \{1, ..., t^*\}$,

$$E\left[\Delta^q Y_{i,t^*+1}^{0}\,|D_i=1\right]=E\left[\Delta^q Y_{i,t^*+1}^{0}\,|D_i=0\right] \tag{5}$$

where $\Delta^q \equiv (1-L)^q$.

For example, Parallel-(1) means that

$$E\left[\Delta Y_{i,t^*+1}^{0}\,|D_i=1\right]=E\left[\Delta Y_{i,t^*+1}^{0}\,|D_i=0\right].$$

Hence, in the absence of treatment, the treated experience from t^* to t^*+1 the same average outcome change as that of the controls. In the simplest case when $T = 2$, Parallel-(1) is Parallel Paths.

Parallel-(2) implies that

$$E\left[\Delta^2 Y^0_{i,t^*+1} \,|D_i = 1\right] = E\left[\Delta^2 Y^0_{i,t^*+1} \,|D_i = 0\right],$$

so that there is a shift from outcome changes to outcome accelerations. For higher q, interpretation of the assumption is similar.

Using Parallel-(q) to build the counterfactual requires data with q pre-treatment periods. Therefore, in terms of information requirements Parallel-(1) is less demanding than Parallel-(2), which is less demanding than Parallel-(3), and so on. Although data requirements are nested, assumptions are not: one Parallel-(q) assumption does not imply any of the others. In Table 1 we illustrate this fact with the simplest case of Parallel-(1) vs. Parallel-(2) with two pre-treatment periods. For three different cases, we give the expected outcomes in the absence of treatment for both treated and controls. In Case (a) Parallel-(1) is satisfied because the expected outcomes of $\Delta Y^0_{i,t^*+1}$ are invariant to treatment, $E\left[\Delta Y^0_{i,t^*+1} \,|D\right] = 1$. In contrast, $E\left[\Delta^2 Y^0_{i,t^*+1} \,|D_i = 1\right] = 0$ while $E\left[\Delta^2 Y^0_{i,t^*+1} \,|D_i = 0\right] = 1$. Thus, Parallel-(2) is violated. This situation arises when the comparability in changes between treated and controls is valid only from the last pre-treatment period, but the dynamics of controls and treated are not comparable in previous pre-treatment periods. In Case (b), Parallel-(1) is violated while Parallel-(2) is satisfied. This situation arises when treated and controls changes are not comparable, but the changes of the changes are. Case (a) and Case (b) show that Parallel-(1) and Parallel-(2) are non-nested. Nonetheless, both assumptions can be simultaneously satis-

fied, as Case (c) illustrates. Given that each higher order Parallel-(q) assumption shifts the variable of interest by taking the first difference from the previous Parallel-(q) assumption, the illustration in Table 1 in fact shows that any two Parallel-(q) assumptions are non-nested.

Table 1: *Non-nested Parallel-(q) assumptions. An illustration with Parallel-*(1) *and Parallel-*(2)

	$\mathrm{E}\left[Y^0_{t^*-1}\vert D\right]$	$\mathrm{E}\left[Y^0_{t^*}\vert D\right]$	$\mathrm{E}\left[Y^0_{t^*+1}\vert D\right]$	$\mathrm{E}\left[\Delta Y^0_{t^*+1}\vert D\right]$	$\mathrm{E}\left[\Delta^2 Y^0_{t^*+1}\vert D\right]$
Case (a)					
$D=1$	1	2	3	1	0
$D=0$	0	0	1	1	1
Case (b)					
$D=1$	1	2	4	2	1
$D=0$	0	0	1	1	1
Case (c)					
$D=1$	1	2	4	2	1
$D=0$	0	1	3	2	1

Note: Expected outcomes in the absence of treatment conditional on treatment. Case (a) illustrates a situation whereby Parallel-(1) is satisfied while Parallel-(2) is violated. Case (b) illustrates the opposite situation. In Case (c), both Parallel assumptions are satisfied.

Furthermore, there is not a general reason to favor a particular value for q. Consider a recursive representation of the potential outcomes under no treatment. Denote ΔY^0_{it} as ϵ^0_{it} so that $Y^0_{it} = Y^0_{i0} + \sum_{j=1}^{t} \epsilon^0_{ij}$. The Parallel-(1) assumption implies that ϵ^0_{i,t^*+1} is mean independent of D. The Parallel-(2) assumption shifts this condition to $\Delta\epsilon^0_{i,t^*+1}$ being mean independent of D, and so on. Parallel-(1) holds but Parallel-(2) does not if ϵ^0_{i,t^*+1} is mean independent of D but ϵ^0_{i,t^*} is not. This happens when shocks in potential outcomes at the time of treatment are mean independent of D but previous shocks are not. In contrast, shocks could fail to be mean independent of D at time of treatment

(and before), but the change in the shocks, $\Delta\epsilon_{i,t}^0$, could satisfy the condition for $t = t^*+1$. In that case, Parallel-(2) would hold while Parallel-(1) would fail.

Identification under each Parallel-(q)

Define $did\,(q)$ as the difference-in-q-differences operator one period ahead,

$$did\,(q) \equiv E\left[\Delta^q Y_{i,t^*+1}\,|D_i = 1\right] - E\left[\Delta^q Y_{i,t^*+1}\,|D_i = 0\right].$$

Theorem 1. *For any $q = 1, ..., t^*$ under Parallel-(q),*

$$\alpha = did\,(q)\,.$$

Proof. See Theorem A1 in Appendix A for the proof of the general case when there may be more than one post-treatment period. □

The treatment effect α is the difference-in-q-differences operator under Parallel-(q). Since the difference-in-q-difference operator is observable, each Parallel-(q) assumption in isolation is sufficient to identify the treatment effect.

Equivalence conditions

Summing up, Parallel-(q) assumptions are non-nested and any Parallel-(q) assumption suffices to identify the treatment effect. By Theorem 1 we have established the connection between the $did\,(q)$ operator and the treatment effect under Parallel-(q). We now study under what conditions consecutive $did\,(q)$ operators—and hence the identification of the treatment effect under consecutive Parallel-(q) assumptions—are equivalent.

Theorem 2. *For any* $q = 2, ..., t^*$,

$$did\,(q) = did\,(q-1)$$

if and only if

$$E\left[\Delta^{q-1} Y_{it^*} \,|D_i = 1\right] = E\left[\Delta^{q-1} Y_{it^*} \,|D_i = 0\right]. \tag{6}$$

Proof. See Theorem A2 in Appendix A for the proof for the general case when there may be more than one post-treatment periods. □

Because by Theorem 1 under Parallel-(q) α is identified by the difference-in-q-differences operator, Theorem 2 sets pre-treatment trend conditions for which assumptions Parallel-(q) and Parallel-($q-1$) are equivalent and gives a testable implication of assuming Parallel-(q) and Parallel-($q-1$) jointly. For example, Parallel-(1) and Parallel-(2) are not equivalent when $E\left[\Delta Y_{it^*} \,|D_i = 1\right] \neq E\left[\Delta Y_{it^*} \,|D_i = 0\right]$, i.e. in the presence of group-specific outcome changes in the last pre-treatment period. More specifically, in the special case when $T = 3$ and $t^* = 2$, Parallel-(1) is equivalent to Parallel-(2) if the difference-in-difference operator at $t = 2$ is equal to zero, i.e. if $E\left[\Delta Y_{i,2} \,|D_i = 1\right] = E\left[\Delta Y_{i,2} \,|D_i = 0\right]$. In the empirical literature, this condition is usually presented as the null in a placebo test of the DID assumption.

Theorem 2 implies that controls and treated have common pre-treatment dynamics if and only if all Parallel-(q) assumptions are equivalent. This result has two implications. First, the comparison of pre-treatment trends for treated and controls frequently implemented in empirical work may be regarded as an informal test for the equivalence of all parallel assumptions between Parallel-(1) and Parallel-(t^*). Second, assuming that common dynamics remain after treatment in the absence of treatment—the *common*

trends assumption—is the same as assuming that $\Delta Y_{it}^0 = \epsilon_{it}^0$ is mean independent of treatment at all pre-treatment periods.

