Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends

Jonathan Roth[*]

April 20, 2021

Abstract

This paper discusses two important limitations of the common practice of testing for pre-existing differences in trends ("pre-trends") when using difference-in-differences and related methods. First, conventional pre-trends tests may have low power. Second, conditioning the analysis on the result of a pre-test can distort estimation and inference, potentially exacerbating bias and under-coverage. I analyze these issues both in theory and in simulations calibrated to a survey of recent papers in leading economics journals, which suggest that these limitations are important in practice. I conclude with practical recommendations for mitigating these issues.

[*] Microsoft. Jonathan.Roth@microsoft.com. I am grateful to Isaiah Andrews, Kirill Borusyak, Gary Chamberlain, Peter Ganong, Ed Glaeser, Nathan Hendren, Peter Hull, Ariella Kahn-Lang, Max Kasy, Larry Katz, Sendhil Mullainathan, Andreas Pick, Ashesh Rambachan, Pedro Sant'Anna, Jesse Shapiro, Jann Spiess, Jim Stock, Sophie Sun, and Elie Tamer for helpful comments, as well as to seminar participants at Harvard University and Erasmus University. This material is based upon work supported by the NSF Graduate Research Fellowship under Grant DGE1144152. An earlier version of this paper circulated under the title "Should We Condition on the Test for Pre-trends in Difference-in-Difference Designs?" in April 2018.

1 Introduction

When using difference-in-differences and related methods, researchers commonly test for pre-treatment differences in trends ("pre-trends") as a way of assessing the plausibility of the parallel trends assumption. These tests are remarkably common: based on my review, 70 recent papers in the journals of the American Economic Association have employed an "event-study plot" to visually test for pre-trends.

This paper highlights two limitations with the practice of pre-testing for pre-trends. First, conventional pre-tests may have low power, meaning that pre-existing trends that produce meaningful bias in the treatment effects estimates may not be detected with substantial probability. Second, conditioning the analysis on the result of a pre-trends test induces distortions to estimation and inference from pre-testing. In other words, the draws of the data that survive a pre-test are a selected sample from the true data-generating process. These distortions can exacerbate bias and under-coverage stemming from an undetected violation of parallel trends.

I begin in Section 2 by illustrating the practical importance of these issues in data-generating processes (DGPs) calibrated to a systematic survey of recent papers in three leading economics journals. I evaluate the power of pre-trends tests by calculating the linear violation of parallel trends that conventional pre-trends tests would detect 50 (or 80) percent of the time. I find that such violations of parallel trends can produce large biases and lead confidence intervals (CIs) to substantially undercover. In the most extreme case, the bias from a trend detected only half the time is larger than the estimated treatment effect and a nominal 95% CI contains the true parameter only 24% of the time. I also analyze the distortions from pre-testing by comparing the bias of estimates in draws of the data where no significant pre-trend is detected to the unconditional bias under the same DGP. The bias conditional on passing the pre-test is larger than the unconditional bias in most specifications – and can be over twice as large – indicating important additional distortions from pre-testing.

In Section 3, I provide a formal theoretical treatment of the distribution of event-study estimates after surviving a pre-test for pre-trends. I derive the bias and variance of conventional estimates *conditional* on surviving the test for pre-existing trends. In general, the bias after surviving a pre-test can be larger or smaller than the unconditional bias. I prove, however, that the bias after pre-testing is necessarily larger in settings with homoskedastic errors and monotone differences in trends, indicating that pre-testing can exacerbate bias in non-pathological cases. I also show under quite general conditions that the variance of conventional estimates is lower conditional on passing the pre-test. As a result, traditional

CIs will tend to over-cover conditional on passing the pre-test when bias is small, but will generally under-cover as the bias grows larger. Finally, a stylized model of the publication process illustrates that requiring an insignificant pre-trend to publish has an ambiguous effect on the bias and coverage in published work, with tradeoffs between the power of the pre-test to prevent biased estimates from being published and the distortions from pre-testing.

I conclude with practical recommendations for applied researchers in Section 4. I describe simple diagnostics that researchers can conduct to evaluate when the limitations of pre-trends testing are likely to be severe, and provide software for their implementation. I also briefly highlight alternative approaches that avoid the pre-test altogether by exploiting economic knowledge about how parallel trends may be violated.

Related Literature. This paper relates to a large literature in econometrics and statistics showing that problems can arise in a variety of contexts if researchers do not account for a pre-testing or model selection step. Concerns about pre-testing date at least to the critiques of Tinbergen (1939) by Keynes (1939) and Friedman (1940). More recent work has examined, for instance, the implications of pre-testing for weak identification (Andrews, 2018), choosing between OLS and IV specifications on the basis of a pre-test (Guggenberger, 2010), using data-driven tuning parameters (Armstrong and Kolesár, 2018), and model selection in high-dimensional settings (Belloni et al., 2014; Farrell, 2015; Belloni et al., 2017); see also, Giles and Giles (1993), Leeb and Pötscher (2005), Lee et al. (2016), and references therein. I show theoretically and empirically that similar issues arise with the common practice of testing for pre-trends in difference-in-differences and related research designs.

This paper also contributes to a large body of work on the econometrics of difference-in-differences and related research designs in particular. A topic of substantial recent interest has been the failure of standard two-way fixed effects models to recover a sensible causal estimand in settings with staggered treatment timing and heterogenous treatment effects, even under a suitable parallel trends assumption (Borusyak and Jaravel, 2016; Sun and Abraham, 2020; Athey and Imbens, 2018; de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2020; Callaway and Sant'Anna, 2020). This paper highlights a conceptually distinct issue: even if we were willing to rule out treatment effect heterogeneity (or use a method robust to it), conventional pre-tests may do a poor job detecting violations of the relevant parallel trends assumption. See Remark 1 for further connection to this literature.

Most closely related to the current paper, recent papers by Freyaldenhoven et al. (2019), Kahn-Lang and Lang (2018), and Bilinski and Hatfield (2020) have warned that traditional pre-tests may have low power to detect meaningful violations of parallel trends. I contribute to this literature by evaluating the power of pre-trends tests in a systematic review of recent papers, and theoretically and empirically analyzing the distortions from pre-testing.[1]

Lastly, this paper relates to the literature on selective publication of scientific results (Rothstein et al. (2005), Christensen and Miguel (2016), and Andrews and Kasy (2019) provide reviews). A particularly relevant paper on selective publication is Snyder and Zhuo (2018), who provide empirical evidence that papers with significant placebo coefficients – which they refer to as "sniff tests" – are less likely to be published. I study tests for pre-trends, a common form of sniff test, and provide theoretical and empirical results on the limitations of these tests in reducing bias and coverage issues.

2 Survey of Recent Papers

I first illustrate the issues with pre-trends testing in data-generating processes calibrated to a systematic review of recent papers published in leading economics journals.

2.1 Selecting the sample of papers

I searched on Google Scholar for occurrences of the phrase "event study" in papers published in the *American Economic Review*, *AEJ: Applied Economics*, and *AEJ: Economic Policy* between 2014 and June 2018.[2] I chose the phrase "event study" since researchers often evaluate pre-trends in an "event study plot."

The search returned 70 total papers that include a figure that the authors describe as an event-study plot. Among these, 27 had available replication data[3], and 15 also reported standard errors for the causal effects estimates (relative to a baseline pre-treatment period). I further require that the authors attribute a causal interpretation to their estimates so that I can benchmark the magnitude of biases from differential trends relative to the estimated causal effects. This yields a final sample of 12 papers. Some of these papers present multiple event-study plots, many of which show robustness checks or heterogeneity analyses. I therefore focus on the first plot presented in the paper that meet the criteria above, which I view as a reasonable proxy for the paper's main specification.

2.2 What pre-tests are researchers using?

The most commonly mentioned criterion for evaluating pre-trends is that none of the pre-period coefficients is individually statistically significant – e.g. "the estimated coefficients of

[1] Relatedly, Daw and Hatfield (2018) and Chabé-Ferret (2015) illustrate that selecting a control group on the basis of pre-period outcomes can induce bias in difference-in-differences.

[2] I include papers that were forthcoming as of June 2018 if data was available on the AEA website.

[3] I omit one paper in which the replication code produced different results from the published paper.

the leads of treatments, i.e. δ_k for all $k \leq -2$, are statistically indifferent from zero" (He and Wang, 2017). However, many papers do not specify the exact criteria that they are using to evaluate pre-trends, and several appeal to a visual inspection of the event-plot without stating a formal criterion. Further, Table 1 makes clear that a statistically significant pre-period coefficient does not necessarily preclude publication: there is at least one statistically significant pre-period coefficient in three of the 12 papers in my final sample, and in two papers the pre-period coefficients are also jointly significant.[4] Although this evidence suggests that not all papers use the individual significance of pre-treatment coefficients as their pre-testing criterion, I nevertheless focus my analysis on this criterion given its prominence in the discussion in applied work.

2.3 Evaluating power and pre-test bias in practice

I now evaluate the power of conventional pre-tests and the distortions from pre-testing in data-generating processes calibrated to my survey of recent papers.

Data-generating processes. All of the papers in the survey plot a vector of coefficients $\hat{\beta}$, which has subvectors $\hat{\beta}_{pre} \in \mathbb{R}^K$ and $\hat{\beta}_{post} \in \mathbb{R}^M$ corresponding with the periods before and after a treatment occurs. In the simulations below, I consider calibrated data-generating processes (DGPs) in which

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \Sigma\right), \tag{1}$$

where the mean β satisfies the causal decomposition

$$\beta = \underbrace{\begin{pmatrix} \delta_{pre} \\ \delta_{post} \end{pmatrix}}_{\delta} + \underbrace{\begin{pmatrix} 0 \\ \tau_{post} \end{pmatrix}}_{\tau}, \tag{2}$$

where τ is a vector of causal effects assumed to be zero in the pre-treatment period, and δ is the bias from a difference in trends. All of the papers report standard errors based on the asymptotic normal approximation (1). I impose that this normal approximation holds exactly in finite-sample so that any biases or coverage issues are the results of issues with violations of parallel trends and/or pre-testing rather than the asymptotic distribution providing a poor approximation in finite sample.

[4] In none of the papers is the slope of the best-fit line through the pre-period coefficients significant at the 5% level. However, no paper mentions this as a criterion of interest, and one case falls just short of significance with a t-statistic of 1.95.

