UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 12-33
Economic Series (48)
December 2012

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Treatment Effect Identification Using Alternative Parallel Assumptions§

Ricardo Mora and Iliana Reggio*

Abstract

The core assumption to identify the treatment effect in difference-in-differences estimators is the so-called Parallel Paths assumption, namely that the average change in outcome for the treated in the absence of treatment equals the average change in outcome for the non-treated. We define a family of alternative Parallel assumptions and show for a number of frequently used empirical specifications which parameters of the model identify the treatment effect under the alternative Parallel assumptions. We further propose a fully flexible model which has two desirable features not present in the usual econometric specifications implemented in applied research. First, it allows for flexible dynamics and for testing restrictions on these dynamics. Second, it does not impose equivalence between alternative Parallel assumptions. We illustrate the usefulness of our approach by revising the results of several recent papers in which the difference-in-differences technique has been applied.

Keywords: difference-in-differences, treatment effect, parallel paths

§ This is an updated version (March 2013) of the December 2012 Working Paper with the same series number and title.

* Ricardo Mora, Universidad Carlos III Madrid, Spain. Iliana Reggio, Universidad Carlos III, Madrid, Spain. Both authors acknowledge financial help from the Spanish government through grant ECO2012-31358.

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. Arguably, two reasons stand behind the popularity of DID. First, its basic implementation under parametric assumptions only requires regression techniques. Second, data requirements are relatively weak. In its simplest version, only data from two periods are needed. In the first period—the pre-treatment period—none of the agents are exposed to the treatment. In the second period—the post-treatment period—those labeled as "treated" are already exposed to treatment while those labeled as "controls" are not. Importantly, although panel data is not required, the technique is robust to some forms of endogeneity arising from unobservable group-specific heterogeneity.

The appropriateness of the technique depends crucially on the validity of several assumptions which have been extensively discussed. For example, several authors focus on the parametric assumption behind the linear regression approach and discussed adjusting for exogenous covariates using propensity score methods (Abadie 2006, Blundell, Dias, Meghir, and Reenen 2004). Other authors generalize the technique in order to identify the entire counter-factual distribution of potential outcomes (Athey and Imbens 2006, Bonhomme and Sauder 2011). Worried by the accuracy of standard inference procedures, Donald and Lang (2007) and Bertrand, Duflo, and Mullainathan (2004) discuss problems with standard methods for computing standard errors. Yet, to our knowledge, little research has been devoted to the study of a critical assumption of the technique, the so-called 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 observed average change in the outcome variable for the controls. This assumption implies that differences between the controls and the treated if untreated are assumed time-invariant. Therefore, Parallel Paths is consistent with unobservable group-specific time-invariant heterogeneity.

We focus on applications in which several pre-treatment periods are available. In this context, Parallel Paths is appealing if trends do not differ before treatment (Angrist and Krueger, 1999). The simplest procedure to check common pre-treatment trends consists on conducting DID on the last pre-treatment period. In the presence of pre-treatment trend differentials, Parallel Paths becomes less attractive as it implies that differing

pre-treatment trends become equal after treatment under no treatment. In practice, researchers who find pre-treatment trend differentials often formulate flexible econometric models to accommodate those trend differentials. It would appear that several empirical strategies are then possible under the Parallel Paths assumption. For example, one could use group-specific invariant linear trends, i.e. group-specific linear trends which survive treatment. Alternatively, one could allow for pre- and post-treatment group-specific trends. After choosing a modeling strategy, the treatment effect is presumably identified as the parameter associated with an interaction of a post-treatment dummy and the treated indicator. This estimator is claimed to be a DID estimator.

We show that whether the interaction parameter identifies a DID estimator will depend both on the trend modeling strategy and the definition of the trend variable. For example, with group-specific pre-treatment linear trends, the trend has to be normalized to be zero in the last pre-treatment period. With group-specific invariant linear trends, however, the interaction term never identifies the policy effect under the Parallel Paths assumption. In this last case, the interaction term does identify the treatment effect under the alternative assumption of Parallel Paths for output first differences (rather than for output levels). We generalize these findings by proposing a family of alternative Parallel assumptions which widen the set of alternative estimators under fully flexible dynamics.

As illustration, we discuss in detail the case of assuming Parallel Paths for outcome first differences rather than for outcome levels. We refer to this assumption as Parallel Growths. We first show that under Parallel Growths the effect of treatment one period after treatment is identified by a difference-in-double-differences operator on outcome levels. We then show that Parallel Growths is equivalent to Parallel Paths in the presence of common pre-treatment trends. In contrast to Parallel Paths, Parallel Growths is consistent with group-specific trends in the post-treatment period unrelated to treatment. Moreover, the counterfactual outcome for the treated if untreated is obtained by adding the average acceleration experienced by the controls to the pre-treatment outcome change of the treated.

We expand this analysis and identify the treatment effect under alternative Parallel assumptions. We show that these alternative assumptions lead in general to differences in the identification of the treatment effect. We then provide the conditions under which different assumptions lead to equivalent identifications of the effect.

In empirical work, treatment effects are frequently obtained using standard linear regression techniques. We discuss several econometric specifications and show identification conditions of the treatment effect under alternative Parallel assumptions. We further present a general additive regression model with fully flexible dynamics. We argue that the fully flexible model has two advantages over usual models proposed in the literature. First, it allows for flexible dynamics and for testing restrictions on these dynamics. Second, it does not impose equivalence between alternative Parallel assumptions—and tests for this equivalence are easy to implement.

Finally, we explore how relevant is our proposal in practice by applying it to data obtained from several recent papers. We study to what extent using a fully flexible model and considering alternative Parallel assumptions modifies the conclusions obtained. We find that results and their significance vary depending on which trend assumption is used and that different Parallel assumptions often lead to significantly different treatment effect estimates.

The rest of the paper is structured as follows. We first define Parallel Paths, Parallel Growths, and present the family of alternative Parallel assumptions in Section 2. Next we discuss several econometric specifications and present a model with fully flexible dynamics in Section 3. In Section 4 we review current practice and explore the practical relevance of our proposal. Finally, we conclude by summarizing our argument and suggesting a change in the implementation of DID estimation in applications in which several pre-treatment periods are available.

2 Alternative Parallel assumptions

In the simplest empirical DID application we have information on the variable of interest in at least two periods: before and after the treatment. More generally, treatment starts sometime after the last pre-treatment period, t^*, and finishes before the first post-treatment period, t^*+1.[1] We have information for $T_0 \geq 2$ periods before treatment and $S \geq 1$ periods after treatment during which the effect of the treatment is to be evaluated

[1] In some empirical applications, treatment date is individual specific and it is not appropriate to define a pre-treatment and a post-treatment period for the controls. Identification of the treatment effect then additionally exploits heterogeneity in timing of treatment. Although we do not address directly this situation, a simple modification in the models accommodates this richer data structure.

(additional post-treatment periods may be available).

Following conventional notation we define Y_t as the observed outcome variable at period t. Let Y_t^0 denote outcome in period t when the individual receives no treatment, and Y_t^1 the outcome in period t when the individual receives treatment. For a given individual either Y_t^0 or Y_t^1 is observed. Let $D = 1$ if the individual receives treatment and $D = 0$ otherwise. Potential and observed outcomes are related to D by $Y_t = Y_t^1 D + Y_t^0 (1 - D)$ for $t > t^*$. For any pre-treatment periods, $Y_t = Y_t^0$. Finally, let $X = \left\{X_{t_1}', ..., X_{t_T}'\right\}'$ where $X_t \in \mathcal{X} \subset \mathbb{R}^k$ is a vector of k individual characteristics.

We study identification conditions for the average treatment effect $s \leq S$ periods after treatment on the treated with individual characteristics X, i.e. identification conditions for

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

where $s = 1, ..., S$.

The estimation of $\alpha(s|X)$ is problematic because $Y_{t^*+s}^0$ is not observable for the treated. In order to estimate the average counterfactual, one can propose an assumption on how the trend behavior of the treated if untreated compares to observed trend behavior of the untreated. The DID estimator, for example, stems from the so-called Parallel Paths assumption.

2.1 The Parallel Paths assumption

At the core of the DID identification strategy for $E\left[Y_{t^*+s}^0 \,|X, D = 1\right]$ lies the so-called Parallel Paths assumption. Let L be the lag operator so that $\Delta \equiv (1 - L)$ denotes the first difference operator and $\Delta_s \equiv (1 - L^s)$, $s \geq 2$, denotes the s-period difference operator. Parallel Paths can be stated as follows.

