UNIVERSIDAD CARLOS III DE MADRID

working papers

dqd: A command for treatment effect estimation under alternative assumptions

Ricardo Mora & Iliana Reggio[1]

Abstract

Conventional difference-in-differences (DID) methods that are used to estimate the effect of a treatment rely on important identifying assumptions. Identification of the treatment effect in a DID framework requires some assumption relating trends for controls and treated in absence of treatment, the most common being the assumption of Parallel Paths. When several pre-treatment periods are available, Mora and Reggio (2012) show that treatment effect identification does not uniquely depend on the Parallel Path assumption, but also on the trend modeling strategy. They further define a family of alternative Parallel assumptions and propose a more flexible model which can be a helpful starting tool to study robustness to alternative Parallel assumptions and trend dynamics. In this paper we introduce a Stata command that implements the fully flexible model presented in Mora and Reggio (2012), producing tests for the equivalence of alternative parallel assumptions and for the dynamic effects of the treatment. The standard DID in model with or without polynomial trends can also be obtained.

Keywords: Difference-in-differences, treatment effect, identification, fully flexible model, dqd.

1 Introduction

Difference-in-differences (DID) methods are widely used to evaluate the impact of policy interventions or other specific treatments on different outcomes of interest. DID estimators require an assumption about how trends for controls and for treated in absence of treatment are related. The most common assumption for that purpose is the Parallel Paths assumption. Parallel Paths requires that in absence of treatment the average change in the outcome variable for the treated equals 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 to be time-invariant.

When several pre-treatment periods are available, Parallel Paths is appealing if trends do not differ between treated and controls before treatment. Many researchers use absence of pre-treatment trend differentials between controls and treated as an argument in favor of the Parallel Paths assumption. In the presence of pre-treatment trend differentials, it is customary to adjust the econometric specification to try to accommodate those differences. Mora and Reggio (2012) show that the inclusion of trend polynomials is not innocuous. Different trend modeling strategies in general imply different parallel assumptions: i.e., assumptions regarding how trends for controls and treated in the absence of treatment are related. The fact that identification of the treatment effect does not uniquely depend on the Parallel Path assumption, but also on the trend modeling strategy is an overlooked issue in studies estimating treatment effects using DID techniques.

More specifically, a very common procedure is to introduce linear trends to account for trend differences between treated and controls. Researchers usually associate the parameter for the interaction of a post-treatment dummy and the treated indicator with the treatment effect. This practice, which is not longer consistent with Parallel Paths, is correct if one assumes that the average acceleration for the treated under no treatment would have been equal to the observed average acceleration for the controls. Mora and Reggio (2012) refer to this assumption as Parallel Growths, or Parallel-2. More generally, they propose a family of alternative Parallel-q assumptions where q is, at most, the number of pre-treatment periods. They further identify the treatment effect under each Parallel-q assumption for a fully flexible dynamic specification. A critical result in Mora and Reggio (2012) is that the treatment effect s periods after treatment

under any given Parallel-q assumption can be expressed as the solution of an equation in differences with the parameters of the fully flexible model.

Implementation of Parallel Paths is very simple with any econometrics package as it only requires standard least squares estimation of a very simple model. As the treatment effect estimate is identified as the parameter of one of the regressors, testing its significance is also straightforward running OLS. In Stata, the user just needs to set the correct variable specification and employ the command **regress**. In contrast, estimation of treatment effects with the fully flexible model under alternative Parallel assumptions requires two steps. In the first step, standard least squares estimation of the fully flexible model is conducted. In the second step, the solution of the equation in differences identifies the estimates. Computation of the standard errors of the treatment effect estimates must take into account that the solution of the equation in differences is a linear combination of the parameters of the fully flexible model.

In this paper we show how the command `dqd`, which is available from SSC, implements this two-step procedure in Stata. The `dqd` command first estimates the fully flexible model and then it computes by default treatment effects under all Parallel assumptions from Parallel-1 (i.e., Parallel Paths) to Parallel-Q, where Q is set by the user. Treatment effects are evaluated for a period set by the user. In addition, `dqd` implements tests for: (a) the equivalence of all Parallel-q assumptions between Parallel-1 and Parallel-Q; (b) for each Parallel-q assumption, the equivalence of Parallel-q and Parallel-$(q-1)$; and (c) for each Parallel-q assumption, the absence of dynamics in treatment effects. In addition to the fully flexible model, `dqd` offers the option to report the DID standard model with flexible common dynamics and extensions that include a linear and a quadratic trend.

