

# Online Appendix

## “Two-way fixed effects estimators with heterogeneous treatment effects”

Clément de Chaisemartin\*      Xavier D’Haultfoeuille†

March 19, 2020

### Abstract

In this web appendix, we first discuss whether common trends necessarily implies homogenous treatment effect. Second, we show that the decomposition in Theorem 1 in the paper extends to fuzzy designs, to regressions with covariates, to regressions with a non-binary treatment, and we derive another decomposition under the supplementary assumption that the treatment effect does not change over time. Third, we extend the  $DID_M$  estimator to non-binary treatments. Fourth, we discuss inference. Fifth, we review all the papers included in our survey of papers published in the AER between 2010 and 2012 (see Section 5 of the paper). Finally, the last section gathers the proofs of all the additional results in this Web Appendix.

## 1 Can common trends hold with heterogeneous treatment effects?

Throughout the paper, we assume that groups experience common trends, but that the effect of the treatment may be heterogeneous between groups and / or over time. We now discuss two examples where this may happen. We then argue that the mechanisms behind these examples are fairly general. Thus treatment effects are often likely to be heterogeneous, even when common trends are plausible.

First, assume one wants to learn the effect of the minimum wage on the employment levels of some US counties. For simplicity, let us assume that the minimum wage can only take two values, a low and a high value. Also, let us assume that there are only two periods, the 90s and the 2000s. Between these two periods, the amount of competition from China for the US

---

\*University of California at Santa Barbara, clementdechaisemartin@ucsb.edu

†CREST, xavier.dhaultfoeuille@ensae.fr

industry increased substantially. Thus, for the common trends assumption to hold for counties A and B, the effect of that increase in competition should be the same on average in those two counties, in the counterfactual state of the world where A and B have a low minimum wage at both dates. For that to be true, the economy of those two counties should be pretty similar. For instance, if A has a very service-oriented economy, while B has a very industry-oriented economy, it is unlikely that their employment levels will react similarly to Chinese competition.

Now, if the economies of A and B are similar, they should also have similar effects of the minimum wage on employment, thus implying that the treatment effect is homogenous between groups. On the other hand, the treatment effect may vary over time. For instance, the drop in the employment levels of A and B due to Chinese competition will probably be higher if their minimum wage is high than if their minimum wage is low. This is equivalent to saying that the effect of the minimum wage on employment diminishes from the first to the second period: due to Chinese competition in the second period, the minimum wage may have a more negative effect on employment then.<sup>1</sup>

Second, assume one wants to learn the effect of a job training program implemented in some US counties on participants' wages. Let us suppose that individuals self-select into the training according to a Roy model:

$$D_{i,g,t} = 1\{Y_{i,g,t}(1) - Y_{i,g,t}(0) > c_{g,t}\}, \quad (4)$$

where  $c_{g,t}$  represents the cost of the training for individuals in county  $g$  and period  $t$ . We consider fuzzy designs such as this one in the next section. Here, the common trends condition requires that average wages without the training follow the same evolution in all counties. As above, for this to hold counties used in the analysis should have similar economies, so let us assume that those counties are actually identical copies of each other: at each period, their distribution of wages without and with the training is the same. Therefore,  $(g, t) \mapsto E(Y_{g,t}(1) - Y_{g,t}(0))$  is constant. However,  $c_{g,t}$  may vary across counties and over time: some counties may subsidize the training more than others, and some counties may change their subsidies over time. Then,  $(g, t) \mapsto E(Y_{i,g,t}(1) - Y_{i,g,t}(0) | D_{i,g,t} = 1)$  will not be constant, despite the fact that all counties in the sample have similar economies and experience similar trends on their wages.

Overall, when the treatment is assigned at the group  $\times$  period level as in the minimum wage example, the economic restrictions underlying the common trends assumption may also imply homogeneous treatment effect between groups. However, those restrictions usually do not imply that the treatment effect is constant over time. Moreover, when the treatment is assigned at the individual level, as in the job training example, the economic restrictions underlying the

---

<sup>1</sup> To simplify our discussion, in this example we consider only two counties. But in order to estimate consistently average treatment effects in the presence of county-specific shocks, the number of groups should tend to infinity, as in Section 5 below.

common trends assumption neither imply homogeneous treatment effects between groups, nor homogeneous treatment effects over time.

## 2 Results in fuzzy designs

In this section, the research design may be fuzzy: the treatment may vary within  $(g, t)$  cells. For instance, Enikolopov et al. (2011) study the effect of having access to an independent TV channel in Russia, and in each Russian region some people have access to that channel while other people do not.

### 2.1 Generalizing the decomposition of $\beta_{fe}$ to fuzzy designs

For any  $(g, t) \in \{1, \dots, G\} \times \{1, \dots, T\}$ , let

$$\Delta_{g,t}^{TR} = \frac{1}{N_{g,t}D_{g,t}} \sum_{i:D_{i,g,t}=1} [Y_{i,g,t}(1) - Y_{i,g,t}(0)]$$

denote the average treatment effect across the treated units of cell  $(g, t)$ . One has

$$\delta^{TR} = E \left( \sum_{g,t} \frac{N_{g,t}D_{g,t}}{N_1} \Delta_{g,t}^{TR} \right),$$

which generalizes (2) to fuzzy designs.<sup>2</sup> Theorem 1 shows that  $\beta_{fe}$  is also equal to the expectation of a weighted sum of the  $\Delta_{g,t}^{TR}$ s. Let

$$w_{g,t}^{TR} = \frac{\varepsilon_{g,t}}{\sum_{g,t} \frac{N_{g,t}D_{g,t}}{N_1} \varepsilon_{g,t}}.$$

**Theorem S1** *Suppose that Assumptions 1 and 3-5 hold. Then,*

$$\beta_{fe} = E \left( \sum_{g,t} \frac{N_{g,t}D_{g,t}}{N_1} w_{g,t}^{TR} \Delta_{g,t}^{TR} \right).$$

Theorem S1 shows that in fuzzy designs,  $\beta_{fe}$  is equal to the expectation of a weighted sum of the ATTs in each  $(g, t)$  cell. Again, some of the weights may be strictly negative. Note that under Assumption 2, Theorem S1 reduces to Theorem 1 in the paper.

The weights have a simple expression in the following special case.

**Assumption S1** (*Heterogenous adoption*)  $T = 2$  and for all  $g \in \{1, \dots, G\}$ ,  $D_{g,2} > D_{g,1} = 0$ .

---

<sup>2</sup>Any equation with a numbering lower than (4) or beginning with an A (e.g., (A1)) refers to an equation in the main paper.

Assumption S1 is satisfied in applications with two time periods, and where all groups are fully untreated at  $t = 1$  and partly treated at  $t = 2$ . This type of design often arises in practice, for instance when an innovation is heterogeneously adopted by various groups.

**Proposition S1** *If Assumptions 1 and S1 hold and  $N_{g,2}/N_{g,1}$  does not vary across  $g$ , then*<sup>3</sup>

$$w_{g,2}^{TR} = \frac{D_{g,2} - D_{.,2}}{\sum_{g=1}^G \frac{N_{g,2}}{N_1} (D_{g,2} - D_{.,2})^2}.$$

Proposition S1 shows that in the heterogeneous adoption design,  $\beta_{fe}$  assigns negative weights to the period-two ATT of groups with a mean treatment lower than the mean treatment in the full population. The reason why negative weights arise is intuitive. With two periods, the FE regression is equivalent to a regression of the first difference of the outcome on the period-two treatment in each group. This regression compares the evolution of the outcome in more- and less-treated groups. Doing so, it subtracts the treatment effect of the less-treated groups, hence the negative weights. Negative weights are a concern if the ATTs of the less- and more-treated groups systematically differ. This could be the case if treatment is determined by a Roy selection model. Then, the groups with the highest proportion of treated units could also be those where the ATT is the highest. On the other hand, if the proportion of treated units is randomly assigned to each group, negative weights are not a concern.<sup>4</sup>

The DID<sub>M</sub> estimator can also be generalized to fuzzy designs, see point 2 of Theorem S1 in the Web Appendix of de Chaisemartin and D’Haultfœuille (2018) for further details.

## 2.2 Application to Enikolopov et al. (2011)

Enikolopov et al. (2011) study the effect of NTV, an independent TV channel introduced in 1996 in Russia, on the share of the electorate voting for opposition parties. NTV’s coverage rate was heterogeneous across subregions: while a large fraction of the population received NTV in urbanized subregions, a smaller fraction received it in more rural subregions. The authors estimate the FE regression: they regress the share of votes for opposition parties in the 1995 and 1999 elections in Russian subregions on subregion fixed effects, an indicator for the 1999 election, and on the share of the population having access to NTV in each subregion at the time of the election. In 1995, the share of the population having access to NTV was equal to 0 in all subregions, while in 1999 it was strictly greater than 0 everywhere. Therefore, the authors’ research design corresponds exactly to the heterogenous adoption design discussed above. Enikolopov et al. (2011) find that  $\widehat{\beta}_{fe} = 6.65$  (s.e. = 1.40). According to this regression, increasing the share of the population having access to NTV from 0 to 100% increases the share

<sup>3</sup>Under Assumption S1,  $D_{g,1} = 0$ , so  $w_{g,1}^{TR}$  does not enter in the decomposition in Theorem 1.

<sup>4</sup>Corollaries 1 and 2 extend directly to fuzzy settings.

of votes for the opposition parties by 6.65 percentage points. Because there are only two time periods in the data and the regression is not weighted by subregions' populations,  $\widehat{\beta}_{fe} = \widehat{\beta}_{fd}$ .

We use the `twowayfeweights` Stata package, downloadable with its help file from the SSC repository, to compute the weights attached to  $\widehat{\beta}_{fe}$ . In 1995, all the weights are equal to zero because NTV does not exist yet. In 1999, 918 weights (47.4%) are strictly positive, while 1,020 (52.6%) are strictly negative. The negative weights sum to -2.26.  $\widehat{\sigma}_{fe} = 0.91$ :  $\beta_{fe}$  and  $\delta^{TR}$  may be of opposite signs if the standard deviation of the effect of NTV across subregions is above 0.91 percentage point.  $\widehat{\sigma}_{fe} = 1.23$ :  $\beta_{fe}$  may be of a different sign than the treatment effect in every subregion if the standard deviation of the effect of NTV across subregions is above 1.23 percentage point, a plausible amount of treatment effect heterogeneity.

Therefore,  $\beta_{fe}$  can only receive a causal interpretation if the effect of NTV is constant across subregions, or if the weights attached to it are uncorrelated with the intensity of that effect in each subregion (Assumption 7). These assumptions are not warranted. First, we estimate  $\widehat{\beta}_{fe}$  again, weighting the regression by subregions' populations. We obtain  $\widehat{\beta}_{fe} = 14.89$ , more than twice its value in the unweighted regression, and the difference between the coefficients is significant (t-stat=2.66). Therefore, we can reject the null that the treatment effect is constant: if the treatment effect was constant across subregions, the weighting would not matter so both the unweighted and the weighted regressions would estimate the same parameter. Second, the weights attached to  $\widehat{\beta}_{fe}$  are correlated with variables that are likely to be themselves associated with the intensity of the effect in each subregion. For instance, the correlation between the weights and subregions' populations is equal to 0.35 (t-stat=14.01). The effect of NTV may be higher in less populated subregions, as those regions are more rural and fewer other sources of information may be available there. This would lead to a violation of Assumption 7.

### 3 Extensions of the decomposition results

We consider hereafter several extensions of our decompositions of  $\beta_{fe}$  and  $\beta_{fd}$  in the paper. First, we consider decompositions under the common trends assumption, and under the assumption that the treatment effect is stable over time. Second, we extend our decompositions to ordered treatments. Third, we investigate the effect of including covariates in the regression. Fourth, we study two-way fixed effects 2SLS regressions. Throughout, we focus on sharp designs to ease the exposition. Nevertheless, all the results generalize to fuzzy designs.

### 3.1 $\beta_{fe}$ and $\beta_{fd}$ as weighted sums of ATEs of switching groups

We first show that under an additional condition,  $\beta_{fe}$  and  $\beta_{fd}$  can be written as a weighted sum of the ATEs of switching groups.

**Assumption S2** (*Stable treatment effect*) For all  $g$  and  $t \geq 2$ ,

$$E(\Delta_{g,t} | \mathbf{D}) D_{g,t-1} = E(\Delta_{g,t-1} | \mathbf{D}) D_{g,t-1}.$$

The stable treatment effect assumption requires that the ATE of every group treated in  $t - 1$  does not change from  $t - 1$  to  $t$ . By iteration, the ATE of a group treated for instance from period  $t_0$  to  $T$  is unrestricted before  $t_0$  but should be constant from  $t_0$  to  $T$ . Assumption S2 rules out the possibility that the treatment effect changes over time. Therefore, it may be implausible and should be carefully discussed.