Assumption 1. *Parallel Paths s Periods Ahead*

$$E\left[\Delta_s Y_{t^*+s}^0 \,|X, D = 1\right] = E\left[\Delta_s Y_{t^*+s}^0 \,|X, D = 0\right], \ \textit{with } s > 0. \tag{2}$$

Parallel Paths states that average changes in output among those treated if untreated are equal to the observed average changes among comparable controls.

Given that for any variable z_t, $z_{t+s} = z_t + \Delta_s z_{t+s}$, using Parallel Paths and the fact that $Y_t^0 = Y_t$ for $t \leq t^*$, we write the counterfactual as:

$$E\left[Y_{t^*+s}^0 \left| X, D=1\right.\right] = E\left[Y_{t^*} \left| X, D=1\right.\right] + E\left[\Delta_s Y_{t^*+s} \left| X, D=0\right.\right] \tag{3}$$

The counterfactual scenario for those treated at $t = t^* + s$ is built by adding the observed average increase in the controls to the last pre-treatment level of the treated. Using this counterfactual, the policy effect in period s, $\alpha(s|X)$, can be expressed as the difference in observed output changes among treated and controls, the difference-in-differences operator s-periods ahead:

$$\alpha(s|X) = E\left[\Delta_s Y_{t^*+s} \left| X, D=1\right.\right] - E\left[\Delta_s Y_{t^*+s} \left| X, D=0\right.\right] \tag{4}$$

In the simple case in which there is only one post-treatment period, $S = 1$, this is the DID operator, $\alpha(1|X) = E\left[\Delta Y_{t^*+1} \left| X, D=1\right.\right] - E\left[\Delta Y_{t^*+1} \left| X, D=0\right.\right]$.

2.2 The Parallel Growths assumption

Consider the three-period situation depicted in Figure 1 in which controls and treated have constant trends before and after treatment. Assume that trends do not change because treatment has no average effect on output.

Parallel Paths identifies $\alpha(1|X)$ by assuming that those treated would have experienced, if untreated, the same average output change as the controls. In a situation like the one described in Figure 1, this assumption would lead to a spurious overestimation of the policy effect due to the underestimation of the counterfactual trend for those treated. A plausible solution in this particular case would be to add to the econometric model for the conditional expectation of the observed outcomes group-specific time-invariant linear trends. Intuitively, these terms would account for the differing trends before and after treatment. In particular, for simplicity assume there is no vector of individual

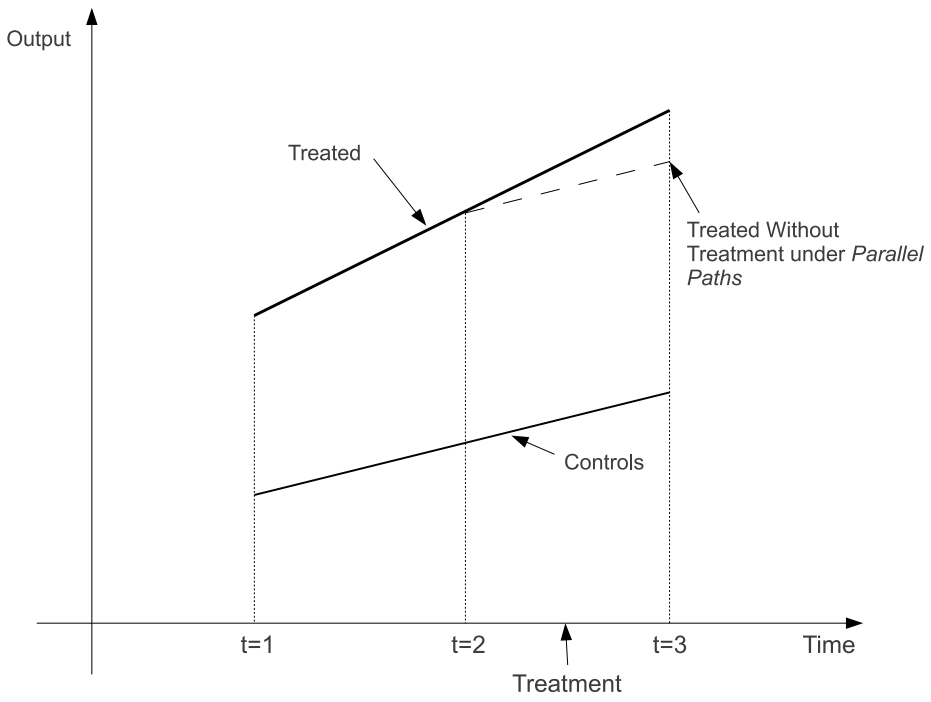

Figure 1: *Difference-in-differences estimation with group-specific time-invariant linear trends.*

controls X so that the conditional expectation takes the form

$$E\left[Y_t \left| D\right.\right] = \delta_0 + \delta_L time_t + \delta_P Post_t + \gamma^D D + \gamma_L^D time_t D + \gamma_P^D Post_t D \tag{5}$$

where $Post_t$ is a step function with value 1 if the observation is from the post-treatment period and 0 otherwise, and $time_t$ is a linear trend such that $time_{t+1} = time_t + 1$. Delta parameters specify common dynamics between controls and treated. Thus, δ_P captures a shift in output after treatment common to all individuals. Given that γ^D and γ_L^D control for group differences in linear trends, one could claim that the parameter of the interaction term, γ_P^D, equals the DID operator after controlling for group-specific linear trends. Note, however, that this statement does not reveal the true identifying assumption for the treatment effect. The treatment effect can only be γ_P^D when the counterfactual for the average growth among the treated if untreated is equal to

$$E\left[\Delta Y_3^0 \left| D = 1\right.\right] = \delta_L + \delta_P + \gamma_L^D \tag{6}$$

so that the identification of the treatment effect in this case is no longer based on the Parallel Paths Assumption, i.e. $E\left[\Delta Y_3^0 \left| D = 1\right.\right] \neq E\left[\Delta Y_3^0 \left| D = 0\right.\right] = \delta_L + \delta_P$.

In this example, one assumption that identifies the treatment effect as γ_P^D is:

Assumption 2. *Parallel Growths*

$$E\left[\Delta_s \Delta Y_{t^*+s}^0 \left| X, D = 1\right.\right] = E\left[\Delta_s \Delta Y_{t^*+s}^0 \left| X, D = 0\right.\right], \; s \in \{1, ..., S\}. \tag{7}$$

One way of thinking about Parallel Growths is that it shifts the variable of interest from the output in levels to the output in first differences: changes in output growth for those treated if untreated would have been equal to the observed changes in output growth for the controls. Parallel Growths implies that in absence of treatment the treated and the controls would have had parallel growth paths .

For the case $s = 1$, under Parallel Growths,

$$\begin{aligned} E\left[Y_{t^*+1}^0 \left| X, D = 1\right.\right] = {} & E\left[Y_{t^*} \left| X, D = 1\right.\right] + \\ & E\left[\Delta Y_{t^*} \left| X, D = 1\right.\right] + E\left[\Delta^2 Y_{t^*+1} \left| X, D = 0\right.\right] \end{aligned} \tag{8}$$

The counterfactual output in period $t^* + 1$ for those treated if untreated is constructed with the average growth for the treated at t^* plus the average acceleration experienced by the controls at $t^* + 1$. In contrast, the counterfactual under Parallel Paths is obtained only with the average growth experienced by the controls at $t^* + 1$. Hence, Parallel Growths allows for group-specific trends before and after treatment while Parallel Paths only allows for different trends before treatment.

It follows from equation (8) that under Parallel Growths the treatment effect the first period after treatment, $\alpha(1|X)$, equals a "difference-in-double-differences" operator,

$$\alpha(1|X) = E\left[\Delta^2 Y_{t^*+1}|X, D=1\right] - E\left[\Delta^2 Y_{t^*+1}|X, D=0\right]. \tag{9}$$

Parallel Growth and Parallel Paths are equivalent if and only if the DID operator equals this difference-in-double-differences operator:

$$\begin{aligned} &E\left[\Delta^2 Y_{t^*+1}|X, D=1\right] - E\left[\Delta^2 Y_{t^*+1}|X, D=0\right] = \\ &\quad E\left[\Delta Y_{t^*+1}|X, D=1\right] - E\left[\Delta Y_{t^*+1}|X, D=0\right] \end{aligned} \tag{10}$$

or, equivalently,

$$E\left[\Delta Y_{t^*}|X, D=1\right] = E\left[\Delta Y_{t^*}|X, D=0\right] \tag{11}$$

Thus, in the presence of pre-treatment group-specific trends, the identification for the treatment effect will be different under Parallel Paths or Parallel Growths.