The rest of the paper is structured as follows. We first define Parallel Paths and Parallel Growths, and present the family of alternative Parallel assumptions in Section 2. Next, we present the fully flexible model in Section 3 and state identification conditions of the treatment effect under alternative Parallel-q assumptions. In Section 4, we describe the syntax of `dqd` and illustrate the use of the command by means of several simulated examples. Section 6 concludes.

2 Alternative Parallel assumptions

In this section, we give an overview of the alternative parallel assumptions in DID applications. A more detailed explanation is found in Mora and Reggio (2012).

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, $t^* + 1$. 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 (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 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 $X = \{X_{t_1}', ..., X_{t_T}'\}'$ where X_t is a vector of k additional controls.

The average treatment effect $s \leq S$ periods after treatment on the treated given X is

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

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

In order to estimate the average counterfactual $E\left[Y_{t^*+s}^0 \,|X, D = 1\right]$, one needs an assumption on how the trend behavior of the treated if untreated compares to the observed trend behavior of the untreated. The DID estimator, in general, relies on the 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 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*

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

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

The counterfactual under Parallel Paths is built by adding to the last pre-treatment outcome level of the treated the average outcome change experienced by the controls. Using this counterfactual, the treatment effect in period s, $\alpha\left(s|X\right)$, is identified as the difference-in-differences operator s-periods ahead:

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

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

2.2   The Parallel Growths assumption

An alternative assumption to Parallel Paths is that the average acceleration for the treated under no treatment would have been equal to the observed average acceleration for the controls. We call this assumption Parallel Growths.

Assumption 2. *Parallel Growths*

$$E\left[\Delta_s \Delta Y^0_{t^*+s}\,|X, D=1\right] = E\left[\Delta_s \Delta Y^0_{t^*+s}\,|X, D=0\right],\ s\in\{1,...,S\}. \tag{4}$$

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

$$\begin{aligned} E\left[Y^0_{t^*+1}\,|X, D=1\right] = {}& E\left[Y_{t^*}\,|X, D=1\right]+ \\ & E\left[\Delta Y_{t^*}\,|X, D=1\right] + E\left[\Delta^2 Y_{t^*+1}\,|X, D=0\right] \end{aligned} \tag{5}$$

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

$$\alpha(1|X) = E\left[\Delta^2 Y_{t^*+1} | X, D = 1\right] - E\left[\Delta^2 Y_{t^*+1} | X, D = 0\right]. \tag{6}$$

Without additional restrictions on pre-treatment dynamics, Parallel Paths does not imply Parallel Growths. They are equivalent if and only if the DID operator equals the difference-in-double-differences operator:

$$\begin{aligned} E\left[\Delta^2 Y_{t^*+1} | X, D = 1\right] - E\left[\Delta^2 Y_{t^*+1} | X, D = 0\right] = \\ E\left[\Delta Y_{t^*+1} | X, D = 1\right] - E\left[\Delta Y_{t^*+1} | X, D = 0\right] \end{aligned} \tag{7}$$

or, equivalently,

$$E\left[\Delta Y_{t^*} | X, D = 1\right] = E\left[\Delta Y_{t^*} | X, D = 0\right] \tag{8}$$

In the presence of pre-treatment group-specific trends, the identification strategy of the treatment effect is different under Parallel Paths and under Parallel Growths.

For the case $s \geq 2$, $\alpha(s|X) = \alpha(s-1|X) + E\left[\Delta Y^1_{t^*+s} - \Delta Y^0_{t^*+s} | X, D = 1\right]$. Given that $E\left[\Delta Y^1_{t^*} - \Delta Y^0_{t^*} | X, D = 1\right] = 0$, the term $E\left[\Delta Y^1_{t^*+s} - \Delta Y^0_{t^*+s} | X, D = 1\right]$ is equivalent to $E\left[\Delta_s \Delta Y^1_{t^*+s} - \Delta_s \Delta Y^0_{t^*+s} | X, D = 1\right]$. Under Parallel Growths, a difference-in-differences operator identifies the change rather than the level of the treatment effect:

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

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, a family of alternative non-nested assumptions can be defined:

Assumption 3. *Parallel-q*

For a given positive integer $q \leq T_0$, and for any $s = 1, ..., S$,

$$E\left[\Delta_s\Delta^{q-1}Y^0_{t^*+s} | X, D=1\right] = E\left[\Delta_s\Delta^{q-1}Y^0_{t^*+s} | X, D=0\right]. \tag{10}$$

Parallel-1 is Parallel Paths, while Parallel-2 is Parallel Growths. Let define $did\left(q,s\right)$ as the difference-in-q-differences operator s periods ahead,

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

Mora and Reggio (2012) show that under Parallel-q,

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

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

This result can be used to obtain $\alpha\left(s|X\right)$ for any value of s under Parallel-q recursively.