The case with more than one post-treatment period

When there are more than one post-treatment period, the effect of the treatment can be evaluated s periods after treatment, where $1 \leq s \leq T - t^*$:

$$\alpha(s) = E\left[Y_{i,t^*+s}^1 - Y_{i,t^*+s}^0 \,|D_i = 1\right]. \tag{7}$$

Identification of treatment after s periods relies on a general version of Parallel-(q) such that for a given pair of positive integers (q, s) with $1 \leq q \leq t^*$ and $1 \leq s \leq T - t^*$,

$$E\left[\Delta_s \Delta^{q-1} Y_{i,t^*+s}^0 \,|D_i = 1\right] = E\left[\Delta_s \Delta^{q-1} Y_{i,t^*+s}^0 \,|D_i = 0\right] \tag{8}$$

where $\Delta_s = (1 - L^s)$. In what follows we refer to Parallel-(q) as the assumption under which equation (8) holds for all $s \in \{1, T - t^*\}$. Define $did(q, s)$ as the difference-in-q-differences operator s periods ahead,

$$did(q, s) \equiv E\left[\Delta_s \Delta^{q-1} Y_{i,t^*+s} \,|D_i = 1\right] - E\left[\Delta_s \Delta^{q-1} Y_{i,t^*+s} \,|D_i = 0\right]. \tag{9}$$

In Theorem A1 in Appendix A we generalize Theorem 1 for the case with possibly several post-treatment periods and show that for a given positive integer $q \leq t^*$, and for any s, $1 \leq s \leq T - t^*$, under Parallel-(q), the treatment effect $\alpha(s)$ is identified as the solution of $\Delta^{q-1}\alpha(s) = did(q, s)$. Moreover, under Parallel-(q) there is an unique solution to this equation such that $\alpha(s) = 0$, denoted by $\alpha^q(s)$ (see Corollary A1 in Appendix). Finally,

the equivalence conditions for the identification of $\alpha(s)$ do not depend on s. To see this, in Theorem A2 in Appendix A we generalize the result from Theorem 2 and show that for any q such that $2 \leq q \leq t^*$ and a given s such that $1 \leq s \leq T - t^*$, $\alpha^q(s) = \alpha^{q-1}(s)$ if and only if

$$E\left[\Delta^{q-1} Y_{it^*} | D_i = 1\right] = E\left[\Delta^{q-1} Y_{it^*} | D_i = 0\right]. \tag{10}$$

3 Regression techniques

In empirical applications, there usually are several pre-treatment and several post-treatment periods. In this section we review the usual empirical strategies focusing the attention for simplicity on the case when there is only one post-treatment period. We show how these strategies imply testable conditions related to the equivalence of alternative Parallel-(q) assumptions.

The standard model

Treatment effects are frequently estimated using standard linear regression techniques. In the simplest case with only two periods, the treatment effect estimate is obtained from a regression that includes a constant, the treated indicator D_i, a dummy variable for the post-treatment period, $Post_t$, and an interaction term, $Post_t \times D_i$. In this set up, the treatment effect is identified by the parameter associated with the interaction term.

When several pre-treatment periods are available, the conditional expectation will include time dummies for all periods and an interaction between the treatment indicator and the post-treatment time dummy (see, for example, Imbens and Wooldridge, 2009).

Assume that the conditional mean function has the following form:

$$E\left[Y_{it}\left|D_i = D\right.\right] = \delta + \sum_{\tau=2}^{T} \delta_\tau I_t^\tau + \gamma D + \gamma_T \times I_t^T \times D \tag{11}$$

where I_t^τ takes value one if $t = \tau$ and zero otherwise and $D \in \{0, 1\}$. Equation (11) imposes that pre-treatment dynamics—captured by time dummies—are identical between the controls and treated. Identifying α with γ_T implies the *common trends* assumption and the equivalence of all Parallel-(q) assumptions. Formally, by applying Δ^q on both sides of equation (11) and taking into account that $\Delta^q \gamma_T = \gamma_T$, we have that

$$did\,(q) = \gamma_T. \tag{12}$$

By Theorem 1, $\alpha = \gamma_T$ under any parallel assumption. When there is more than one post-treatment period, the usual practice is to interact the treatment variable with a post-treatment dummy. In that case, the long-term effect of treatment is already present at $t^* + 1$. In some applications where the aim is to identify dynamic effects, researchers include interactions between the treatment dummy and post-treatment time dummies (see, for example, Furman and Stern, 2011). These specifications still impose the common trends assumption if the treatment effects are identified as the interaction parameters.

A polynomial trend for the differences in group dynamics

An essential aspect in the previous model is that pre-treatment dynamics are identical for controls and treated. Researchers accommodate group-specific trends by adding in equation (11) an interaction between the treatment indicator and a polynomial time

trend of order $R < t^*$. Let us consider the leading case of a linear polynomial:

$$E\left[Y_{it}\left|D_i = D\right.\right] = \delta + \sum_{\tau=2}^{T} \delta_\tau I_t^\tau + \gamma D + \gamma_T \times I_t^T \times D + \gamma_1^p \times t \times D \tag{13}$$

In equation (11) pre-treatment dynamics are identical for controls and treated. By contrast, in equation (13) the term $\gamma_1^p \times t$ captures differences in group dynamics that predate treatment and remain after treatment.

By Theorem 1 we have that under Parallel-(q),

$$\alpha = \gamma_T + \gamma_1^p \Delta^q \left(t^* + 1\right). \tag{14}$$

so that the treatment effect will generally not be identified as the interaction term γ_T. Because $\Delta^q \left(t^* + 1\right) = 0$ for all $q > 1$, $\alpha = \gamma_T$ if $q > 1$. Hence, the interaction term γ_T identifies the treatment effect α if we assume all assumptions between Parallel-(2) and Parallel-(t^*).

In some empirical applications, researchers include higher order polynomials:

$$E\left[Y_{it}\left|D_i = D\right.\right] = \delta + \sum_{\tau=2}^{T} \delta_\tau I_t^\tau + \gamma D + \gamma_T \times I_t^T \times D + \sum_{r=1}^{R} \gamma_r^p \times t^r \times D. \tag{15}$$

Now the term $\sum_{r=1}^{R} \gamma_r^D \times t^r$ captures differences in group dynamics that predate treatment and remain after treatment. Equation (14) now becomes

$$\alpha = \gamma_T + \sum_{r=1}^{R} \gamma_r^p \Delta^q \left(t^* + 1\right)^r$$

so that the treatment effect will generally not be identified as the interaction term γ_T.

Because $\Delta^q (t^* + 1)^r = 0$ for all $q > r$, $\alpha = \gamma_T$ if $q > R$. For example, if we include a quadratic polynomial interacted with treatment when there are more than two pre-treatment periods, $t^* > 2$, then $\alpha = \gamma_T$ if we assume all parallel assumptions between Parallel-(3) and Parallel-(t^*). More generally, in a model with a polynomial of order R, $\alpha = \gamma_T$ if we assume all parallel assumptions between Parallel-$(R + 1)$ and Parallel-(t^*).[2]

To summarize our results, the inclusion of group-specific trends in the standard model relaxes the *common trends* assumption by not imposing at least one Parallel assumption. When a linear trend is introduced, Parallel-(1) is not imposed but the validity of all other parallel assumptions is assumed. When a quadratic trend is included, then both Parallel-(1) and Parallel-(2) are disregarded but all other parallel assumptions are still imposed. In the presence of many pre-treatment periods, this strategy may still impose too many equivalence conditions. In the next subsection, we present a model that allows for more flexible identification strategies.

Fully flexible group–specific dynamics

All econometric specifications discussed so far identify the treatment effect imposing the equivalence of several Parallel-(q) assumptions. We know from Theorem 1 that every Parallel-(q) assumption can independently identify the treatment effect. Thus, all these econometric models are imposing more restrictions than necessary to identify the effect. We propose a model that allows separate identification of the treatment effect under each Parallel-(q) assumption and testing of equivalence conditions.

[2]When there are more than one post-treatment period, the same result applies to the identification of $\alpha(s)$ with parameters γ_s associated to interactions of treatment with post-treatment time dummies. By Theorem A1 in Appendix A we have that under Parallel-(q), $\Delta^{q-1}\alpha^q(s) = \Delta^{q-1}\gamma_s + \sum_{r=1}^{R} \gamma_r^p \Delta_s \Delta^{q-1} (t^* + s)^r$. The identification of the treatment effect s periods ahead is complex to evaluate for $q \leq R$ because there is no closed-form solution for $\Delta_s \Delta^{q-1} (t^* + s)^r$. For $q > R$, $\Delta_s \Delta^{q-1} (t^* + s)^r = 0$ for all r and $\alpha^q(s) = \gamma_s$. Thus, the identification of the treatment effect with γ_s imposes the equivalence of all parallel assumptions between Parallel-$(R + 1)$ and Parallel-(t^*).

Consider a general additive model with fully flexible groups–specific dynamics:[3]

$$E\left[Y_{it}\left|D_i = D\right.\right] = \delta + \sum_{\tau=2}^{T} \delta_\tau I_t^\tau + \gamma D + \sum_{\tau=2}^{T} \gamma_\tau \times I_t^\tau \times D. \tag{16}$$

The standard model in equation (11) is a particular case where $\gamma_\tau = 0$ for all $\tau < T$. For brevity, we refer to equation (16) as the fully flexible model. Equation (16) does not impose parametric assumptions on trends for treated or controls. This flexibility allows identifying the treatment effect separately under each Parallel-(q) assumption:

Theorem 3. *Under Parallel-(q) and equation* (16)*:*

$$\alpha = \Delta^q \gamma_T \tag{17}$$

Proof. See Theorem A3 in Appendix A for the proof in the general case when there may be more than one post-treatment period. □

Theorem 3 identifies the treatment effect as a linear combination of the original parameters of the model. The effect of treatment will generally differ under alternative Parallel-(q) assumptions. For example, under Parallel-(1), $\alpha = \gamma_T - \gamma_{T-1}$ while under Parallel-(2), $\alpha = \gamma_T - 2\gamma_{T-1} + \gamma_{T-2}$.