For the case $s \geq 2$, from the definition of $\alpha(s|X)$, it can be shown that $\alpha(s|X) = \alpha(s-1|X) + E\left[\Delta Y^1_{t^*+s} - \Delta Y^0_{t^*+s}|X, D=1\right]$. Given that $E\left[\Delta Y^1_{t^*} - \Delta Y^0_{t^*}|X, D=1\right] = 0$ we have that

$$E\left[\Delta Y^1_{t^*+s} - \Delta Y^0_{t^*+s}|X, D=1\right] = E\left[\Delta_s \Delta Y^1_{t^*+s} - \Delta_s \Delta Y^0_{t^*+s}|X, D=1\right]. \tag{12}$$

Under Parallel Growths, the variable on which the Parallel Paths assumption is applied

is the variable in first differences. Thus, instead of the treatment effect, a difference-in-differences operator identifies the change in the treatment effect:

$$\Delta\alpha\left(s|X\right)=E\left[\Delta_{s}\Delta Y_{t^{*}+s}\left|X,D=1\right.\right]-E\left[\Delta_{s}\Delta Y_{t^{*}+s}\left|X,D=0\right.\right],\ s\geq2 \tag{13}$$

where $\Delta\alpha\left(s|X\right)\equiv\alpha\left(s|X\right)-\alpha\left(s-1|X\right)$.

2.3 A general family of Parallel assumptions

Generalizing from the discussion on Parallel Paths and Parallel Growths, we propose a family of alternative non-nested assumptions:

Assumption 3. *Parallel-*(q,S)

For a given positive integer $q\leq T_{0}$*, and for any* $s=1,...,S$*,*

$$E\left[\Delta_{s}\Delta^{q-1}Y_{t^{*}+s}^{0}\left|X,D=1\right.\right]=E\left[\Delta_{s}\Delta^{q-1}Y_{t^{*}+s}^{0}\left|X,D=0\right.\right]. \tag{14}$$

For $q=1$, Parallel-(q,S) is Parallel Paths, while for $q=2$, Parallel-(q,S) is Parallel Growths. In these two particular cases, we have already established the link between difference-in-differences operators and the treatment effects $\alpha\left(s|X\right)$. To generalize this link to the (q,S) case, we first define $did\left(q,s\right)$ as the difference-in-q-differences operator s periods ahead,

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

The next theorem characterizes the treatment effect under Parallel-(q,S).

Theorem 1. *Under Parallel-*(q,S)*,*

$$\Delta^{q-1}\alpha\left(s|X\right)=did\left(q,s\right)$$

where $\Delta^{q-1}\equiv\left(1-L\right)^{q-1}$ *and* $L^{r}\alpha\left(s|x\right)=0$ *for all* $r\geq s$*.*

Proof. See the Appendix. □

Theorem 1 can be used to obtain $\alpha(s|X)$ for any value of s under Parallel-(q,S) recursively. For example, for $s=1$, $\Delta^{q-1}\alpha(1|X)=\alpha(1|X)$ so that Theorem (1) states that the treatment effect is the difference-in-q-differences operator one period ahead, $did(q,1)$.[2]

When $s>1$ the link between the treatment effect and the difference-in-q-differences operator s periods ahead will depend on q. For example, under Parallel Paths (i.e. when $q=1$) $\alpha(s|X)$ equals $did(1,s)$ for any $s=1,...,S$. In contrast, under Parallel Growths (i.e. $q=2$), $\alpha(s|X)$ is the cumulative effect of the difference-in-double-differences operators up to s, $\sum_{j=1}^{s} did(2,j)$.

Define the operator $\alpha^q(s|X)$ as the mapping on $did(q,s)$ that identifies the true effect of treatment under Parallel-(q,s). The conditions under which $\alpha^q(s|X)$ equals $\alpha^{q-1}(s|X)$ are given by the next theorem.

Theorem 2. *For any* $q\in\{2,...,T_0\}$ *and* $s\in\{1,...,S\}$,

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

if and only if

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

Proof. See the Appendix. □

Theorem 2 sets pre-treatment trend conditions under which assumptions Parallel-(q,s) and Parallel-$(q-1,s)$ are equivalent.

For the important case in which $q=2$, it states that in the presence of pre-treatment group-specific trends $\alpha^1(s|X)$ cannot be equal to $\alpha^2(s|X)$ for any $s\leq S$. The frequent comparison in empirical work of pre-treatment trends between treated and controls can be seen as an informal test for the equivalence of Parallel Paths and Parallel Growths.

[2]Since under Parallel-$(q,1)$ the counterfactual output in period t^*+1 equals

$$E\left[Y^0_{t^*+1}|X,D=1\right]=\sum_{r=0}^{q-1}E\left[\Delta^r Y_{t^*}|X,D=1\right]+E\left[\Delta^q Y_{t^*+1}|X,D=0\right], \quad (15)$$

the treated if untreated and the controls may differ in their average output differences up to order $q-1$.

3 Regression techniques

In empirical work, treatment effects are frequently obtained using standard linear regression techniques. In the simplest case with only two periods, the treatment effect can be estimated from a regression that includes a constant, the treated indicator D, a dummy variable for the post-treatment period, $Post_t$, and an interaction term, $Post_t \times D$.[3] In this set up, the treatment effect is identified by the parameter associated with the interaction term.

On applications in which several pre-treatment periods are available, the standard model allows for time fixed effects δ_t (Bertrand, Duflo, and Mullainathan, 2004):

$$E\left[Y_t \left| D\right.\right] = \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \gamma_P^D Post_t \times D \tag{16}$$

where $I_{\tau,t}$ is a dummy for period τ. The specification in equation (16) is restrictive in two ways. The first restriction is that pre-treatment dynamics—captured by time fixed effects—are identical for controls and treated. By Theorem 2, this implies that all Parallel assumptions are equivalent. In other words, average first and higher order differences in output levels are equal for the two groups in the absence of treatment. The second restriction is that there is a permanent shift in output of size γ_P^D in the first period after treatment. Hence, the long-term effect of treatment is already present at $t^* + 1$.[4]

In the presence of group-specific trends and when treatment has different short-run and long-run effects, the specification of the conditional expectation in equation (16) is inappropriate. Consequently, in empirical studies where these considerations arise, more flexible econometric specifications are frequently proposed. We revise several of these econometric specifications and show under which Parallel assumptions, if any, an interaction parameter identifies the treatment effect.

[3] A set of controls X is usually included although we omit it in this section for simplicity.

[4] Given that all Parallel assumptions are equivalent, the easiest way to see the effect s periods ahead is by noting that $did\left(1, s\right) = \gamma_P^D$.

3.1 A polynomial trend for the differences in group dynamics

One way to extend the standard model in equation (16) to accommodate group-specific trends is by including an interaction between D and a polynomial time trend:[5]

$$E\left[Y_t \left| D\right.\right] = \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \gamma_P^D Post_t \times D + \sum_{r=1}^{R} \gamma_r^D t^r \times D. \tag{17}$$

In the standard model, pre-treatment dynamics are identical for controls and treated. In contrast, in equation (17) the polynomial $\sum_{r=1}^{R} \gamma_r^D t^r$ captures differences in group dynamics which predate treatment and remain after treatment. The equivalence of equations (16) and (17) depends on the values for γ_r^D. It is misleading to state that the inclusion of group-specific trends makes a given Parallel assumption, say Parallel Paths, more plausible. Next, we show that including these trends rather changes the Parallel assumption under which the interaction term identifies the treatment effect.

By applying Theorem 1 to equation (17) we have that

$$\Delta^{q-1}\alpha^q\left(s\right) = \gamma_P^D \Delta^{q-1} Post_{t^*+s} + \sum_{r=1}^{R} \gamma_r^D \Delta_s \Delta^{q-1}\left(t^* + s\right)^r \tag{18}$$

so that the treatment effect after s periods will generally not be identified as the interaction term γ_P^D. Consider, for example, the case $s = 1$:

$$\alpha^q\left(1\right) = \gamma_P^D + \sum_{r=1}^{R} \gamma_r^D \Delta^q\left(t^* + 1\right)^r. \tag{19}$$

Since $\Delta^q\left(t^* + 1\right)^r = 0$ for all $q > r$, the interaction term γ_P^D does identify the treatment effect one period after treatment if $q > R$. If this condition is not satisfied, then γ_P^D may not identify the effect. For example, since $\Delta^R\left[\left(t^* + 1\right)^R\right] = R$, then when $q = R$, $\alpha^R\left(1\right) = \gamma_P^D + R\gamma_R^D$. In the particular case of a linear polynomial, $R = 1$, the interaction term identifies the treatment effect one period ahead if we assume at least Parallel Growths ($q \geq 2$). In contrast, under Parallel Paths, $\Delta\left(t^* + 1\right) = 1$ and the treatment

[5] When the cross-section is small and identification of the time fixed effects is poor, it is frequently assumed that dynamics for controls follow a polynomial, usually linear or quadratic, trend. For brevity, we do not discuss in detail this case but the main identification implications of the model are similar.

effect is compounded by both the permanent shift and the linear trend differential, $\alpha^1(1) = \gamma_P^D + \gamma_1^D$.