If trends do not differ between treated and controls before treatment, Parallel Paths is equivalent to Parallel Growths. More generally, define the operator $\alpha^q\left(s|X\right)$ as the mapping on $did\left(q,s\right)$ that identifies the effect of treatment under Parallel-q. Mora and Reggio, 2012 show that 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

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

This result sets pre-treatment trend conditions under which assumptions Parallel-q and Parallel-$(q-1)$ are equivalent.

3 Alternative modeling strategies

The conventional DID estimator is 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 treatment indicator D, a dummy variable for the post-treatment period, $Post_t$, and an interaction term, $Post_t \times D$. In this set up, the treatment effect is identified by the parameter associated with the interaction term.

The Standard model

On applications in which several pre-treatment periods are available, the Standard model allows for time fixed effects δ_t common to treated and controls (see, for example, Bertrand, Duflo, and Mullainathan 2004):

$$E\left[Y_t \,|X, D\right] = \beta X_t + \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \gamma_P^D Post_t \times D \qquad (12)$$

where $I_{\tau,t}$ is a dummy for period τ. The specification in equation (12) identifies the treatment effect exploiting two restrictions. The first restriction is that pre-treatment dynamics—captured by time fixed effects—are identical between the controls and the treated. This condition implies that all Parallel assumptions are equivalent. The second restriction is that there is a permanent shift in output of size γ_P^D in the first period after treatment. Therefore, the long-term effect of treatment is already present at $t^* + 1$.

The Linear Trend model

One way to extend the standard model in equation (12) to accommodate group-specific trends is by including an interaction between D and a linear time trend:

$$E\left[Y_t \,|X, D\right] = \beta X_t + \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \gamma_P^D Post_t \times D + \gamma_1^D t \times D. \qquad (13)$$

In equation (13) the terms $\gamma_1^D t$ captures differences in group dynamics before and after

treatment. Including this trend, the Parallel assumption under which the interaction term γ_P^D identifies the treatment effect changes: identifying the treatment effect with γ_P^D implies a departure from Parallel Paths but is still consistent with any other Parallel assumption.

The Quadratic Trend model

A more flexible approach adds a quadratic polynomial instead of a linear trend,

$$E\left[Y_t \,|X, D\right] = \beta X_t + \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \gamma_P^D Post_t \times D + \gamma_1^D t \times D + \gamma_2^D t^2 \times D, \ (14)$$

where the terms $\gamma^D + \gamma_1^D t + \gamma_2^D t^2$ capture differences between controls and treated before and after treatment. In this case, identifying the treatment effect with γ_P^D implies a departure from Parallel Paths and Parallel Growths but is still consistent with any other Parallel assumption.

The Fully Flexible model

Consider the Fully Flexible model with group-specific, fully-flexible pre- and post-treatment dynamics:

$$E\left[Y_t \,|X, D\right] = \beta X_t + \delta + \sum_{\tau=t_2}^{T} \delta_\tau I_{\tau,t} + \gamma^D D + \sum_{\tau=t_2}^{T} \gamma_\tau^D \times I_{\tau,t} \times D. \qquad (15)$$

Equation (15) imposes no parametric assumptions on the dynamics under no treatment neither for the treated nor the controls. Still, the treatment effect is identified under any Parallel-q assumption:

In the Fully Flexible model, under Parallel-q (Mora and Reggio, 2012):

$$\Delta^{q-1}\alpha\left(s\right) = \Delta_s \Delta^{q-1}\gamma_{t^*+s}^D \qquad (16)$$

Equation (16) implies that the identification strategy of the treatment effect will generally differ under alternative Parallel assumptions. Only when $\gamma_\tau^D = 0$ for all $\tau \leq t^*$, i.e., only when pre-treatment trends are equal between treated and controls, then $\alpha(s) = \gamma_{t^*+s}^D$ under any Parallel-q assumption. Therefore, 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 assumptions.

The inclusion in equation (15) of fully flexible pre-treatment trend differentials between treated and controls allows for the comparison of any two consecutive Parallel-q assumptions. Using condition 11 we see that 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 and Parallel-$(q-1)$. In the case of Parallel Paths and Parallel Growths, the test would be $H_0 : \gamma_{t^*}^D = \gamma_{t^*-1}^D$ vs. $H_1 : \gamma_{t^*}^D \neq \gamma_{t^*-1}^D$.