From Theorems 2 and 3 we can also identify the restrictions on the model parameters that imply the same Parallel-(q) assumptions. Consider the equivalence of two consecutive parallel assumptions Parallel-(q) and Parallel-($q+1$). They are equivalent if $\Delta^q \gamma_T = \Delta^{q+1} \gamma_T$ which—given that $\Delta^{q+1} \gamma_T = \Delta^q \gamma_T - \Delta^q \gamma_{T-1}$—implies that $\Delta^q \gamma_{T-1} = 0$. In the case of equivalence between Parallel-(1) and Parallel-(2), the null hypothesis is H_0 :

[3]We are aware of only one empirical work—Reber, 2005—that proposes a flexible model by introducing interactions of treatment with time dummies for a limited number of periods before treatment and several periods after treatment.

$\gamma_{T-1} = \gamma_{T-2}$. Following a similar argument, Parallel-(1) is equivalent to Parallel-(2) and Parallel-(3) if and only if $\gamma_{T-1} = \gamma_{T-2} = \gamma_{T-3}$. More generally, all parallel assumptions between Parallel-(1) and Parallel -(q) are equivalent if and only if $\gamma_{T-1} = \gamma_{T-2} = ... = \gamma_{T-q}$. Hence, a fully flexible model that imposes all Parallel assumptions between Parallel-(1) and Parallel-(t^*) must satisfy that $\gamma_\tau = 0$ for all $\tau \in \{2, ..., t^*\}$. Thus, only when pre-treatment dynamics are equal between treated and controls $\alpha = \gamma_T$. Therefore, the test of the null hypothesis of pre-treatment *common trends* ($H_0 : \gamma_\tau = 0$ for all $\tau \in \{2, ..., t^*\}$) is a test for the simultaneous equivalence of all Parallel-(q) assumptions. Given that the standard model is the fully flexible model with $\gamma_\tau = 0$ for all $\tau \in \{2, ..., t^*\}$, we can interpret it as a fully flexible model that satisfies all Parallel assumptions.

The following result states the restrictions on the parameters in the fully flexible model that imply the same Parallel-(q) assumptions as introducing polynomial trends in a standard model.

Theorem 4. *For any value $R \in \{1, t^* - 1\}$, a fully flexible model as in equation* (16) *that satisfies equivalence of Parallel-$(R+1)$, Parallel-$(R+2)$, ..., and Parallel-(t^*) is equivalent to a standard model with a polynomial of order R as in equation* (15).

Proof. See Theorem A4 in Appendix A. □

For example, using a linear trend in equation (13) and identifying the treatment effect with γ_T implies the equivalence of all Parallel-(q) assumptions from Parallel-(2) until Parallel-(t^*). Using a quadratic polynomial implies the equivalence of all parallel assumptions from Parallel-(3) until Parallel-(t^*).

If a researcher has appealing reasons to prefer one assumption over the rest, then the sensible strategy is to impose it. One can also test whether the imposition of additional

Parallel-(q) assumptions would lead to similar results. This is easy to implement using an F statistic of the joint significance of the corresponding equivalence conditions.[4]

We saw in Table 1 that in Case (a) Parallel-(1) holds but Parallel-(2) does not. This case highlights that the validity of Parallel-(1) does not depend on the dynamics before treatment. Researchers already know that the usual placebo test on pre-treatment data is never a test on Parallel-(1). We stress here that rejection of the equivalence between Parallel-(1) and Parallel-(2) is not evidence in favor of any of the two. Similarly, non-rejection is evidence that both Parallel-(q) assumptions yield similar results, but not evidence in favor of the two. Ultimately, tests on pre-treatment periods do not test the validity of identifying Parallel-(q) assumptions, that are fundamentally untestable.

Two final remarks relate to the modeling of dynamic effects and to the computation of the standard errors in the fully flexible model. Consider an applied researcher who correctly identifies the effect under Parallel-(q) but wrongly imposes no dynamic treatment effects. The result is a consistent estimate of a weighted average of the effects in the post-treatment period. On the other hand, if Parallel-(q) is false, the result will generally not be related with any meaningful treatment effect parameter. Overall, failing to select the correct identifying assumption is a potentially more serious issue than failing to account for dynamic treatment effects.[5] Regarding the computation of the standard errors in the fully flexible model, as with any DID regression, in the fully flexible model we regress outcomes at the individual level on controls that are variant at group level. In this situation, cross-sectional correlation within groups as well as correlations over time of the units in the different groups (in panel data applications) has been shown to substantially affect the validity of traditional inference (Moulton, 1990, Bertrand

[4]See Mora and Reggio (2015) for an implementation of the fully flexible model and several equivalence tests in Stata.

[5]We thank an anonymous referee for pointing this out.

et al., 2004, Donald and Lang, 2007, Hansen, 2007a, and Hansen, 2007b). The model in equation (16) is subject to the same potential pitfalls regarding the computation of standard errors as any other DID regression method.

4 A Monte Carlo Example

Under the appropriate set of Parallel-(q) assumptions, standard regression techniques provide consistent estimates of the treatment effects. The fully flexible model permits to explore which identifying assumptions can be used together. In contrast, the models used in applied research directly impose the equivalence of many Parallel-(q) assumptions. We illustrate by means of a Monte Carlo simulation the potential relative performance of four alternative models when the data generation process is a particular specification of the fully flexible model in equation (16).

The data, with five periods before treatment and two periods after treatment, are generated from:

$$y_{it} = \sum_{\tau=2}^{7} \delta_\tau I_t^\tau + \gamma D_i + \sum_{\tau=2}^{7} \gamma_\tau \times I_t^\tau \times D_i + u_{it} \tag{18}$$

where $\Pr(D_i = 1) = 0.5$ and $u_{it} \sim N(0, 0.25)$. The sequence δ_τ is the Fibonacci sequence $\{1, 1, 2, 3, 5, 8\}$, $\gamma = 3$, and the sequence γ_τ is $\{4, 4, 5, 6, 8, 9\}$. We consider four different sample sizes ($N \times T = 250$, 750, 2000, 5000) and conduct 10,000 replications.

In addition to the fully flexible model from equation (16), we consider the standard model in equation (11) where we add an interaction between D_i and the dummy variable for the second post treatment period as well as a model with a linear trend polynomial as in equation (13) and a model with a quadratic polynomial. For the fully flexible model,

Table 2: *Monte Carlo:* $H_0 : \alpha(1) = 1$

	$N \times T = 250$	$N \times T = 750$	$N \times T = 2000$	$N \times T = 5000$
Standard Model	1.000	1.000	1.000	1.000
Linear Model	0.827	1.000	1.000	1.000
Quadratic Model	0.410	0.878	1.000	1.000
Fully flexible, $q = 1$	0.973	1.000	1.000	1.000
Fully flexible, $q = 2$	0.050	0.054	0.051	0.052
Fully flexible, $q = 3$	0.053	0.050	0.050	0.049
Fully flexible, $q = 4$	0.108	0.208	0.471	0.850
Fully flexible, $q = 5$	0.683	0.991	1.000	1.000

Note: Monte Carlo results using 10,000 replications. Results show the proportion of rejections of the null at 5% significance level. T is fixed at 7 and N slightly changes per period.

we report results considering separately Parallel-(1) to Parallel-(5).

The modeling of the conditional expectation as in equation (18) is not enough to identify the treatment effect. We must make additional assumptions. Under Parallel-(1) and Parallel-(4) $\alpha = 2$, under Parallel-(2) and Parallel-(3) $\alpha = 1$, and under Parallel-(5) $\alpha = 8$. Table 2 shows the proportion of rejections of the null $H_0 : \alpha = 1$ versus $H_1 : \alpha \neq 1$ using the 5% significance level. For the fully flexible model both under Parallel-(2) and under Parallel-(3), the null is rejected in approximately the same proportion as the significance level. The standard, the linear, and the quadratic models identify the treatment effect wrongly imposing the equivalence of alternative identifying assumptions and over-reject the null for all sample sizes in the case that the null is true. When the null is not true, the power of the flexible model is large, except when the true treatment is identified with $q = 4$, a result arguably related to the fact that the true value is close to the null and that we identify the effect using a high q.

The values of parameters γ_τ are such that Parallel-(2) and Parallel-(3) are equivalent while no other equivalence condition is satisfied. In Table 3 we present the proportion of rejections of several equivalence tests at 5% significance levels. We focus only on the equivalence of sets of consecutive Parallel-(q) assumptions. The standard model

imposes the *common trends* assumption (Parallel-(1) to Parallel-(5)). The Linear and Quadratic models assume the equivalence of Parallel-(2) to Parallel-(5) and Parallel-(3) to Parallel-(5), respectively. For even small samples, the procedure leads to rejection of the equivalence conditions required for each of these models. The equivalence test is rejected in 5% of cases in the test for equivalence of Parallel-(2) and Parallel-(3). These results suggest that no other equivalence restriction should be imposed in the identification of the treatment effect.

With this data generation process, what would be a typical approach to the identification of the effect? The *common trends* assumption would be rejected (most of the time). Instead, a linear (or quadratic) trend would be included without testing the validity of the equivalence conditions imposed. In this illustration, this would imply the imposition of invalid assumptions to identify the effect.

Equivalence tests do not indicate which Parallel-(q) assumption is the right one. Instead they identify sets of assumptions that can be used together. Ultimately, researchers have to use their own judgment to select the identifying assumption most likely to be true. In this illustration, equivalence tests will normally indicate that the only admissible identifying assumptions are jointly Parallel-(2) and Parallel-(3) or separately each one of the others.