We now show that under the common trend and stable treatment effects assumptions,  $\beta_{fe}$  and  $\beta_{fd}$  may identify weighted averages of ATEs. Let  $N_S = \sum_{(g,t):D_{g,t} \neq D_{g,t-1}} N_{g,t}$  and, for all  $g$  and  $t \geq 2$ ,

$$w_{g,t}^S = \frac{(D_{g,t} - D_{g,t-1}) \sum_{t' \geq t} \frac{N_{g,t'}}{N_{g,t}} \varepsilon_{g,t'}}{\sum_{(g,t):t \geq 2} \left[ \frac{N_{g,t}}{N_S} (D_{g,t} - D_{g,t-1}) \sum_{t' \geq t} \frac{N_{g,t'}}{N_{g,t}} \varepsilon_{g,t'} \right]},$$

$$w_{fd,g,t}^S = \frac{(D_{g,t} - D_{g,t-1}) \varepsilon_{fd,g,t}}{\sum_{(g,t):t \geq 2} \frac{N_{g,t}}{N_S} (D_{g,t} - D_{g,t-1}) \varepsilon_{fd,g,t}}.$$

**Theorem S2** *Suppose that Assumptions 1-5 and S2 hold. Then,*

$$\beta_{fe} = E \left( \sum_{(g,t):D_{g,t} \neq D_{g,t-1}, t \geq 2} \frac{N_{g,t}}{N_S} w_{g,t}^S \Delta_{g,t} \right),$$

$$\beta_{fd} = E \left( \sum_{(g,t):D_{g,t} \neq D_{g,t-1}, t \geq 2} \frac{N_{g,t}}{N_S} w_{fd,g,t}^S \Delta_{g,t} \right).$$

Moreover,  $w_{fd,g,t}^S \geq 0$  for all  $g$  and  $t \geq 2$ . If Assumption 6 holds and  $N_{g,t}/N_{g,t-1}$  does not vary across  $g$  for all  $t \geq 2$ ,  $w_{g,t}^S \geq 0$  for all  $g$  and  $t \geq 2$ .

Theorem S2 shows that in sharp designs, under the common trends and stable treatment effect assumptions,  $\beta_{fe}$  and  $\beta_{fd}$  identify weighted sums of ATEs of switching cells. The weights differ from those in Theorems 1 and 2. Now the weights attached to  $\beta_{fe}$  are all positive in staggered adoption designs, while the weights attached to  $\beta_{fd}$  are all positive in all sharp designs. Therefore, in staggered adoption (resp. sharp) designs,  $\beta_{fe}$  (resp.  $\beta_{fd}$ ) relies on the assumption that the treatment effect is stable over time, but it does not require that treatment effects be homogeneous between groups.

### 3.2 Non-binary, ordered treatment

We now consider the case where the treatment takes a finite number of ordered values,  $D_{i,g,t} \in \{0, 1, \dots, \bar{d}\}$ , and show that Theorem 1 can easily be extended to this case.<sup>5</sup> We need to define potential outcomes for all the possible treatment values. For instance  $Y_{i,g,t}(\bar{d})$  is the counterfactual outcome of observation  $i$  in cell  $(g, t)$  if she receives treatment value  $\bar{d}$ . We also need to modify the treatment effect parameters we consider. In lieu of  $\delta^{TR}$ , we consider the average causal response (ACR) on the treated,

$$\delta^{ACR} = E \left( \frac{1}{N_1} \sum_{i,g,t} Y_{i,g,t}(D_{g,t}) - Y_{i,g,t}(0) \right).$$

Similarly, for all  $(g, t)$  such that  $D_{g,t} \neq 0$ , we consider, instead of  $\Delta_{g,t}$ ,

$$\Delta_{g,t}^{ACR} = \frac{1}{N_{g,t} D_{g,t}} \sum_{i=1}^{N_{g,t}} [Y_{i,g,t}(D_{g,t}) - Y_{i,g,t}(0)].$$

Then, similarly to (2), the following decomposition holds:

$$\delta^{ACR} = E \left( \sum_{(g,t):D_{g,t} \neq 0} \frac{N_{g,t} D_{g,t}}{N_1} \Delta_{g,t}^{ACR} \right).$$

Let  $w_{g,t}^O = \frac{\varepsilon_{g,t}}{\sum_{g,t} \frac{N_{g,t} D_{g,t}}{N_1} \varepsilon_{g,t}}$ . Note that if the treatment is binary,  $w_{g,t}^O = w_{g,t}$ .

**Theorem S3** *Suppose that Assumptions 1-5 hold and  $D_{i,g,t} \in \{0, \dots, \bar{d}\}$ . Then,*

$$\beta_{fe} = E \left( \sum_{(g,t):D_{g,t} \neq 0} \frac{N_{g,t} D_{g,t}}{N_1} w_{g,t}^O \Delta_{g,t}^{ACR} \right).$$

Theorem S3 shows that under Assumption 5, when the treatment is not binary  $\beta_{fe}$  identifies a weighted sum of the ACRs in all the  $(g, t)$  cells that are not untreated. Then, since the proof of Corollary 1 does not rely on the nature of the treatment, Corollary 1 directly applies to ordered treatments as well, by just replacing  $w_{g,t}$  and  $N_{g,t}$  by  $w_{g,t}^O$  and  $N_{g,t} D_{g,t}$ , respectively. Corollary 2 extends as well to this set-up, by simply modifying the no-correlation condition appropriately.

---

<sup>5</sup>Theorem 2 can also be extended to the case of a non-binary treatment.

### 3.3 Including covariates

Often times, researchers also include a vector of covariates  $X_{g,t}$  as control variables in their regression. In this section, we show that our Theorem 1 can be extended to this case.<sup>6</sup> We start by redefining Regression 1 in this context.

**Regression 1X** (*Fixed-effects regression with covariates*)

Let  $\widehat{\beta}_{fe}^X$  denote the coefficient of  $D_{g,t}$  in an OLS regression of  $Y_{i,g,t}$  on group and period fixed effects,  $D_{g,t}$ , and  $X_{g,t}$ . Let  $\beta_{fe}^X = E\left(\widehat{\beta}_{fe}^X\right)$ .

Then, we need to modify Assumptions 3-5. Hereafter, we let  $\mathbf{X}_g = (X_{g,1}, \dots, X_{g,t})$ .

**Assumption S3** (*Independent groups with covariates*) The vectors  $(Y_{g,t}(0), Y_{g,t}(1), D_{g,t}, X_{g,t})_{1 \leq t \leq T}$  are mutually independent.

**Assumption S4** (*Strong exogeneity and common trends with covariates*) There is a vector  $\gamma$  of same dimension as  $X_{g,t}$  such that

$$E(Y_{g,t}(0) - Y_{g,t-1}(0) - (X_{g,t} - X_{g,t-1})'\gamma | \mathbf{D}_g, \mathbf{X}_g) = E(Y_{g,t}(0) - Y_{g,t-1}(0) - (X_{g,t} - X_{g,t-1})'\gamma)$$

and  $E(Y_{g,t}(0) - Y_{g,t-1}(0) - (X_{g,t} - X_{g,t-1})'\gamma)$  does not vary across  $g$ .

Rearranging, Assumption S4 requires that

$$E(Y_{g,t}(0) | \mathbf{D}_g, \mathbf{X}_g) - E(Y_{g,t-1}(0) | \mathbf{D}_g, \mathbf{X}_g) = (X_{g,t} - X_{g,t-1})'\gamma + \lambda_t,$$

for some constant  $\lambda_t$ . Then, Assumption S4 allows for the possibility that groups experience different evolutions of their  $Y_{g,t}(0)$  over time, but it requires that those differential evolutions are fully accounted for by a linear model in  $X_{g,t} - X_{g,t-1}$ , the change in a group's covariates. Assumption S4 is implied by the linear model that is often invoked to justify the use of the FE regression with covariates. For instance, the use of Regression 1X is often justified by the following model:

$$Y_{g,t}(0) = \gamma_g + \lambda_t + X_{g,t}'\gamma + \eta_{g,t}, \quad E(\eta_{g,t} | \mathbf{D}_g, \mathbf{X}_g) = 0. \quad (5)$$

Equation (5) implies Assumption S4, but it does not imply Assumption 5.

An interesting special case is when the control variables are group-specific linear trends. Then, Assumption S4 requires that for all  $t \geq 2$ ,

$$E(Y_{g,t}(0) | \mathbf{D}_g, \mathbf{X}_g) - E(Y_{g,t-1}(0) | \mathbf{D}_g, \mathbf{X}_g) = \gamma_g + \lambda_t,$$

---

<sup>6</sup>Theorem 2 can also be extended to regressions with covariates.

for some constants  $\gamma_g$  and  $\lambda_t$ . From  $t - 1$  to  $t$ , the evolution of  $Y(0)$  in group  $g$  should deviate from its group-specific linear trend  $\gamma_g$  by an amount  $\lambda_t$  common to all groups. Then, Assumption S4 is a “common deviation from linear trends” assumption, which may be more plausible than the standard common trends assumption.

Let  $\varepsilon_{g,t}^X$  denote the residual of observations in cell  $(g, t)$  in the regression of  $D_{g,t}$  on group and period fixed effects and  $X_{g,t}$ . One can show that if the regressors in Regression 1 are not collinear, the average value of  $\varepsilon_{g,t}^X$  across all treated  $(g, t)$  cells differs from 0:  $\sum_{(g,t):D_{g,t}=1} (N_{g,t}/N_1) \varepsilon_{g,t}^X \neq 0$ . Then, let

$$w_{g,t}^X = \frac{\varepsilon_{g,t}^X}{\sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} \varepsilon_{g,t}^X}.$$

**Theorem S4** *Suppose that Assumptions 1-2 and S3-S4 hold. Then,*

$$\beta_{fe}^X = E \left[ \sum_{(g,t):D_{g,t}=1} \frac{N_{g,t}}{N_1} w_{g,t}^X \Delta_{g,t} \right].$$

Theorem S4 shows that under a modified version of the common trends assumption accounting for the covariates,  $\beta_{fe}^X$  identifies a weighted sum of the  $\Delta_{g,t}^{TR}$ s, as  $\beta_{fe}$  in Theorem 1, with different but still potentially negative weights.<sup>7</sup> Assumption S4 may be more plausible than Assumption 5, but adding covariates may increase the prevalence of negative weights, or the correlation between the weights and the  $\Delta_{g,t}$ s, thus making  $\beta_{fe}^X$  less robust to heterogeneous effects than  $\beta_{fe}$ .

### 3.4 2SLS regressions

Researchers have sometimes estimated 2SLS versions of Regressions 1 and 2. Our main conclusions also apply to those regressions. Let  $\widehat{\beta}_{fe}^{2SLS}$  denote the coefficient of  $D_{i,g,t}$  in a 2SLS regression of  $Y_{g,t}$  on group and period fixed effects and  $D_{i,g,t}$ , using a variable  $Z_{g,t}$  constant within each group  $\times$  period as the instrument for  $D_{i,g,t}$ .  $Z_{g,t}$  typically represents an incentive for treatment allocated at the group  $\times$  period level. For instance, Duflo (2001) studies the effect of years of schooling on wages in Indonesia, using a primary school construction program as an instrument. Specifically, she estimates a 2SLS regression of wages on cohort and district of birth fixed effects and years of schooling, using the interaction of belonging to a cohort entering primary school after the program was completed and the number of schools constructed in one’s district of birth as the instrument for years of schooling.

---

<sup>7</sup>In a previous version of this paper, we had shown that under a different, and arguably less natural, common trends assumption,  $\beta_{fe}^X$  identifies a weighted sum of the  $\Delta_{g,t}^{TR}$ , with the same weights as in Theorem 1. We thank an anonymous referee for pointing out issues with the common trends assumption we had previously proposed.

Remark that  $\widehat{\beta}_{fe}^{2SLS} = \widehat{\beta}_{fe}^Y / \widehat{\beta}_{fe}^D$ , where  $\widehat{\beta}_{fe}^Y$  (resp.  $\widehat{\beta}_{fe}^D$ ) is the coefficient of  $Z_{g,t}$  in the reduced-form regression of  $Y_{g,t}$  (resp. the first-stage regression of  $D_{g,t}$ ) on group and period fixed effects and  $Z_{g,t}$ . Then let  $\beta_{fe}^{2SLS} = E[\widehat{\beta}_{fe}^Y] / E[\widehat{\beta}_{fe}^D]$ .<sup>8</sup> Following Imbens and Angrist (1994), for any  $z \in \text{Supp}(Z)$  let  $D_{i,g,t}(z)$  denote the potential treatment of unit  $i$  in  $(g, t)$  if  $Z_{i,g,t} = z$ . It follows from Theorem 1 that under a common trends assumption on  $D_{i,g,t}(0)$ ,  $E[\widehat{\beta}_{fe}^D]$  is equal to a weighted sum of the average effects of the instrument on the treatment in each group and time period, with potentially many negative weights. Similarly, under a common trends assumption on  $Y_{i,g,t}(D_{i,g,t}(0))$  instead of  $Y_{i,g,t}(0)$ ,  $E[\widehat{\beta}_{fe}^Y]$  is equal to a weighted sum of the average effects of the instrument on the outcome, again with potentially many negative weights. For instance, in Duflo (2001), under a common trends assumption on  $D_{i,g,t}(0)$ , the number of years of schooling individuals would complete if zero new schools were constructed in their district, the first stage coefficient identifies a weighted sum of the effect of one new school on years of schooling in every district, with many negative weights.<sup>9</sup>

Hence, it is only if the average effects of  $Z_{g,t}$  on  $Y_{i,g,t}$  and  $D_{i,g,t}$  are constant across groups and periods, or if the weights are uncorrelated to treatment effects as in Assumption 7, that the reduced-form and first-stage coefficients respectively identify the average effect of  $Z_{i,g,t}$  on  $Y_{i,g,t}$  and  $D_{i,g,t}$ . Then, this implies that  $\beta_{fe}^{2SLS}$  identifies, under the conditions in Imbens and Angrist (1994), the LATE of  $D_{i,g,t}$  on  $Y_{i,g,t}$  among units that comply with the instrument.<sup>10</sup>

## 4 Extending the DID<sub>M</sub> estimator

Theorem 3 can be extended to the case where  $D_{i,g,t}$  is not binary but takes values in  $\mathcal{D} = \{0, \dots, \bar{d}\}$ . The causal effect we consider is the switchers' causal response

$$\delta^{SCR} = E \left( \frac{1}{N_{D,S}} \sum_{(i,g,t): t \geq 2, D_{g,t} \neq D_{g,t-1}} [Y_{i,g,t}(\max(D_{g,t}, D_{g,t-1})) - Y_{i,g,t}(\min(D_{g,t}, D_{g,t-1}))] \right),$$

where  $N_{D,S} = \sum_{(g,t): t \geq 2} N_{g,t} |D_{g,t} - D_{g,t-1}|$ . Note that  $\delta^{SCR} = \delta^S$  when  $D_{i,g,t}$  is binary.