Calibrating the model. For each paper in my survey, I calibrate the finite-sample normal model (1) so that the number of pre-treatment and post-treatment periods matches that in the original paper. I set Σ to be the estimated variance-covariance matrix from the specification in the original paper, using whatever clustering method was specified by the authors. I set τ_{post} equal to the estimated coefficients $\hat{\beta}_{post}$, although this choice has no impact on the results.[5] The bias from the difference in trends δ is calibrated based on the power calculations described below.

Power calculations. For each study in my sample, I evaluate the power of common pre-trends tests to detect linear violations of parallel trends. In light of the emphasis in published work on the individual statistical significance of the pre-period coefficients, I base my calculations on pre-tests that check this criterion for all pre-treatment coefficients (using 95% CIs). Specifically, I consider linear violations of parallel trends with a slope of γ, so that the element of δ corresponding with relative time t is $\delta_t = \gamma \cdot t$. I then compute the value of γ for which the probability of passing the pre-test is 50 or 80 percent.[6] I choose 80 percent since this is often used as a benchmark for the minimum detectable effect in power analyses (Cohen, 1988). I refer to the resulting values, $\gamma_{0.5}$ and $\gamma_{0.8}$, as the slopes against which pre-tests have 50 or 80 percent power.[7]

Target Parameter and Estimator. For simplicity, I focus on estimation of a scalar estimand of the form $\tau_* = l'\tau_{post}$ ($l \in \mathbb{R}^M$). Researchers are often interested in the average treatment effect across all post-treatment periods, and so in the main text I focus on estimation of $\tau_* = \frac{1}{M}(\tau_1 + ... + \tau_M) =: \bar{\tau}$. I also consider estimation of the effect for the first period after treatment, $\tau_* = \tau_1$. I focus on the natural plug-in estimate of τ_* under the assumption that $\delta_{post} = 0$ (parallel trends), i.e. $\hat{\tau} = l'\hat{\beta}_{post}$, and the associated CIs $CI_{\tau_*} = \hat{\tau} \pm 1.96\sigma_{\hat{\tau}}$, where $\sigma^2_{\hat{\tau}} = l'\Sigma l$.

Bias and size calculations. I evaluate the performance of these estimators and CIs under data-generating processes with linear violations of parallel trends with slopes $\gamma_{0.5}$ or $\gamma_{0.8}$.

[5] Specifically, the distribution of $\hat{\beta}_{post}$ conditional on a pre-test of $\hat{\beta}_{pre}$ is equivariant with respect to τ_{post}, and thus has no impact on bias or coverage for τ_{post}.

[6] Formally, this is the probability that $\hat{\beta}_{pre} \in B_{NIS}(\Sigma)$, where $B_{NIS}(\Sigma) = \{\beta \in \mathbb{R}^K : |\beta_t| \leq 1.96\sigma_t, \text{ for all t}\}$, where σ_t is the standard error of $\hat{\beta}_{pre,t}$.

[7] The power of the pre-test under a slope γ could easily be calculated via simulation. However, under the normality assumption, these probabilities can actually be calculated analytically using results from Cartinhour (1990) and Manjunath and Wilhelm (2012), which I implement using the R package `tmvtnorm`. A similar approach is applied for the bias and coverage calculations described below. I have verified that simulations yield similar results to the analytical approach.

Specifically, I calculate the unconditional bias $\mathbb{E}\left[\hat{\tau} - \tau_*\right]$, and the bias conditional on passing the pre-test $\mathbb{E}\left[\hat{\tau} - \tau_* \mid \hat{\beta}_{pre} \in B_{NIS}(\Sigma)\right]$, where $B_{NIS}(\Sigma)$ denotes the set of realizations for $\hat{\beta}_{pre}$ such that there is no individually significant coefficient at the 95% level. Analogously, I calculate the size (i.e. null rejection probability) of $CI_{\tau *}$ both unconditionally and conditionally, $\mathbb{P}\left(\tau_* \notin CI_{\tau_*}\right)$ and $\mathbb{P}\left(\tau_* \notin CI_{\tau_*} \mid \beta_{pre} \in B_{NIS}(\Sigma)\right)$.

Results. My results indicate that pre-trends tests often have low power against violations of parallel trends that would produce meaningful bias in the treatment effects estimates. The green triangles in Figure 1 show the bias for the average effect ($\bar{\tau}$) from a linear difference in trends which would be detected 80% of the time ($\gamma_{0.8}$). These biases are benchmarked relative to the magnitude of the treatment effect estimate in the original paper (plotted in blue). The bias from such a trend is often of a magnitude comparable to, and in some cases larger than, the estimated treatment effect! As a result of these biases, traditional CIs exhibit substantial undercoverage under these violations of parallel trends, as shown in Table 2. Although the true parameter should nominally fall outside a 95% confidence interval no more than 5% of the time, in several specifications this occurs over 50% of the time. Results for the first period after treatment (τ_1) and using a 50% power threshold ($\gamma_{0.5}$) show qualitatively similar patterns, although somewhat less extreme, and are presented in Appendix D.

I also find substantial distortions from pre-testing. The red squares in Figure 1 show the bias for $\bar{\tau}$ conditional on surviving the pre-test. As can be seen, the conditional bias can be substantially different from, and in most cases worse than, the unconditional bias. Table 3 summarizes the additional bias from pre-testing as a fraction of the unconditional bias: for the trend against which pre-tests have 50 percent power, the pre-test bias can be as much as 103 percent of the unconditional bias for the first period after treatment, and as much as 48 percent for the average post-treatment effect.[8] Moreover, the pre-test bias and the bias from trend go in the same direction in all but two of the studies in the sample when the estimand is $\bar{\tau}$, and all but three of the studies when it is τ_1. Thus, in most cases the bias from pre-testing exacerbates the bias from the underlying trend. Similarly, Table 2 shows that the null rejection rates for 95% CIs conditional on passing the pre-test can differ substantially from the unconditional null rejection rates, and are worse in many cases.

[8] We expect the bias from pre-testing to be a larger fraction of the unconditional bias for periods closer to treatment, since the unconditional bias from the differential trend grows linearly in the number of periods after treatment, whereas the pre-test bias need not grow over time (whether it does depends on the covariance between the pre-period and post-period coefficients).

2.4 Caveats and Discussion

An important caveat to these results is that by construction my sample only includes papers that made it through the publication process at leading economics journals and reported an event-study plot in the published manuscript. To the extent that papers with insignificant pre-trends are more likely to be published, or that analyses with significant pre-trends are not reported in the final manscript, the sample may be biased towards papers where the power of pre-tests is low.

A second important caveat is that these results only directly provide evidence about the power of pre-trends tests against *linear* violations of parallel trends. The fact that researchers worried about differential trends often include parametric linear controls (e.g., Wolfers (2006); Dobkin et al. (2018); Goodman-Bacon (2018)) indicates that authors perceive linear violations of parallel trends to be relevant in many cases. Nevertheless, one may be worried about non-linear violations of parallel trends as well.[9] Although these results do not directly provide evidence on non-linear differences in trends, it is worth noting that they do provide a lower bound on the worst-case power over any set of possible violations of parallel trends that includes linear violations, e.g., sets of "smooth" violations with bounded second derivative (Rambachan and Roth, 2020). Heuristically, we expect the power of the pre-test against differential trends that produce a given bias to be even worse if the difference in trends can be approximately linear in the pre-treatment period and then become steeper post-treatment. One may also worry that the treatment and control group are subject to common stochastic shocks that lead to a violation of parallel trends. In Appendix C, I conduct a similar exercise under such common stochastic shocks, and again find poor performance of standard pre-testing methods in controlling size distortions from the differential trends.

A final issue is that several of the papers in my survey use two-way fixed effects (TWFE) methods in settings with staggered treatment timing (see Table 1 in Sun and Abraham (2020)). As pointed out in Sun and Abraham (2020), the coefficient β_{post} from a TWFE model may not have a sensible causal interpretation and β_{pre} may be non-zero under parallel trends if there is treatment effect heterogeneity across adoption cohorts. These issues with heterogeneity are important but distinct from those considered here. My analysis suggests that even if one were willing to impose homogeneity across adoption cohorts, so that β has a sensible interpretation under parallel trends, standard pre-tests may do a poor job of detecting violations of the relevant parallel trends assumption. Moreover, as discussed in Remark 1 below, similar pre-testing issues can arise with newly-introduced methods that are robust to treatment effect heterogeneity.

[9] Indeed, if linear violations of parallel trends were the only concern, one could include parametric controls and avoid the pre-test altogether.

3 Theoretical Analysis

I now provide a theoretical analysis of the distribution of event-study estimates after pre-testing for pre-trends.

3.1 Model

I analyze the normal model introduced in equations (1) and (2) in the simulation section above. The main goal of our analysis will be to analyze the distribution of the post-treatment coefficients $\hat{\beta}_{post}$ conditional on passing a pre-test based on the pre-treatment estimates $\hat{\beta}_{pre}$, i.e. conditional on the event $\hat{\beta}_{pre} \in B(\Sigma)$ for some (measurable) set $B(\Sigma)$ potentially depending on the covariance matrix (e.g. individual or joint tests of significance). For ease of notation, I consider the case where there is one post-treatment period ($M = 1$) unless noted otherwise; all of the results for $M = 1$ will then apply to each individual post-period coefficient (or linear combinations thereof) in the case when $M > 1$.

Remark 1. The finite-sample normal model (1) can be be thought of as an asymptotic approximation to a variety of estimators which yield asymptotically normal coefficients, $\sqrt{N}(\hat{\beta}_n - \beta_n) \rightarrow_d \mathcal{N}(0, \Sigma)$. Estimators yielding event-study coefficients of this form (under suitable regularity conditions) include dynamic two-way fixed effects (TWFE) estimators, the GMM estimator of Freyaldenhoven et al. (2019), and methods for difference-in-differences conditional on covariates (Abadie, 2005; Heckman et al., 1997; Sant'Anna and Zhao, 2020). The recent proposals by Callaway and Sant'Anna (2020) and Sun and Abraham (2020) for constructing event-study estimates that have a sensible interpretation under staggered treatment timing and treatment effect heterogeneity also yield asymptotically normal coefficients. The results here are thus directly applicable to these estimators, which highlights that the issues surrounding pre-testing are distinct from those related to the interpretation of TWFE models under heterogeneity. ■

Appendix B shows that the results derived in the finite sample normal model hold uniformly over a wide range of data-generating processes under which the probability of passing the pre-test does not disappear asymptotically.[10] The asymptotics also allow for the pre-test to depend on a consistently estimated covariance matrix, $\hat{\Sigma} \rightarrow_p \Sigma$.