In general, identification of the treatment effect s periods ahead is complex to evaluate because there is no closed form solution for the operator $\Delta_s \Delta^{q-1}$ onto the term $(t^* + s)^r$. However, for the case $q > R$, $\Delta_s \Delta^{q-1}(t^* + s)^r = 0$, for all $r \leq R$, and, by Theorem 1, $\Delta^{q-1}\alpha^q(s) = \gamma_P^D \Delta^{q-1} Post_{t^*+s}$. Given that $\Delta^{q-1}\alpha^q(s) = \Delta^{q-1} Post_{t^*+s}\alpha^q(s)$, the difference operators cancel out and $\alpha^q(s) = \gamma_P^D$ for any $s \geq 1$. For the case $q \leq R$, we comment the inclusion of a linear or a quadratic time trend (that are the most frequent empirical cases as in Friedberg, 1998, and Wolfers, 2006).

A linear time trend: $\sum_{r=1}^{R} \gamma_r^D t^r = \gamma_1^D t$. Under Parallel-$(q, s)$ with $q \geq 2$, the treatment effect after $s \geq 1$ periods is γ_P^D. Under Parallel Paths the treatment effect is linear in s, $\alpha^1(s) = \gamma_P^D + \gamma_1^D s$. Parallel Paths states that in the absence of treatment, treated and controls are comparable in changes. Any difference in observed trends after treatment are assumed to arise because of treatment. Hence, the parameter for the linear time trend that captures differences in group dynamics (before and, crucially, after treatment) is included in the effect. In sum, the identification of the treatment effect with γ_P^D implies a departure from Parallel Paths but it is still consistent with any other Parallel assumption.

A quadratic time trend: $\sum_{r=1}^{R} \gamma_r^D t^r = \gamma_1^D t + \gamma_2^D t^2$. Under Parallel-$(q, s)$ with $q \geq 3$, $\alpha^q(s) = \gamma_P^D$. Under Parallel Paths, $\alpha^1(s) = \gamma_P^D + \left(2t^*\gamma_2^D + \gamma_1^D\right) s + \gamma_2^D s^2$ and under Parallel Growths, $\alpha^2(s) = \gamma_P^D + \gamma_2^D (s+1)\, s$. Parallel Paths implies that group differentials in output first differences arise because of treatment. Since these differentials follow a quadratic polynomial, the treatment effect track these differentials and is also quadratic in s. Under Parallel Growths, treated and controls are comparable in acceleration rates in the absence of treatment. Hence, the parameter associated with acceleration in the quadratic time trend also appears in the treatment effect. In contrast to $\alpha^1(s)$, $\alpha^2(s)$ does not include the parameter γ_1^D since controls and treated are not comparable in first differences. In line with Theorem 2, $\alpha^1(s)$ and $\alpha^2(s)$ are the same only if pre-treatment changes are equal among treated and controls, i.e. if $\gamma_1^D - \gamma_2^D + 2t^*\gamma_2^D = 0$.[6] To summa-

[6] If the time trend is normalized so that $t^* = 0$, $\alpha^1(s) = \gamma_P^D + \gamma_1^D s + \gamma_2^D s^2$ and $\alpha^1(s) = \alpha^2(s)$ only if $\gamma_1^D = \gamma_2^D$.

rize the quadratic case, identification of the treatment effect with γ_P^D implies a departure from both Parallel Paths and Parallel Growths but it is still consistent with any other Parallel assumption.

3.2 Modeling flexible dynamics for the treatment effect

One way to extend the standard model to accommodate flexible dynamics for the treatment effect is by adding interactions between D and the time dummies after treatment:

$$E\left[Y_t\left|D\right.\right] = \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \sum_{s=1}^{S} \gamma_s^D \times I_{t^*+s,t} \times D. \tag{20}$$

By applying the difference operators $\Delta_s \Delta^{q-1}$ on both sides of equation (20) and taking into account that $\Delta_s \gamma_s^D = \gamma_s^D$, we have that

$$E\left[\Delta_s \Delta^{q-1} Y_{t^*+s}\left|D\right.\right] = \Delta_s \Delta^{q-1} \delta_{t^*+s} + \Delta^{q-1} \gamma_s^D \times D. \tag{21}$$

Applying Theorem 1, we have that $\Delta^{q-1}\alpha^q\left(s\right) = \Delta^{q-1}\gamma_s^D$ and, given that $\alpha^q\left(s\right) = 0$ for $s \leq 0$, the difference operators cancel out and $\alpha^q\left(s\right) = \gamma_s^D$ for all q and s. In contrast to the standard specification and also to equation (17), equation (20) provides a fully flexible form to capture the response function to treatment (for an application, see Wolfers, 2006). A less flexible version of dynamic effects often used in empirical work (see, for example, Groen and Polivka, 2008) assumes a linear time trend for the effect of treatment, $\gamma_s^D = \gamma_p^D + \gamma_L^D\left(t^* + s\right)$, so that:

$$E\left[Y_t\left|D\right.\right] = \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \gamma_p^D Post_t \times D + \gamma_L^D Post_t \times t \times D. \tag{22}$$

An essential aspect in equations (20) and (22) is that pre-treatment dynamics are identical for controls and treated. Hence, as already stated, all Parallel assumptions are equivalent. Moreover, as with the standard specification from equation (16), these models are inappropriate in the presence of group specific pre-treatment trends. In two of his three specifications, Wolfers (2006) includes group specific linear and quadratic time trends, i.e. he combines equations similar to equations (17) and (20). The additive com-

bination of the two models affects the identification of the treatment effect as expected. First, for the case $q > R$, $\alpha^q(s) = \gamma_s^D$ for any $s \geq 1$. Second, for the quadratic case, $\alpha^1(s) = \gamma_s^D + \left(2t^*\gamma_2^D + \gamma_1^D\right)s + \gamma_2^D s^2$ and $\alpha^2(s) = \gamma_s^D + \gamma_2^D(s+1)s$. Hence, identification of the treatment effect s periods after treatment with γ_s^D implies a departure from both Parallel Paths and Parallel Growths but it is still consistent with any other Parallel assumption. Finally, for the linear case, identification of the treatment effect with γ_s^D implies a departure from Parallel Paths but it is still consistent with any other Parallel assumption.

3.3 A fully flexible model

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

$$E[Y_t|D] = \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \sum_{\tau=t_2}^{T} \gamma_\tau^D \times I_{\tau,t} \times D. \tag{23}$$

Theorem 3. *Under Parallel-(q, S) and equation (23):*

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

Proof. See the Appendix. □

The operators Δ^{q-1} at each side of the equation do not cancel out because while $L^r\alpha(s) = 0$ for all $r \geq s$ by definition, γ_τ^D may be different from 0 for some $\tau \leq t^*$. This implies that the effect of treatment will generally differ under alternative Parallel assumptions. Only in the case in which $\gamma_\tau^D = 0$ for all $\tau \leq t^*$, i.e. only when pre-treatment trends are equal on average between treated and controls, then $\alpha(s) = \gamma_{t^*+s}^D$. Hence, the test of the null hypothesis of common pre-treatment trends ($H_0 : \gamma_\tau^D = 0$ for all $\tau \leq t^*$) is a test for the simultaneous equivalence of all Parallel$-(q, S)$ assumptions.

The inclusion in equation (23) of fully flexible pre-treatment trend differentials between treated and controls allows for the comparison of any two consecutive Parallel-(q, S) assumptions. From Theorem 2, testing the null $H_0 : \Delta^{q-1}\gamma_{t^*}^D = 0$ vs. the alternative $H_1 : \Delta^{q-1}\gamma_{t^*}^D \neq 0$ with $1 < q \leq T_0$ is a test for the equivalence of Parallel-(q, S) and

Parallel-$(q-1, S)$. In the leading case of Parallel Paths and Parallel Growths, the test would be $H_0 : \gamma^D_{t^*} = \gamma^D_{t^*-1}$ vs. $H_1 : \gamma^D_{t^*} \neq \gamma^D_{t^*-1}$.

As in equation (20), the inclusion of fully flexible post-treatment trend differentials allows us to implement tests on the dynamics of the treatment effect. For example, under $q = 1$ testing the null $H_0 : \gamma^D_{t^*+s} = \gamma^D_{t^*+s+1}$ with $s = 1, ..., S-1$ is a test for the effect to be constant in the post-treatment period.