5 Practical relevance

In this section, we assess briefly the practical relevance of our analysis by discussing some empirical contributions to the DID literature in ten major journals for the $2009 - 2012$ period. We do not attempt to review the main results of the papers, which in some cases are not derived from the DID analysis, and instead focus on the DID implementations.

Table 3: *Monte Carlo: Equivalence tests*

	$N \times T = 250$	$N \times T = 750$	$N \times T = 2000$	$N \times T = 5000$
Parallel-1 to Parallel-5	1.000	1.000	1.000	1.000
Parallel-2 to Parallel-5	1.000	1.000	1.000	1.000
Parallel-3 to Parallel-5	1.000	1.000	1.000	1.000
Parallel-1 vs. Parallel-2	0.971	1.000	1.000	1.000
Parallel-2 vs. Parallel-3	0.054	0.048	0.051	0.049
Parallel-3 vs. Parallel-4	0.238	0.575	0.939	1.000
Parallel-4 vs. Parallel-5	0.976	1.000	1.000	1.000

Note: Monte Carlo results using 10,000 replications. Results show the proportion of rejections of the null at 5% significance level. T is fixed at 7.

We find nine papers that include an application with several pre-treatment periods and for which the data are publicly available.[6] In three of them, the estimation strategy involves exploiting panel data information by including individual-specific time trends. As we cannot compare their results with those obtained using the fully flexible model from Equation (16), we do not consider them in the review.

In the six remaining papers we can estimate the fully flexible model. In Table B2 of Appendix B we present the list of 13 regressions that we revise from these six papers. In all specifications the treatment effect is identified as the interaction term. Currie and Walker (2011) use the standard model with a linear polynomial interacted with the treatment variable as in equation (13). Furman and Stern (2011) estimate an extension of the standard model in equation (11) which allows for dynamic treatment effects. All other implementations use the standard model. When there is a discussion of the *common trends* assumption, sometimes it involves plotting a graph of pre-treatment dynamics while in other cases it involves applying the model on a placebo year or the last pre-treatment period.[7] The authors generally motivate the inclusion of trends as a

[6]The details on how we selected the papers and the list of papers can be found in Appendix B.

[7]Moser and Voena (2012) estimate pre-treatment differences in time dummies. They do not implement the joint test for common pre-treatment trends that we propose in Section (3) (i.e., $H_0 : \gamma_\tau^D = 0$ for all $\tau \leq t^*$). Instead they plot each pair $\left(\hat{\delta}_t + \hat{\gamma}_t^D, \hat{\delta}_t\right)$ and their confidence intervals and report that

robustness check of the technique or as a relaxation of the *common trends* assumption. Altogether the number of papers is small but we think that they reflect the usual practice of DID in the empirical literature. In the remaining part of this section, we discuss how the DID results from these papers are robust to the implementation of the fully flexible model (see Table B2 of Appendix B for details).

We compare the original results with results using the fully flexible model under Parallel-(1) and Parallel-(2). We provide two sets of estimates. For both Parallel-(1) and Parallel-(2) we first report the effect using the model in Equation (16) restricted by imposing that the effect does not depend on s—the columns labeled as "Restricted". This effect can be interpreted as an estimate of the average effect throughout the post-treatment period. Second we report the effect at $s = 1$ in a fully flexible model that allows for different effects at different s—the columns labeled as "Unrestricted".[8] We also test for the equivalence of Parallel-(1) and Parallel-(2) and for the *common trends* and *linear trend* assumptions.

In 11 out of 13 treatment effect estimates, the original papers report significant effects. In the fully flexible model with unrestricted dynamics the estimated effects remain significant and with the same sign at 10% in six out of the 11 cases under $q = 1$. Under $q = 2$, only three estimates remain significant.[9] We interpret these results as anecdotal evidence that, in empirical work, the identification of the treatment effect usually relies on restrictions involving parallel assumptions beyond Parallel-(1) and/or Parallel-(2). These results are generally similar if we restrict the post-treatment dynamics in the fully flexible model to the same dynamics as in the original papers. Hence, differences

"(the) test reveals no systematic differences in pre-trends across treated and untreated subclasses."

[8] In Kotchen and Grant (2011) there is only one post-treatment period, so there is no difference between the Restricted and Unrestricted models and the effects in the two columns are exactly the same by definition.

[9] Generally, standard errors under $q = 1$ are larger than under the original model specification and under $q = 2$ they are larger than under $q = 1$.

between the results in the fully flexible model and the original estimates are not likely to be due to the modeling of post-treatment dynamics.

In four out of the 13 cases, we reject at the 5% significance level that Parallel-(1) and Parallel-(2) lead to equivalent results. These rejections are evidence of pre-treatment trend differentials. Hence, either Parallel-(1) or Parallel-(2) or both are inappropriate. The *common trends* assumption implies not only the equivalence of Parallel-(1) and Parallel-(2), but also the equivalence of all Parallel assumptions. In nine out of the 10 cases from the four papers in which we can test the common trends assumption we reject *common trends*, the assumption exploited in the original estimates. The usual practice after rejection of the *common trends* assumption is to include a linear trend interacted with the treatment dummy. This strategy is equivalent to assuming the equivalence of all Parallel assumptions between Parallel-(2) and Parallel-(t^*). We provide the test for this assumption, the *linear trend* test, in the last column of the Table. We reject *the linear trend* assumption in the same nine cases. We interpret this result as anecdotal evidence that in empirical work identification of the effects is usually based on a too large set of parallel assumptions.

Finding different results after imposing different assumptions is not surprising. It is also not surprising that more flexible models provide less accuracy than the standard model. Our tests results suggest that both the common trends assumption and controlling for a linear trend differential are too restrictive strategies in most applications. However, in nine out of the 26 fully flexible effect estimates, we obtain results that coincide in sign with the original results and are significant as in the original results. These outcomes suggest that the models usually implemented in empirical studies can be unduly restrictive. More generally, our results highlight that strategies that combine alternative Parallel assumptions in the fully flexible model can lead to more accurate

estimates without imposing too strong assumptions. In the next subsection we discuss how to implement two alternative strategies with an illustration using two of the selected papers.

A practical guidance to applied researchers

Researchers have to use their own judgment to select the identifying assumption most likely to be true within the family of parallel assumptions. Estimation can be carried out using the fully flexible model. Still, there might be gains in basing estimation on a more parsimonious model. For clarity, we discuss the cases in which the researcher takes either Parallel-(1) or Parallel-(t^*) as the most plausible identifying assumption and considers adding consecutive assumptions.[10] For example, consider the case with three pre-treatment periods. Starting from Parallel-(1) relies on the comparability (at t^*+1) under non-treatment between controls and treated average output changes. Starting from Parallel-(3) relies on the comparability between controls and treated in changes in output accelerations. If one suspects some type of pre-treatment dynamic differentials so that growth comparisons are potentially inadequate, it looks safer to start assuming Parallel-(3). If one suspects that comparability may hold only in the last pre-treatment periods, then it looks safer to start with Parallel-(1).

If Parallel-(1) is chosen, Parallel-(2) can be added if the test for equivalence between Parallel-(2) and Parallel-(1) , i.e. $H_0 : \gamma_{T-1} = \gamma_{T-2}$, is not rejected. The equivalence of Parallel-(3) with Parallel-(2) and Parallel-(1) can be tested by testing $H_0 : \gamma_{T-1} =$

[10]There are no *a priori* grounds on which assumptions should be added to the preferred one, but it seems reasonable to focus this discussion on identification strategies that involve the sequential equivalence of alternative parallel assumptions. This implies, for example, that if Parallel-(1) and Parallel-(2) are not equivalent, then Parallel-(1) and Parallel-(3) will also not be equivalent. In the general case, the researcher chooses q^* and considers either adding lower or higher Parallel-(q) assumptions. For clarity, here we consider the two particular cases when $q^* = 1$ and $q^* = t^*$, but it is straightforward to apply our discussion to the general case.

$\gamma_{T-2} = \gamma_{T-3}$. Parallel-(3) would be added if the null is not rejected. One can proceed adding higher parallel assumptions until the equivalence test is rejected. As an illustration, consider Moser and Voena (2012). Using annual data from 1875 until 1939, they study the effect of the Trading with the Enemy Act (TWEA) in 1918 on the number of patents by US inventors. We focus on the DID result reported in column 1 of Table 2 of the original paper. Their reported 0.151 estimate can be interpreted as the average of additional patents produced by domestic inventors per year after 1919 due to the TWEA. To obtain this estimate, they use the standard model, therefore accepting the common trends assumption, i.e. the equivalence of all parallel assumptions between Parallel-(1) and Parallel-(43). We reject this assumption (see column "Common Trends" in our Table B2 in Appendix B). Which parallel assumptions should we exploit to identify the effect? If we assume that Parallel-(1) is the most credible assumption, since we do not reject its equivalence with Parallel-(2) (see column "Equiv. Test") we add Parallel-(2). The equivalence of Parallel-(1), Parallel-(2), and Parallel-(3) is not rejected (p-value of 0.18) while we strongly reject the equivalence of all parallel assumptions between Parallel-(1) and Parallel-(4) (p-value of 0.0047). We would in this case advise to exploit Parallel-(1), Parallel-(2), and Parallel-(3) only. This can be done by estimating equation (16) imposing $\gamma_{T-1} = \gamma_{T-2} = \gamma_{T-3} = \theta$ and estimating the treatment effect as $\widehat{\alpha} = \widehat{\gamma}_T - \widehat{\theta}$. Assuming no dynamic effects, the estimated effect is 0.298, approximately doubling the original estimate. The standard error is larger, 0.043 vs. 0.036, but the p-value is lower. Hence, following this strategy, we find stronger evidence of a stronger effect using a more robust identification strategy.