---

<sup>8</sup>We do not consider here  $E[\widehat{\beta}_{fe}^{2SLS}]$ , as the 2SLS estimator may not have an expectation. Moreover, under conditions similar to those imposed in Section 5 below,  $\beta_{fe}^{2SLS}$  is the probability limit of  $\widehat{\beta}_{fe}^{2SLS}$ , which makes  $\beta_{fe}^{2SLS}$  the proper estimand here.

<sup>9</sup>New schools were constructed in every district, so this application falls into the heterogeneous adoption case.

<sup>10</sup>In the special case with two groups and two periods, a binary incentive for treatment, and where only group 1 in period 1 receives the incentive, de Chaisemartin (2010) and Hudson et al. (2015) show that in a 2SLS regression of  $Y_{i,g,t}$  on  $1\{g = 2\}$ ,  $1\{t = 2\}$  and  $D_{i,g,t}$ , using  $Z_{g,t} = 1\{g = 2\}1\{t = 2\}$  as the instrument, the coefficient of  $D_{i,g,t}$  identifies a LATE under common trends assumptions on  $Y_{i,g,t}(D_{i,g,t}(0))$  and  $D_{i,g,t}(0)$ . However, the discussion above shows that this result does not generalize to applications with multiple groups and periods or a non-binary instrument, as in Duflo (2001) where the number of new schools constructed varies across districts.

We identify  $\delta^{SCR}$  under the following two conditions, which generalize Assumptions 4-5 and 9-12 to non-binary treatments.

**Assumption S5** (*Mean independence between a group's outcome and other groups treatments*)  
For all  $(d, g, t) \in \mathcal{D} \times \{1, \dots, G\} \times \{1, \dots, T\}$ ,  $E(Y_{g,t}(d)|\mathbf{D}) = E(Y_{g,t}(d)|\mathbf{D}_g)$ .

**Assumption S6** (*Strong exogeneity*) For all  $(d, g, t) \in \mathcal{D} \times \{1, \dots, G\} \times \{2, \dots, T\}$ ,  $E(Y_{g,t}(d) - Y_{g,t-1}(d)|\mathbf{D}_g) = E(Y_{g,t}(d) - Y_{g,t-1}(d))$ .

**Assumption S7** (*Common trends*) For every  $d$ , for all  $t \geq 2$  and  $g$ ,  $E(Y_{g,t}(d) - Y_{g,t-1}(d))$  does not vary across  $g$ .

**Assumption S8** (*Existence of "stable" groups*) For all  $t \in \{2, \dots, T\}$ , for all  $(d, d') \in \mathcal{D}^2$ ,  $d \neq d'$ , if there is at least one  $g \in \{1, \dots, G\}$  such that  $D_{g,t-1} = d$  and  $D_{g,t} = d'$ , then there exists at least one  $g' \neq g$ ,  $g' \in \{1, \dots, G\}$  such that  $D_{g',t-1} = D_{g',t} = d$ .

When the treatment takes a large number of values, Assumption S8 may be violated. A solution, then, is to consider a modified treatment variable  $\tilde{D}_{g,t} = h(D_{g,t})$  that groups together several values of  $D_{g,t}$ , to ensure that Assumption S8 holds for  $\tilde{D}_{g,t}$ . For instance, if the treatment can be equal to 0, 1, 2, or 3, and there is a group whose treatment switches from 2 to 3 between periods 1 and 2, but no group whose treatment remains equal to 2 between those two dates, one may define  $\tilde{D}_{g,t} = \min(D_{g,t}, 2)$  if there is a group whose treatment is equal to 3 at periods 1 and 2. Then, Theorem S5 below still holds, after replacing  $D_{g,t}$  by  $\tilde{D}_{g,t}$  in the DID $_{d,d',t}$  estimators defined below, and if Assumption S7 is replaced by the requirement that  $E(Y_{g,t}(d) - Y_{g,t-1}(d))$  only depends on  $t$  and  $h(d)$ .

In order to define DID $_M$  in this context, let us introduce, for all  $(d, d', t) \in \mathcal{D}^2 \times \{2, \dots, T\}$ ,

$$\text{DID}_{d,d',t} = [1\{d < d'\} - 1\{d' < d\}] \left[ \sum_{(g,t): D_{g,t}=d', D_{g,t-1}=d, t \geq 2} \frac{N_{g,t}}{N_{d,d',t}} [Y_{g,t} - Y_{g,t-1}] - \sum_{(g,t): D_{g,t}=D_{g,t-1}=d, t \geq 2} \frac{N_{g,t}}{N_{d,d,t}} [Y_{g,t} - Y_{g,t-1}] \right],$$

where  $N_{d,d',t}$  is defined as in (3) for any  $(d, d') \in \mathcal{D}^2$ . Then

$$\text{DID}_M = \sum_{t=2}^T \sum_{(d,d') \in \mathcal{D}^2, d \neq d'} \frac{N_{d,d',t}}{N_{D,S}} \text{DID}_{d,d',t}.$$

If the treatment is binary, the DID $_M$  estimator defined above is equal to that defined in Section 3 of the paper.

**Theorem S5** *Suppose that  $D_{i,g,t} \in \mathcal{D}$  and Assumptions 1-2 and S5-S8 hold. Then  $E[\text{DID}_M] = \delta^{SCR}$ .*

Theorem S5 generalizes Theorem 3 to non-binary treatments. We can also extend Theorem 4 in the same way to construct placebo tests of Assumption S7.

Finally, Theorem 3 can also be extended to the case with covariates. Under versions of Assumptions 10 and 11 written conditional on  $X$ , a conditional version of the DID<sub>M</sub> estimator is consistent for  $\delta^S$  under the common support condition  $\text{Supp}(X_{d,g,t}) = \text{Supp}(X)$ . We refer to de Chaisemartin and D'Haultfœuille (2018) for further details.

## 5 Statistical properties of DID<sub>M</sub> and inference on $\delta^S$

In this section, we establish the asymptotic properties of DID<sub>M</sub> and construct confidence intervals on  $\delta^S$  based on DID<sub>M</sub>. We consider an asymptotic framework where the number of groups  $G$  tends to infinity. To define the confidence intervals, let  $P_{d,d',t} = N_{d,d',t}/G$  and

$$Q_{d,d',t} = \frac{1}{G} \sum_g N_{g,t} \mathbb{1}\{D_{g,t} = d, D_{g,t-1} = d'\} (Y_{g,t} - Y_{g,t-1}).$$

Then, let  $\hat{\sigma}^2 = \sum_g \hat{\psi}_g^2/G$ , with

$$\begin{aligned} \hat{\psi}_g &= \frac{G}{N_S} \sum_{t>1} N_{g,t} \left[ \mathbb{1}\{D_{g,t} \neq D_{g,t-1}\} (Y_{g,t} - Y_{g,t-1} - \text{DID}_M) - \hat{\psi}_{g,t}^B \right], \\ \hat{\psi}_{g,t}^B &= \frac{1}{P_{0,0,t}} \left[ \mathbb{1}\{D_{g,t} > D_{g,t-1}\} Q_{0,0,t} + P_{1,0,t} \mathbb{1}\{D_{g,t} = D_{g,t-1} = 0\} \left( Y_{g,t} - Y_{g,t-1} - \frac{Q_{0,0,t}}{P_{0,0,t}} \right) \right] \\ &\quad + \frac{1}{P_{1,1,t}} \left[ \mathbb{1}\{D_{g,t} < D_{g,t-1}\} Q_{1,1,t} + P_{0,1,t} \mathbb{1}\{D_{g,t} = D_{g,t-1} = 1\} \left( Y_{g,t} - Y_{g,t-1} - \frac{Q_{1,1,t}}{P_{1,1,t}} \right) \right]. \end{aligned}$$

We consider confidence intervals of the form

$$\text{CI}_{1-\alpha}(\delta^S) = \left[ \text{DID}_M - z_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{G}}, \text{DID}_M + z_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{G}} \right],$$

where  $z_{1-\alpha/2}$  denotes the quantile of order  $1 - \alpha/2$  of a standard normal variable.

We now establish the asymptotic properties of DID<sub>M</sub> and  $\text{CI}_{1-\alpha}(\delta^S)$  under the following assumptions. Hereafter, we denote  $\mathbf{U} = (P_{0,0,1}, Q_{0,0,1}, \dots, P_{1,1,T}, Q_{1,1,T})$ .

**Assumption S9** (*Existence of moments and limits*)  $\sup_{g,t} N_{g,t} < +\infty$  and  $\sup_{(d,g,t)} E(Y_{g,t}^4(d)) < +\infty$ .  $\lim_G E[\mathbf{U}]$  and  $\lim_G G \times V(\mathbf{U})$  exist.

**Assumption S10** (*Positive probability of “stable” groups and existence of switchers*) For all  $(d, g, t) \in \{0, 1\} \times (\mathbb{N} \setminus \{0\}) \times \{2, \dots, T\}$ ,  $\Pr(D_{g,t} = 1-d, D_{g,t-1} = d) > 0$  implies  $\lim_G E[P_{1-d,d,t}] > 0$  and  $\lim_G E[P_{d,d,t}] > 0$ . Moreover,  $\lim_G E[P_{0,1,t} + P_{1,0,t}] > 0$  for at least one  $t$ .

Assumption S9 imposes the (uniform) existence of moments of order 4 of  $Y_{g,t}(d)$ , and that some non-random averages converge as  $G$  tends to infinity. These assumptions ensure that we can apply law of large numbers and central limit theorems in our set-up where groups are independent but not necessarily identically distributed. Assumption S10 imposes that when at least one group switches from  $d$  to  $1 - d$  with a positive probability on a given period, then on average over all groups, the limit probabilities of switching from  $d$  to  $1 - d$  will be positive as  $G \rightarrow \infty$ . The limit probability of remaining at  $d$  will also be positive. This latter condition may be seen as a weaker version of Assumption 11, as it imposes the existence of “stable” groups only with probability tending to one as  $G \rightarrow \infty$ . The last condition in Assumption S10 simply states that asymptotically, the proportion of switchers is strictly positive.

The following result shows that under these conditions,  $DID_M$  is asymptotically normal, and that  $CI_{1-\alpha}(\delta^S)$  is asymptotically conservative.

**Theorem S6** *Suppose that Assumptions 1-5, 9-10 and S9-S10 hold. Then, as  $G \rightarrow \infty$ ,*

$$\sqrt{G} (DID_M - \delta^S) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

*with  $\sigma^2$  defined in (19) below. Moreover,*

$$\limsup_{G \rightarrow \infty} \Pr(\delta^S \in CI_{1-\alpha}(\delta^S)) \geq 1 - \alpha.$$

Theorem S6 shows that  $DID_M$  is an asymptotically normal estimator of  $\delta^S$  when the number of groups tends to infinity, provided the outcomes and treatments are independent across groups. As is usually the case for estimators constructed using independent but not identically distributed random variables (see e.g. Liu and Singh, 1995), the asymptotic variance of  $DID_M$  can only be conservatively estimated. As a result, the confidence interval we propose is asymptotically conservative.

## 6 Detailed literature review

We now review the 33 papers that use two-way fixed effects or closely related regressions that we found in our literature review. For each paper, we use the following presentation:

**Authors (year), Title.** *Where the two-way fixed effects estimator is used in the paper.*

Description of the two-way fixed effects estimator used in the paper, and how it relates to Regression 1 or 2. Assessment of whether the stable groups assumption holds in this paper. Assessment of whether the research design is sharp or fuzzy.

1. **Chandra et al. (2010), Patient Cost-Sharing and Hospitalization Offsets in the Elderly.** *First line of Tables 2 and 3.*

In the regressions in the first line of Tables 2 and 3, the outcomes (e.g. a measure of utilization for plan  $p$  in month  $t$ ) are regressed on plan fixed effects, month fixed effects, and an indicator of whether plan  $p$  had increased copayments in month  $t$  (see regression equation at the bottom of page 198). This regression corresponds to Regression 1. The period analyzed runs from January 2000 to September 2003. The stable groups assumption is satisfied until January 2002, when the HMO plans also become treated. This is a sharp design.