Finally, the inclusion of fully flexible post-treatment trend differentials also allows us to implement tests on the dynamics of the treatment effect under any Parallel-q assumption. For example, under Parallel-1 testing the null $H_0 : \gamma_{t^*+s}^D = \gamma_{t^*+s+1}^D$ with $s = 1, ..., S-1$ is a test for the effect to be constant in the post-treatment period.

4   Stata Implementation

In this section we describe the Stata command `dqd` that performs diff-in-diffs estimates under alternative Parallel-q assumptions. There are some minimum data requirements so that the command can be executed. First, the data must contain at least one observation per group and period combinations. Second, there must be at least one period before treatment starts and one period after treatment ends. Clearly, the data set must contain a dependent variable on which the effects are to be estimated. It must also have a variable that identifies the period from which each observation is drawn, and a time-invariant treatment variable that signals treatment. Additional controls can be added.

Command `dqd` first estimates an auxiliary regression using Stata command `regress` and then it computes—in Mata—the treatment effects and test statistics as linear combinations of the estimates of the auxiliary regression. `dqd` is by-able, admits weights, and is an r-class ado. In addition to the treatment effects and their standard errors, `dqd` also saves the vector of coefficient estimates of the auxiliary regression and their variance

covariance matrix.

4.1 Syntax

`dqd` *depvar* `[`*indepvars*`]` `[`*if*`]` `[`*in*`]` `[`*weight*`]` `,` <u>`treat`</u>`ed(`*treatvar*`)` <u>`time`</u>`(`*timevar*`)` `[`*options*`]`

depvar is the variable on which the effects are to be estimated and it must be numeric. *indepvars* is an optional variable list for the inclusion of controls X_{it} in the model.

<u>`treat`</u>`ed(`*treatvar*`)` specifies the variable that signals treatment. *treatvar* is time invariant and must take value 0 for observations from the control group and value 1 for observations from the treated group.

<u>`time`</u>`(`*timevar*`)` sets numeric variable *timevar* as the time variable. *timevar* are the discrete periods from which the observations are taken and is assumed to be an integer variable (i.e., byte, int, or long). Two consecutive periods differ by 1.

4.2 Options

The command `dqd` offers seven model options (`begin`, `end`, `q`, `ff`, `standard`, `linear`, and `quadratic`), two standard errors options (`robust` and `cluster`), and two reporting options (`detail` and `level`).

4.2.1 Model options

`begin(#)` and `end(#)` set the first and the last post-treatment periods on which we want to evaluate the effects, $t^* + 1$ and $t^* + S$ respectively. They only take integers as arguments. At $timevar = t^* + 1$, $s = 1$. Values $t^* + 1$ and $t^* + S$ must be such that

$$\min\left(timevar\right) < t^* + 1 \leq t^* + S \leq \max\left(timevar\right).$$

By default, `begin()` and `end()` are set equal to $\max\left(timevar\right)$, i.e., the last period in *timevar* is assumed to be the only post-treatment period. If one of the two options is not specified, then the missing option is set equal to the one specified.

`q(#)` sets the highest Parallel-q assumption, q^{max}, to be used in the estimations. There-fore, the integer chosen must lie between 1 and the number of pre-treatment periods, Q. For example, with only two pre-treatment periods, q^{max} may be, at most, equal to 2. If we want to compute estimates under both Parallel Paths and Parallel Growths, we must set $q^{max} = 2$. Otherwise, if we set $q^{max} = 1$ and only the estimates under Parallel Paths will be obtained. If option `q()` is not specified or q^{max} is set equal to a value larger than Q, then q^{max} is set equal to Q.

`ff`, `standard`, `linear`, and `quadratic` all refer to the model to be used. Hence, only one option may be specified. Option `ff` estimates the fully flexible model from equation (15). This is the default if no model is specified. When the fully flexible model is chosen, `dqd` displays all estimates $\hat{\alpha}(q, s)$, $q = 1, ..., q^{max}$ and $s = 1, ..., S$, and their standard errors. In addition, three types of tests are conducted. First, the test for the equivalence of all Parallel-q assumptions between 1 and q^{max}. Second, for any $q \in \{2, ..., q^{max}\}$, the equivalence of Parallel-q and Parallel-$(q-1)$. Finally, if $S > 1$, for any $q \in \{1, ..., q^{max}\}$, the test of absence of dynamics in treatment effects, i.e., $H_0 : \alpha(q, s) = \alpha(q, s - 1)$, for $s = 2, ..., S$.