Suppose that Parallel-(t^*) is chosen. Given that Parallel-(t^*) uses the Δ^{t^*} operator to identify the effect, in practice the estimator is likely to be less accurate than using lower q assumptions. One simple strategy to circumvent this issue is to search for the

largest set of parallel assumptions that are equivalent to Parallel-(t^*). We can do this by imposing all parallel assumptions between Parallel-(1) and Parallel-(t^*) and then sequentially dropping parallel assumptions starting from Parallel-(1). This can be done by adding higher order polynomial trends. Hence, in the first iteration one would test the equivalence of all parallel assumptions, that is, implementing the common trends test, i.e. $H_0 : \gamma_\tau = 0$ for all $\tau \in \{2, ..., t^*\}$. If the test is rejected, we test the equivalence of all parallel assumptions between Parallel-(2) and Parallel-(t^*), $H_0 : \Delta\gamma_\tau = \gamma_2$ for all $\tau \in \{3, ..., t^*\}$ (see Lemma A3 in Appendix A). In case of rejection, we then test the equivalence of all parallel assumptions between Parallel-(3) and Parallel-(t^*). We stop this testing sequence when the test is not rejected or when only Parallel-(t^*) remains.[11] As an illustration, consider Abramitzky et al. (2011). The authors investigate the effect of male scarcity due to military mortality during World War I on marriage market outcomes in France. They use a sample with six pre-treatment periods. As output they use three alternative definitions of a bad marriage outcome for men. We focus here on the class of the bride minus the class of the groom (column 1 from their Table 3 and the second row in our Table B2 in Appendix B). They assume common trends, i.e. the equivalence of all parallel assumptions between Parallel-(1) and Parallel-(6). Their estimate is negative and significant at the 10% level suggesting that higher regional mortality led to better marriage outcomes for men. We reject the equivalence of Parallel-(1) and Parallel-(2) (see Table B2 in the Appendix). If we assume that Parallel-(6) is the most credible assumption, we can implement the described strategy. First, we reject common trends, finding evidence against the standard model. We also reject the linear trend assumptions. However we cannot reject the equivalence of all parallel assumptions between Parallel-(3) and Parallel-(6) (with a p-value of .583). The estimate

[11] The usual practice of starting with the standard model and adding a linear or quadratic polynomial interaction is in the same spirit with this strategy.

under Parallel-(3) to Parallel-(6) is -0.030, approximately 50% larger in absolute value than the original estimate. As in the previous illustration, the standard error is slightly larger, 0.013 vs. 0.010, and the resulting p-value is smaller, 0.029. Hence, in this case we also find stronger evidence of a stronger effect using a more robust identification strategy.

6 Conclusions

This paper studies identification of treatment effects using DID methods when several pre-treatment periods are available. We define a family of non-nested, identifying Parallel-(q) assumptions. Furthermore, we can determine for any econometric model whether it imposes the equivalence of several of these assumptions for identification of the treatment effect. Two models are, in practice, very important. The standard DID model imposes the equivalence of all Parallel-(q) assumptions (usually referred to as the *common trends* assumption). Including a linear trend differential for the dynamics between treated and controls imposes the equivalence of all Parallel-(q) assumptions except Parallel-(1) (i.e., the assumption that changes from the last pre-treatment period are equal for treated and controls in the absence of treatment). Since each Parallel-(q) assumption identifies the treatment effect, assuming *common trends* with or without polynomials implies over-identification and testable restrictions. In current practice, informal tests—such as graphical evidence—for *common trends* are sometimes presented. When polynomials are included, the restrictions implied in the new specifications are routinely not tested.

Failure to test the validity of equivalence conditions may lead to imposition of invalid assumptions and inconsistent estimates of the effect. We propose a model—which we

refer to as the fully flexible model—that does not impose equivalence restrictions of alternative Parallel-(q) assumptions. Using the estimates of the fully flexible model, the treatment effect can be estimated under any Parallel-(q) assumption. Because the estimators of the treatment effects turn out to be linear combinations of the original parameters of the model, tests for the equivalence of Parallel-(q) assumptions are easy to implement.

Equivalence tests do not indicate which set of Parallel-(q) assumptions is the right one. Instead they identify which assumptions can be used together. Ultimately, researchers have to use their own judgment to select the identifying assumption most likely to be true. We revise the results of several recent papers in which the DID technique has been applied and reach two main conclusions. In most papers we find evidence against the *common trends* assumption and, at the same time, the *linear trend* assumption, which is the alternative to *common trends* usually advocated. In addition, we find evidence that the models can be unduly restrictive in the sense that in about a third of the specifications the empirical results would remain under one Parallel assumption without imposing any equivalence condition. Finally, we provide two testing strategies to impose equivalence conditions to improve the efficiency of the treatment effect estimates.

References

Aaronson, D. and B. Mazumder (2011). The impact of Rosenwald Schools on Black achievement. *Journal of Political Economy 119*(5), 821.

Abramitzky, R., A. Delavande, and L. Vasconcelos (2011). Marrying Up: The Role of Sex Ratio in Assortative Matching. *American Economic Journal: Applied Economics 3*(3), 124–57.

Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, pp. 1277–1366. Elsevier.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics 119*(1), 249–275.

Blundell, R., M. Dias, C. Meghir, and J. Reenen (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association 2*(4), 569–606.

Currie, J. and R. Walker (2011). Traffic Congestion and Infant Health: Evidence from E-ZPass. *American Economic Journal: Applied Economics 3*(1), 65–90.

De Jong, P., M. Lindeboom, and B. Van der Klaauw (2011). Screening disability insurance applications. *Journal of the European Economic Association 9*(1), 106–129.

Donald, S. and K. Lang (2007). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics 89*(2), 221–233.

Friedberg, L. (1998). Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data. *The American Economic Review 88*(3), 608–627.

Furman, J. and S. Stern (2011). Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *The American Economic Review 101*(5), 1933–1963.

Hansen, C. B. (2007a). Asymptotic properties of a robust variance matrix estimator for panel data when t is large. *Journal of Econometrics 141*(2), 597–620.

Hansen, C. B. (2007b). Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics 140*(2), 670–694.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *The American Economic Review 83*(4), pp. 685–709.

Jayachandran, S., A. Lleras-Muney, and K. Smith (2010). Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. *American Economic Journal: Applied Economics 2*(2), 118–146.

Kotchen, M. and L. Grant (2011). Does daylight saving time save energy? evidence from a natural experiment in indiana. *The Review of Economics and Statistics 93*(4), 1172–1185.

Mora, R. and I. Reggio (2012). Treatment effect identification using alternative parallel assumptions. Technical report, WP 12-33. UC3M.

Mora, R. and I. Reggio (2015). didq: A command for treatment-effect estimation under alternative assumptions. *Stata Journal 15*(3), 796–808(13).

Moser, P. and A. Voena (2012). Compulsory Licensing: Evidence from the Trading with the Enemy Act. *The American Economic Review 102*(1), 396–427.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The review of Economics and Statistics*, 334–338.

Reber, S. J. (2005). Court-ordered desegregation successes and failures integrating american schools since brown versus board of education. *Journal of Human Resources 40*(3), 559–590.

Redding, S., D. Sturm, and N. Wolf (2011). History and industry location: Evidence from German airports. *The Review of Economics and Statistics 93*(3), 814–831.

Wolfers, J. (2006). Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *The American Economic Review 96*(5), 1802–1820.

Appendix A

In this Appendix, we present the proofs of the results presented in sections 2 and 3. We consider the general case in which there may be several periods before and after treatment and we estimate the effect of treatment s periods after treatment.