2. **Duggan and Morton (2010), The Effect of Medicare Part D on Pharmaceutical Prices and Utilization.** *Tables 2 and 3.*

In regression Equation (1), the dependent variable is the change in the price of drug  $j$  between 2003 and 2006, the explanatory variables are the Medicare market share for drug  $j$  in 2003, and some control variables. This regression corresponds to Regression 2, with some control variables. The stable groups assumption is presumably not satisfied: it seems unlikely that there are drugs whose Medicare market share in 2003 is equal to 0. This is a sharp design.

3. **Aizer (2010), The Gender Wage Gap and Domestic Violence.** *Table 2.*

In regression Equation (2), the dependent variable is the log of female assaults among females of race  $r$  in county  $c$  and year  $t$ , and the explanatory variables are race, year, county, race  $\times$  year, race  $\times$  county, and county  $\times$  year fixed effects, as well as the gender wage gap in county  $c$ , year  $t$ , and race  $r$ , and some control variables. This regression is a “three-way fixed effects” version of Regression 1, with some control variables. The stable groups assumption is presumably satisfied: it seems likely that between each pair of consecutive years, there are counties where the gender wage gap does not change. This is a fuzzy design: the treatment of interest is the gender wage gap in a couple (see the bargaining model in Appendix 1), which varies within (year, county) cells.

4. **Algan and Cahuc (2010), Inherited Trust and Growth.** *Figure 4.*

Figure 4 presents a regression of changes in income per capita from 1935 to 2000 on changes in inherited trust over the same period and a constant. This regression corresponds to Regression 2. The stable groups assumption is satisfied: there are countries where inherited trust does not change from 1935 to 2000. This is a sharp design.

5. **Ellul et al. (2010), Inheritance Law and Investment in Family Firms.** *Table 7.*

In the regressions presented in Table 7, the dependent variable is the capital expenditure of firm  $j$  in year  $t$ , and the explanatory variables are firm fixed effects, an indicator for whether year  $t$  is a succession period for firm  $j$ , some controls, and three treatment variables: the interaction of the succession indicator with the level of investor protection in the country where firm  $j$  is located, the interaction of the succession indicator with the level of

inheritance laws permissiveness in the country where firm  $j$  is located, and the interaction of the succession indicator with the level of inheritance laws permissiveness and the level of investor protection in the country where firm  $j$  is located. This regression is similar to Regression 1 with controls, except that it has three treatment variables. The stable groups assumption is presumably not satisfied: for instance, it seems unlikely that there are countries with no investor protection at all. This is a sharp design.

6. **Bustos (2011), Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms.** *Tables 3 to 12.*

In regression Equation (11), the dependent variable is the change in exporting status of firm  $i$  in sector  $j$  between 1992 and 1996, and the explanatory variables are the change in trade tariffs in Brasil for products in sector  $j$  over the same period, and some control variables. This regression corresponds to Regression 2, with some controls. The stable groups assumption is presumably satisfied: it seems likely that there are sectors where trade tariffs in Brasil did not change between 1992 and 1996. This is a sharp design.

7. **Anderson and Sallee (2011), Using Loopholes to Reveal the Marginal Cost of Regulation: The Case of Fuel-Economy Standards.** *Table 5 Column 2.*

In the regression in Table 5 Column (2), the dependent variable is an indicator for whether a car sold is a flexible fuel vehicle, and the explanatory variables are state and month fixed effects, the percent of gas stations that have ethanol fuel in each month  $\times$  state, and some controls. This regression corresponds to Regression 1. The stable groups assumption is presumably satisfied: it seems likely that between each pair of consecutive months, there are states where the percent ethanol availability does not change. This a fuzzy design: the treatment of interest is whether a car buyer has access to ethanol fuel, which varies within (month,state) cells.

8. **Bagwell and Staiger (2011), What Do Trade Negotiators Negotiate About? Empirical Evidence from the World Trade Organization.** *Table 3, OLS columns.*

In regression equations (15a) and (15b), the dependent variable is the ad valorem tariff level bound by country  $c$  on product  $g$ , while the explanatory variables are country and product fixed effects, and two treatment variables which vary at the country  $\times$  product level. These regressions are similar to Regression 1, except that they have two treatment variables. The stable groups assumption is not applicable here, as none of the two sets of fixed effects included in the regression correspond to an ordered variable. This is a sharp design.

9. **Zhang and Zhu (2011), Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia.** *Tables 3 and 4, Columns 4-6.*

In the regression in, say, Table 3 Column (4), the dependent variable is the total number

of contributions to Wikipedia by individual  $i$  at period  $t$ , regressed on individual fixed effects, an indicator for whether period  $t$  is after the Wikipedia block, the interaction of this indicator and a measure of social participation by individual  $i$ , and some controls. This regression corresponds to Regression 1 with some controls. The stable groups assumption is satisfied: there are individuals with a social participation measure equal to 0. This is a sharp design.

10. **Hotz and Xiao (2011), The Impact of Regulations on the Supply and Quality of Care in Child Care Markets.** *Table 7, Columns 4 and 5.*

In Regression Equation (1), the dependent variable is the outcome for market  $m$  in state  $s$  and year  $t$ , and the explanatory variables are state and year fixed effects, various measures of regulations in state  $s$  in year  $t$ , and some controls. This regression corresponds to Regression 1 with several treatment variables and with some controls. The stable groups assumption is presumably satisfied: between each pair of consecutive years, it is likely that there are states whose regulations do not change. This is a sharp design.

11. **Mian and Sufi (2011), House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis.** *Tables 2 and 3.*

In Regression Equation (1), the dependent variable is the change in homeowner leverage from 2002 to 2006 for individual  $i$  living in zip code  $z$  in MSA  $m$ , and the dependent variable is the change in the house price for that individual, instrumented by MSA-level housing supply elasticity. This regression is the 2SLS version of Regression 2, with some controls. The stable groups assumption is presumably not satisfied: it is unlikely that some MSAs have an housing supply elasticity equal to 0. This is a sharp design.

12. **Wang (2011), State Misallocation and Housing Prices: Theory and Evidence from China.** *Table 5, Panel A.*

In regression Equation (15), the dependent variable is the quantity of housing services in household  $i$ 's residence in year  $t$ , while the explanatory variables are an indicator for period  $t$  being after the reform, a measure of mismatch in household  $i$ , the interaction of the measure of mismatch and the time indicator, and some controls. This regression is similar to Regression 1 with some controls, except that it has a measure of mismatch in household  $i$  instead of household fixed effects. The stable groups assumption is presumably satisfied: it is likely that some households have a mismatch equal to 0. This is a sharp design.

13. **Duranton and Turner (2011), The Fundamental Law of Road Congestion: Evidence from US Cities.** *Table 5.*

In the regressions presented in, say, the first column of Table 5, the dependent variable is the change in vehicle kilometers traveled in MSA  $s$  between decades  $t$  and  $t-1$ , and the

explanatory variables are the change in kilometers of roads in MSA  $s$  between decades  $t$  and  $t-1$ , and decade effects. This regression corresponds to Regression 2. The stable groups assumption is presumably satisfied: it is likely that between each pair of consecutive decades, there are some MSAs where the kilometers of roads do not change. This is a sharp design.

14. **Acemoglu et al. (2011), The Consequences of Radical Reform: The French Revolution.** *Table 3.*

In regression Equation (1), the dependent variable is urbanization in polity  $j$  at time  $t$ , while the explanatory variables are time and polity fixed effects, and the number of years of French presence in polity  $j$  interacted with the time effects. This regression corresponds to Regression 1. The stable groups assumption is satisfied as there are several polities that did not experience any year of French presence. This is a sharp design.

15. **Baum-Snow and Lutz (2011), School Desegregation, School Choice, and Changes in Residential Location Patterns by Race.** *Tables 2 to 6.*

In regression Equation (1), the dependent variable is, say, whites public school enrolment in MSA  $j$  in year  $t$ , while the explanatory variables are MSA and region  $\times$  time fixed effects, and an indicator for whether MSA  $j$  is desegregated. This regression corresponds to Regression 1 with controls. The stable groups assumption is satisfied: between each pair of consecutive years, there are MSAs whose desegregation status does not change. This is a sharp design.

16. **Dinkelman (2011), The Effects of Rural Electrification on Employment: New Evidence from South Africa.** *Tables 4 and 5 Columns 5-8, Table 8 Columns 3-4, Table 9 Column 2, and Table 10 Columns 2, 4, and 6.*

In regression Equation (3), the dependent variable is, say, the first difference of the female employment rate for community  $j$  between periods 1 and 2, and the explanatory variables are district fixed effects, the change of electrification status of community  $j$  between periods 1 and 2, and some statistical controls. The land gradient in community  $j$  is used as an instrument for the change in electrification. This regression corresponds to the 2SLS version of Regression 2 with some controls. The stable groups assumption is presumably satisfied: it is likely that there are communities whose land gradient is 0. This is a sharp design.

17. **Enikolopov et al. (2011), Media and Political Persuasion: Evidence from Russia.** *Table 3.*

In regression Equation (5), the dependent variable is the share of votes for party  $j$  in election-year  $t$  and subregion  $s$ , and the explanatory variables are subregion and election fixed effects, and the share of people having access to NTV in subregion  $s$  in election-year  $t$ . This regression corresponds to Regression 1. The stable groups assumption is not satisfied:

the share of people having access to NTV strictly increases in all regions between 1995 and 1999, the two elections used in the analysis. This a fuzzy design: the treatment of interest is whether a person has access to NTV, which varies within (subregion,year) cells.

18. **Fang and Gavazza (2011), Dynamic Inefficiencies in an Employment-Based Health Insurance System: Theory and Evidence.** *Tables 2, 3, 5, and 6, Column 3.*

In regression Equation (7), the dependent variable is the health expenditures of individual  $j$  working in industry  $i$  in period  $t$  and region  $r$ , and the explanatory variables are individual effects, region specific time effects, and the job tenure of individual  $j$ . The death rate of establishments in industry  $i$  in period  $t$  and region  $r$  is used as an instrument for the job tenure of individual  $j$ . This regression is the 2SLS version of Regression 2 with controls. The stable groups assumption is presumably satisfied: between each pair of consecutive years, it is likely that there are some industry  $\times$  region pairs where the death rate of establishments does not change. This a fuzzy design: the instrument of interest is whether a person's former employee closed down over the current year, which varies within (industry,year) cells.

19. **Gentzkow et al. (2011), The Effect of Newspaper Entry and Exit on Electoral Politics.** *Tables 2 and 3.*

In regression Equation (2), the dependent variable is the change in voter turnout in county  $c$  between elections year  $t$  and  $t-1$ , and the explanatory variables are state  $\times$  year effects, and the change in the number of newspapers in county  $c$  between  $t$  and  $t-1$ . This regression corresponds to Regression 2 with controls. The stable groups assumption is satisfied: between each pair of consecutive years, there are some counties where the number of newspapers does not change. This is a sharp design.

20. **Bloom et al. (2012), Americans Do IT Better: US Multinationals and the Productivity Miracle.** *Table 2, Columns 6-8.*

In the regression in, say, Column 6 of Table 2, the dependent variable is the log of output per worker in firm  $i$  in period  $t$ , while the explanatory variables are firms and time fixed effects, the log of the amount of IT capital per employee  $\ln(C/L)$ , the interaction of  $\ln(C/L)$  and an indicator for whether the firm is owned by a US multinational, the interaction of  $\ln(C/L)$  and an indicator for whether the firm is owned by a non-US multinational, and some controls. This regression is similar to Regression 1 with some controls, except that it has three treatment variables. The stable groups assumption is presumably satisfied: between each pair of consecutive years, it is likely that there are some firms where the amount of IT capital per employee  $\ln(C/L)$  does not change. This is a sharp design.

21. **Simcoe (2012), Standard Setting Committees: Consensus Governance for Shared Technology Platforms.** *Table 4, Columns 1-3.*

In regression Equation (5), the dependent variable is a measure of time to consensus for project  $i$  submitted to committee  $j$ , while the explanatory variables are an indicator for projects submitted to the standards track, a measure of distributional conflict, the interaction of the standards track and distributional conflict, and some controls variables. This regression is similar to Regression 1 with some controls, except that it has a measure of distributional conflict instead of committee fixed effects. The stable groups assumption is presumably not satisfied: it is unlikely that there is any committee where the measure of distributional conflict is equal to 0. This is a sharp design.

22. **Moser and Voena (2012), Compulsory Licensing: Evidence from the Trading with the Enemy Act.** *Table 2.*

In the regression equation in the beginning of Section III, the dependent variable is the number of patents by US inventors in patent class  $c$  at period  $t$ , and the explanatory variables are patent class and time fixed effects, the interaction of period  $t$  being after the trading with the enemy act and the number of licensed patents in class  $c$ , and some control variables. This regression corresponds to Regression 1 with some controls. The stable groups assumption is satisfied: there are patent classes where no patent was licensed. This is a sharp design.

23. **Forman et al. (2012), The Internet and Local Wages: A Puzzle.** *Tables 2 and 4.*

In regression Equation (1), the dependent variable is the difference between log wages in 2000 and 1995 in county  $i$ , and the explanatory variables are the proportion of businesses using Internet in county  $i$  in 2000, and control variables. This regression corresponds to Regression 2 with some controls. The stable groups assumption is satisfied: there are counties with no Internet investment in 2000. This a fuzzy design: the treatment of interest is whether a business uses Internet, which varies within (county,year) cells.