[10] The condition that the probability of passing the pre-test does not disappear asymptotically requires that the pre-treatment trend δ_{pre} be shrinking with the sample size. This local-to-0 approximation captures the fact that in finite samples the pre-trend may be of a similar order of magnitude as the sampling uncertainty in the data (as with $\gamma_{0.5}$ and $\gamma_{0.8}$). In a model with fixed δ_{pre}, the probability of rejecting the pre-test would be either 0 or 1 asymptotically, which does not capture the fact that in practice we are often uncertain whether the pre-trend is zero or not.

3.2 Bias After Pre-testing

I begin by analyzing the bias of $\hat{\beta}_{post}$ for τ_{post} conditional on passing the pre-test. The following result, which follows from standard arguments using the conditional distributions of multivariate normals, provides a formula for the conditional bias.

Proposition 3.1. *For any conditioning set $B(\Sigma)$,*

$$\mathbb{E}\left[\hat{\beta}_{post} \,|\, \hat{\beta}_{pre} \in B(\Sigma)\right] = \tau_{post} + \delta_{post} + \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B(\Sigma)\right] - \beta_{pre}\right),$$

where $\mathbb{V}ar\left[\begin{pmatrix} \hat{\beta}_{post} \\ \hat{\beta}_{pre} \end{pmatrix}\right] = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.

The formula in Proposition 3.1 illustrates that the expectation of $\hat{\beta}_{post}$ conditional on passing the pre-test is the sum of i) the treatment effect of interest τ_{post}, ii) the unconditional bias δ_{post}, and iii) an additional "pre-test bias" term, which depends on the distortion to the mean of the pre-treatment coefficients from pre-testing, as well as on the covariance between the pre-treatment and post-treatment coefficients.

An immediate implication of Proposition 3.1 is that when parallel trends holds, $\hat{\beta}_{post}$ remains unbiased for τ_{post} after pre-testing so long as the pre-test is symmetric in the sense that we reject the hypothesis of parallel pre-trends for $\hat{\beta}_{pre}$ if and only if we reject the hypothesis for $-\hat{\beta}_{pre}$, a property which holds for two-sided tests of significance.

Corollary 3.1 (No pre-test bias under parallel trends)**.** *Suppose that parallel trends holds in both the pre-treatment and post-treatment periods, so that $\delta_{pre} = \delta_{post} = 0$. If the pre-test $B(\Sigma)$ is such that $\hat{\beta}_{pre} \in B(\Sigma)$ if and only if $-\hat{\beta}_{pre} \in B(\Sigma)$, then*

$$\mathbb{E}\left[\hat{\beta}_{post} \,|\, \hat{\beta}_{pre} \in B(\Sigma)\right] = \tau_{post}.$$

3.2.1 Sufficient conditions for bias exacerbation

In the simulations in Section 2, we saw that for most specifications, the bias of $\hat{\beta}_{post}$ for τ_{post} was worse conditional on passing the pre-test when there were linear violations of parallel trends. I now show that this is necessarily the case under monotone trends and homoskedastic errors.

Assumption 1. *Σ has a common term σ^2 on the diagonal and a common term $\rho > 0$ off of the diagonal, with $\sigma^2 > \rho$.*[11]

[11] If $K = 1$, it suffices to impose that $Cov(\hat{\beta}_{pre}, \hat{\beta}_{post}) > 0$.

Assumption 1 is implied by a suitable homoskedasticity assumption in the canonical two-way fixed effects difference-in-differences model with non-staggered timing. To see this, suppose that the data is generated from the model

$$y_{it} = \alpha_i + \phi_t + \sum_{s \neq 0} \underbrace{\beta_s}_{\tau_s + \delta_s} \times D_i + \epsilon_{it}, \tag{3}$$

where D_i is an indicator for whether i is first treated at $t = 1$ or never treated. If the researcher estimates β_s via OLS, then the estimated coefficients will be given by

$$\hat{\beta}_s = \beta_s + \Delta\bar{\epsilon}_s - \Delta\bar{\epsilon}_0,$$

where $\Delta\bar{\epsilon}_t$ is the difference in the average residuals for the treatment and control groups in period t. It follows immediately that if the ϵ_{it} are homoskedastic, $\mathbb{V}\text{ar}\left[\hat{\beta}_k\right] = 2\mathbb{V}\text{ar}\left[\Delta\bar{\epsilon}_0\right] =: \sigma^2$ and $\text{Cov}(\hat{\beta}_k, \hat{\beta}_j) = \sigma^2/2 =: \rho$, so Assumption 1 holds.

We now show that under Assumption 1, the bias after testing for significant pre-treatment coefficients is worse than the unconditional bias under arbitrary monotone violations of parallel trends.

Proposition 3.2 (Sign of bias under monotone trend)**.** *Suppose that there is an upward pre-trend in the sense that $\delta_{pre} < 0$ (elementwise) and $\delta_{post} > 0$.*[12] *If Assumption 1 holds, then*

$$\mathbb{E}\left[\hat{\beta}_{post} \,|\, \hat{\beta}_{pre} \in B_{NIS}(\Sigma)\right] > \beta_{post} > \tau_{post}.$$

The analogous result holds replacing ">" with "<" and vice versa.

Remark 2. Monotonicity of δ is often implied in the discussion of violations of parallel trends in applied work. For instance, Lovenheim and Willen (2019) argue that violations of parallel trends cannot explain their results because "pre-[treatment] trends are either zero or in the wrong direction (i.e., opposite to the direction of the treatment effect)." Likewise, Greenstone and Hanna (2014) estimate upward-sloping pre-existing trends and argue that their estimates would be upward biased "if the pre-trends had continued." Nonetheless, there are economic settings in which we do not expect monotonicity to hold, with the "Ashenfelter's dip" expected in job-training programs as a notable example (Ashenfelter, 1978). ■

Remark 3. The homoskedasticity assumption required to obtain the bias exacerbation in Proposition 3.2 is of course strong and unlikely to hold exactly in most practical applications.

[12] Technically, the restriction that $\delta_{pre} < 0$ and $\delta_{post} > 0$ is somewhat weaker than monotonicity. It allows, for instance, for $\delta_{-3} > \delta_{-2}$, so long as both are less than 0.

Nonetheless, the fact that bias is exacerbated under homoskedasticty and arbitrary monotone violations of parallel trends indicates that pre-testing can exacerbate bias in non-pathological cases. Further, the survey of papers in Section 2 suggests that although homoskedasticity typically does not hold exactly in practice, the pre-test bias typically goes in the direction predicted for the homoskedastic case. ■

3.3 Variance after pre-testing

Having analyzed the properties of the mean of the treatment effect estimate conditional on passing a pre-test for parallel trends, we now turn to analyzing its variance. We begin with a general formula, which expresses the conditional variance of the treatment effect in terms of its unconditional variance and the distortion to the variance of the pre-period coefficients.

Proposition 3.3.

$$\mathbb{V}ar\left[\hat{\beta}_{post}|\hat{\beta}_{pre} \in B(\Sigma)\right] = \mathbb{V}ar\left[\hat{\beta}_{post}\right] + (\Sigma_{12}\Sigma_{22}^{-1})\left(\mathbb{V}ar\left[\hat{\beta}_{pre}\,|\,\hat{\beta}_{pre} \in B(\Sigma)\right] - \mathbb{V}ar\left[\hat{\beta}_{pre}\right]\right)(\Sigma_{12}\Sigma_{22}^{-1})'.$$

Proposition 3.3 implies that the variance of $\hat{\beta}_{post}$ will typically be smaller after conditioning on the result of the pre-test. Indeed, this is the case when the acceptance region for the pre-test is convex, a property which holds for most tests of individual or joint significance.

Proposition 3.4 (Pre-testing reduces variance)**.** *Suppose that* $B(\Sigma)$ *is a convex set. Then* $\mathbb{V}ar\left[\hat{\beta}_{post}\,|\,\hat{\beta}_{pre} \in B(\Sigma)\right] \leq \mathbb{V}ar\left[\hat{\beta}_{post}\right]$.

Since $\hat{\beta}_{post}$ is unbiased conditional on passing the pre-test under parallel trends (Corollary 3.1), Proposition 3.4 suggests that typical confidence intervals will tend to over-cover conditional on passing the pre-test under parallel trends.[13] When parallel trends is violated, however, $\hat{\beta}_{post}$ will be biased, and thus conventional CIs will tend to under-cover if the bias is sufficiently large, as shown in the simulations in Section 2.

3.4 Implications for Published Studies

So far we have considered the properties of event-study estimates for a fixed violation of parallel trends δ. In practice, however, researchers consider a variety of studies with different biases. We thus consider a simple model of publication in which researchers conduct studies with different values of δ. The model highlights that the effectiveness of pre-testing in

[13] This is not formally implied by the proposition, since the conditional distribution of $\hat{\beta}_{post}$ may be non-normal. It is, however, always the case in simulations based on the survey of papers in Section 2; see Table 2.

reducing bias (and undercoverage) in published work will depend on both the power of the pre-test and the statistical distortions from pre-testing.[14]

For simplicity, consider the setting where parallel trends holds ($\delta = 0$) in fraction $1-\theta$ of studies and in the remaining θ fraction of studies $\delta = \bar{\delta} \neq 0$. If all studies were published, regardless of pre-trends, then the expected bias in published work (assuming the pre-test is symmetric in the sense of Corollary 3.1) would be

$$Bias^{Notest} = P(\delta = \bar{\delta})\bar{\delta}_{post} = \theta\bar{\delta}_{post}.$$

On the other hand, if we only published the studies without a significant pre-trend ($\hat{\beta}_{pre} \in B(\Sigma)$), the expected bias in published work would be

$$Bias^{Pre-test} = P(\delta = \bar{\delta} \,|\, \hat{\beta}_{pre} \in B(\Sigma))\mathbb{E}\left[\hat{\beta}_{post} - \tau_{post} \,|\, \hat{\beta}_{pre} \in B(\Sigma)\right].$$

Comparing the biases under the two publication regimes, we have

$$\frac{Bias^{Test}}{Bias^{Notest}} = \underbrace{\frac{P(\delta = \bar{\delta} \,|\, \hat{\beta}_{pre} \in B(\Sigma))}{P(\delta = \bar{\delta})}}_{\substack{\text{Relative fraction of biased} \\ \text{studies}}} \cdot \underbrace{\frac{\mathbb{E}\left[\hat{\beta}_{post} - \tau_{post} \,|\, \delta = \bar{\delta}, \hat{\beta}_{pre} \in B(\Sigma)\right]}{\bar{\delta}_{post}}}_{\text{Ratio of bias when publish biased design}}. \tag{4}$$

The first term represents the relative fraction of published studies with a biased design ($\delta = \bar{\delta}$) across the two regimes. This will tend to be less than 1, since the pre-test will reject less frequently conditional on $\delta = 0$. By contrast, the second term represents the ratio of the bias conditional on surviving the pre-test to the unconditional bias when $\delta = \bar{\delta}$, which will often be greater than 1 (see Proposition 3.2).