Despite the flexibility gained in equation (23), we are aware of only one empirical work which uses a close specification (Reber, 2005). In some cases, not using equation (23) is justified because data requirements are simply not met. When data requirements are met, but results do not change with additional flexibility, there might be a reason to use a more parsimonious model. In the next section, we explore how empirical results obtained from less flexible models are robust to the fully flexible specification in equation (23).

4 A brief review of current practice

In this section we explore to what extent using a fully flexible model and considering alternative Parallel assumptions modifies conclusions obtained in several recent papers. 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 (see Table 1 for the list). We look at the last four years as we are primarily interested in current practice. The journals chosen are characterized by being among the highest ranked economic journals on several criteria and also by having the policy of allowing access to the data sets used in the published papers (at least during some part of the searched period).

The second condition is that the paper must include an application of DID. We identify these papers by a search on the terms "difference-in-differences" or "diff-in-diff" on the paper (with the exception of the bibliography section) and, for those papers which include these terms, by checking that the DID application does exist. We find 59 papers which 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 a request to access the data was required. Therefore, 22 papers also satisfy this additional requirement.

The final condition is that the data must include more than one pre-treatment period and that controls and treated must have observations before and after treatment. If there were only two periods, then the only implementable Parallel-(q, S) assumption is Parallel-$(1, S)$. We are especially interested in the comparison between Parallel-$(1, S)$ and Parallel-$(2, S)$, so that at least two pre-treatment periods are required. There are 13 papers which do not satisfy this condition.

Table 1: *List of Selected Papers*

Author	Year	Journal	Title
Aaronson and Mazumder	2011	JPE	The impact of Rosenwald Schools on Black achievement
Abramitzky, Delavande, and Vasconcelos	2011	AEJ:AE	Marrying Up: The Role of Sex Ratio in Assortative Matching
Currie and Walker	2011	AEJ:AE	Traffic Congestion and Infant Health: Evidence from E-ZPass
De Jong, Lindeboom, and Van der Klaauw	2011	JEEA	Screening disability insurance applications
Jayachandran, Lleras-Muney, and Smith	2010	AEJ:AE	Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs
Furman and Stern	2011	AER	Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research
Kotchen and Grant	2011	REStat	Does Daylight Saving Time Save Energy? Evidence from a Natural Experiment in Indiana
Moser and Voena	2012	AER	Compulsory Licensing: Evidence from the Trading with the Enemy Act
Redding, Sturm, and Wolf	2011	REStat	History and industry location: Evidence from German airports

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 is 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.

In total, nine papers, listed in Table 1, meet all requirements. In five papers, the DID econometric specification is similar to the standard model described in equation (16). In one of them, the dependent variable is the output variable in first differences. In two other papers where panel data is used, individual-specific linear time trends for differences in dynamics are used. One of the two remaining papers presents two models: a model with group-specific linear time trends, equation (17) with a linear time trend, and an extension of this model to allow for a linear time trend for the effect of treatment, a linear version of equation (22). The other paper considers a short-run treatment effect for a window of three periods, and a permanent treatment effect for later periods.

In all specifications but one in Jayachandran, Lleras-Muney, and Smith (2010), the treat-

ment effect is identified as the interaction term. Clearly, there are no discussions about which Parallel-(q, s) assumption is used for the identification of the counterfactual and the treatment effect. This discussion is not relevant—in the sense that the treatment effect is identified by the interaction term for all q—when there are no group-specific pretreatment trends, as in the standard model in equation (16).[7] However, when the econometric specification allows for group-specific pre-treatment trends, as in the linear time trends models proposed in some of the papers, the discussion becomes relevant because the parameters that identify the treatment effect depend on the Parallel assumption. For example, only $q \geq 2$ is consistent with the interaction term as the treatment effect when comparing the estimates from the standard and the linear time trend model.

We use the two econometric models considered in Jayachandran, Lleras-Muney, and Smith (2010) to illustrate the need to state the identifying Parallel assumption for some specifications. The authors study the contribution of sulfa drugs, a groundbreaking medical innovation in the 1930s, to declines in US mortality. Their first specification includes linear time trends for both treated and controls:

$$y_{it} = \beta_0 + \beta_1 D_i \times Post_t + \beta_2 D_i \times year_t + \beta_3 D_i + \beta_4 year_t + \beta_5 Post_t + \epsilon_{it} \tag{24}$$

where y_{it} is the yearly average log of the U.S. mortality rate of illness i at year t, D_i is an indicator for illness combated with sulfa drugs after 1937, $year_t$ is a year trend normalized to 0 in 1937, and $Post_t$ is a dummy indicator for year later than 1936.

From the discussion in section 3, $\alpha^1(s) = \beta_1 + \beta_2 s$ and $\alpha^q(s) = \beta_1$ for $q \geq 2$. Jayachandran, Lleras-Muney, and Smith (2010) focus on β_1 as the coefficient of interest, which is consistent with $q \geq 2$. In their second specification they add an interaction between D_i, $Post_t$, and $year_t$ to allow for "a change in both the intercept and the slope after 1937":

$$\begin{aligned} y_{it} =& \beta_0 + \beta_1 D_i \times Post_t \times year_t + \beta_2 D_i \times Post_t + \beta_3 D_i \times year_t \\ &+ \beta_4 D_i + \beta_5 year_t + \beta_6 Post_t + \epsilon_{it}. \end{aligned} \tag{25}$$

They claim that "(i)n this model, the statistical question of interest is whether β_1 and β_2 are jointly significantly different from zero." Policy effects under this specification are more complex under both Parallel-$(1, S)$ and Parallel-$(2, S)$. First, under Parallel-$(1, S)$

[7] In Redding, Sturm, and Wolf (2011), $q > 1$ since the dependent variable is the output in first differences.

and taking into account the normalization in $year_t$, $\alpha^1(s) = \beta_2 + \beta_1(s-1) + \beta_3 s$ for any $s \geq 1$. Under Parallel-(q, S) with $q \geq 2$, $\alpha^q(s) = \beta_2 + \beta_1(s-1)$ so that in 1937 (i.e. $year_{1937} = 0$) the effect depends only on β_2 while after 1937, it depends both on β_1 and β_2.[8] Thus, the dynamic effects are captured by β_1 and β_3 under $q = 1$ and by β_1 under $q = 2$.

In the remaining part of this section, our goal is to discuss how sensitive DID techniques are in practice to alternative Parallel assumptions and model specifications. We do not attempt to review the main results of the papers, which in some cases are not derived from the DID application implemented. We must include all applications that satisfy the criteria regardless of whether the DID application is the most important evidence provided. Hence, it is important to stress that our revision exercise should not be used to question the main results of the original papers. In two papers (Jayachandran, Lleras-Muney, and Smith 2010 and Redding, Sturm, and Wolf 2011) it is not possible to use the flexible specification in equation (23) because there is only one treated agent and one control. In another paper which uses a panel (De Jong, Lindeboom, and Van der Klaauw 2011), we cannot estimate a fully flexible version because the authors include individual-specific linear trends. For these three papers, we discuss how alternative assumptions may yield different results, but a direct comparison of results using the flexible model with their own reported ones is not possible.

De Jong, Lindeboom, and Van der Klaauw (2011): The authors analyze the impact of stricter screening of disability insurance applications on long-term sickness absenteeism and disability insurance applications. Using a standard DID model with two years of data (2002 and 2003) for 26 Dutch regions, they find that stricter screening significantly reduces long-term sickness absenteeism reports. They also obtain a negative estimate on disability insurance applications, but the estimate is not significant. They implement a DID estimate at the last period before treatment to "test for conditional random assignment". This is in fact the equivalence test for $q = 2$ and $q = 1$. They cannot reject that both assumptions provide the same estimated treatment effects. They claim to relax the Parallel Paths assumption by allowing for region-specific time trends. They then estimate the treatment effect by regressing the second difference in 2003 in the outcome variable on the treatment dummy. This is not a relaxation of the Parallel

[8] With an alternative normalization, the effect in the first period would also depend on β_1, but it would still be different from the effect in later periods.