24. **Besley and Mueller (2012), Estimating the Peace Dividend: The Impact of Violence on House Prices in Northern Ireland.** *Table 1, Columns 3 and 5-7.*

In regression Equation (1), the dependent variable is the price of houses in region  $r$  at time  $t$ , while the explanatory variables are region and time fixed effects, and the number of people killed because of the civil war in region  $r$  at time  $t-1$ . This regression corresponds to Regression 1. The stable groups assumption is presumably satisfied: between each pair of consecutive years, it is likely that there are some regions where the number of people killed because of the civil war does not change. This is a sharp design.

25. **Dafny et al. (2012), Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry.** *Table 3.*

In regression Equation (3), the dependent variable is the the concentration of the hospital industry in market  $m$  and year  $t$ , and explanatory variables are time fixed effects, market

fixed effects, and the change in concentration in market  $m$  induced by a merger interacted with an indicator for  $t$  being after the merger. This regression corresponds to Regression 1. The stable groups assumption is satisfied: there are many markets where the merger did not change concentration. This is a sharp design.

26. **Hornbeck (2012), The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe.** *Table 2.* In regression Equation (1), the dependent variable is, say, the change in log land value in county  $c$  between period  $t$  and 1930, and the explanatory variables are state  $\times$  year fixed effects, the share of county  $c$  in high erosion regions, the share of county  $c$  in medium erosion regions, and some control variables. This regression is similar to Regression 1 with controls, except that it has two treatment variables. The stable groups assumption is satisfied: many counties have 0% of their land situated in medium or high erosion regions. This a fuzzy design: the treatments of interest are whether a piece of land is in high or in medium erosion regions, which varies within (county,year) cells.

27. **Bajari et al. (2012), A Rational Expectations Approach to Hedonic Price Regressions with Time-Varying Unobserved Product Attributes: The Price of Pollution.** *Table 5.*

In, say, the first regression equation in the bottom of page 1915, the dependent variable is the change in the price of house  $j$  between sales 2 and 3, and the explanatory variables are the change in various pollutants in the area around house  $j$  between sales 2 and 3, and some controls. This regression is similar to Regression 2 with controls, except that it has several treatment variables. The stable groups assumption is presumably satisfied: it is likely that for each pair of consecutive sales, there are houses where the level of each pollutant does not change. This is a sharp design.

28. **Dahl and Lochner (2012), The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit.** *Table 3.*

In regression Equation (4), the dependent variable is the change in test scores for child  $i$  between years  $a$  and  $a-1$ , while the explanatory variables are the change in the EITC income of her family and some controls, and the change in the expected EITC income of her family based on her family income in year  $a-1$  is used to instrument for the actual change of her family's EITC income. This regression is a 2SLS version of Regression 2 with controls, except that it does not have years fixed effects. The stable groups assumption is presumably satisfied: it is likely that for each pair of consecutive years, there are children whose family's expected EITC income does not change. This is a sharp design.

29. **Imberman et al. (2012), Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees.** *Tables 3-6.*

In regression Equation (1), the dependent variable is the test score of student  $i$  in school  $j$  in grade  $g$  and year  $t$ , and the explanatory variables are school and grade  $\times$  year fixed effects, the fraction of Katrina evacuee students received by school  $j$  in grade  $g$  and year  $t$ , and some controls. This regression is a three-way fixed effects version of Regression 1. The stable groups assumption is satisfied: there are schools that did not receive any Katrina evacuee. This a fuzzy design: the treatment of interest is the proportion of evacuees in one's class, which varies within (school,grade,year) cells.

30. **Chaney et al. (2012), The Collateral Channel: How Real Estate Shocks Affect Corporate Investment.** *Table 5.*

In regression Equation (1), the dependent variable is the value of investment in firm  $i$  and year  $t$  divided by the lagged book value of properties, plants, and equipments (PPE), and the explanatory variables are firm and time fixed effects and the market value of firm  $i$  in year  $t$  divided by its lagged PPE, and some controls. This regression corresponds to Regression 1, with some controls. The stable groups assumption is presumably satisfied: it is likely that between each pair of consecutive years, there are firms whose market value divided by their lagged PPE does not change. This is a sharp design.

31. **Aaronson et al. (2012), The Spending and Debt Response to Minimum Wage Hikes.** *Tables 1, 2, and 5.*

In regression Equation (1), the outcome variable is, say, income of household  $i$  at period  $t$ , and the explanatory variables are household and time fixed effects, and the minimum wage in the state where household  $i$  lives in period  $t$ . This regression corresponds to Regression 1. The stable groups assumption is satisfied: between each pair of consecutive periods, there are states where the minimum wage does not change. This is a sharp design.

32. **Brambilla et al. (2012), Exports, Export Destinations, and Skills.** *Table 5.*

In the regression in, say, the first column of Table 2, the dependent variable is a measure of skills in the labor force employed by firm  $i$  in industry  $j$  at period  $t$ , and the explanatory variables are firm and industry  $\times$  period fixed effects, the ratio of exports to sales in firm  $i$  at period  $t$ , and some controls. This regression corresponds to Regression 1, with some controls. The stable groups assumption is presumably satisfied: it is likely that between each pair of consecutive periods, there are firms whose ratio of exports to sales does not change. This is a sharp design.

33. **Faye and Niehaus (2012), Political Aid Cycles.** *Table 3, Columns 4 and 5, and Tables 4 and 5.*

In regression Equation (2), the dependent variable is the amount of donations received by receiver  $r$  from donor  $d$  in year  $t$ , and the explanatory variables are donor  $\times$  receiver fixed effects, an indicator for whether there is an election in country  $r$  in year  $t$ , a measure of

alignment between the ruling political parties in countries  $r$  and  $d$  at  $t$ , and the interaction of the election indicator and the measure of alignment. This regression corresponds to Regression 1. The stable groups assumption is presumably not satisfied: it is unlikely that there are donor-receiver pairs that are perfectly unaligned. This is a sharp design.

## 7 Proofs

Theorem S1 relies on the following lemma.

**Lemma S1** *If Assumptions 1 and 3-5 hold, for all  $(g, g', t, t') \in \{1, \dots, G\}^2 \times \{1, \dots, T\}^2$ ,*

$$\begin{aligned} & E(Y_{g,t}|\mathbf{D}) - E(Y_{g,t'}|\mathbf{D}) - (E(Y_{g',t}|\mathbf{D}) - E(Y_{g',t'}|\mathbf{D})) \\ &= D_{g,t}E(\Delta_{g,t}^{TR}|\mathbf{D}) - D_{g,t'}E(\Delta_{g,t'}^{TR}|\mathbf{D}) - (D_{g',t}E(\Delta_{g',t}^{TR}|\mathbf{D}) - D_{g',t'}E(\Delta_{g',t'}^{TR}|\mathbf{D})). \end{aligned}$$

*Proof of Lemma S1*

For all  $(g, t) \in \{1, \dots, G\} \times \{1, \dots, T\}$ ,

$$\begin{aligned} E(Y_{g,t}|\mathbf{D}) &= E\left(\frac{1}{N_{g,t}} \sum_{i=1}^{N_{g,t}} (Y_{i,g,t}(0) + D_{i,g,t}(Y_{i,g,t}(1) - Y_{i,g,t}(0))) \middle| \mathbf{D}\right) \\ &= E(Y_{g,t}(0)|\mathbf{D}) + D_{g,t}E(\Delta_{g,t}^{TR}|\mathbf{D}). \end{aligned}$$

The end of the proof is the same as that of Lemma 1.

## Proof of Theorem S1

The proof of that result is very similar to the proof of Theorem 1.

$$\begin{aligned}
\beta_{fe} &= E \left( \frac{\sum_{g,t} N_{g,t} \varepsilon_{g,t} Y_{g,t}}{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t}} \right) \\
&= E \left( \frac{\sum_{g,t} N_{g,t} \varepsilon_{g,t} E(Y_{g,t} | \mathbf{D})}{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t}} \right) \\
&= E \left( \frac{\sum_{g,t} N_{g,t} \varepsilon_{g,t} (E(Y_{g,t} | \mathbf{D}) - E(Y_{g,1} | \mathbf{D}) - E(Y_{1,t} | \mathbf{D}) + E(Y_{1,1} | \mathbf{D})))}{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t}} \right) \\
&= E \left( \frac{\sum_{g,t} N_{g,t} \varepsilon_{g,t} (D_{g,t} E(\Delta_{g,t}^{TR} | \mathbf{D}) - D_{g,1} E(\Delta_{g,1}^{TR} | \mathbf{D}) - D_{1,t} E(\Delta_{1,t}^{TR} | \mathbf{D}) + D_{1,1} E(\Delta_{1,1}^{TR} | \mathbf{D})))}{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t}} \right) \\
&= E \left( \frac{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t} E(\Delta_{g,t}^{TR} | \mathbf{D})}{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t}} \right) \\
&= E \left( \frac{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t} \Delta_{g,t}^{TR}}{\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t}} \right).
\end{aligned}$$

The second equality follows from the law of iterated expectations. The third and fifth equalities follow from Equations (A2) and (A3). The fourth equality follows from Lemma S1. The last equality follows from the law of iterated expectations.

## Proof of Proposition S1

Assuming that  $N_{g,2}/N_{g,1}$  does not vary across  $g$  ensures that there exists a strictly positive real number  $\phi$  such that  $N_{g,2}/N_{g,1} = \phi$ . Then,

$$\begin{aligned}
\varepsilon_{g,2} &= D_{g,2} - D_{g,\cdot} - D_{\cdot,2} + D_{\cdot,\cdot} \\
&= D_{g,2} - \left( \frac{N_{g,1}}{N_{g,\cdot}} D_{g,1} + \frac{N_{g,2}}{N_{g,\cdot}} D_{g,2} \right) - D_{\cdot,2} + \left( \frac{N_{\cdot,1}}{N} D_{\cdot,1} + \frac{N_{\cdot,2}}{N} D_{\cdot,2} \right) \\
&= D_{g,2} - \left( \frac{1}{1+\phi} D_{g,1} + \frac{\phi}{1+\phi} D_{g,2} \right) - D_{\cdot,2} + \left( \frac{1}{1+\phi} D_{\cdot,1} + \frac{\phi}{1+\phi} D_{\cdot,2} \right) \\
&= \frac{1}{1+\phi} (D_{g,2} - D_{g,1} - D_{\cdot,2} + D_{\cdot,1}), \tag{6}
\end{aligned}$$

where the first and third equalities follow from the fact  $N_{g,2}/N_{g,1}$  does not vary across  $g$ .

Then, the definition of  $w_{g,2}^{TR}$ , Equation (6) and Assumption 1 imply that

$$\begin{aligned} w_{g,2}^{TR} &= \frac{(D_{g,2} - D_{.,2})}{\sum_{g=1}^G \frac{N_{g,2}}{N_1} (D_{g,2} - D_{.,2}) D_{g,2}} \\ &= \frac{(D_{g,2} - D_{.,2})}{\sum_{g=1}^G \frac{N_{g,2}}{N_1} (D_{g,2} - D_{.,2})^2}. \end{aligned}$$

## Proof of Theorem S2

The proof relies on the lemma below, which we start by proving, before proving the theorem.

**Lemma S2** *If Assumptions 1-5 and S2 hold,*

$$\begin{aligned} &E(Y_{g,t}|\mathbf{D}) - E(Y_{g,t-1}|\mathbf{D}) - (E(Y_{g',t}|\mathbf{D}) - E(Y_{g',t-1}|\mathbf{D})) \\ &= (D_{g,t} - D_{g,t-1})E(\Delta_{g,t}|\mathbf{D}) - (D_{g',t} - D_{g',t-1})E(\Delta_{g',t}|\mathbf{D}). \end{aligned}$$

*Proof of Lemma S2*

By Lemma 1 and Assumption S2,

$$\begin{aligned} &E(Y_{g,t}|\mathbf{D}) - E(Y_{g,t-1}|\mathbf{D}) - (E(Y_{g',t}|\mathbf{D}) - E(Y_{g',t-1}|\mathbf{D})) \\ &= D_{g,t}E(\Delta_{g,t}|\mathbf{D}) - D_{g,t-1}E(\Delta_{g,t-1}|\mathbf{D}) - D_{g',t}E(\Delta_{g',t}|\mathbf{D}) + D_{g',t-1}E(\Delta_{g',t-1}|\mathbf{D}) \\ &= (D_{g,t} - D_{g,t-1})E(\Delta_{g,t}|\mathbf{D}) - (D_{g',t} - D_{g',t-1})E(\Delta_{g',t}|\mathbf{D}). \end{aligned}$$

*Proof of the decomposition for the fixed-effect regression*

First, we have

$$\begin{aligned} &\sum_{g,t} N_{g,t} \varepsilon_{g,t} E(Y_{g,t}|\mathbf{D}) \\ &= \sum_{g,t} N_{g,t} \varepsilon_{g,t} [E(Y_{g,t}|\mathbf{D}) - E(Y_{1,t}|\mathbf{D})] \\ &= \sum_g \sum_{t=2}^T \left[ \sum_{t' \geq t} N_{g,t'} \varepsilon_{g,t'} \right] [E(Y_{g,t}|\mathbf{D}) - E(Y_{g,t-1}|\mathbf{D}) - (E(Y_{1,t}|\mathbf{D}) - E(Y_{1,t-1}|\mathbf{D}))] \\ &= \sum_g \sum_{t=2}^T \left[ \sum_{t' \geq t} N_{g,t'} \varepsilon_{g,t'} \right] [(D_{g,t} - D_{g,t-1})E(\Delta_{g,t}|\mathbf{D}) - (D_{1,t} - D_{1,t-1})E(\Delta_{1,t}|\mathbf{D})] \\ &= \sum_{(g,t): D_{g,t} \neq D_{g,t-1}, t \geq 2} \left[ N_{g,t} (D_{g,t} - D_{g,t-1}) \sum_{t' \geq t} \frac{N_{g,t'}}{N_{g,t}} \varepsilon_{g,t'} \right] E(\Delta_{g,t}|\mathbf{D}). \end{aligned} \tag{7}$$

The first equality follows by (A3). The second equality follows from summation by part and (A3). The third equality follows from Lemma S2. The fourth equality stems from the fact that by (A3), the terms with  $g = 1$  vanish.