The effect of requiring an insignificant pre-test to publish on the bias in published work is thus ambiguous, and depends on the relative magnitude of these two factors. It is straightforward to show that the first term converges to 1 if either i) $\theta \to 1$, so that all studies are equally biased, or ii) the Bayes Factor, $P(\hat{\beta}_{pre} \in B(\Sigma) \,|\, \delta = \bar{\delta})/P(\hat{\beta}_{pre} \in B(\Sigma) \,|\, \delta = 0)$ converges to 1, so that the pre-test has no power to distinguish between a biased and unbiased design.

The pre-testing regime is thus least effective in reducing bias in published work when either the ex ante credibility of studies (as proxied by $1-\theta$) is low, or the pre-test is underpowered (meaning the Bayes Factor is low). A similar analysis applies to the null

[14] A similar conclusion would be reached if we considered one researcher choosing between many specifications.

rejection probability in published studies.

4 Practical Recommendations

This paper highlights important issues with the standard practice of pre-testing for pre-trends in difference-in-differences and related research designs. What, then, should applied researchers do when worried about pre-existing differences in trends? My results suggest that the issues with pre-testing are more severe when either the power of the pre-test or the ex ante credibility of the research design is low. Researchers relying on pre-trends tests should therefore pay close attention to the power of the pre-test against relevant violations of parallel trends and the ex ante credibility of the research design.

As a diagnostic, researchers relying on pre-trends tests could report power calculations against economically relevant violations of parallel trends, similar to the exercise in Section 2. I provide the R package `pretrends` and an accompanying `Shiny application` to facilitate such analyses. Figure 2 displays the user-interface of the Shiny application, which allows the user to visualize and conduct power analyses for any (potentially non-linear) hypothesized violation of parallel trends, and to analyze the expected statistical distortions from pre-testing.

It is important to note that in order to calculate the power of a pre-trends test, it is necessary to place restrictions on the possible violations of parallel trends (e.g., the analysis in Section 2 restricts to linear violations). Without any such restrictions, the difference in trends can change arbitrarily between the pre- and post-treatment periods, and thus no pre-test can have more than trivial power against a violation of parallel trends that would produce bias of an arbitrary magnitude.

If the researcher is willing to place ex ante restrictions on the violations of parallel trends, however, there are alternative approaches that yield valid inference while avoiding the pre-test altogether. I briefly highlight three such approaches. First, if the researcher has knowledge about the likely functional form of the difference in trends, then the bias can be removed via parametric controls (e.g., Wolfers (2006); Dobkin et al. (2018); Goodman-Bacon (2018)).[15] Second, if researchers are not sure about the right functional form for the difference in trends, Rambachan and Roth (2020) provide tools to assess the sensitivity of results under different sets of non-parametric assumptions that impose that the difference in trends not change "too much" between the pre-treatment and post-treatment periods. Fi-

[15] In the longer working paper version of this paper, I show how tools for correcting for publication bias from Andrews and Kasy (2019) can be adapted to apply these methods retrospectively to published papers that have been screened based on pre-trends.

nally, Freyaldenhoven et al. (2019) provide a method for settings where there is a covariate assumed to be affected by the relevant confounding factors but not by the treatment itself. Each of these approaches applies different restrictions on the way that parallel trends can be violated, and thus the most appropriate tool will depend on context.

Regardless of the exact method, I urge researchers to use context-specific economic knowledge to inform the discussion of possible violations of parallel trends. Bringing economic knowledge to bear on how parallel trends might plausibly be violated in a given context will yield stronger, more credible inferences than relying on the statistical significance of pre-trends tests alone.

References

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1):1–19.

Andrews, I. (2018). Valid Two-Step Identification-Robust Confidence Sets for GMM. *The Review of Economics and Statistics*, 100(2):337–348.

Andrews, I. and Kasy, M. (2019). Identification of and Correction for Publication Bias. *American Economic Review*, 109(8):2766–2794.

Armstrong, T. B. and Kolesár, M. (2018). A Simple Adjustment for Bandwidth Snooping. *The Review of Economic Studies*, 85(2):732–765.

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics*, 60(1):47–57.

Athey, S. and Imbens, G. (2018). Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption. *arXiv:1808.05293 [cs, econ, math, stat]*.

Bailey, M. J. and Goodman-Bacon, A. (2015). The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans. *American Economic Review*, 105(3):1067–1104.

Belloni, A., Chernozhukov, V., Fernández-Val, I., and Hansen, C. (2017). Program Evaluation and Causal Inference With High-Dimensional Data. *Econometrica*, 85(1):233–298. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA12723.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2):29–50.

Bilinski, A. and Hatfield, L. A. (2020). Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions. *arXiv:1805.03273 [stat]*. arXiv: 1805.03273.

Borusyak, K. and Jaravel, X. (2016). Revisiting Event Study Designs. SSRN Scholarly Paper ID 2826228, Social Science Research Network, Rochester, NY.

Bosch, M. and Campos-Vazquez, R. M. (2014). The Trade-Offs of Welfare Policies in Labor Markets with Informal Jobs: The Case of the "Seguro Popular" Program in Mexico. *American Economic Journal: Economic Policy*, 6(4):71–99.

Callaway, B. and Sant'Anna, P. H. C. (2020). Difference-in-Differences with multiple time periods. *Journal of Econometrics*.

Cartinhour, J. (1990). One-dimensional marginal density functions of a truncated multivariate normal density function. *Communications in Statistics-Theory and Methods*, 19:197–203.

Chabé-Ferret, S. (2015). Analysis of the bias of Matching and Difference-in-Difference under alternative earnings and selection processes. *Journal of Econometrics*, 185(1):110–123.

Christensen, G. S. and Miguel, E. (2016). Transparency, Reproducibility, and the Credibility of Economics Research. Working Paper 22989, National Bureau of Economic Research.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Academic Press. Google-Books-ID: YleCAAAAIAAJ.

Daw, J. R. and Hatfield, L. A. (2018). Matching and Regression to the Mean in Difference-in-Differences Analysis. *Health Services Research.*

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.

Deryugina, T. (2017). The Fiscal Cost of Hurricanes: Disaster Aid versus Social Insurance. *American Economic Journal: Economic Policy*, 9(3):168–198.

Deschênes, O., Greenstone, M., and Shapiro, J. S. (2017). Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program. *American Economic Review*, 107(10):2958–2989.

Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. 108(2):308–352.

Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, 189(1):1–23.

Fitzpatrick, M. D. and Lovenheim, M. F. (2014). Early retirement incentives and student achievement. *American Economic Journal: Economic Policy*, 6(3):120–154.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019). Pre-event Trends in the Panel Event-Study Design. *American Economic Review*, 109(9):3307–3338.

Friedman, M. (1940). Review of Jan Tinbergen. Statistical testing of business cycle theories, II: Business cycles in the United States of America. *American Economic Review*, 30.

Gallagher, J. (2014). Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the United States. *American Economic Journal: Applied Economics*, 6(3):206–233.

Giles, J. A. and Giles, D. E. A. (1993). Pre-Test Estimation and Testing in Econometrics: Recent Developments. *Journal of Economic Surveys*, 7(2):145–197.

Goodman-Bacon, A. (2018). Public insurance and mortality: Evidence from medicaid implementation. *Journal of Public Economics*, 126(1):216–262.

Goodman-Bacon, A. (2020). Difference-in-Differences with Variation in Treatment Timing. Working paper.

Greenstone, M. and Hanna, R. (2014). Environmental Regulations, Air and Water Pollution, and Infant Mortality in India. *American Economic Review*, 104(10):3038–3072.

Guggenberger, P. (2010). The Impact of a Hausman Pretest On The Asymptotic Size Of A Hypothesis Test. *Econometric Theory*, 26(2):369–382.

He, G. and Wang, S. (2017). Do College Graduates Serving as Village Officials Help Rural China? *American Economic Journal: Applied Economics*, 9(4):186–215.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4):605–654. Publisher: Oxford Academic.

Kahn-Lang, A. and Lang, K. (2018). The Promise and Pitfalls of Differences-in-Differences: Reflections on '16 and Pregnant' and Other Applications. Working Paper 24857, National Bureau of Economic Research.

Keynes, J. M. (1939). Professor Tinbergen's Method. *The Economic Journal*, 49(195):558–577.

Kuziemko, I., Meckel, K., and Rossin-Slater, M. (2018). Does Managed Care Widen Infant Health Disparities? Evidence from Texas Medicaid. *American Economic Journal: Economic Policy*, 10(3):255–283.

Lafortune, J., Rothstein, J., and Schanzenbach, D. W. (2018). School Finance Reform and the Distribution of Student Achievement. *American Economic Journal: Applied Economics*, 10(2):1–26.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927.

Leeb, H. and Pötscher, B. M. (2005). Model Selection and Inference: Facts and Fiction. *Econometric Theory*, 21(1):21–59.

Lovenheim, M. F. and Willen, A. (2019). The long-run effects of teacher collective bargaining. *American Economic Journal: Economic Policy*, 11(3):292–324.

Manjunath, B. and Wilhelm, S. (2012). Moments Calculation For the Doubly Truncated Multivariate Normal Density. *arXiv:1206.5387 [stat]*.

Markevich, A. and Zhuravskaya, E. (2018). The Economic Effects of the Abolition of Serfdom: Evidence from the Russian Empire. *American Economic Review*, 108(4-5):1074–1117.

Rambachan, A. and Roth, J. (2020). An honest approach to parallel trends.

Rothstein, H. R., Sutton, A. J., and Borenstein, M. (2005). Publication Bias in Meta-Analysis. In Co-Chair, H. R. R., Co-Author, A. J. S., and PI, M. B. D. A. L., editors, *Publication Bias in Meta-Analysis*, pages 1–7. John Wiley & Sons, Ltd.