`standard` estimates the Standard model with common flexible dynamics from equation (12). `linear` estimates the Linear Trend model from equation (13). `quadratic` estimates the Quadratic Trend model from equation (14). The output displayed in the three options includes the estimate and standard error of the diff-in-diff estimator as well as tests for the absence of dynamics in treatment effects between $t^* + 1$ and $t^* + S$ under the respective model. In addition, it includes the test of the equivalence of all Parallel-q assumptions implicitly assumed in each model. Hence, it reports the test of the equivalence of all parallel assumptions between Parallel-1 and Parallel-Q with `standard`, between Parallel-2 and Parallel-Q with `linear`, and between Parallel-3 and Parallel-Q with `quadratic`.

4.2.2   SE/Robust and Reporting options

`robust`, the default, uses the Huber/White/sandwich estimator whereas option `cluster(`*varname*`)` uses the clustered sandwich estimator.

`detail` is only relevant when the fully flexible model is estimated. When this option is chosen, `dqd` additionally displays t-ratios, p-values, and confidence intervals of all effect

estimates.

`level(#)` sets the confidence level, as a percentage, for confidence intervals. The default is `level(95)` and it is only relevant with option `detail`.

4.3 Some preliminary examples

Consider the simulated data dqd_examples.dta that is also available from SSC. Variable t records the observation period and ranges from 1 to 5. Variable *output* is the output on which we want to estimate the effects and D is the treatment indicator. The data, with 250 observations in each of 5 periods, were generated from a particular case of standard model:

$$y_{it} = \delta + x + \sum_{\tau=2}^{5} \delta_\tau I_{\tau,t} + \gamma^D D_i + \gamma_P^D Post_t \times D_i + u_{it} \tag{17}$$

where $x \sim N(0, .25)$, $I_{\tau,t} = I(t = \tau)$, $Post_t = I(t \geq 4)$, $u \sim N(0, 1)$, $\Pr(D_i = 1) = 0.5$, $D \perp x$, $\delta = \gamma^D = \gamma_P^D = 1$, and $\delta_t = t$ for all $t = 2, ..., 5$. In this model, conditional on exogenous x, controls and treated outputs are subject to a common linear trend. The treated differ on average from controls before by a constant γ^D and additionally after treatment by $\gamma^D + \gamma_P^D$. Under all Parallel-q, $q = 1, 2, 3$, the treatment effect is identified as γ_P^D.

Example 1:

Assume that you want to use only observations from periods 3 and 4. The following example estimates the treatment effect at period 4 with the standard model without additional controls under the Parallel Paths assumption:

. dqd *output* if (*t*==3 |*t*==4), treated(*D*) time(*t*) standard

```
Unconditional Standard Model
Output: output                                    Number of obs =     500
Sample Period: 3:4                    H0: Common Pre-dynamics =     n/a
Treatment Period: 4:4
```

	All s	H0: s=s-1
All q	1.083984 (0.2005)	n/a

```
Robust Standard Errors in parenthesis
```

The heading of the output display provides basic information on the model, the dependent variable, and the sample. Because there is only one pre-treatment period, no test for common pre-treatment dynamics is applicable. The Standard model assumes the equivalence of all Parallel-q assumptions and that the effect has no dynamics. The estimated effect, 1.08, is thus presented under "All q" and "All s" categories. Because there is only one post-treatment period, the test for dynamics of the treatment effect is also not applicable.

Example 2:

Suppose that we also want to use the data from period 2. In the Fully Flexible model with two pre-treatment periods, there are two alternative assumptions, Parallel-1 and Parallel-2, that lead to two alternative estimates:

```
.  dqd output if (t>1 & t<5), treated(D) time(t)

       Unconditional Fully Flexible Model
       Output: output                              Number of obs =     750
       Sample Period: 2:4                 H0: Common Pre-dynamics =   .6931
       Treatment Period: 4:4                          p-value =   .4051
       ─────────────────────────────────────────────────────────────────
                      │      s=1               H0: q=q-1    H0: s=s-1
       ────────────────────────────────────────────────────────────────
            q=1       │  1.083984                                  n/a
                      │  (0.2005)
            q=2       │  1.250342             -.1663581            n/a
                      │  (0.3453)             [0.4051]
       ─────────────────────────────────────────────────────────────────

       Robust Standard Errors in parenthesis
       p-values in brackets
```