Similarly,

$$\begin{aligned}
\sum_{g,t} N_{g,t} \varepsilon_{g,t} D_{g,t} &= \sum_g \sum_{t=2}^T \left[ \sum_{t' \geq t} N_{g,t'} \varepsilon_{g,t'} \right] [D_{g,t} - D_{g,t-1}] \\
&= \sum_{(g,t): D_{g,t} \neq D_{g,t-1}, t \geq 2} N_{g,t} (D_{g,t} - D_{g,t-1}) \sum_{t' \geq t} \frac{N_{g,t'}}{N_{g,t}} \varepsilon_{g,t'}.
\end{aligned} \tag{8}$$

The result follows by combining (A1), (7), (8), and the law of iterated expectations.

*Proof of the decomposition for the first-difference regression*

First, we have

$$\begin{aligned}
&\sum_{(g,t): t \geq 2} N_{g,t} \varepsilon_{fd,g,t} (E(Y_{g,t} | \mathbf{D}) - E(Y_{g,t-1} | \mathbf{D})) \\
&= \sum_{(g,t): t \geq 2} N_{g,t} \varepsilon_{fd,g,t} (E(Y_{g,t} | \mathbf{D}) - E(Y_{g,t-1} | \mathbf{D}) - (E(Y_{1,t} | \mathbf{D}) - E(Y_{1,t-1} | \mathbf{D}))) \\
&= \sum_{(g,t): t \geq 2} N_{g,t} \varepsilon_{fd,g,t} [(D_{g,t} - D_{g,t-1}) E(\Delta_{g,t} | \mathbf{D}) - (D_{1,t} - D_{1,t-1}) E(\Delta_{1,t} | \mathbf{D})] \\
&= \sum_{(g,t): t \geq 2} N_{g,t} \varepsilon_{fd,g,t} (D_{g,t} - D_{g,t-1}) E(\Delta_{g,t} | \mathbf{D}).
\end{aligned}$$

The first equality follows from (A11). The second equality follows from Lemma S2. The third equality follows from (A11) again. The result follows by combining (A10) with the last display, and using the law of iterated expectations.

*Proof that  $w_{g,t}^S \geq 0$  under Assumption 6 and if  $N_{g,t}/N_{g,t-1}$  does not depend on  $g$*

Under Assumption 6, one has that  $D_{g,t} = 1\{t \geq a_g\}$ , with  $a_g \in \{1, \dots, T+1\}$ . Therefore, given the form of  $w_{g,t}^S$ , we just have to prove that for all  $g$ ,

$$\sum_{t \geq a_g} N_{g,t} \varepsilon_{g,t} \geq 0. \tag{9}$$

Because  $N_{g,t}/N_{g,t-1}$  does not vary across  $g$  for all  $t \geq 2$ , we have  $N_{g,t} = N_{g,0} \gamma_t$  for some  $\gamma_t \geq 0$ . Moreover,  $\varepsilon_{g,t} = D_{g,t} - D_{g,\cdot} - D_{\cdot,t} + D_{\cdot,\cdot}$ . Let  $\tilde{\gamma}_t = \gamma_t / \sum_{t \geq 0} \gamma_t$ , then  $D_{g,\cdot} = \sum_{t \geq a_g} \tilde{\gamma}_t$ , and  $D_{\cdot,\cdot} = \sum_{t \geq 0} \tilde{\gamma}_t D_{\cdot,t}$ . Hence,

$$\begin{aligned}
\frac{1}{N_{g,0} \sum_t \gamma_t} \sum_{t \geq a_g} N_{g,t} \varepsilon_{g,t} &= D_{g,\cdot} (1 - D_{g,\cdot} + D_{\cdot,\cdot}) - \sum_{t \geq a_g} \tilde{\gamma}_t D_{\cdot,t} \\
&= D_{g,\cdot} \left( 1 - D_{g,\cdot} + \sum_{t < a_g} \tilde{\gamma}_t D_{\cdot,t} \right) - \left( \sum_{t \geq a_g} \tilde{\gamma}_t D_{\cdot,t} \right) (1 - D_{g,\cdot}).
\end{aligned} \tag{10}$$

Now, because  $D_{.,t} \leq 1$ ,

$$\sum_{t \geq a_g} \tilde{\gamma}_t D_{.,t} \leq \sum_{t \geq a_g} \tilde{\gamma}_t = D_{g,.}$$

Hence, in view of (10),

$$\frac{1}{N_{g,0} \sum_t \gamma_t} \sum_{t \geq a_g} N_{g,t} \varepsilon_{g,t} \geq D_{g,.} \sum_{t < a_g} \tilde{\gamma}_t D_{.,t} \geq 0.$$

Therefore, (9) and the result follows.

*Proof that  $w_{fd,g,t}^S \geq 0$*

We just have to focus on the cases where  $D_{g,t} \neq D_{g,t-1}$ . Note that  $\varepsilon_{fd,g,t} = D_{g,t} - D_{g,t-1} - (D_{.,t} - D_{.,t-1})$ . Then, if  $D_{g,t} - D_{g,t-1} = 1$ , the numerator of  $w_{fd,g,t}^S$  has the same sign as  $1 - (D_{.,t} - D_{.,t-1})$ , which is positive. If  $D_{g,t} - D_{g,t-1} = -1$ , the numerator of  $w_{fd,g,t}^S$  has the same sign as  $1 + (D_{.,t} - D_{.,t-1})$ , which is also positive. Because the denominator sums terms that are always positive, it is positive as well. The result follows.

### Proof of Theorem S3

The reasoning is exactly the same as in Theorem 1, except that we rely on Lemma S3 below, instead of Lemma 1. We thus only prove Lemma S3.

**Lemma S3** *If Assumptions 1-5 hold and  $D_{i,g,t} \in \{0, \dots, \bar{d}\}$ .*

$$\begin{aligned} & E(Y_{g,t} | \mathbf{D}) - E(Y_{g,t'} | \mathbf{D}) - (E(Y_{g',t} | \mathbf{D}) - E(Y_{g',t'} | \mathbf{D})) \\ &= D_{g,t} E(\Delta_{g,t}^{ACR} | \mathbf{D}) - D_{g,t'} E(\Delta_{g,t'}^{ACR} | \mathbf{D}) - (D_{g',t} E(\Delta_{g',t}^{ACR} | \mathbf{D}) - D_{g',t'} E(\Delta_{g',t'}^{ACR} | \mathbf{D})). \end{aligned}$$

*Proof of Lemma S3*

Under Assumption 2, we have  $E(Y_{g,t} | \mathbf{D}) = E(Y_{g,t}(0) | \mathbf{D}) + E(Y_{g,t}(D_{g,t}) - Y_{g,t}(0) | \mathbf{D})$ . The result follows by decomposing similarly the three other terms  $E(Y_{g,t'} | \mathbf{D})$ ,  $E(Y_{g',t} | \mathbf{D})$ , and  $E(Y_{g',t'} | \mathbf{D})$ , using Assumptions 3-5, and finally using the definition of  $\Delta_{g,t}^{ACR}$ .

### Proof of Theorem S4

The proof relies on the following lemma, that resembles Lemma 1 and that we do not prove.

**Lemma S4** *If Assumptions 1, 2, and S3-S4 hold, for all  $(g, g', t, t') \in \{1, \dots, G\}^2 \times \{1, \dots, T\}^2$ ,*

$$\begin{aligned} & E(Y_{g,t} - X'_{g,t} \gamma | \mathbf{X}, \mathbf{D}) - E(Y_{g,t'} - X'_{g,t'} \gamma | \mathbf{X}, \mathbf{D}) \\ & - (E(Y_{g',t} - X'_{g',t} \gamma | \mathbf{X}, \mathbf{D}) - E(Y_{g',t'} - X'_{g',t'} \gamma | \mathbf{X}, \mathbf{D})) \\ &= D_{g,t} E(\Delta_{g,t} | \mathbf{X}, \mathbf{D}) - D_{g,t'} E(\Delta_{g,t'} | \mathbf{X}, \mathbf{D}) - (D_{g',t} E(\Delta_{g',t} | \mathbf{X}, \mathbf{D}) - D_{g',t'} E(\Delta_{g',t'} | \mathbf{X}, \mathbf{D})). \end{aligned}$$

It follows from the Frisch-Waugh theorem and the definition of  $\varepsilon_{g,t}^X$  that

$$E\left(\widehat{\beta}_{fe}\middle|\mathbf{X}, \mathbf{D}\right) = \frac{\sum_{g,t} N_{g,t} \varepsilon_{g,t}^X E(Y_{g,t}|\mathbf{X}, \mathbf{D})}{\sum_{g,t} N_{g,t} \varepsilon_{g,t}^X D_{g,t}}. \quad (11)$$

Now, by definition of  $\varepsilon_{g,t}^X$  again,

$$\sum_{t=1}^T N_{g,t} \varepsilon_{g,t}^X = 0 \text{ for all } g \in \{1, \dots, G\}, \quad (12)$$

$$\sum_{g=1}^G N_{g,t} \varepsilon_{g,t}^X = 0 \text{ for all } t \in \{1, \dots, T\}, \quad (13)$$

$$\sum_{g,t} N_{g,t} \varepsilon_{g,t}^X X_{g,t} = 0. \quad (14)$$

Then,

$$\begin{aligned} & \sum_{g,t} N_{g,t} \varepsilon_{g,t}^X E(Y_{g,t}|\mathbf{X}, \mathbf{D}) \\ &= \sum_{g,t} N_{g,t} \varepsilon_{g,t}^X E(Y_{g,t} - X'_{g,t} \lambda|\mathbf{X}, \mathbf{D}) \\ &= \sum_{g,t} N_{g,t} \varepsilon_{g,t}^X (E(Y_{g,t} - X'_{g,t} \lambda|\mathbf{X}, \mathbf{D}) - E(Y_{g,1} - X'_{g,1} \lambda|\mathbf{X}, \mathbf{D}) \\ & \quad - E(Y_{1,t} - X'_{1,t} \lambda|\mathbf{X}, \mathbf{D}) + E(Y_{1,1} - X'_{1,1} \lambda|\mathbf{X}, \mathbf{D})) \end{aligned}$$

The first equality follows from (14). The second follows from Equations (12) and (13). Hence,

$$\begin{aligned} \sum_{g,t} N_{g,t} \varepsilon_{g,t}^X E(Y_{g,t}|\mathbf{X}, \mathbf{D}) &= \sum_{g,t} N_{g,t} \varepsilon_{g,t}^X (D_{g,t} E(\Delta_{g,t}|\mathbf{X}, \mathbf{D}) - D_{g,1} E(\Delta_{g,1}|\mathbf{X}, \mathbf{D}) \\ & \quad - D_{1,t} E(\Delta_{1,t}|\mathbf{X}, \mathbf{D}) + D_{1,1} E(\Delta_{1,1}|\mathbf{X}, \mathbf{D})) \\ &= \sum_{g,t} N_{g,t} \varepsilon_{g,t}^X D_{g,t} E(\Delta_{g,t}|\mathbf{X}, \mathbf{D}) \\ &= \sum_{(g,t): D_{g,t}=1} N_{g,t} \varepsilon_{g,t}^X E(\Delta_{g,t}|\mathbf{X}, \mathbf{D}). \end{aligned} \quad (15)$$

The first equality follows from Lemma 4. The second follows from Equations (12) and (13). The third follows from Assumption 2. Finally, Assumption 2 implies that

$$\sum_{g,t} N_{g,t} \varepsilon_{g,t}^X D_{g,t} = \sum_{(g,t): D_{g,t}=1} N_{g,t} \varepsilon_{g,t}^X. \quad (16)$$

Combining (11), (15), (16), and the law of iterated expectations yields the result.

## Proof of Theorem S5

Reasoning as in the proof of Theorem 3, we get that for all  $t \geq 2$ ,

$$\begin{aligned} N_{d,d',t} E(\text{DID}_{d,d',t} | \mathbf{D}) &= \sum_{g: D_{g,t}=d', D_{g,t-1}=d} N_{g,t} E(Y_{g,t}(\max(d, d')) - Y_{g,t}(\min(d, d')) | \mathbf{D}) \\ &= \sum_{(i,g): D_{g,t}=d', D_{g,t-1}=d} E(Y_{i,g,t}(\max(D_{g,t}, D_{g,t-1})) - Y_{i,g,t}(\min(D_{g,t}, D_{g,t-1})) | \mathbf{D}). \end{aligned}$$

For all  $(g, t)$ , there exists one  $(d, d') \in \mathcal{D}^2$  such that  $D_{g,t} = d'$  and  $D_{g,t-1} = d$ . Hence,

$$\begin{aligned} &\sum_{t=2}^T \sum_{(d,d') \in \mathcal{D}^2: d \neq d'} N_{d,d',t} E(\text{DID}_{d,d',t} | \mathbf{D}) \\ &= \sum_{t=2}^T \sum_{(d,d') \in \mathcal{D}^2} N_{d,d',t} E(\text{DID}_{d,d',t} | \mathbf{D}) \\ &= \sum_{(i,g,t): t \geq 2} E(Y_{i,g,t}(\max(D_{g,t}, D_{g,t-1})) - Y_{i,g,t}(\min(D_{g,t}, D_{g,t-1})) | \mathbf{D}). \end{aligned}$$

The result follows by definition of  $\text{DID}_M$  and  $\delta^{SCR}$ , and the law of iterated expectations.