Theorem A1. *For a given positive integer $q \leq t^*$, and for any s, $1 \leq s \leq T - t^*$, under Parallel-(q), the treatment effect $\alpha(s)$ is identified as the solution of the following equation:*

$$\Delta^{q-1}\alpha(s) = did(q,s).$$

Proof. By definition, $\alpha(s) \equiv E\left[Y^1_{i,t^*+s}|D_i=1\right] - E\left[Y^0_{i,t^*+s}|D_i=1\right]$. Taking q differences,

$$\Delta^q\alpha(s) \equiv E\left[\Delta^q Y^1_{i,t^*+s}|D_i=1\right] - E\left[\Delta^q Y^0_{i,t^*+s}|D_i=1\right].$$

Since, for any variable z_t

$$\begin{aligned}\Delta^q z_{t+s} &= \Delta^{q-1}z_{t+s} - \Delta^{q-1}z_{t+s-1}\\ &= \left(\Delta^{q-1}z_{t+s} - \Delta^{q-1}z_t\right) - \left(\Delta^{q-1}z_{t+s-1} - \Delta^{q-1}z_t\right)\\ &= \Delta_s\Delta^{q-1}z_{t+s} - \Delta_{s-1}\Delta^{q-1}z_{t+s-1},\end{aligned}$$

then

$$\Delta^q Y^\rho_{i,t^*+s} = \Delta_s\Delta^{q-1}Y^\rho_{i,t^*+s} - \Delta_{s-1}\Delta^{q-1}Y^\rho_{i,t^*+s-1}$$

where $\rho = 0, 1$. Hence, under Parallel-(q)

$$\begin{aligned}\Delta^q\alpha(s) = &\left(E\left[\Delta_s\Delta^{q-1}Y^1_{i,t^*+s}|D_i=1\right] - E\left[\Delta_{s-1}\Delta^{q-1}Y^1_{i,t^*+s-1}|D_i=1\right]\right)\\ &- \left(E\left[\Delta_s\Delta^{q-1}Y^0_{i,t^*+s}|D_i=0\right] - E\left[\Delta_{s-1}\Delta^{q-1}Y^0_{i,t^*+s-1}|D_i=0\right]\right)\end{aligned}$$

and, by definition of $did\,(q,s)$,

$$\Delta^q \alpha\,(s) = did\,(q,s) - did\,(q,s-1) = \Delta did\,(q,s)\,. \tag{19}$$

The theorem is then proved if we show that this condition holds for all s if and only if

$$\Delta^{q-1} \alpha\,(s) = did\,(q,s)\,. \tag{20}$$

The "if" part of the statement is straightforward. To prove the "only" part, we only need to prove that $\Delta^{q-1}\alpha\,(s) = did\,(q,s)$ for some s, for instance $s=1$. We know that $\Delta^{q-1}\alpha\,(1) = \alpha\,(1)$, because—by definition—$\alpha\,(s) = 0$ for all $s \leq 0$. Hence, all is left to show is that under Parallel-(q) $\alpha\,(1) = did\,(q,1)$. Since $Y^0_{i,t^*+1} = \sum_{r=1}^{q} \Delta^{r-1} Y^0_{it^*} + \Delta^q Y^0_{i,t^*+1}$, then

$$\begin{aligned}
\alpha\,(1) &= E\left[Y^1_{i,t^*+1} | D_i = 1\right] - E\left[\sum_{r=1}^{q} \Delta^{r-1} Y^0_{it^*} + \Delta^q Y^0_{i,t^*+1} | D_i = 1\right] \\
&= E\left[Y_{i,t^*+1} | D_i = 1\right] - E\left[\sum_{r=1}^{q} \Delta^{r-1} Y_{it^*} + \Delta^q Y^0_{i,t^*+1} | D_i = 1\right]
\end{aligned}$$

where in the second equality we have used the fact that $Y_{i,t} = Y^0_{i,t} = Y^1_{i,t}$ for any $t \leq t^*$. Therefore, under Parallel-(q),

$$\begin{aligned}
\alpha\,(1) &= E\left[Y_{i,t^*+1} - \sum_{r=1}^{q} \Delta^{r-1} Y_{it^*} | D_i = 1\right] - E\left[\Delta^q Y_{i,t^*+1} | D_i = 0\right] \\
&= E\left[\Delta^q Y_{i,t^*+1} | D_i = 1\right] - E\left[\Delta^q Y_{i,t^*+1} | D_i = 0\right] \\
&= did\,(q,1)
\end{aligned}$$

and the theorem is proved. □

Corollary A1. *For a given positive integer $q \leq t^*$, and for any s, $1 \leq s \leq T - t^*$, the unique solution to equation (??) such that $\alpha(s) = 0$ for all $s \leq 0$ under Parallel-(q), denoted by $\alpha^q(s)$, is:*

$$\alpha^q(s) = \begin{cases} did(q,s) \ \ if\ q = 1 \\ \sum_{j=1}^{s} did(q,j) \ \ if\ q = 2 \\ \sum_{j_2=1}^{s} \sum_{j=1}^{j_2} did(q,j) \ \ if\ q = 3 \\ \sum_{j_{q-1}=1}^{s} \sum_{j_{q-2}=1}^{j_{q-1}} \cdots \sum_{j=1}^{j_2} did(q,j) \ \ if\ q \geq 4 \end{cases}$$

Proof. The result for $q = 1$ follows directly from Theorem A1 as $\Delta^0 = 1$. For $q = 2$ by Theorem A1 we have that $\alpha(s) = \alpha(s-1) + did(2,s)$ for $s > 1$ and that $\alpha(1) = did(2,1)$ since $\alpha(0) = 0$ by definition. Hence, $\alpha(s) = \sum_{j=1}^{s} did(2,j)$. For $q \geq 3$, note that taking a first difference on $\sum_{j_{q-1}=1}^{s} \cdots \sum_{j=1}^{j_2} did(q,j)$ gives $\sum_{j_{q-2}=1}^{s} \cdots \sum_{j=1}^{j_2} did(q,j)$. More generally, if we apply the Δ^d operator, we have $\sum_{j_{q-(d+1)}=1}^{s} \cdots \sum_{j=1}^{j_2} did(q,j)$ so that $\sum_{j_{q-1}=1}^{s} \cdots \sum_{j=1}^{j_2} did(q,j)$ is the solution of $\Delta^{q-1}\alpha(s) = did(q,s)$.A. □

Theorem A2. *For any $q = 2, ..., t^*$ and a given $s = 1, ..., T - t^*$,*

$$\alpha^q(s) = \alpha^{q-1}(s)$$

if and only if

$$E\left[\Delta^{q-1} Y_{i,t^*} | D_i = 1\right] = E\left[\Delta^{q-1} Y_{i,t^*} | D_i = 0\right].$$

Proof. We first prove the theorem for $q = 2$. For $s = 1$, $\alpha^2(1) = \alpha^1(1)$ if and only if $did(2,1) = did(1,1)$ (by Theorem A1). By definition of the $did(q,s)$ operator, this condition is equivalent to

$$\begin{aligned} &E\left[\Delta^2 Y_{i,t^*+1} | D_i = 1\right] - E\left[\Delta^2 Y_{i,t^*+1} | D_i = 0\right] \\ &= E\left[\Delta Y_{i,t^*+1} | D_i = 1\right] - E\left[\Delta Y_{i,t^*+1} | D_i = 0\right] \end{aligned}$$

or, given that $\Delta^2 Y_{i,t^*+1} = \Delta Y_{i,t^*+1} - \Delta Y_{it^*}$,

$$E\left[\Delta Y_{i,t^*} | D_i = 1\right] = E\left[\Delta Y_{it^*} | D_i = 0\right].$$

For $s > 1$,

$$\begin{aligned}\alpha^2(s) &= \sum_{j=1}^{s} did(2,j) \\ &= \sum_{j=1}^{s} \{E[\Delta_j \Delta Y_{i,t^*+j} | D_i = 1] - E[\Delta_j \Delta Y_{i,t^*+j} | D_i = 0]\} \\ &= \sum_{j=1}^{s} \{E[\Delta Y_{i,t^*+j} - \Delta Y_{it^*} | D_i = 1] - E[\Delta Y_{i,t^*+j} - \Delta Y_{it^*} | D_i = 0]\} \\ &= \sum_{j=1}^{s} \{E[\Delta Y_{i,t^*+j} | D_i = 1] - E[\Delta Y_{i,t^*+j} | D_i = 0]\} \\ &\quad - s\{E[\Delta Y_{it^*} | D_i = 1] - E[\Delta Y_{it^*} | D_i = 0]\}.\end{aligned}$$

Taking into account that $\sum_{j=1}^{s} E[\Delta Y_{i,t^*+j} | D] = E[\Delta_s Y_{i,t^*+s} | D]$ and the definition of $did(1,s)$, we have that

$$\begin{aligned}\alpha^2(s) &= \alpha^1(s) \\ &\quad - s\{E[\Delta Y_{it^*} | D_i = 1] - E[\Delta Y_{it^*} | D_i = 0]\}.\end{aligned}$$

Thus, for $s > 1$, $\alpha^2(s) = \alpha^1(s)$ if and only if $E[\Delta Y_{it^*} | D_i = 1] - E[\Delta Y_{it^*} | D_i = 0]$.

To prove the theorem for $q > 2$, we use the following two lemmas.

Lemma A1. *For any $q = 2, ..., t^*$ and a given $s = 1, ..., T - t^*$,*

$$\Delta^{q-1} \alpha_Y^q(s) = \alpha_{\Delta^{q-1} Y}(s)$$

where

$$\alpha_{\Delta^{q-1} Y}(s) \equiv E\left[\Delta^{q-1} Y_{i,t^*+s}^1 | D_i = 1\right] - E\left[\Delta^{q-1} Y_{i,t^*+s}^0 | D_i = 1\right].$$

Proof. From the linear properties of the Δ_s operator, it follows that

$$\begin{aligned}\Delta^{q-1} \alpha_Y^q(s) &= E\left[\Delta_s \Delta^{q-1} Y_{i,t^*+s} | D_i = 1\right] - E\left[\Delta_s \Delta^{q-1} Y_{i,t^*+s} | D_i = 0\right] \\ &= \alpha_{\Delta^{q-1} Y}(s).\end{aligned}$$

□

Lemma A2. *For any $q = 2, ..., t^*$ and a given $s = 1, ..., T - t^*$,*

$$\alpha_Y^q(s) = \alpha_Y^{q-1}(s)$$

if and only if

$$\alpha_{\Delta^{q-1}Y}(s) = \Delta\alpha_{\Delta^{q-2}Y}(s).$$

Proof. We first prove sufficiency. By applying the $(q-1)$th difference, we have that if $\alpha_Y^q(s) = \alpha_Y^{q-1}(s)$ then $\Delta^{q-1}\alpha_Y^q(s) = \Delta^{q-1}\alpha_Y^{q-1}(s)$. Therefore, by Lemma A1, $\alpha_{\Delta^{q-1}Y}(s) = \Delta\alpha_{\Delta^{q-2}Y}(s)$.