Sant'Anna, P. H. C. and Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122.

Snyder, C. and Zhuo, R. (2018). Sniff Tests in Economics: Aggregate Distribution of Their Probability Values and Implications for Publication Bias. Working Paper 25058, National Bureau of Economic Research.

Sun, L. and Abraham, S. (2020). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, (Forthcoming).

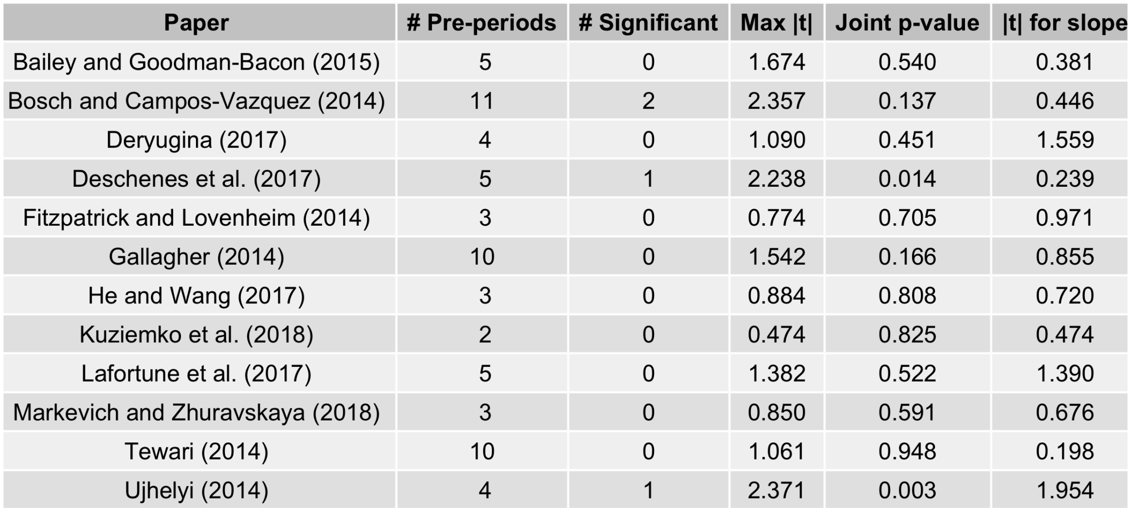
Tewari, I. (2014). The Distributive Impacts of Financial Development: Evidence from Mortgage Markets during US Bank Branch Deregulation. *American Economic Journal: Applied Economics*, 6(4):175–196.

Tinbergen, J. (1939). *Statistical Testing of Business Cycle Theories: Part II: Business Cycles in the United States of America, 1919-1932.* Agaton Press, New York. Publication Title: Books (Jan Tinbergen).

Ujhelyi, G. (2014). Civil Service Rules and Policy Choices: Evidence from US State Governments. *American Economic Journal: Economic Policy*, 6(2):338–380.

Wolfers, J. (2006). Did unilateral divorce laws raise divorce rates? a reconciliation and new results. *American Economic Review*, 96:1802–1820.

5 Figures and Tables

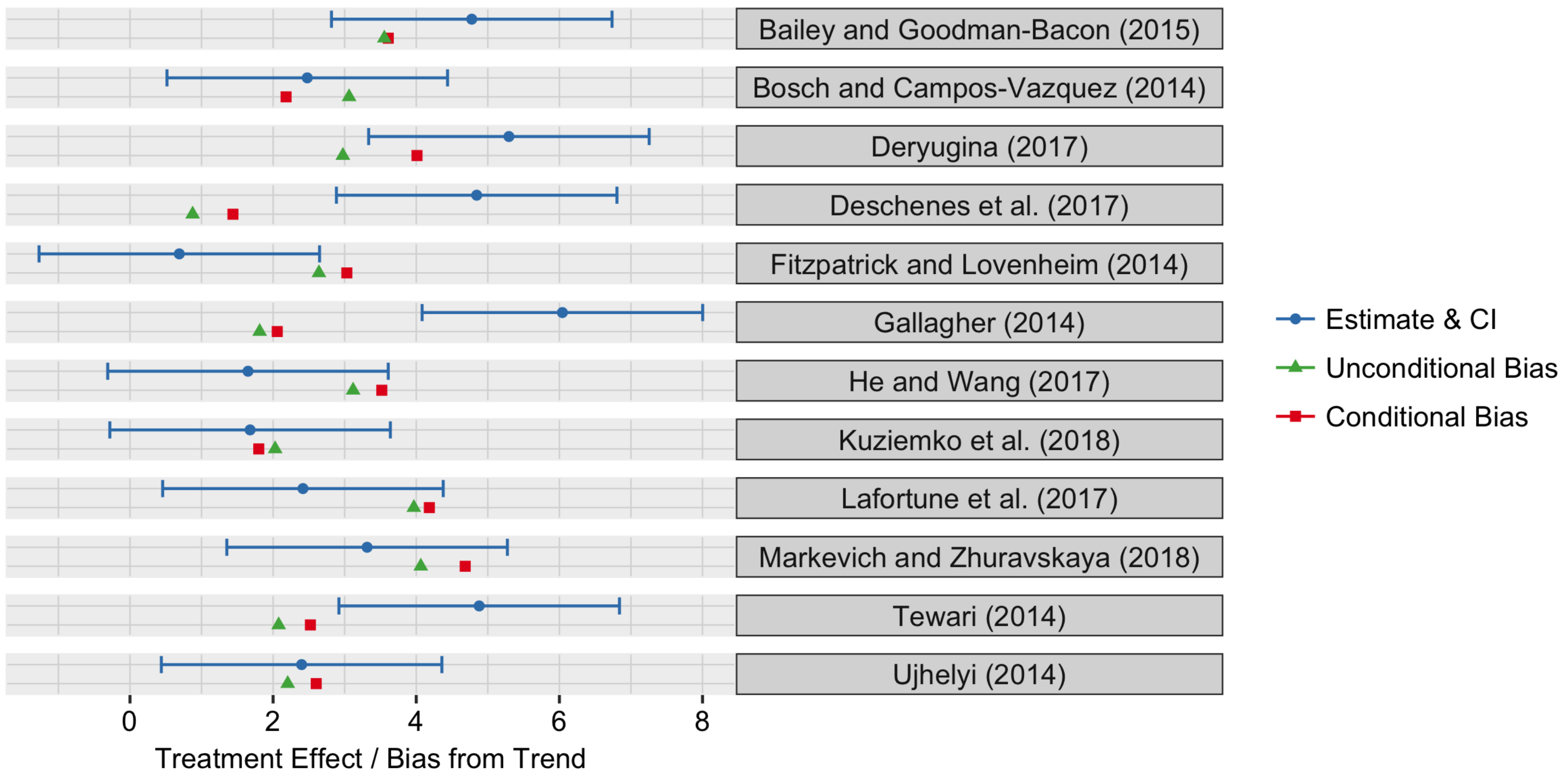


Paper	# Pre-periods	# Significant	Max |t|	Joint p-value	|t| for slope
Bailey and Goodman-Bacon (2015)	5	0	1.674	0.540	0.381
Bosch and Campos-Vazquez (2014)	11	2	2.357	0.137	0.446
Deryugina (2017)	4	0	1.090	0.451	1.559
Deschenes et al. (2017)	5	1	2.238	0.014	0.239
Fitzpatrick and Lovenheim (2014)	3	0	0.774	0.705	0.971
Gallagher (2014)	10	0	1.542	0.166	0.855
He and Wang (2017)	3	0	0.884	0.808	0.720
Kuziemko et al. (2018)	2	0	0.474	0.825	0.474
Lafortune et al. (2017)	5	0	1.382	0.522	1.390
Markevich and Zhuravskaya (2018)	3	0	0.850	0.591	0.676
Tewari (2014)	10	0	1.061	0.948	0.198
Ujhelyi (2014)	4	1	2.371	0.003	1.954

Table 1: Summary of Pre-period Event-Study Coefficients

Note: This table provides information about the pre-period event-study coefficients in the papers reviewed. The table shows the number of pre-periods in the event-study, the fraction of the pre-period coefficients that are significant at the 95% level, the maximum t-stat among those coefficients, the p-value for a chi-squared test of joint significance, and the t-stat for the slope of the linear trend through the pre-period coefficients. See Section 2 for more detail on the sample of papers reviewed.

Figure 1: OLS Estimates and Bias from Linear Trends for Which Pre-tests Have 80 Percent Power – Average Treatment Effect



Note: I calculate the linear trend against which conventional pre-tests would reject 80 percent of the time ($\gamma_{0.8}$). The red squares show the bias that would result from such a trend conditional on passing the pre-test ($\mathbb{E}\left[\hat{\tau} - \tau_* \,|\, \hat{\beta}_{pre} \in B_{NIS}(\Sigma)\right]$); the green triangles show the unconditional bias from such a trend ($\mathbb{E}\left[\hat{\tau} - \tau_*\right]$). As a benchmark, I plot in blue the original OLS estimates and 95% CIs from the paper. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The estimand is the average of the treatment effects in all periods after treatment began, $\tau_* = \bar{\tau}$.