By default, the beginning and end of the treatment period is the last period in the estimating sample. Since there are two pre-treatment periods, the test for common pre-treatment dynamics is displayed. Each alternative estimate of the treatment effect is displayed under the corresponding q line. Line q=1 corresponds to assuming Parallel-1 and 1.08 is the estimate under Parallel-1. This is by construction the same estimate (and the same standard error) than the estimate of the treatment effect with the Standard model using only periods 3 and 4. The estimate under Parallel-2, which is displayed in line q=1, is slightly larger, 1.25. The display also includes the test of the equivalence between Parallel-2 and Parallel-1 (at line q=2 and column H0:q=q-1). In the case with two pre-treatment periods, this test is equivalent to the test on common pre-dynamics, so that the p-value is the same.[1] The conclusion of the test is that both Parallel assumptions are equivalent. In other words, the controls and the treated have common pre-treatment dynamics.

Example 3:

With only three periods, the treatment effect estimate under Parallel-2 is equivalent to the estimate of the treatment effect with the standard model and linear deterministic trends:

---

[1] The statistic of the equivalence of the Parallel assumptions is the estimated effect on the last pre-treatment period under Parallel-2. The test statistic on the common dynamics is the Wald test of the joint significance of all pre-treatment γ_t^D in equation (15).

```
.  dqd output if (t>1 & t<5), treated(D) time(t) linear

        Unconditional Linear Trend Model
        Output: output                          Number of obs =    750
        Sample Period: 2:4            H0: Common Pre-dynamics =    n/a
        Treatment Period: 4:4
        ─────────────────────────────────────────────────────────────
                     │    All s                            H0: s=s-1
        ─────────────────────────────────────────────────────────────
           All q     │  1.250342                                 n/a
                     │ (0.3453)
        ─────────────────────────────────────────────────────────────

        Robust Standard Errors in parenthesis
```

Given that the Linear model implies Parallel-2 and beyond, the test for common pre-dynamics requires at least three pre-treatment periods. In this example, there are only two pre-treatment periods so that the test is not applicable.

Example 4:

Consider the full sample, i.e. three pre-treatment periods ($t = 1, 2, 3$) and two post-treatment periods ($t = 4, 5$). With more than one post-treatment period, options `begin` and `end` should be used to identify the interval in which to obtain the effects estimates. Under the Fully Flexible Model, we can obtain three alternative estimates for the effect in period 4 and three alternative estimates for the effect in period 5:

```
.  dqd output, treated(D) time(t) begin(4) end(5)

        Unconditional Fully Flexible Model
        Output: output                          Number of obs =    1250
        Sample Period: 1:5            H0: Common Pre-dynamics =   1.359
        Treatment Period: 4:5                          p-value =   .507
        ─────────────────────────────────────────────────────────────
                  │    s=1       s=2           H0: q=q-1   H0: s=s-1
        ─────────────────────────────────────────────────────────────
           q=1    │ 1.083984  1.002169                      .1756106
                  │ (0.2005)  (0.1926)                      [0.6752]
           q=2    │ 1.250342  1.334885         -.1663581     .0915826
                  │ (0.3453)  (0.5245)         [0.4051]      [0.7622]
           q=3    │ 1.365318  1.679811         -.1149753     .1244636
                  │ (0.6308)  (1.4784)         [0.7388]      [0.7242]
        ─────────────────────────────────────────────────────────────

        Robust Standard Errors in parenthesis
        p-values in brackets
```

The three alternative estimates for the effect at period 4 are shown under the heading `s=1`

while those for period 5 are shown under the heading s=1.[2] With three pre-treatment periods, the test for common pre-treatment dynamics is a test for the joint equivalence of Parallel-1, Parallel-2, and Parallel-3. Two tests for the equivalence of Parallel assumptions are additionally shown under column H0:q=q-1: (a) the equivalence of Parallel-1 and Parallel-2 in line q=2; and (b) the equivalence of Parallel-2 and Parallel-3 in line q=3. Since there are more than one post-treatment period, we can conduct, for any given Parallel assumption, a test on the equality of the effect on all post-treatment periods. These tests are shown in column H0:s=s-1.