Paths assumption, but assuming $q \geq 2$ with regional specific linear time trends. They obtain significant and larger effects on both outcomes.[9]

Jayachandran, Lleras-Muney, and Smith (2010): As already said, they study the contribution of sulfa drugs to declines in US mortality. They provide national and regional evidence using aggregates of US mortality rates of three different conditions treated with sulfa drugs: maternal complications after child birth, pneumonia, and scarlet fever. We comment for brevity and data availability the results concerning scarlet fever and national aggregates. The estimate for β_2 in their first specification—equation (24)—is not significant (p-value 0.75) so that Parallel Paths and Parallel Growths cannot be rejected to be equivalent. By construction the effect is constant and captured by $\hat{\beta}_1 = -0.877$. For their second specification—equation (25)—$\hat{\beta}_3 = 0.049$ is strongly significant (t-ratio 7.27) so that Parallel Paths and Parallel Growths are not equivalent. Moreover, $\hat{\beta}_1 + \hat{\beta}_3 = -0.205$ and $\hat{\beta}_1 = -0.254$ are strongly significant (t-ratios -5.71 and -6.99, respectively) so that dynamic effects are present for all q. Hence, under Parallel Paths $\hat{\alpha}^1(s) = -0.256 - 0.205s$ and under $q \geq 2$, $\hat{\alpha}^q(s) = -0.256 - 0.254s$.

Redding, Sturm, and Wolf (2011): The authors study how industry location is not uniquely determined by fundamentals. In particular, they estimate the effects of the division of Germany after World War II and the reunification after 1989 on airport passenger shares in Berlin and Frankfurt. They find a change in passenger shares after division and no effect after reunification. For brevity, we discuss the results concerning division. Let γ_1^i be the slope of the linear time trend for airport i during the pre-war period and γ_2^i the slope after the war. For the authors' model, it can be shown that under Parallel Paths the expected gain in passenger shares of Frankfurt relative to Berlin is $\alpha^1(s) = \left(\gamma_2^F - \gamma_2^B\right) \times s$.[10] Hence, under parallel paths the effect of the

[9] As already said, we cannot apply the fully flexible model with region-specific trends. Nevertheless, we can estimate the fully flexible model assuming the same dynamics for all controls and the same dynamics for the two regions treated. Since there is only one year after treatment, we are able to estimate the impact of the treatment in only one period: $s = 1$. The effect of the treatment under $q = 2$ is not significantly different from the effect under $q = 1$ when looking at sickness absenteeism. However, for disability insurance applications the effects under alternative assumptions are statistically different, although both of them are negative and strongly significant. These results and those reported in footnote 11 are available upon request.

[10] Note that the authors estimate a model similar to equation (25) with $\beta_2 = 0$. The pre-war period goes from 1926 until 1938. The post-war period before reunification is 1950 : 1989. It is implicitly

change in Germany's hub is an annual relative increase of 1.25 percentage points in Frankfurt's passenger shares. In a given year during the 40-year post-war period, the relative accumulated gain for Frankfurt is, on average, 25.61 percentage points. For $q \geq 2$, $\alpha^2(s) = \left[\left(\gamma_2^F - \gamma_2^B\right) - \left(\gamma_1^F - \gamma_1^B\right)\right] \times s$. The authors identify the estimate of this expression, 3.07, as the effect of the division of Germany. It implies an average relative accumulated gain for Frankfurt of 62.98 percentage points after 40 years. The difference of the estimated effects under the two assumptions is strongly significant: the t-ratio for the equivalence test between Parallel Paths and Parallel Growths is 6.41.[11]

For the rest of the papers, in Table 2 we compare the original results with results from the flexible model under Parallel-$(1,1)$ and Parallel-$(2,1)$. We also test in the fully flexible model for the equivalence of Parallel-$(1,1)$ and Parallel-$(2,1)$ and for constant treatment effects under Parallel-$(1,s)$ and Parallel-$(2,s)$ with $s \geq 1$.

Aaronson and Mazumder (2011): The authors study the educational gains of rural southern blacks in the U.S. brought about by the construction of nearly $5,000$ schools—known as the Rosenwald Rural Schools Initiative—from 1914 to 1931. The authors find evidence of the effects using several datasets and output measures. They also use urban blacks and rural whites as additional controls in a diff-in-diff-in-diffs framework. We comment on school attendance results using cross-sectional samples drawn from the 1900 to 1930 decennial censuses. We also restrict the discussion to the simplest DID case, i.e. the identification of the treatment effect by comparison to rural blacks in counties without any Rosenwald schools. In some of the counties there was never a Rosenwald school. In some counties, the first Rosenwald school was constructed between 1910 and 1920 while in some other counties it was constructed between 1920 and 1930. As some units change their treatment status, identification of the treatment effect exploits the timing in which this change takes place. It is still possible to estimate a fully flexible

assumed that the periods between 1938 and 1950 did not exist so that the year after 1938 is 1950. The pre-war time trend and the post-war time trend must equal zero in the last pre-treatment period, 1938.

[11] Since there is only one treated unit (Frankfurt's airport) and one control (Berlin's airport), it is not possible to estimate the fully flexible model from equation (23). We instead run the fully flexible model on Frankfurt relative to all other airports but Berlin and on Berlin relative to all other airports except Frankfurt. The effects of division are of the expected sign (positive for Frankfurt and negative for Berlin) but the size is markedly lower. Moreover, while the equivalence test for $q = 2$ and $q = 1$ is strongly rejected for Berlin, it cannot be rejected for Frankfurt. Finally, we find strong evidence of significant changes in the accumulated effects after several periods, although these dynamics do not seem to follow a linear trend.

model in which the dynamics of the controls is driven by decade dummies while the dynamics of the treated are also influenced by the time distance to treatment. We report the results for the model with additional covariates and county fixed effects (i.e. Column 4 in their Table 1). Our estimates of the treatment effect are similar as those reported in the paper, regardless of the Parallel assumption made. We do not reject equivalence between Parallel Paths and Parallel Growths and we find evidence at 5% of dynamic effects under $q = 2$.

Abramitzky, Delavande, and Vasconcelos (2011): The authors investigate the effect of male scarcity due to military mortality during World War I on marriage market outcomes in France. The authors use the class of the bride minus the class of the groom, a dummy variable for the groom marrying a bride of lower class, and a dummy variable for the bride being of low social class as three alternative definitions of a bad marriage outcome for men. For all three definitions, they find that decreases in bad marriages—compared to prewar years—were significantly larger the larger the regional male mortality rates during the war. They conclude that higher regional mortality led to better marriage outcomes. The authors use a postwar dummy variable interacted with mortality rate in each region to identify the effect of mortality rates on marriage outcomes. We apply a fully flexible model for the three alternative outcomes for the full sample of grooms (columns 1 to 3 from their Table 3). Although the estimated effects are not significant, we find some evidence of dynamic effects in all cases and, in one case, we reject that Parallel Paths is equivalent to Parallel Growths.

Currie and Walker (2011): The authors study the effect of the introduction of electronic toll collection (E-ZPass) on infant health via vehicle emissions near highway toll plazas. As their data on infant health are not available, we discuss the effect of E-ZPass on pollution levels. The authors compare the effects on NO_2 levels (for which cars are an important source) and on SO_2 (for which cars are not an important source). More specifically, they look at differences between pollution at one monitor near (< 2 km) to a toll plaza and pollution at all monitors farther than 2 km from a toll plaza, before and after E-ZPass. Hence, there is only one treated agent and many controls. The empirical model is a restricted version of the model in equation (17) with monitor specific

Table 2: *Fully flexible model results and reported results from selected papers*

	Reported	Fully Flexible Model				
	Estimated	$q=1$		$q=2$		Equivalence
Article	Effect	Effect	Dynamics	Effect	Dynamics	Test
Aaronson and Mazumder	0.072***	0.039***	3.337	0.053***	6.488	1.420
(2011)	(0.007)	(0.012)	[0.068]	(0.017)	[0.011]	[0.234]
Abramitzky, Delavande, and	-0.020**	0.036	22.651	0.106	23.428	-0.069
Vasconcelos (2011) - 1	(0.010)	(0.039)	[0.012]	(0.073)	[0.009]	[0.059]
Abramitzky, Delavande, and	-0.010***	0.008	15.983	0.010	16.205	-0.003
Vasconcelos (2011) - 2	(0.004)	(0.016)	[0.100]	(0.030)	[0.094]	[0.435]
Abramitzky, Delavande, and	-0.017***	0.003	26.989	0.031	28.664	-0.028
Vasconcelos (2011) - 3	(0.005)	(0.013)	[0.003]	(0.022)	[0.001]	[0.021]
Currie and Walker (2011) - 1	-0.208***	-0.506***	13.748	-0.386	13.796	-0.121
	(0.028)	(0.198)	[0.132]	(0.395)	[0.131]	[0.300]
Currie and Walker (2011) - 2	-0.090***	-0.582***	33.123	-1.071***	30.811	0.489
	(0.024)	(0.198)	[0.000]	(0.353)	[0.000]	[0.006]
Currie and Walker (2011) - 3	-0.065***	0.029	13.304	0.136	15.950	-0.107
	(0.017)	(0.101)	[0.149]	(0.128)	[0.068]	[0.040]
Currie and Walker (2011) - 4	-0.181***	-0.191*	25.404	-0.380*	27.992	0.189
	(0.023)	(0.108)	[0.003]	(0.204)	[0.001]	[0.051]
Currie and Walker (2011) - 5	0.018	-0.421	20.420	-0.592	14.565	0.171
	(0.038)	(0.374)	[0.016]	(0.736)	[0.104]	[0.357]
Furman and Stern (2011)	0.535***	0.471***	1.605	0.666	1.562	0.262
	(0.142)	(0.123)	[0.071]	(0.417)	[0.083]	[0.610]
Kotchen and Grant (2011) -1	0.009***	0.006**		-0.002		0.008
	(0.003)	(0.003)		(0.005)		[0.003]
Kotchen and Grant (2011) -2	-0.003	-0.006**		-0.013***		0.007
	(0.003)	(0.003)		(0.005)		[0.004]
Moser and Voena (2012)	0.151***	0.075	4.606	0.006	3.995	2.362
	(0.036)	(0.046)	[0.000]	(0.081)	[0.000]	[0.124]