	Conditional on passing pre-test?					
	No			Yes		
	Slope of differential trend:					
	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.61	0.94	0.05	0.62	0.95
Bosch and Campos-Vazquez (2014)	0.05	0.49	0.86	0.03	0.28	0.61
Deryugina (2017)	0.05	0.49	0.84	0.01	0.75	1.00
Deschenes et al. (2017)	0.05	0.09	0.14	0.03	0.10	0.25
Fitzpatrick and Lovenheim (2014)	0.05	0.41	0.75	0.05	0.50	0.87
Gallagher (2014)	0.05	0.19	0.44	0.04	0.22	0.54
He and Wang (2017)	0.05	0.54	0.88	0.05	0.63	0.95
Kuziemko et al. (2018)	0.05	0.28	0.53	0.04	0.21	0.42
Lafortune et al. (2017)	0.05	0.71	0.98	0.05	0.75	0.99
Markevich and Zhuravskaya (2018)	0.05	0.76	0.98	0.04	0.87	1.00
Tewari (2014)	0.05	0.20	0.55	0.04	0.25	0.72
Ujhelyi (2014)	0.05	0.29	0.60	0.04	0.36	0.76

Table 2: Null Rejection Probabilities for Nominal 5% Test of Average Treatment Effect Under Linear Trends Against Which Pre-tests Have 50 or 80% Power

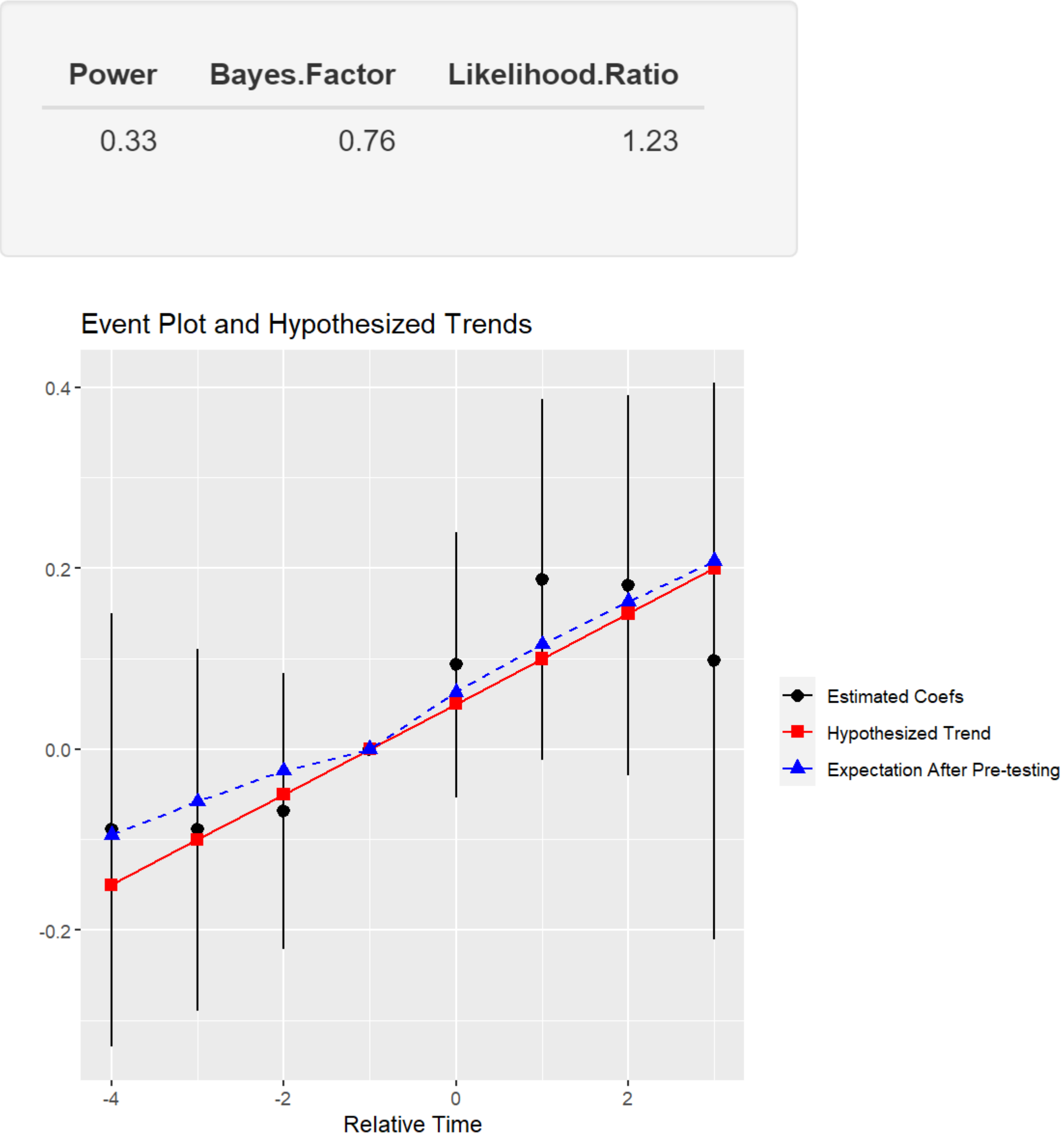
Note: This table shows null rejection probabilities, i.e. the probability that the true parameter falls outside a nominal 95% confidence interval, under data-generating processes in which parallel trends holds (slope of differential trend = 0) or in which there are linear violations of parallel trends that conventional pre-tests would detect 50 or 80% of the time ($\gamma_{0.5}$ and $\gamma_{0.8}$). The first three columns show unconditional null rejection probabilities, whereas the latter three columns condition on passing the pre-test. The estimand is the average of the post-treatment causal effects, $\bar{\tau}$.

	Estimand:			
	τ_1		$\bar{\tau}$	
	Slope of differential trend:			
Paper	$\gamma_{0.5}$	$\gamma_{0.8}$	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	51	56	1	2
Bosch and Campos-Vazquez (2014)	-29	-34	-25	-29
Deryugina (2017)	103	120	30	35
Deschenes et al. (2017)	88	119	48	64
Fitzpatrick and Lovenheim (2014)	25	30	12	15
Gallagher (2014)	57	62	11	14
He and Wang (2017)	29	34	11	13
Kuziemko et al. (2018)	-16	-20	-9	-11
Lafortune et al. (2017)	-9	-10	5	5
Markevich and Zhuravskaya (2018)	52	62	13	15
Tewari (2014)	90	102	19	21
Ujhelyi (2014)	51	59	15	18

Table 3: Percent Additional Bias Conditional on Passing Pre-test

Note: This table shows the additional bias from conditioning on none of the pre-period coefficients being statistically significant as a percentage of the unconditional bias, i.e. $100 \cdot$ (Conditional Bias $-$ Unconditional Bias)/(Unconditional Bias). Biases are calculated under linear violations of parallel trends with slopes $\gamma_{0.5}$ and $\gamma_{0.8}$, against which conventional pre-tests have 50 or 80% power. The estimand in the first two columns is the treatment effect in the first period (τ_1), and in the last two columns it is the average effect across all post-treatment periods ($\bar{\tau}$).

Figure 2: Screen-shot of Shiny Application



Power	Bayes.Factor	Likelihood.Ratio
0.33	0.76	1.23

Note: this figure shows a screen-shot of the Shiny application accompanying this paper using the event-study from He and Wang (2017). Given the results of an event-study and a user-inputted hypothesized violation of parallel trends ($\bar{\delta}$), the application calculates the power of the pre-test, the Bayes Factor (see Section 3.4), and the likelihood ratio $l(\hat{\beta}_{pre}|\delta_{pre} = \bar{\delta}_{pre})/l(\hat{\beta}_{pre}|\delta_{pre} = 0)$, where $l(\cdot)$ is the likelihood function. It also displays an event-plot with the estimated coefficients and confidence intervals, the hypothesized trend, and the expectation of $\hat{\beta}$ conditional on passing the pre-test, $\mathbb{E}\left[\hat{\beta} \,|\, \hat{\beta}_{pre} \in B_{NIS}(\Sigma), \delta = \bar{\delta}\right]$.

Supplement to the paper

Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends

For online publication

Jonathan Roth

April 20, 2021

This supplement contains proofs and additional results for the paper "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends." Section A provides proofs for the results in the main text. Section B states and proves asymptotic results. Section C provides additional simulation results in which the treatment and control group receive stochastic common shocks. Finally, Section D contains additional tables and figures.

A Proofs for Results in the Main Text

This section collects proofs for the results in the main text, as well as some auxiliary lemmas. For ease of notation, I leave the dependence of B on Σ implicit unless needed for clarity. We begin with a lemma, which will be useful in the following proofs.

Lemma A.1. *Let $\tilde{\beta}_{post} = \hat{\beta}_{post} - \Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre}$. Then $\tilde{\beta}_{post}$ and $\hat{\beta}_{pre}$ are independent.*

Proof. Note that by assumption, $\hat{\beta}_{post}$ and $\hat{\beta}_{pre}$ are jointly normal. Since $\tilde{\beta}_{post}$ is a linear combination of $\hat{\beta}_{post}$ and $\hat{\beta}_{pre}$, it follows that $\hat{\beta}_{pre}$ and $\tilde{\beta}_{post}$ are jointly normal. It thus suffices to show that $\hat{\beta}_{pre}$ and $\tilde{\beta}_{post}$ are uncorrelated. We have

$$\begin{aligned}\text{Cov}\left(\hat{\beta}_{pre}\,,\,\tilde{\beta}_{post}\right) &= \mathbb{E}\left[\left(\hat{\beta}_{pre} - \beta_{pre}\right)\left((\hat{\beta}_{post} - \beta_{post}) - \Sigma_{12}\Sigma_{22}^{-1}(\hat{\beta}_{pre} - \beta_{pre})\right)'\right] \\ &= \Sigma_{12}' - \Sigma_{22}\Sigma_{22}^{-1}\Sigma_{12}' \\ &= 0\end{aligned}$$

which completes the proof.

□

Proof of Proposition 3.1 Note that by construction, $\hat{\beta}_{post} = \tilde{\beta}_{post} + \Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre}$. It follows that

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}_{post} \,|\, \hat{\beta}_{pre} \in B\right] &= \mathbb{E}\left[\tilde{\beta}_{post} \,|\, \hat{\beta}_{pre} \in B\right] + \Sigma_{12}\Sigma_{22}^{-1}\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] \\
&= \mathbb{E}\left[\tilde{\beta}_{post}\right] + \Sigma_{12}\Sigma_{22}^{-1}\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] \\
&= \mathbb{E}\left[\hat{\beta}_{post} - \Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre}\right] + \Sigma_{12}\Sigma_{22}^{-1}\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] \\
&= \beta_{post} - \Sigma_{12}\Sigma_{22}^{-1}\beta_{pre} + \Sigma_{12}\Sigma_{22}^{-1}\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] \\
&= \beta_{post} + \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] - \beta_{pre}\right)
\end{aligned}$$

where the second line uses the independence of $\tilde{\beta}_{post}$ and $\hat{\beta}_{pre}$ from Lemma A.1, and the third and fourth use the definition of $\tilde{\beta}_{post}$, β_{post}, and β_{pre}. Since $\beta_{post} = \tau_{post} + \delta_{post}$ by definition, the result follows. □

Definition 1 (Symmetric Truncation About 0)**.** We say that $B \subset \mathbb{R}^K$ is a symmetric truncation around 0 if $\beta \in B$ iff $-\beta \in B$.