Example 5:

Additional controls can be added to improve the accuracy of the estimates:

. dqd *output* $x1$, treated(D) time(t) begin(4) end(5)

```
Conditional Fully Flexible Model
Output: output                                 Number of obs =   1250
Sample Period: 1:5               H0: Common Pre-dynamics =   2.721
Treatment Period: 4:5                             p-value =   .2566
```

	s=1	s=2	H0: q=q-1	H0: s=s-1
q=1	1.123735	1.081014		.0608667
	(0.1820)	(0.1723)		[0.8051]
q=2	1.325644	1.484831	-.2019087	.400746
	(0.3148)	(0.4776)	[0.2678]	[0.5267]
q=3	1.445622	1.844765	-.1199781	.2430778
	(0.5743)	(1.3449)	[0.7015]	[0.6220]

```
Robust Standard Errors in parenthesis
p-values in brackets
```

Example 6:

The tests in examples 4 and 5 suggest that the Standard Model is appropriate for this simulated data because we cannot reject common pre-treatment dynamics and equal dynamic effects.

---

[2]If there are more than three post-treatment periods, the default display only reports the effects for s=1, s=2, and s=3. To display all the effects, option detail should be used.

. dqd *output x*1, treated(D) time(t) begin(4) end(5) standard

```
Conditional Standard Model
Output: output                              Number of obs =   1250
Sample Period: 1:5                 HO: Common Pre-dynamics =  2.721
Treatment Period: 4:5                              p-value =  .2566
```

	All s	HO: s=s-1
All q	.9404042	.0611879
	(0.1131)	[0.8046]

```
Robust Standard Errors in parenthesis
p-values in brackets
```

Using the Standard Model with the full sample and additional controls, we obtain a reduction in the standard error of the estimated effect. In contrast with Example 1, it is possible to test both pre-treatment dynamics and equal dynamic effects. The common pre-dynamics test is the test of the joint equivalence of Parallel-1, Parallel-2, and Parallel-3 and is the same as the test in Example 5. The test of equal dynamic effects (in column HO:s=s-1) is a Wald test of $H_0 : \gamma_4^D = \gamma_5^D$ in a Standard Model where the treatment effects can differ by period:

$$y_{it} = \delta + x + \sum_{\tau=2}^{5} \delta_\tau I_{\tau,t} + \gamma^D D_i + \gamma_4^D I_4 \times D_i + \gamma_5^D I_5 \times D_i + u_{it} \qquad (18)$$

5   A Monte Carlo Example

In the examples of the previous section, obtained with simulated data from a Standard Model, we have seen that the most precise estimates are obtained when using all Parallel assumptions simultaneously (i.e. using the Standard Model). Using the Fully Flexible Model does not lead to inconsistent estimates of the effect, but as illustrated by the examples, results in a loss of accuracy.

We explore in this section by means of a Monte Carlo simulation with 10000 replications the relative performance of the four alternative models for a particular specification of equation (15). The data, with seven periods, are generated from:

$$y_{it} = \delta + \sum_{\tau=2}^{7} \delta_\tau I_{\tau,t} + \gamma^D D_i + \sum_{\tau=2}^{7} \gamma_\tau^D \times I_{\tau,t} \times D_i + u_{it} \qquad (19)$$

18

Table 1: *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 10000 replications. Results show the proportion of rejections of the null at 5% significance level. T is fixed at 7.

where $\Pr(D_i = 1) = 0.5$, $u \sim N(0, 0.25)$, $\delta = 0$. The sequence δ_τ is the Fibonacci sequence $\{1, 1, 2, 3, 5, 8\}$, $\gamma^D = 3$, and $\gamma_t^D = \{4, 4, 5, 6, 8, 9\}$. Treatment starts at period 6 and ends at period 7, the last period of the sample. Parameters γ_t^D are such that, both under Parallel-2 and under Parallel-3, $\alpha(1) = \alpha(2) = 1$. Because both assumptions are equivalent, under Parallel-2 there is null effect in the last pre-treatment period. Under any other Parallel assumption, the treatment effects would be different. Assuming $\alpha(1) = \alpha(2) = 1$, we now study the relative performance of the alternative models using 4 different sample sizes ($N \times T = 250, 750, 2000, 5000$).

Table 1 shows the proportion of rejections of the null $H_0 : \alpha(1) = 1$ using 5% significance level and the t-statistic (which can be computed from the default display and, for the Fully Flexible model is displayed using the `detail` option). 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. All other models identify the treatment effects using alternative assumptions that are incorrect and over-reject the null for all sample sizes.

In Table 2 we further present the proportion of rejections of $H_0 : \alpha(1) = \alpha(2)$ in each model (these tests are displayed in `dqd` in column `H0: s=s-1`). Again we find the expected rejection rate only for the fully flexible model both under Parallel-2 and under Parallel-3.