Note: Reported Estimated Effect refers to the results originally published. Effect for $q=1$ and $q=2$ reports the estimated effects at $s=1$ in the fully flexible model under $q=1$ and $q=2$, respectively. Standard errors in parenthesis. Dynamics tests whether effects are constant for S periods. Equivalence tests for the equivalence of Parallel Paths and Parallel Growths. p values 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, Delavande, and Vasconcelos (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 DST period while Kotchen and Grant (2011) 2 reports the effect during non DST period (column d in Tables 4 and 5 in the original paper). Moser and Voena (2012) reports results comparable with those of their column 1 in Table 2.

linear time trends.[12] Note that for this model, the interaction term, γ_P^D, identifies the treatment effect for $q \geq 2$. For $q=1$, identification of the treatment effect is complex. First, given that the time trend is an annual step function, we must define s to represent any day of the calendar year s years after introduction of E-ZPass. Second, estimation of the treatment effect requires the estimation of the average monitor linear trend for the controls, $E\left[\gamma_1^i | D_i = 0\right]$, because $\alpha^1(s) = \gamma_P^D + \left(\gamma_1^D - E\left[\gamma_1^i | D_i = 0\right]\right) s$. The estimate of γ_P^D for NO_2 is negative (-0.108) and significant. We cannot reject the equivalence between $q=1$ and $q=2$ (p-value 0.20). For SO_2 the interaction term is positive (0.053) and not significant. In contrast to the NO_2 case, we reject the equivalence test between $q=1$ and $q=2$. In practice, however, the estimate of the effect under both assumptions is rather

[12]The variable *Post* takes value 1 the first day after the introduction of E-ZPass in the nearest toll plaza. The time trend is annual.

similar: Under $q = 1$, the effect is slowly increasing in s, $\alpha^1(s) = 0.053 + 0.00012s$, and becomes significant only after 14 years. As a robustness check, Currie and Walker (2011) also report for NO_2 5 additional estimates, each obtained using as control a randomly chosen monitor. We estimate the fully flexible model using weeks as the periods in equation (23) and setting the same randomly chosen monitors as controls. In 5 out of the 10 estimations we find negative and significant effects one week after the introduction of E-ZPass at the 5% significance level. In 4 of them we find that these effects change within the first 10 weeks. Finally, in 2 out of 5 cases, we reject at the 5% significance level that the effects under $q = 1$ and $q = 2$ are equivalent.

Furman and Stern (2011): They study the effect of biological resource centers (BRC) on knowledge accumulation. BRCs collect, certify and distribute biological organisms and they have the objective of enhancing cumulative knowledge production. The authors study if access to biomaterials through a BRC amplifies the impact of scientific research. The authors find evidence of the effects using both a DID linear model and a nonlinear negative binomial that the authors consider more appropriate for inference given the highly skewed nature of the dependent variable. Their DID model includes the interaction of treatment (being a BRC-article) with two dummies: the first dummy equals one during the year immediately prior to, the year of, and the year immediately after treatment. The second dummy equals one since the second year after treatment. Hence, they distinguish between a "window period" effect and a "post deposit" effect. Regarding the post deposit effect, they find that treatment causes over a 50% increase to the citation rate. We estimate a fully flexible model in which the controls dynamics are governed by year dummies while treated dynamics are additionally determined by the time distance to treatment (second column in their Table 3). We concentrate on the post deposit effects after two years. We do not reject at 5% the absence of dynamic effects for both $q = 1$ and $q = 2$. Moreover, we also find a positive and significant effect under $q = 1$. Although under $q = 2$ the effect is not significant, the difference between the effects under $q = 1$ and $q = 2$ is not significant.

Kotchen and Grant (2011): They estimate the effect of daylight saving time (DST) on residential electricity consumption using monthly microdata from the majority of households in southern Indiana spanning from January 2004 through December 2006. To identify the effect the authors exploit that, prior to 2006, some of the counties in

southern Indiana did not practice DST while the other counties did. Starting in 2006, all counties were required by state law to practice DST. The authors provide evidence using several model specifications that account for unobservable tenant fixed effects and differ in the way they control for weather and the dynamics of the monthly effects during the DST period. Their main finding is that DST increases electricity demand in the DST period but there are no significant effects outside the DST period. For brevity, we comment on their simplest specification in terms of the dynamic effects (i.e. the effect is the same across months within the DST period) with the most flexible specification in the weather effects (columns d in Tables 4 and 5 in the original paper). Estimating the fully flexible model only adds to the authors' specification an interaction term of treatment with a year dummy for 2005. We find that under $q = 1$ the effect during the DST period is significant (although smaller than the initial estimate) but under $q = 2$, the effect is not significant. Moreover, we find that the difference between the effects for $q = 1$ and $q = 2$ is significant. During the non-DST period, however, our results show a negative and significant effect both under $q = 1$ and $q = 2$. Since there is only one post-treatment period, we cannot run tests for the dynamics.

Moser and Voena (2012): Using data from 1875 until 1939, they study the effect of the Trading with the Enemy Act (TWEA) in 1918 on the number of patents (aggregated into $7,248$ classes across chemicals) by US inventors. The authors use several specifications and estimation strategies. For brevity, we comment on their basic conditional DID specification in which a class is treated if it contains at least one enemy-owned patent that was licensed to a US firm. We report on results comparable with those of their column 1 in Table 2. They find a significant and positive impact of TWEA on the number of patents. Using a flexible model we only find a marginally significant impact for $q = 1$ (p-value 10.4%). We find significant dynamic effects and we do not reject at 5% that $q = 1$ and $q = 2$ are equivalent.

In sum, in 11 out of 13 treatment effect estimates, the original papers report significant effects. By estimating the fully flexible model, we can evaluate to what extent the significance of these estimates relies on more assumptions than Parallel Paths or Parallel Growths. We find that under $q = 1$ in 5 out of the 11 cases the estimated effects remain significant. Under $q = 2$ only 3 estimates remain significant. It is generally true that standard errors for the effect estimates under $q = 1$ and, specially, under $q = 2$, are larger

than under the original model specification. In fact, in many cases we cannot reject that the estimated effects under $q = 1$ and $q = 2$ are significantly different from the original estimates. We interpret these results as anecdotal evidence than, in empirical work, the identification of the treatment effect usually relies on restrictions involving dynamics beyond the stated Parallel assumption.

In 11 out of 13 models we can test whether effects are constant as in the standard model and we reject at 5% the absence of dynamic effects in 6 cases both under $q = 1$ and $q = 2$. Finally, the fully flexible model also allows us to test for the equivalence of results under Parallel Paths and Parallel Growths. Given that the effects are not significant in a number of cases, one could expect that this test would fail to find significant differences between estimates under $q = 1$ and $q = 2$. However, in 5 out of the 13 cases we reject at the 5% significance level that the two assumptions lead to equivalent results.

The fully flexible model can be used as a guide to more parsimonious models. We do not attempt to study how more parsimonious versions of the fully flexible model would help recover some of the results of the original papers. However, we believe that imposing restrictions on the parameters of the fully flexible model may help in improving the accuracy of the estimates. For example, guided by dynamic tests using the results from the fully flexible model, one could impose some restrictions on the dynamics after treatment. Also, guided by equivalence tests, one could impose linear restrictions on the interaction parameters before treatment.