Lemma A.2. *Suppose $Y \sim \mathcal{N}(0, \Sigma)$ is a K-dimensional multivariate normal, and B is a symmetric truncation around 0. Then $\mathbb{E}[Y \,|\, Y \in B] = 0$.*

Proof. Note that if $Y \sim \mathcal{N}(0, \Sigma)$, then $-Y$ is also distributed $\mathcal{N}(0, \Sigma)$. Using this, combined with the fact that $(-Y) \in B$ iff $Y \in B$ by assumption, we have

$$\begin{aligned}
\mathbb{E}[Y \,|\, Y \in B] &= \mathbb{E}[-Y \,|\, (-Y) \in B] \\
&= \mathbb{E}[-Y \,|\, Y \in B] \\
&= -\mathbb{E}[Y \,|\, Y \in B],
\end{aligned}$$

which implies that $\mathbb{E}[Y \,|\, Y \in B] = 0$.

□

Proof of Corollary 3.1 From Proposition 3.1, it suffices to show that $\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] - \beta_{pre} = 0$. However, $\beta_{pre} = 0$ by the assumption of parallel trends, and $\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] = 0$ by Lemma A.2. □

We now prove a series of Lemmas leading up to the proof of Proposition 3.2.

Lemma A.3. *Suppose Y is a k-dimensional multivariate normal, $Y \sim \mathcal{N}(\mu, \Sigma)$, and let $B \subset \mathbb{R}^k$ be a convex set such that $\mathbb{P}(Y \in B) > 0$. Letting D_μ denote the Jacobian operator with respect to μ, we have*

1. $D_\mu \mathbb{E}[Y \mid Y \in B, \mu] = \mathbb{V}ar[Y \mid Y \in B, \mu]\, \Sigma^{-1}$*.*

2. $\mathbb{V}ar[Y \mid Y \in B] - \Sigma$ *is negative semi-definite.*

Proof.[16]

Define the function $H : \mathbb{R}^k \to \mathbb{R}$ by

$$H(\mu) = \int_B \phi_\Sigma(y - \mu) dy$$

for $\phi_\Sigma(x) = (2\pi)^{-\frac{k}{2}} det(\Sigma)^{-\frac{1}{2}} exp(-\frac{1}{2} x'\Sigma^{-1} x)$ the PDF of the $\mathcal{N}(0, \Sigma)$ distribution. We now argue that H is log-concave in μ. Note that we can write $H(\mu) = \int_{\mathbb{R}^k} g_1(y, \mu) g_2(y, \mu) dy$ for $g_1(y, \mu) = \phi_\Sigma(y - \mu)$ and $g_2(y, \mu) = 1[y \in B]$. The normal PDF is log-concave, and g_1 is the composition of the normal PDF with a linear function, and hence log-concave as well. Likewise, g_2 is log-concave since B is a convex set. The product of log-concave functions is log-concave, and the marginalization of a log-concave function with respect to one of its arguments is log-concave by Prekopa's theorem (see, e.g. Theorem 3.3 in Saumard and Wellner (2014)), from which it follows that H is log-concave in μ.

Now, applying Leibniz's rule and the chain rule, we have that the $1 \times k$ gradient of $\log H$ with respect to μ is equal to

$$\begin{aligned} D_\mu \log H &= \frac{\int_B D_\mu \phi_\Sigma(y - \mu) dy}{\int_B \phi_\Sigma(y - \mu) dy} \\ &= \frac{\int_B \phi_\Sigma(y - \mu)(y - \mu)'\Sigma^{-1} dy}{\int_B \phi_\Sigma(y - \mu) dy} \\ &= (\mathbb{E}[Y \mid Y \in B] - \mu)'\Sigma^{-1}. \end{aligned}$$

where the second line takes the derivative of the normal PDF, $D_\mu \phi_\Sigma(y - \mu) = \phi_\Sigma(y - \mu) \cdot (y - \mu)'\Sigma^{-1}$, and the third uses the definition of the conditional expectation. It follows that

$$\mathbb{E}[Y \mid Y \in B, \mu] = \mu + \Sigma(D_\mu \log H)'.$$

Differentiating again with respect to μ, we have that the $k \times k$ Jacobian of $\mathbb{E}[Y \mid Y \in B, \mu]$ with respect to μ is given by

$$D_\mu \mathbb{E}[Y \mid Y \in B, \mu] = I + \Sigma D_\mu (D_\mu \log H)'. \tag{5}$$

[16] I am grateful to Alecos Papadopolous, whose answer on StackOverflow to a related question inspired this proof.

Since H is log-concave, $D_\mu(D_\mu \log H)'$ is the Hessian of a concave function, and thus is negative semi-definite. Next, note that by definition,

$$\mathbb{E}\left[Y \,|\, Y \in B, \mu\right] = \frac{\int_B y \, \phi_\Sigma(y-\mu)\, dy}{\int_B \phi_\Sigma(y-\mu)\, dy}.$$

Thus, applying Leibniz's rule again along with the product rule,

$$\begin{aligned} D_\mu \mathbb{E}\left[Y \,|\, Y \in B, \mu\right] =& \frac{\int_B y \, D_\mu \phi_\Sigma(y-\mu)\, dy}{\int_B \phi_\Sigma(y-\mu)\, dy} + \\ & \left[\int_B y \, \phi_\Sigma(y-\mu)\, dy\right] \cdot D_\mu \left[\int_B \phi_\Sigma(y-\mu)\, dy\right]^{-1}. \end{aligned} \tag{6}$$

Recall that

$$D_\mu \phi_\Sigma(y-\mu) = \phi_\Sigma(y-\mu) \cdot (y-\mu)'\Sigma^{-1}.$$

The first term on the right-hand side of (6) thus becomes

$$\begin{aligned} \frac{\int_B y(y-\mu)'\phi_\Sigma(y-\mu)\, dy}{\int_B \phi_\Sigma(y-\mu)\, dy}\Sigma^{-1} &= \\ \left(\mathbb{E}\left[YY' \,|\, Y \in B, \mu\right] - \mathbb{E}\left[Y \,|\, Y \in B, \mu\right] \mu'\right)\Sigma^{-1}&. \end{aligned}$$

Applying the chain-rule, the second term on the right-hand side of (6) becomes

$$\begin{aligned} -\frac{\int_B y \, \phi_\Sigma(y-\mu)\, dy \cdot \int_B (y-\mu)' \, \phi_\Sigma(y-\mu)\, dy}{\left[\int_B \phi_\Sigma(y-\mu)\, dy\right]^2}\Sigma^{-1} &= \\ \left(-\mathbb{E}\left[Y \,|\, Y \in B, \mu\right] \mathbb{E}\left[Y \,|\, Y \in B, \mu\right]' + \mathbb{E}\left[Y \,|\, Y \in B, \mu\right] \mu'\right)\Sigma^{-1}&. \end{aligned}$$

Substituting the expressions in the previous two displays back into (6), we have

$$\begin{aligned} D_\mu \mathbb{E}\left[Y \,|\, Y \in B, \mu\right] &= \left(\mathbb{E}\left[YY' \,|\, Y \in B, \mu\right] - \mathbb{E}\left[Y \,|\, Y \in B, \mu\right] \mathbb{E}\left[Y \,|\, Y \in B, \mu\right]'\right)\Sigma^{-1} \\ &= \mathbb{V}\mathrm{ar}\left[Y \,|\, Y \in B, \mu\right]\Sigma^{-1}, \end{aligned} \tag{7}$$

which establishes the first result. Additionally, combining (5) and (7), we have that

$$\mathbb{V}\mathrm{ar}\left[Y \,|\, Y \in B, \mu\right]\Sigma^{-1} = I + \Sigma D_\mu (D_\mu \log H)', \tag{8}$$

which implies that

$$\mathbb{V}\text{ar}\left[Y \mid Y \in B, \mu\right] - \Sigma = \Sigma\, D_\mu (D_\mu \log H)'\, \Sigma. \tag{9}$$

Thus, for any vector $x \in \mathbb{R}^k$,

$$\begin{aligned} x' \left(\mathbb{V}\text{ar}\left[Y \mid Y \in B, \mu\right] - \Sigma\right) x &= x' \left(\Sigma\, D_\mu (D_\mu \log H)'\, \Sigma\right) x \\ &= (\Sigma x)' \; \left(D_\mu (D_\mu \log H)'\right) \; (\Sigma x) \\ &\leq 0 \end{aligned}$$

where the inequality follows from the fact that $D_\mu(D_\mu \log H)'$ is negative semi-definite. Since $\mathbb{V}\text{ar}\left[Y \mid Y \in B, \mu\right] - \Sigma$ is symmetric, it follows that it is negative semi-definite, as we desired to show. □

Lemma A.4. *Suppose that Σ satisfies Assumption 1. Then for ι the vector of ones and some $c_1 > 0$, $\iota'\Sigma_{22}^{-1} = c_1\iota'$. Additionally, $\Sigma_{12}\Sigma_{22}^{-1} = c_2\iota'$, for a constant $c_2 > 0$.*

Proof. First, note that if $K = 1$, then Σ_{12} and Σ_{22} are each positive scalars, and the result follows trivially. For the remainder of the proof, we therefore consider $K > 1$. Note that we can write $\Sigma_{22} = \Lambda + \rho\iota\iota'$, where $\Lambda = (\sigma^2 - \rho)I$. It follows from the Sherman-Morrison formula that

$$\begin{aligned} \Sigma_{22}^{-1} &= \Lambda^{-1} - \frac{\rho^2 \Lambda^{-1}\iota\iota'\Lambda^{-1}}{1 + \rho^2\iota'\Lambda^{-1}\iota} \\ &= (\sigma^2 - \rho)^{-1} I - \frac{\rho^2(\sigma^2-\rho)^{-2}\iota\iota'}{1 + \rho^2(\sigma^2-\rho)^{-1}\iota'\iota}. \end{aligned}$$

Thus:

$$\begin{aligned} &\iota'\Sigma_{22}^{-1} = \\ &\iota' \left((\sigma^2 - \rho)^{-1} I - \frac{\rho^2(\sigma^2-\rho)^{-2}\iota\iota'}{1 + \rho^2(\sigma^2-\rho)^{-1}\iota'\iota} \right) = \\ &(\sigma^2-\rho)^{-1} \left(1 - \frac{\rho^2(\sigma^2-\rho)^{-1}\iota'\iota}{1 + \rho^2(\sigma^2-\rho)^{-1}\iota'\iota} \right) \iota' = \\ &\underbrace{(\sigma^2-\rho)^{-1} \left(\frac{1}{1 + \rho^2(\sigma^2-\rho)^{-1}\iota'\iota} \right)}_{:= c_1} \iota'. \end{aligned}$$

Since $\sigma^2 - \rho > 0$, all of the terms in c_1 are positive, and thus $c_1 > 0$, as needed. Finally, note that Assumption 1 implies that $\Sigma_{12} = \rho\iota'$. It follows that $\Sigma_{12}\Sigma_{22}^{-1} = \rho c_1 \iota' = c_2\iota'$ for

$c_2 = \rho c_1 > 0$.