Finally, in Table 3 we present the proportion of rejections of several Equivalence tests at 5% significance levels. The Standard Model imposes that all Parallel assumptions

Table 2: *Monte Carlo: Absence of Dynamics in Treatment Effects*

	$N \times T = 250$	$N \times T = 750$	$N \times T = 2000$	$N \times T = 5000$
Standard Model	0.974	1.000	1.000	1.000
Linear Model	0.199	0.501	0.893	0.999
Quadratic Model	0.998	1.000	1.000	1.000
Fully flexible, $q = 1$	0.973	1.000	1.000	1.000
Fully flexible, $q = 2$	0.052	0.051	0.051	0.048
Fully flexible, $q = 3$	0.056	0.049	0.051	0.048
Fully flexible, $q = 4$	0.147	0.321	0.700	0.975
Fully flexible, $q = 5$	0.828	0.999	1.000	1.000

Note: Monte Carlo results using 10000 replications. Results show the proportion of rejections of the null at 5% significance level. T is fixed at 7.

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 10000 replications. Results show the proportion of rejections of the null at 5% significance level. T is fixed at 7.

(Parallel-1 to Parallel-5 in this case) are equivalent. The Linear and Quadratic models assume the equivalence of Parallel-2 to Parallel-5 and Parallel-3 to Parallel-5, respectively (these tests are computed with command `dqd` when the relevant model is chosen as option and displayed as the `H0:  Common Pre-dynamics` test). For even small samples, the procedure leads to rejection of the equivalence tests required for each models and suggests to be a powerful tool to guide the choice of identifying Parallel assumptions. The Equivalence test "Parallel-1 vs. Parallel-2"—displayed in column `H0:  q=q-1`— corresponds to the test on pre-treatment common trends frequently used in the literature to justify Parallel-Paths. The test is overwhelmingly rejected, suggesting that, when faced with this data generation process, the usual practice would lead to the rejection of the Standard model.

Which model should be preferred? Including a deterministic trend implies the Linear

Model and it is not appropriate. Assuming Parallel-2 or Parallel-3 is equivalent (and the test is rejected, as expected, at around 5% of the cases). As the "Parallel-2 vs Parallel-3" test is a test on Parallel Growths in the last pre-treatment period, following the usual argument we could use the result of the test to justify the assumption of Parallel-2. Following this argument, no other Parallel assumption—with the obvious exception of Parallel-3—should be assumed.

6   Conclusions

Identification of treatment effects using cross–sections when the data set contains more than one pre-treatment period depends on specific assumptions about pre-treatment dynamics, and how they inform the counterfactual for the treated in the absence of treatment. Mora and Reggio (2012) discuss the most popular models used in the empirical literature and present the Fully Flexible model. For all these models, they derive the identification conditions of the treatment effect in terms of alternative assumptions.

In this article we present a new command dqd that performs difference-in-differences estimations under alternative assumptions as proposed by Mora and Reggio (2012). We illustrate how to use the command dqd by means of several simulated examples. We additionally perform a Monte Carlo simulation to asses the relative performance of four alternative models when the data generation process does not fit the models most frequently used in the literature.

We argue that command dqd is a helpful tool to analyze the robustness of estimated effects to alternative identifying assumptions and dynamic specifications. Moreover, equivalence and dynamics tests can be used to validate alternative models.

7   Saved results

Command dqd saves the following in r():

- r(N): the number of observations used in the estimation of the auxiliary model

- r(alpha): a $q^{max}xS$ matrix where element $alpha\,(q,s)$ corresponds to $\hat{\alpha}\,(q,s)$.

- r(std_alp): a $q^{max}xS$ matrix where element $std\_alp\,(q,s)$ corresponds to $\hat{std}\,(\hat{\alpha}\,(q,s))$.

- r(beta): vector of estimates in auxiliary regression. The first elements of r(beta) are the estimates of the coefficients for the interactions between the treatment variable and the time dummies in the fully flexible model. For the standard, linear, and quadratic model, the first elements are the estimates of the coefficients of the interactions between the treatment variable and the corresponding polynomial elements (i.e., constant, linear, and quadratic terms). In all models, the estimate of the coefficient of the treatment dummy is next. The estimates for the coefficients for the common time dummies follow. Finally, r(beta) includes estimates for the coefficients for the additional controls (when available) and the constant

- r(Vbeta) (co)variance estimates in auxiliary regression

- r(tests) equivalence and common dynamics tests

- r(p_values) p-values for the equivalence and common dynamics tests

References

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.

MORA, R., AND I. REGGIO (2012): "Treatment effect identification using alternative parallel assumptions," Discussion paper, WP 12-33. UC3M.