5 Conclusions

In applications in which several pre-treatment periods are available, Parallel Paths is appealing if trends do not differ between treated and controls before treatment. In practice, researchers who find pre-treatment trend differentials often formulate flexible econometric models to accommodate those trend differentials. We start by noting that identification of the treatment effect does not uniquely depend on the Parallel Path assumption, but also on the trend modeling strategy and the definition of the trend variable. As inclusion of trend differentials between treated and controls does change the identifying assumption of the treatment effect, it is important to characterize the set of alternative assumptions which lead to identification of the effects. With this

purpose, we define a family of alternative Parallel assumptions and show, for a number of frequently used empirical specifications, which parameters of the model identify the treatment effect under the alternative Parallel assumptions.

We further present a fully flexible model where treatment effects under any Parallel assumption are shown to be linear combinations of the original parameters. The fully flexible model has two advantages. First, it allows for flexible dynamics and for testing restrictions on these dynamics. Second, it does not impose equivalence between alternative Parallel assumptions though tests for this equivalence are easy to implement.

We view the fully flexible model as a benchmark model to analyze the robustness of estimated effects to alternative Parallel assumptions and dynamic specifications. We revise the results of several recent papers in which the difference-in-differences technique has been applied. We find anecdotal evidence that the identification of the treatment effect usually relies on restrictions involving dynamics beyond the Parallel assumption stated.

For example, it is usually imposed that the full long-term effect of treatment takes place immediately after treatment. Our results suggest that this is an important restriction for the identification of the effects. Moreover, another usual restriction is to add a step or a linear trend differential between treated and controls. The step differential implies the equivalence of all Parallel assumptions while the linear trend differential implies the equivalence of all Parallel assumptions beyond Parallel Paths. Our results suggest that these restrictions play an important role in the identification of the effects.

In view of these considerations, we advocate a change in current practice. When the data structure allows for it, we think that the fully flexible model can be a helpful starting tool to study robustness to alternative Parallel assumptions and dynamics. Equivalence and dynamics tests can be used to validate the standard model or more flexible versions of it. Moreover, the fully flexible model can be used as a guide to more parsimonious models.

References

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

Abadie, A. (2006): "Semiparametric Difference-in-Differences Estimators," *The Review of Economic Studies*, 72, 1–19.

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 *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3A, pp. 1277–1366. Elsevier.

Athey, S., and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74(2), 431–497.

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.

Bonhomme, S., and U. Sauder (2011): "Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling," *The Review of Economics and Statistics*, 93(2), 479–494.

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.

Groen, J., and A. Polivka (2008): "The effect of Hurricane Katrina on the labor market outcomes of evacuees," *The American Economic Review*, 98(2), 43–48.

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.

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

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.

A Appendix

Theorem 1. *Under Parallel-(q, S),*

$$\Delta^{q-1}\alpha\left(s|X\right) = did\left(q,s\right)$$

where $\Delta^{q-1} \equiv (1-L)^{q-1}$ and $L^r\alpha\left(s|x\right) = 0$ for all $r \geq s$.

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

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

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

$$\Delta^q Y^i_{t^*+s} = \Delta_s\Delta^{q-1}Y^i_{t^*+s} - \Delta_{s-1}\Delta^{q-1}Y^i_{t^*+s-1}.$$

Hence, under Parallel-(q, S)

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

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

$$\Delta^q\alpha\left(s|X\right) = did\left(q,s\right) - did\left(q,s-1\right). \tag{26}$$

Since $did\left(q,0\right) = 0$, for $s=1$ and $q \geq 1$ equation (26) immediately implies that

$$\Delta^q\alpha\left(1|X\right) = did\left(q,1\right).$$

Hence, the theorem is proved by induction if we show that if it is true for any $s-1 > 0$, it must be true for s. By equation (26),

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

If the Theorem is true for $s-1$, then $\Delta^{q-1}\alpha\left(s-1|X\right)=did\left(q,s-1\right)$. Hence,

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

□

Theorem 2. *For any* $q\in\{2,...,T_0\}$ *and* $s\in\{1,...,S\}$,

$$\alpha^{q}\left(s|X\right)=\alpha^{q-1}\left(s|X\right)$$

if and only if

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

Proof. We first prove the Theorem for $q=2$. For $s=1$, $\alpha^{2}\left(1|X\right)=\alpha^{1}\left(1|X\right)$ if and only if $did\left(2,1\right)=did\left(1,1\right)$. By definition of the $did\left(q,s\right)$ operator, this condition is equivalent to

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

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

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

For $s>1$,

$$\begin{aligned}\alpha^{2}\left(s|X\right)&=\sum_{j=1}^{s}did\left(2,j\right)\\&=\sum_{j=1}^{s}\left\{E\left[\Delta_{j}\Delta Y_{t^*+j}|X,D=1\right]-E\left[\Delta_{j}\Delta Y_{t^*+j}|X,D=0\right]\right\}\\&=\sum_{j=1}^{s}\left\{E\left[\Delta Y_{t^*+j}-\Delta Y_{t^*}|X,D=1\right]-E\left[\Delta Y_{t^*+j}-\Delta Y_{t^*}|X,D=0\right]\right\}\\&=\sum_{j=1}^{s}\left\{E\left[\Delta Y_{t^*+j}|X,D=1\right]-E\left[\Delta Y_{t^*+j}|X,D=0\right]\right\}\\&-s\left\{E\left[\Delta Y_{t^*}|X,D=1\right]-E\left[\Delta Y_{t^*}|X,D=0\right]\right\}.\end{aligned}$$

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

$$\begin{aligned}\alpha^2\left(s|X\right) &= \alpha^1\left(s|X\right) \\ &\quad - s\left\{E\left[\Delta Y_{t^*}|X,D=1\right] - E\left[\Delta Y_{t^*}|X,D=0\right]\right\}.\end{aligned}$$

Thus, for $s>1$, $\alpha^2\left(s|X\right) = \alpha^1\left(s|X\right)$ if and only if $E\left[\Delta Y_{t^*}|X,D=1\right] - E\left[\Delta Y_{t^*}|X,D=0\right]$ and this ends the proof of the Theorem for $q=2$.

To prove the theorem for $q>2$, we can make use of the following two lemmata.

Lemma 1. *For any* $q \in \{2,...,T_0\}$ *and* $s \in \{1,...,S\}$,

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

where

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

Proof. It follows directly from the linear properties of the Δ_s operator that

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

□

Lemma 2. *For any* $q \in \{2,...,T_0\}$ *and* $s \in \{1,...,S\}$,

$$\alpha_Y^q\left(s|X\right) = \alpha_Y^{q-1}\left(s|X\right)$$

if and only if

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

Proof. We first prove sufficiency. By applying the $(q-1)$th difference we have that if $\alpha_Y^q\left(s|X\right) = \alpha_Y^{q-1}\left(s|X\right)$ then it follows that $\Delta^{q-1}\alpha_Y^q\left(s|X\right) = \Delta^{q-1}\alpha_Y^{q-1}\left(s|X\right)$. By Lemma 1, this implies that $\alpha_{\Delta^{q-1}Y}\left(s|X\right) = \Delta\alpha_{\Delta^{q-2}Y}\left(s|X\right)$.

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

$$\Delta^{q-1}\alpha_Y^q\left(s|X\right) = \Delta^{q-1}\alpha_Y^{q-1}\left(s|X\right)$$

for all s. This implies that both $\alpha_Y^q\left(s|X\right)$ and $\alpha_Y^{q-1}\left(s|X\right)$ satisfy the same initial conditions and have the same differential equations. Hence they must be the same.

□

To prove the Theorem for $q > 2$, define $z_t = \Delta^{q-2} Y_t$. By Lemma 2, the Theorem is proved if $\alpha_{\Delta z}(s|X) = \Delta\alpha_z(s|X)$ is true if and only if $E[\Delta z_{t^*}|D=1] = E[\Delta z_{t^*}|D=0]$. By Lemma 1, we then need to prove that $\Delta\alpha_z^2(s|X) = \Delta\alpha_z(s|X)$ for all s is true if and only if $E[\Delta z_{t^*}|D=1] = E[\Delta z_{t^*}|D=0]$. Given that $\Delta\alpha_z^2(s|X) = \Delta\alpha_z(s|X)$ for all s is true if and only if $\alpha_z^2(s|X) = \alpha_z(s|X)$ for all s, then the Theorem is proved if we prove that it is true for $q = 2$, which we did.

□

Theorem 3. *Consider a general additive model with group-specific, fully-flexible pre- and post-treatment trends:*

$$E[Y_t|D] = \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_\tau + \gamma^D D + \sum_{\tau=t_2}^{T} \gamma_\tau^D \times I_\tau \times D.$$

Under Parallel-(q, S):

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

Proof. Given that

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

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

□