## Proof of Theorem S6

### 1. Asymptotic normality

Let us define  $P^S = N_S/G$  and

$$T^S = \frac{1}{G} \sum_{(g,t): t \geq 2} N_{g,t} \mathbf{1}\{D_{g,t} \neq D_{g,t-1}\} [Y_{g,t}(1) - Y_{g,t}(0)].$$

We prove the result in two steps. First, we prove that  $\sqrt{G}(\text{DID}_M - E(T^S)/E(P^S))$  is asymptotically normal. Second, we show that the difference between  $E(T^S)/E(P^S)$  and  $\delta^S$  is asymptotically negligible.

### Convergence of $\sqrt{G}(\text{DID}_M - E(T^S)/E(P^S))$

By Assumption S9,

$$\sup_{g,t} N_{g,t}^4 E[\mathbf{1}\{D_{g,t} = d, D_{g,t-1} = d'\} (Y_{g,t} - Y_{g,t-1})^4] < +\infty.$$

Thus, Lyapunov's condition for the central limit theorem holds, and because  $\Sigma = \lim_g G \times V(\mathbf{U})$  exists,

$$\sqrt{G}(\mathbf{U} - E[\mathbf{U}]) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

By Assumption S9,  $P_{d,d',t}^\infty = \lim_{G \rightarrow \infty} E(P_{d,d',t})$  exists. Then for  $d \in \{0, 1\}$ , define  $\mathcal{T}_d = \{t : P_{1-d,d,t}^\infty > 0\}$  and the event  $\mathcal{D} = \{N_S > 0\} \cap \mathcal{D}_0 \cap \mathcal{D}_1$ , with

$$\mathcal{D}_d = \{t \in \mathcal{T}_d \iff \min(N_{d,d,t}, N_{1-d,d,t}) > 0\}.$$

By the law of large numbers and Assumption S10,  $N_S > 0$  with probability approaching one. Next, fix  $d \in \{0, 1\}$ . If  $t \in \mathcal{T}_d$ , then, by the law of large numbers,  $N_{1-d,d,t} > 0$  with probability approaching one. Moreover, for such a  $t$ , there exists  $g$  such that  $\Pr(D_{g,t} = 1-d, D_{g,t-1} = d) > 0$  and thus, by Assumption S10,  $P_{d,d,t}^\infty > 0$ . Then, by the law of large numbers again,  $N_{d,d,t} > 0$  with probability approaching one. Conversely, if  $\min(N_{d,d,t}, N_{1-d,d,t}) > 0$ , then there exists  $g$  such that  $\Pr(D_{g,t} = 1-d, D_{g,t-1} = d) > 0$ . Hence, by Assumption S10 again,  $P_{1-d,d,t}^\infty > 0$ . This shows that  $\mathcal{D}_d$ , and thus  $\mathcal{D}$ , holds with probability approaching one.

Now, by definition of DID<sub>M</sub>, under  $\mathcal{D}$  we have DID<sub>M</sub> =  $f(\mathbf{U})$ , with

$$f(p_{0,0,1}, q_{0,0,1}, \dots, p_{1,1,T}, q_{1,1,T}) = \frac{\sum_{d=0}^1 (-1)^d \sum_{t \in \mathcal{T}_d} q_{1-d,d,t} - (p_{1-d,d,t}/p_{d,d,t})q_{d,d,t}}{\sum_{t=1}^T p_{1,0,t} + p_{0,1,t}}, \quad (17)$$

for all  $(p_{d,d',t})_{(d,d',t)}$  such that all denominators are strictly positive. By Assumption S9 again,  $E[\mathbf{U}]$  converges to  $\mathbf{U}^\infty$ . Furthermore,  $f$  is continuously differentiable in a neighborhood of  $\mathbf{U}^\infty$ . Thus, by the uniform delta method (see, e.g. van der Vaart, 2000, Theorem 3.8),

$$\sqrt{G}(\text{DID}_M - f(E[\mathbf{U}])) = J_f(\mathbf{U}^\infty) \times \sqrt{G}(\mathbf{U} - E[\mathbf{U}]) + o_P(1). \quad (18)$$

Finally,  $f(E(\mathbf{U})) = f_N/f_D$ , with

$$f_N = \frac{1}{G} \sum_{d=0}^1 (-1)^d \sum_{t \in \mathcal{T}_d} \sum_g N_{g,t} \left[ E(\mathbb{1}\{D_{g,t} = 1-d, D_{g,t-1} = d\}(Y_{g,t} - Y_{g,t-1})) \right. \\ \left. - \frac{E(P_{1-d,d,t})}{E(P_{d,d,t})} E(\mathbb{1}\{D_{g,t} = D_{g,t-1} = d\}(Y_{g,t} - Y_{g,t-1})) \right] \\ f_D = \sum_{t \geq 2} E[P_{1,0,t} + P_{0,1,t}].$$

Reasoning as in the proof of Theorem 3 (see in particular Equations (A17)-(A18)) and noting that if  $t \notin \mathcal{T}_d$ , then, by Assumption S10  $\Pr(D_{g,t} = 1-d, D_{g,t-1} = d) = 0$ , we get  $f_N = E(T^S)$ . Moreover,  $f_D = E(P^S)$ . Hence,

$$\sqrt{G} \left( \text{DID}_M - \frac{E(T^S)}{E(P^S)} \right) \xrightarrow{d} \mathcal{N}(0, \sigma^2), \quad \text{with } \sigma^2 = J_f(\mathbf{U}^\infty) \Sigma J_f(\mathbf{U}^\infty)'. \quad (19)$$

**Convergence to 0 of  $\sqrt{G} (E(T^S)/E(P^S) - \delta^S)$**

Let us define  $\mathbf{D}_G = (D_{g,t})_{(g,t):g \leq G, t=1 \dots T}$ ,  $\tilde{T} = E[T^S | \mathbf{D}_G]$  and  $I = \mathbb{1}\{|P^S - E(P^S)| < \varepsilon_G\}$ ,

where  $\varepsilon_G > 0$  will be specified below. By Assumption S10,  $\lim_{G \rightarrow \infty} E(P^S) > 0$ . Thus, it suffices to prove that  $\sqrt{G} (E(P^S)\delta^S - E(T^S)) \rightarrow 0$ . Because  $\delta^S = E[T^S/P^S]$ , we have

$$\begin{aligned} E(P^S)\delta^S - E(T^S) &= E \left[ \tilde{T}(E(P^S) - P^S)/P^S \right] \\ &= E \left[ \tilde{T}I(E(P^S) - P^S)/P^S \right] + E \left[ \tilde{T}(1 - I)(E(P^S) - P^S)/P^S \right]. \end{aligned} \quad (20)$$

First, consider the second term on the right-hand side. By applying twice the Cauchy-Schwarz inequality, we get

$$\sqrt{G} |E[\tilde{T}(1 - I)(E(P^S) - P^S)/P^S]| \leq \sqrt{GV}(P^S)^{1/2} \left[ E[(\tilde{T}/P^S)^4] \Pr(I = 0) \right]^{1/4}$$

By Assumption S9,  $\sqrt{GV}(P^S)^{1/2}$  converges towards a finite limit. Thus, it suffices to show that the term into brackets tends to 0. To this end, note first that

$$E[T^S | \mathbf{D}_G] = \frac{1}{G} \sum_{(g,t):t \geq 2} N_{g,t} \mathbb{1}\{D_{g,t} \neq D_{g,t-1}\} E[Y_{g,t}(1) - Y_{g,t}(0) | \mathbf{D}_G].$$

Now, let  $A_{g,t} = N_{g,t} \mathbb{1}\{D_{g,t} \neq D_{g,t-1}\}$  and  $B_{g,t} = E[Y_{g,t}(1) - Y_{g,t}(0) | \mathbf{D}_G]$ . By Assumption S9 and Jensen's inequality,  $\sup_{(g,G,t):g \leq G} E[|B_{g,t}|^4] < +\infty$ . Then

$$\begin{aligned} \left( \frac{\tilde{T}}{P^S} \right)^4 &\leq \left( \frac{\sum_{g,t} A_{g,t} |B_{g,t}|}{\sum_{g,t} A_{g,t}} \right)^4 \\ &\leq \left( \max_{g,t} |B_{g,t}|^4 \right) \\ &\leq \sum_{g,t} |B_{g,t}|^4. \end{aligned}$$

Hence,  $E[(\tilde{T}/P^S)^4] \Pr(I = 0) \leq K_1 G \Pr(I = 0)$  for some constant  $K_1 > 0$ . Moreover, by Hoeffding's inequality,

$$\Pr(I = 0) \leq 2 \exp \left( \frac{-2G^2 \varepsilon_G^2}{\sum_g N_{g,\cdot}^2} \right).$$

By Assumption S9, there exists  $c > 0$  such that for all  $G$ ,  $1/G \sum_{g=1}^G N_{g,\cdot}^2 < 2c$ . Then,

$$\frac{-2G^2 \varepsilon_G^2}{\sum_g N_{g,\cdot}^2} \leq -\frac{G \varepsilon_G^2}{c}.$$

Let  $\varepsilon_G = (C \ln(G)/G)^{1/2}$ , for some  $C > c$ . Then, by what precedes,  $G \Pr(I = 0) \rightarrow 0$  and the second term of the right-hand side of (20) tends to zero.

Now, let us move to the first term of the right-hand side of (20). We have

$$\begin{aligned}
& |E[\tilde{T}I(E(P^S) - P^S)/P^S]| \\
&= \left| E[I(\tilde{T} - E(\tilde{T}))(E(P^S) - P^S)/P^S] + E(\tilde{T}) (E(P^S)E(I/P^S) - E(I)) \right| \\
&\leq E(P^S)V(\tilde{T})^{1/2} \left( E[I(1/P^S - 1/E(P^S))^2] \right)^{1/2} \\
&\quad + |E(\tilde{T})|E(P^S)|E(I/P^S) - E(I)/E(P^S)|.
\end{aligned} \tag{21}$$

We now prove that both terms on the right-hand side tend to zero. First, by Taylor expansions of  $x \mapsto 1/x$  around  $E(P^S)$ , there exist  $(P_1^S, P_2^S)$  in the interval between  $P^S$  and  $E(P^S)$  such that

$$\frac{1}{P^S} = \frac{1}{E(P^S)} - \frac{P^S - E(P^S)}{P_1^{S2}}, \tag{22}$$

$$\frac{1}{P^S} = \frac{1}{E(P^S)} - \frac{P^S - E(P^S)}{E(P^S)^2} + \frac{(P^S - E(P^S))^2}{P_2^{S3}}. \tag{23}$$

When  $I = 1$ ,  $|P_1^S - E(P^S)| < \varepsilon_G$  and  $|P_2^S - E(P^S)| < \varepsilon_G$ . Recall also that  $\lim_G E(P^S) > 0$  and  $\varepsilon_G \rightarrow 0$ . Then, in view of (22) and by Assumption S9,

$$E[I(1/P^S - 1/E(P^S))^2] \leq \frac{V(P^S)}{(E(P^S) - \varepsilon_G)^4} \rightarrow 0.$$

Moreover, by definition of  $\tilde{T}$ ,  $GV(\tilde{T}) \leq GV(T^S)$ , and the latter is bounded by Assumption S9. Therefore, the first term of the right-hand side of (21) tends to 0.