Now we prove necessity. By Lemma A1, if $\alpha_{\Delta^{q-1}Y}(s) = \Delta\alpha_{\Delta^{q-2}Y}(s)$ then

$$\Delta^{q-1}\alpha_Y^q(s) = \Delta^{q-1}\alpha_Y^{q-1}(s)$$

for all s. Therefore, both $\alpha_Y^q(s)$ and $\alpha_Y^{q-1}(s)$ satisfy the same initial conditions and have the same differential equations so they must be the same.

□

To prove the theorem for $q > 2$, define $z_t = \Delta^{q-2}Y_{it}$. By Lemma A2, the theorem is proved if $\alpha_{\Delta z}(s) = \Delta\alpha_z(s)$ is true if and only if $E[\Delta z_{t^*}|D_i = 1] = E[\Delta z_{t^*}|D_i = 0]$. By Lemma A1, we need to prove that $\Delta\alpha_z^2(s) = \Delta\alpha_z(s)$ is true for all s if and only if $E[\Delta z_{t^*}|D_i = 1] = E[\Delta z_{t^*}|D_i = 0]$. Given that $\Delta\alpha_z^2(s) = \Delta\alpha_z(s)$ is true for all s if and only if $\alpha_z^2(s) = \alpha_z(s)$ for all s. We know this condition to be true because is the theorem for $q = 2$, which we have already proved.

□

Theorem A3. *Consider a general additive model with group-specific, fully-flexible pre- and post-treatment trends:*

$$E[Y_{it}|D_i = D] = \delta + \sum_{\tau=2}^{T}\delta_\tau I_\tau + \gamma D + \sum_{\tau=2}^{T}\gamma_\tau \times I_t^\tau \times D.$$

Under Parallel-(q):

$$\Delta^{q-1}\alpha(s) = \Delta_s\Delta^{q-1}\gamma_{t^*+s}.$$

Proof. Given that

$$E\left[\Delta_s\Delta^{q-1}Y_{i,t^*+s}|D_i = D\right] = \Delta_s\Delta^{q-1}\delta_{t^*+s} + \Delta_s\Delta^{q-1}\gamma_{t^*+s} \times D,$$

from the definition of $did(q,s)$, it follows that $did(q,s) = \Delta_s\Delta^{q-1}\gamma_{t^*+s}$. By Theorem A1, the theorem is proved.

□

Theorem A4. *For any value $R \in \{1, t^* - 1\}$, the fully flexible model*

$$E\left[Y_{it} \left| D_i = D\right.\right] = \delta + \sum_{\tau=2}^{T} \delta_\tau I_t^\tau + \gamma D + \sum_{\tau=2}^{T} \gamma_\tau \times I_t^\tau \times D$$

that satisfies equivalence of Parallel-$(R+1)$, Parallel-$(R+2)$, ..., and Parallel-(t^*) is equivalent to the standard model with a polynomial of order R

$$E\left[Y_{it} \left| D_i = D\right.\right] = \delta + \sum_{\tau=2}^{T} \delta_\tau I_t^\tau + \gamma D + \gamma_T \times I_t^T \times D + \sum_{r=1}^{R} \gamma_r^p \times t^r \times D.$$

Proof. We prove this result in two steps. In the first step we show that in the fully flexible model, for any $q_l \in \{1, ..., t^* - 1\}$, all parallel assumptions between Parallel-(q_l) and Parallel-(t^*) are equivalent if and only if $\Delta^{q_l - 1}\gamma_{t^*} = \Delta^{q_l - 1}\gamma_{t^*-1} = ... = \Delta^{q_l - 1}\gamma_{q_l}$. In the second step we show that given this condition, R values γ_2,...,γ_{R+1}, and the normalization $\gamma_1 = 0$, the fully flexible model is equivalent to the standard model with a polynomial of order R.

Lemma A3. *In the fully flexible model, for any $q_l \in \{1, ..., t^* - 1\}$, all parallel assumptions between Parallel-(q_l) and Parallel-(t^*) are equivalent if and only if $\Delta^{q_l - 1}\gamma_{t^*} = \Delta^{q_l - 1}\gamma_{t^*-1} = ... = \Delta^{q_l - 1}\gamma_{q_l}$.*

Proof. By Theorem A2, it is true for $t^* = q_l + 1$. Now we must prove that if it is true for $t^* = \tau$, it must be true for $t^* = \tau + 1$. If it is true for $t^* = \tau$, then $\Delta^{q_l - 1}\gamma_\tau = \Delta^{q_l - 1}\gamma_{\tau-1} = ... = \Delta^{q_l - 1}\gamma_{q_l}$. For $t^* = \tau + 1$, the assumption is that all parallel assumptions between Parallel-(q_l) and Parallel-$(\tau + 1)$ are equivalent. This implies that Parallel-(τ) is equivalent to Parallel-$(\tau + 1)$, that is, $\Delta^\tau \gamma_{\tau+1} = 0$ or, equivalently, that $\Delta^{\tau-1}\gamma_{\tau+1} = \Delta^{\tau-1}\gamma_\tau$. Since the result is true for $t^* = \tau$ and $\tau - 1 > q_l - 1$, $\Delta^{\tau-1}\gamma_\tau = 0$ so that $\Delta^{\tau-2}\gamma_{\tau+1} = \Delta^{\tau-2}\gamma_\tau$. Again the result is true for $t^* = \tau$, so $\Delta^{\tau-2}\gamma_\tau = 0$ so that $\Delta^{\tau-3}\gamma_{\tau+1} = \Delta^{\tau-3}\gamma_\tau$. Using this argument, we can set differences to zero until $q_l - 1$ so that $\Delta^{q_l - 1}\gamma_{\tau+1} = \Delta^{q_l - 1}\gamma_\tau$. This, together with the fact that $\Delta^{q_l - 1}\gamma_\tau = \Delta^{q_l - 1}\gamma_{\tau-1} = ... = \Delta^{q_l - 1}\gamma_{q_l}$ proves the first step.

□

For the second step, we must show that the polynomial can be re-parameterized as the restricted fully flexible model and vice versa. To see that the polynomial implies the restricted fully flexible, we note first that without loss of generality the polynomial can be re-parameterized as

$$E\left[Y_{it}\left|D_i\right.\right] = \delta + \sum_{\tau=2}^{T} \delta_\tau I_t^\tau + \gamma D_i + \gamma_T \times I_t^T \times D_i + \sum_{r=1}^{R} \gamma_r^p \times (t-1)^r \times D_i.$$

Hence, for the first $R+1$ periods we have

$$\begin{aligned}
E\left[Y_{i1}\left|D\right.\right] &= \delta + \gamma D \\
E\left[Y_{i2}\left|D\right.\right] &= \delta + \delta_2 + \gamma D + \sum_{r=1}^{R} \gamma_r^p \times D \\
&\vdots \\
E\left[Y_{i,R+1}\left|D\right.\right] &\delta + \delta_{R+1} + \gamma D + \sum_{r=1}^{R} \gamma_r^p \times R^r \times D.
\end{aligned}$$

Thus, defining

$$\begin{aligned}
\gamma_2 &= \sum_{r=1}^{R} \gamma_r^p \\
&\vdots \\
\gamma_{R+1} &= \sum_{r=1}^{R} \gamma_r^p \times R^r
\end{aligned}$$

results in the restricted fully flexible for the first $R+1$ periods. Since by construction for any $t > R+1$ the polynomial satisfies $\Delta^R \gamma_{t^*} = \Delta^R \gamma_{t^*-1} = ... = \Delta^R \gamma_{q_l}$, the polynomial implies the restricted fully flexible model for any t.

To show that the restricted fully flexible can be re-parameterized as the polynomial, we consider the set of $R+1$ points (t, γ_t) (with $\gamma_1 = 0$,γ_2,...,γ_{R+1}) where no two t are the same and we look for a polynomial of degree R, $p_R(t)$, such that $p_R(t) = \gamma_t$ for $t \in \{1, ..., R+1\}$. The unisolvence theorem states that such a polynomial exists and is unique. Since this polynomial is of degree R, it must also satisfy $\Delta^R \gamma_{t^*} = \Delta^R \gamma_{t^*-1} = ... = \Delta^R \gamma_{q_l}$.

□

Appendix B

In this appendix, we describe the conditions that led to the selection of papers reviewed in section 5, we list the papers, and we present the results of the implementation of the flexible model. In the working paper version (Mora and Reggio 2012) we include a more detailed analysis of each paper.