Now, multiplying (23) by  $I$  and taking the expectation on both sides, we obtain:

$$\begin{aligned}
\sqrt{G}|E(I/P^S) - E(I)/E(P^S)| &\leq \sqrt{G} \left| \frac{E[I(P^S - E(P^S))]}{E(P^S)^2} \right| + \frac{\sqrt{G}V(P^S)}{(E(P^S) - \varepsilon_G)^3} \\
&\leq \left| \frac{[G \Pr(I = 0)V(P^S)]^{1/2}}{E(P^S)^2} \right| + \frac{\sqrt{G}V(P^S)}{(E(P^S) - \varepsilon_G)^3} \\
&\rightarrow 0.
\end{aligned}$$

## 2. Validity of the confidence intervals

**There exists  $\bar{\sigma}_G^2$  such that  $\sum_g \hat{\psi}_g^2/G - \bar{\sigma}_G^2 \xrightarrow{\mathbb{P}} 0$ , with  $\liminf_G \bar{\sigma}_G^2 \geq \sigma^2$ .**

Let  $Q_{d,d',t}^\infty = \lim_{G \rightarrow \infty} E[Q_{d,d',t}]$  and  $\boldsymbol{\lambda}_g$  be the column vector such that  $\mathbf{U} = \sum_{g=1}^G \boldsymbol{\lambda}_g/G$ . Some tedious algebra show that  $J_f(\mathbf{U}^\infty) \times \boldsymbol{\lambda}_g = \psi_g$ , with

$$\psi_g = \frac{1}{P_S^\infty} \sum_{t>1} N_{g,t} [\mathbf{1}\{D_{g,t} \neq D_{g,t-1}\}(Y_{g,t} - Y_{g,t-1} - f(\mathbf{U}^\infty)) - \psi_{g,t}^B]$$

and

$$\begin{aligned} \psi_{g,t}^B &= \frac{1}{P_{0,0,t}^\infty} \left[ \mathbb{1}\{D_{g,t} > D_{g,t-1}\} Q_{0,0,t}^\infty + P_{1,0,t}^\infty \mathbb{1}\{D_{g,t} = D_{g,t-1} = 0\} \left( Y_{g,t} - Y_{g,t-1} - \frac{Q_{0,0,t}^\infty}{P_{0,0,t}^\infty} \right) \right] \\ &\quad + \frac{1}{P_{1,1,t}^\infty} \left[ \mathbb{1}\{D_{g,t} < D_{g,t-1}\} Q_{1,1,t}^\infty + P_{0,1,t}^\infty \mathbb{1}\{D_{g,t} = D_{g,t-1} = 1\} \left( Y_{g,t} - Y_{g,t-1} - \frac{Q_{1,1,t}^\infty}{P_{1,1,t}^\infty} \right) \right]. \end{aligned}$$

Hence, in view of (18)-(19),  $\sigma^2 = \lim_G \sum_g V(\psi_g)/G$ . Next, we show that  $\sum_g \widehat{\psi}_g^2/G$  is asymptotically larger than  $\sigma^2$ . For that purpose, let  $\bar{\sigma}_G^2 = \sum_g E(\psi_g^2)/G$  and remark that  $\widehat{\psi}_g = J_f(\mathbf{U}) \times \boldsymbol{\lambda}_g$ . Then

$$\frac{1}{G} \sum_g \widehat{\psi}_g^2 - \bar{\sigma}_G^2 = \frac{1}{G} \sum_g \left[ \widehat{\psi}_g^2 - \psi_g^2 \right] + \left[ \frac{1}{G} \sum_g \psi_g^2 - \bar{\sigma}_G^2 \right]. \quad (24)$$

Let  $\lambda_{k,g}$  denote the  $k$ th coordinate of  $\boldsymbol{\lambda}_g$ . Assumption S9 ensures that

$$\sup_{k,g} E[\lambda_{k,g}^4] < +\infty. \quad (25)$$

Thus, we also have  $\sup_g E(\psi_g^2) < +\infty$ . Therefore, by the weak law of large numbers, the second term on the right-hand side of (24) converges to 0. Next,

$$\frac{1}{G} \sum_g \left[ \widehat{\psi}_g^2 - \psi_g^2 \right] = (J_f(\mathbf{U}) - J_f(\mathbf{U}^\infty)) \left[ \frac{1}{G} \sum_g \boldsymbol{\lambda}_g \boldsymbol{\lambda}_g' \right] (J_f(\mathbf{U}) + J_f(\mathbf{U}^\infty))'. \quad (26)$$

By (25) again and the weak law of large numbers,  $\mathbf{U} \xrightarrow{\mathbb{P}} \mathbf{U}^\infty$ . Moreover,  $f$  is continuously differentiable in a vicinity of  $\mathbf{U}^\infty$ , Thus, by the continuous mapping theorem,  $J_f(\mathbf{U}) - J_f(\mathbf{U}^\infty) \xrightarrow{\mathbb{P}} 0$ . By (25) once again and the weak law of large numbers,

$$\frac{1}{G} \sum_g (\boldsymbol{\lambda}_g \boldsymbol{\lambda}_g' - E[\boldsymbol{\lambda}_g \boldsymbol{\lambda}_g']) \xrightarrow{\mathbb{P}} 0.$$

Moreover, by the Cauchy-Schwarz inequality and (25), we have, for all  $k, \ell$ ,

$$\left| \frac{1}{G} \sum_g E[\lambda_{k,g} \lambda_{\ell,g}] \right| \leq \sup_{k,\ell,g} E[|\lambda_{k,g} \lambda_{\ell,g}|] \leq \sup_{k,\ell,g} E[|\lambda_{k,g}|^2] < +\infty.$$

As a result,  $\sum_g \boldsymbol{\lambda}_g \boldsymbol{\lambda}_g'/G = O_P(1)$ . Finally, because  $J_f(\mathbf{U})$  converges in probability,  $J_f(\mathbf{U}) + J_f(\mathbf{U}^\infty) = O_P(1)$ . Thus, in view of (26), the first term on the right-hand side of (24) converges in probability to 0. Hence, we have proven that  $\sum_g \widehat{\psi}_g^2/G - \bar{\sigma}_G^2 \xrightarrow{\mathbb{P}} 0$ . Finally,  $E(\psi_g^2) \geq V(\psi_g)$  and thus  $\bar{\sigma}_G^2 - \sum_g V(\psi_g)/G \geq 0$ . Therefore,  $\liminf_G \bar{\sigma}_G^2 \geq \sigma^2$ .

**CI $_{1-\alpha}(\delta^S)$  is asymptotically conservative.**

By (19), the convergence to 0 of  $\sqrt{G} (E(T^S)/E(P^S) - \delta^S)$ , (18), and since  $\psi_g = J_f(\mathbf{U}^\infty) \times \boldsymbol{\lambda}_g$ ,

$$\sqrt{G} \frac{\text{DID}_M - \delta^S}{\widehat{\sigma}} = \frac{\sigma}{\bar{\sigma}_G} \left[ \frac{\bar{\sigma}_G}{\widehat{\sigma}} \frac{1}{\sqrt{G}} \sum_g \psi_g + o_P(1) \right].$$

Let  $Z_G$  denote the term into brackets. By the first step above and Slutsky's lemma  $Z_G \xrightarrow{d} \mathcal{N}(0, 1)$ . Fix  $\eta > 0$ . Because  $\liminf_G \bar{\sigma}_G^2 \geq \sigma^2$ , there exists  $G_0$  such that for every  $G \geq G_0$ ,

$$\sqrt{G} \left| \frac{\text{DID}_M - \delta^S}{\hat{\sigma}} \right| \leq (1 + \eta) |Z_G|.$$

Then, letting  $\Phi$  denote the cdf of the standard normal distribution, we get

$$\limsup_G \Pr(\delta^S \in \text{CI}_{1-\alpha}) \geq \Phi\left(\frac{z_{1-\alpha/2}}{1+\eta}\right) - \Phi\left(\frac{z_{\alpha/2}}{1+\eta}\right).$$

The result follows by letting  $\eta$  tend to zero.

## References

- Aaronson, D., Agarwal, S. and French, E. (2012), ‘The spending and debt response to minimum wage hikes’, *American Economic Review* **102**(7), 3111–3139.
- Acemoglu, D., Cantoni, D., Johnson, S. and Robinson, J. A. (2011), ‘The consequences of radical reform: The french revolution’, *American Economic Review* **101**(7), 3286–3307.
- Aizer, A. (2010), ‘The gender wage gap and domestic violence’, *American Economic Review* **100**(4), 1847–1859.
- Algan, Y. and Cahuc, P. (2010), ‘Inherited trust and growth’, *American Economic Review* **100**(5), 2060–2092.
- Anderson, S. T. and Sallee, J. M. (2011), ‘Using loopholes to reveal the marginal cost of regulation: The case of fuel-economy standards’, *American Economic Review* **101**(4), 1375–1409.
- Bagwell, K. and Staiger, R. W. (2011), ‘What do trade negotiators negotiate about? empirical evidence from the world trade organization’, *American Economic Review* **101**(4), 1238–1273.
- Bajari, P., Fruehwirth, J. C., Kim, K. I. and Timmins, C. (2012), ‘A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution’, *American Economic Review* **102**(5), 1898–1926.
- Baum-Snow, N. and Lutz, B. F. (2011), ‘School desegregation, school choice, and changes in residential location patterns by race’, *American Economic Review* **101**(7), 3019–3046.
- Besley, T. and Mueller, H. (2012), ‘Estimating the peace dividend: The impact of violence on house prices in northern ireland’, *American Economic Review* **102**(2), 810–833.
- Bloom, N., Sadun, R. and Van Reenen, J. (2012), ‘Americans do it better: Us multinationals and the productivity miracle’, *American Economic Review* **102**(1), 167–201.
- Brambilla, I., Lederman, D. and Porto, G. (2012), ‘Exports, export destinations, and skills’, *American Economic Review* **102**(7), 3406–3438.
- Bustos, P. (2011), ‘Trade liberalization, exports, and technology upgrading: Evidence on the impact of mercosur on argentinian firms’, *American Economic Review* **101**(1), 304–340.
- Chandra, A., Gruber, J. and McKnight, R. (2010), ‘Patient cost-sharing and hospitalization offsets in the elderly’, *American Economic Review* **100**(1), 193–213.
- Chaney, T., Sraer, D. and Thesmar, D. (2012), ‘The collateral channel: How real estate shocks affect corporate investment’, *American Economic Review* **102**(6), 2381–2409.

- Dafny, L., Duggan, M. and Ramanarayanan, S. (2012), ‘Paying a premium on your premium? consolidation in the us health insurance industry’, *American Economic Review* **102**(2), 1161–1185.
- Dahl, G. B. and Lochner, L. (2012), ‘The impact of family income on child achievement: Evidence from the earned income tax credit’, *American Economic Review* **102**(5), 1927–1956.
- de Chaisemartin, C. (2010), A note on instrumented difference in differences. Working Paper.
- de Chaisemartin, C. and D’Haultfœuille, X. (2018), ‘Fuzzy differences-in-differences’, *The Review of Economic Studies* **85**(2), 999–1028.
- Dinkelman, T. (2011), ‘The effects of rural electrification on employment: New evidence from south africa’, *American Economic Review* **101**(7), 3078–3108.
- Duflo, E. (2001), ‘Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment’, *American Economic Review* **91**(4), 795–813.
- Duggan, M. and Morton, F. S. (2010), ‘The effect of medicare part d on pharmaceutical prices and utilization’, *American Economic Review* **100**(1), 590–607.
- Duranton, G. and Turner, M. A. (2011), ‘The fundamental law of road congestion: Evidence from us cities’, *American Economic Review* **101**(6), 2616–2652.
- Ellul, A., Pagano, M. and Panunzi, F. (2010), ‘Inheritance law and investment in family firms’, *American Economic Review* **100**(5), 2414–2450.
- Enikolopov, R., Petrova, M. and Zhuravskaya, E. (2011), ‘Media and political persuasion: Evidence from russia’, *American Economic Review* **101**(7), 3253–3285.
- Fang, H. and Gavazza, A. (2011), ‘Dynamic inefficiencies in an employment-based health insurance system: Theory and evidence’, *American Economic Review* **101**(7), 3047–77.
- Faye, M. and Niehaus, P. (2012), ‘Political aid cycles’, *American Economic Review* **102**(7), 3516–3530.
- Forman, C., Goldfarb, A. and Greenstein, S. (2012), ‘The internet and local wages: A puzzle’, *American Economic Review* **102**(1), 556–575.
- Gentzkow, M., Shapiro, J. M. and Sinkinson, M. (2011), ‘The effect of newspaper entry and exit on electoral politics’, *American Economic Review* **101**(7), 2980–3018.
- Hornbeck, R. (2012), ‘The enduring impact of the american dust bowl: Short-and long-run adjustments to environmental catastrophe’, *American Economic Review* **102**(4), 1477–1507.

- Hotz, V. J. and Xiao, M. (2011), ‘The impact of regulations on the supply and quality of care in child care markets’, *American Economic Review* **101**(5), 1775–1805.
- Hudson, S., Hull, P. and Liebersohn, C. (2015), Interpreting instrumented difference-in-differences, Technical report, Working Paper (available upon request).
- Imbens, G. W. and Angrist, J. D. (1994), ‘Identification and estimation of local average treatment effects’, *Econometrica* **62**(2), 467–475.
- Imberman, S. A., Kugler, A. D. and Sacerdote, B. I. (2012), ‘Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees’, *American Economic Review* **102**(5), 2048–2082.
- Liu, R. Y. and Singh, K. (1995), ‘Using iid bootstrap inference for general non-iid models’, *Journal of statistical planning and inference* **43**(1-2), 67–75.
- Mian, A. and Sufi, A. (2011), ‘House prices, home equity-based borrowing, and the us household leverage crisis’, *American Economic Review* **101**(5), 2132–2156.
- Moser, P. and Voena, A. (2012), ‘Compulsory licensing: Evidence from the trading with the enemy act’, *American Economic Review* **102**(1), 396–427.
- Simcoe, T. (2012), ‘Standard setting committees: Consensus governance for shared technology platforms’, *American Economic Review* **102**(1), 305–336.
- van der Vaart, A. (2000), *Asymptotics Statistics*, Cambridge University Press.
- Wang, S.-Y. (2011), ‘State misallocation and housing prices: Theory and evidence from china’, *American Economic Review* **101**(5), 2081–2107.
- Zhang, X. M. and Zhu, F. (2011), ‘Group size and incentives to contribute: A natural experiment at chinese wikipedia’, *American Economic Review* **101**(4), 1601–1605.