The papers are selected by imposing several conditions. The first condition is that the paper must have been published in the period 2009 : 2012 in one of 10 Economics journals: AEJ:AE, AER, JAppEcon, JEcon, JEEA, JLabEc, JPE, QJE, REStat, and REStud. The journals chosen are characterized by being among the highest-ranked economic journals on several criteria and also by allowing access to the data sets.

The paper must also include an application of DID. We search the terms "difference-in-differences" or "diff-in-diff" in the paper (with the exception of the bibliography section). For those papers that include these terms, we verify that a DID application exists. Overall, 59 papers satisfy this condition.

The next condition is that the data for the DID application are publicly available online by the publishing journal. There are 37 papers for which data are not available and one paper for which access to the data was granted after we made a formal request. Thus, out of the original 59 DID applications, only in 22 papers the data were available.

The final condition is that the data must include more than one pre-treatment period because with only two periods the only implementable Parallel-(q) assumption is Parallel-(1). There are 13 papers that do not satisfy this condition.

In total, nine papers, listed in Table B1, meet all requirements. In one of the nine papers (Jayachandran et al. 2010), it is not possible to use the fully flexible model in equation

Table B1: *List of Selected Papers*

Author	Year	Journal	Title	No. Pre	No. Post
Aaronson and Mazumder	2011	JPE	The impact of Rosenwald Schools on Black achievement	2	2
Abramitzky et al.	2011	AEJ:AE	Marrying Up: The Role of Sex Ratio in Assortative Matching	6	11
Currie and Walker	2011	AEJ:AE	Traffic Congestion and Infant Health: Evidence from E-ZPass	300	168
De Jong et al.	2011	JEEA	Screening disability insurance applications	2	1
Jayachandran et al.	2010	AEJ:AE	Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs	12	7
Furman and Stern	2011	AER	Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research	14	18
Kotchen and Grant	2011	REStat	Does Daylight Saving Time Save Energy? Evidence from a Natural Experiment in Indiana	2	1
Moser and Voena	2012	AER	Compulsory Licensing: Evidence from the Trading with the Enemy Act	43	22
Redding et al.	2011	REStat	History and industry location: Evidence from German airports	12	40

Note: Papers are listed by the alphabetical order obtained from the author's name. The papers selected satisfy the following conditions: (a) There is an application of DID; (b) the sample includes more than one period before treatment; (c) data are publicly available; and (d) the paper is published in the period 2009:2012 in one of the following 10 Economics journals: AEJ:AE, AER, JAppEcon, JEcon, JEEA, JLabEc, JPE, QJE, REStat, and REStud. "No. Pre." refers to the number of pre-treatment periods and "No. Post." refers to the number of post-treatment periods.

(16) because there is only one treated agent and one control. In two other papers, Redding et al. (2011) and De Jong et al. (2011), we cannot estimate a fully flexible version because the authors include individual-specific linear trends. For these three papers, we discuss in Mora and Reggio (2012) how alternative assumptions may yield different results and find evidence that Parallel-(1) and Parallel-(2) are not equivalent and that there are dynamic effects.

In each of the remaining six papers there is at least one application for which we can estimate the fully flexible model from Equation (16). Specifically, in the case of Aaronson and Mazumder (2011) we compare results from the fully flexible model with results reported in column 4 in their Table 1. Regarding Abramitzky et al. (2011) we present results for each of the three alternative definitions of outcome reported in the original paper in columns 1 to 3 of Table 3. We also report the estimates for five specifications whose original results are reported in Currie and Walker (2011) in columns 3 to 7 of their Table 7. For Furman and Stern (2011) our estimates are to be compared with those in the second column in their Table 3. Finally, in the case of Kotchen and Grant

(2011) we report estimates comparable with those in columns d of Tables 4 and 5 in the original paper.

In the first column of Table B2 we print the original estimates from the papers. In the second column we report results assuming Parallel-(1) within the fully flexible model from Equation (16). In the third column we allow for dynamic effects after treatment and report the effects at $s = 1$. In Kotchen and Grant (2011) there is only one post-treatment period, so there is no difference between the Restricted and Unrestricted models and the effects in the two columns are exactly the same by definition.In columns fourth and fifth we conduct the same analysis under Parallel(2). In the sixth column, we report the equivalence test of Parallel-(1) and Parallel-(2). In the last two columns we report the *common trends* test and the *linear trend* test for all cases in which there are more than two pre-treatment periods. Standard errors are computed using the same method as in the original papers.

Table B2: *Fully flexible model results and reported results from selected papers*

	Reported	Fully Flexible Model						
	Estimated	$q=1$		$q=2$		Equiv.	Common	Linear
Article	Effect	Restricted	Unrestricted	Restricted	Unrestricted	Test	Trends	Trend
Aaronson and	0.072***	0.034***	0.039***	0.054***	0.053***	1.420		
Mazumder	(0.007)	(0.012)	(0.012)	(0.017)	(0.017)	[0.234]		
Abramitzky et al. - 1	-0.020*	0.025	0.036	0.095	0.106	-0.069	13.41	13.287
	(0.010)	(0.024)	(0.039)	(0.064)	(0.073)	[0.118]	[0.020]	[0.010]
Abramitzky et al. - 2	-0.010***	-0.009	0.008	-0.006	0.010	-0.003	4.339	2.633
	(0.004)	(0.013)	(0.016)	(0.028)	(0.030)	[0.870]	[0.502]	[0.621]
Abramitzky et al. - 3	-0.017***	0.0005	0.003	0.028	0.031	-0.028	11.27	10.76
	(0.005)	(0.009)	(0.013)	(0.021)	(0.022)	[0.042]	[0.046]	[0.029]
Currie and Walker - 1	-0.208***	-0.298*	-0.506***	-0.178	-0.386	-0.121	652.85	323.79
	(0.028)	(0.175)	(0.198)	(0.376)	(0.395)	[0.600]	[0.000]	[0.000]
Currie and Walker - 2	-0.090***	-0.039	-0.582***	-0.528*	-1.071***	0.489	173.23	172.45
	(0.024)	(0.149)	(0.198)	(0.320)	(0.353)	[0.013]	[0.000]	[0.000]
Currie and Walker - 3	-0.065***	0.117***	0.029	0.224***	0.136	-0.107	351.47	282.06
	(0.017)	(0.036)	(0.101)	(0.084)	(0.128)	[0.079]	[0.000]	[0.000]
Currie and Walker - 4	-0.181***	-0.238***	-0.191*	-0.428***	-0.380*	0.189	581.98	316.51
	(0.023)	(0.089)	(0.108)	(0.190)	(0.204)	[0.100]	[0.000]	[0.000]
Currie and Walker - 5	0.018	-0.240	-0.421	-0.411	-0.592	0.171	268.15	246.61
	(0.038)	(0.294)	(0.374)	(0.681)	(0.736)	[0.714]	[0.000]	[0.000]
Furman and Stern	0.535***	0.446***	0.471***	0.643	0.666	0.262	69.26	69.83
	(0.142)	(0.133)	(0.123)	(0.425)	(0.417)	[0.610]	[0.000]	[0.000]
Kotchen and Grant -1	0.009***	0.006*	0.006*	-0.002	-0.002	7.28		
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	[0.007]		
Kotchen and Grant -2	-0.003	-0.006**	-0.006**	-0.013***	-0.013***	3.97		
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	[0.0471]		
Moser and Voena	0.151***	0.272***	0.075	0.204**	0.006	2.362	6.84	2.89
	(0.036)	(0.052)	(0.046)	(0.082)	(0.081)	[0.124]	[0.000]	[0.000]

Note: Reported Estimated Effect refers to the results originally published. Restricted reports the effect under $q = 1$ and $q = 2$ in the fully flexible model in Equation (16). Unrestricted reports the effects at $s = 1$ allowing for dynamic effects after treatment. Standard errors in parenthesis are computed using the same method as in the original paper. Equivalence tests for the equivalence of Parallel-(1) and Parallel-(2). Common Trends tests for the equivalence of all Parallel assumptions, except in Currie and Walker (2011), where we test the equivalence of all Parallel for $q \leq 52$. Linear Trend tests for the equivalence of all Parallel assumptions except Parallel-(1), again with the exception of Currie and Walker (2011). For all tests, p-values are shown in square brackets. Aaronson and Mazumder (2011) refers to estimates for Black rural using additional controls and county fixed effects (column 4 in their Table 1). Abramitzky et al. (2011) -1,-2, and -3 correspond to the three alternative definitions of a bad marriage outcome for the full sample of grooms (columns 1 to 3 in their Table 3). Each of Currie and Walker (2011) - 1 to -5 reports the estimates using as controls 1 of 5 randomly chosen monitors (columns 3 to 7 in their Table 7). Furman and Stern (2011) reports results comparable with those in the second column in their Table 3. Kotchen and Grant (2011) -1 reports the effect during the DST period while Kotchen and Grant (2011) -2 reports the effect during the non-DST period (column d in Tables 4 and 5 in the original paper). Moser and Voena (2012) refers to their results reported in their column 1 in Table 2. Levels of significance: * p<0.10, ** p<0.05, *** p<0.01.