□

Lemma A.5. *Suppose $Y \sim N(0, \Sigma)$ is K-dimensional normal, with Σ satisfying the requirements on Σ_{22} imposed by Assumption 1. Let $B = \{y \in \mathbb{R}^K \,|\, a_j \leq y \leq b_j \text{ for all } j\}$, where $-b_j < a_j < b_j$ for all j. Then for ι the vector of ones, $\mathbb{E}[\iota' Y \,|\, Y \in B] = \mathbb{E}[Y_1 + \ldots + Y_k \,|\, Y \in B]$ is elementwise greater than 0.*

Proof. For any $x \in \mathbb{R}^K$ such that $x_j \leq b_j$ for all j, define $B^X(x) = \{y \in \mathbb{R}^K \,|\, x_j \leq y \leq b_j \text{ for all } j\}$. Let $b = (b_1, \ldots, b_K)$. Note that $B^X(-b)$ is a symmetric rectangular truncation around 0, so from Lemma A.2, we have that $\mathbb{E}\left[Y \,|\, Y \in B^X(-b)\right] = 0$. Now, define

$$g(x) = \mathbb{E}\left[\iota' Y \,|\, Y \in B^X(x)\right].$$

From the argument above, we have that $g(-b) = 0$, and we wish to show that $g(a) > 0$. By the mean-value theorem, for some $t \in (0,1)$,

$$\begin{aligned} g(a) &= g(-b) + (a - (-b))\, \nabla g\left(ta + (1-t)(-b)\right) \\ &= (a+b) \nabla g\left(ta + (1-t)(-b)\right) \\ &=: (a+b) \nabla g(x^t). \end{aligned}$$

By assumption, $(a+b)$ is elementwise greater than 0. It thus suffices to show that all elements of $\nabla g\left(x^t\right)$ are positive. Without loss of generality, we show that $\dfrac{\partial g(x^t)}{\partial x_K} > 0$.

Using the definition of the conditional expectation and Leibniz's rule, we have

$$\begin{aligned} &\frac{\partial g(x^t)}{\partial x_K} = \\ &\frac{\partial}{\partial x_K}\left[\left(\int_{x_1^t}^{b_1} \cdots \int_{x_K^t}^{b_K} (y_1 + \ldots + y_K)\, \phi_\Sigma(y)\, dy_1 \ldots dy_K\right)\left(\int_{x_1^t}^{b_1} \cdots \int_{x_K^t}^{b_K} \phi_\Sigma(y)\, dy_1 \ldots dy_K\right)^{-1}\right] = \\ &\left(\int_{x_1^t}^{b_1} \cdots \int_{x_K^t}^{b_K} (y_1 + \ldots + y_K)\, \phi_\Sigma(y)\, dy_1 \ldots dy_K \times \int_{x_1^t}^{b_1} \cdots \int_{x_{K-1}^t}^{b_{K-1}} \phi_\Sigma\left(\left(\begin{array}{c} y_{-K} \\ x_K^t \end{array}\right)\right) dy_1 \ldots dy_{K-1}\right. \\ &\left. - \int_{x_1^t}^{b_1} \cdots \int_{x_{K-1}^t}^{b_{K-1}} (y_1 + \ldots + y_{K-1} + x_K^t)\, \phi_\Sigma\left(\left(\begin{array}{c} y_{-K} \\ x_K^t \end{array}\right)\right) dy_1 \ldots dy_{K-1} \times \int_{x_1^t}^{b_1} \cdots \int_{x_K^t}^{b_K} \phi_\Sigma(y)\, dy_1 \ldots dy_K\right) \\ &\times \left(\int_{x_1^t}^{b_1} \cdots \int_{x_K^t}^{b_K} \phi_\Sigma(y)\, dy_1 \ldots dy_K\right)^{-2} \end{aligned} \tag{10}$$

where $\phi_\Sigma(y)$ denotes the PDF of a multivariate normal with mean 0 and variance Σ, and the second line uses the quotient rule. It follows from (10) that $\dfrac{\partial g(x^t)}{\partial x_K} > 0$ if and only if

$$\frac{\int_{x_1^t}^{b_1}\cdots\int_{x_k^t}^{b_K}(y_1+\ldots+y_K)\,\phi_\Sigma(y)\,dy_1\ldots dy_K}{\int_{x_1^t}^{b_1}\cdots\int_{x_k^t}^{b_K}\;\phi_\Sigma(y)\,dy_1\ldots dy_K} >$$
$$\frac{\int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}(y_1+\cdots+y_{K-1}+x_K^t)\,\phi_\Sigma\left(\left(\begin{array}{c} y_{-K} \\ x_K^t \end{array}\right)\right)\,dy_1\ldots dy_{K-1}}{\int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}\;\phi_\Sigma\left(\left(\begin{array}{c} y_{-K} \\ x_K^t \end{array}\right)\right)\,dy_1\ldots dy_{K-1}}$$

or equivalently,

$$\mathbb{E}\left[Y_1+\ldots+Y_K\,|\,x_j^t\le Y_j\le b_j,\forall j\right] > \mathbb{E}\left[Y_1+\ldots+Y_K\,|\,x_j^t\le Y_j\le b_j,\text{ for } j<K,\,Y_K=x_K^t\right].$$

It is clear that $\mathbb{E}\left[Y_K\,|\,x_j^t\le Y_j\le b_j,\forall j\right] > x_K^t$, since $x_K^t < b_K$ and the Kth marginal density of the rectangularly-truncated normal distribution is positive for all values in $[x_K^t, b_K]$ (see Cartinhour (1990)). This completes the proof for the case where $K=1$. For $K>1$, it suffices to show that

$$\mathbb{E}\left[Y_1+\ldots+Y_{K-1}\,|\,x_j^t\le Y_j\le b_j,\forall j\right] \ge \mathbb{E}\left[Y_1+\ldots+Y_{K-1}\,|\,x_j^t\le Y_j\le b_j,\text{ for } j<K,\,Y_K=x_K^t\right]. \tag{11}$$

To see why (11) holds, let $\tilde{Y}_{-K} = Y_{-K}-\Sigma_{-K,K}\Sigma_{K,K}^{-1}Y_K$, where a "$-K$" subscript denotes all of the indices except for K. By an argument analogous to that in the Proof of Lemma A.1 for $\tilde{\beta}_{post}$, one can easily verify that $\tilde{Y}_{-K}$ is independent of Y_K and $\tilde{Y}_{-K}\sim\mathcal{N}\left(0,\,\tilde{\Sigma}\right)$ for $\tilde{\Sigma}=\Sigma_{-K,-K}-\Sigma_{-K,K}\Sigma_{K,K}^{-1}\Sigma_{K,-K}$. By construction, $Y_{-K}=\tilde{Y}_{-K}+\Sigma_{-K,K}\Sigma_{K,K}^{-1}Y_K$, from which it follows that

$$Y_{-K}\,|\,Y_K=y_K\sim\mathcal{N}\left(\Sigma_{-K,K}\Sigma_{K,K}^{-1}y_K,\,\tilde{\Sigma}\right).$$

We now argue that $\Sigma_{-K,K}\Sigma_{K,K}^{-1}y_K = c\,y_K\,\iota$ for a positive constant c. If $K=2$, then by Assumption 1, $\Sigma_{-K,K}\Sigma_{K,K}^{-1}=\rho/\sigma^2$ is the product of two positive scalars, and can thus be trivially written as $c\iota$. For $K>2$, we verify that $\tilde{\Sigma}$ meets the requirements that Assumption 1 places on Σ_{22}, and then apply Lemma A.4 to obtain the desired result. To do this, note that by Assumption 1, Σ has common terms σ^2 on the diagonal and ρ on the off-diagonal, and thus the same holds for $\Sigma_{-K,-K}$. Additionally, under Assumption 1, $\Sigma_{-K,K}=\rho\iota$ and $\Sigma_{K,K}^{-1}=\frac{1}{\sigma^2}$, so $\Sigma_{-K,K}\Sigma_{K,K}^{-1}\Sigma_{K,-K}$ equals ρ^2/σ^2 times $\iota\iota'$, the matrix of ones. The diagonal terms of $\tilde{\Sigma} = \Sigma_{-K,-K} - \Sigma_{-K,K}\Sigma_{K,K}^{-1}\Sigma_{K,-K}$ are thus equal to $\tilde{\sigma}^2 = \sigma^2 - \rho^2/\sigma^2$, and the off-diagonal terms are equal to $\tilde{\rho} = \rho - \rho^2/\sigma^2$, or equivalently $\tilde{\rho} = \rho(1 - \rho/\sigma^2)$. Since by Assumption 1, $0 < \rho < \sigma^2$, it is clear that $\tilde{\sigma}^2 > \tilde{\rho}$. Additionally, $0 < \rho < \sigma^2$ implies that $1 - \rho/\sigma^2 > 0$, and hence $\tilde{\rho} > 0$, which completes the proof that $\tilde{\Sigma}$ satisfies the requirements of Assumption 1 for Σ_{22}. Hence, $\Sigma_{-K,K}\Sigma_{K,K}^{-1}y_K = c\, y_K\, \iota$ by Lemma A.4. We can therefore write

$$Y_{-K} \,|\, Y_K = y_K \sim \mathcal{N}\left(c\, y_K\, \iota,\, \tilde{\Sigma}\right).$$

Let $h(\mu) = \mathbb{E}\left[X | X \in B_{-K},\, X \sim \mathcal{N}\left(\mu,\, \tilde{\Sigma}\right)\right]$ for $B_{-K} = \{\tilde{x} \in \mathbb{R}^{K-1} | x_j^t \leq \tilde{x}_j \leq b_j, \text{ for } j = 1, \ldots, K-1\}$. Then the previous display implies $\mathbb{E}\left[\iota' Y_{-K} \,|\, x_j^t \leq Y_j \leq b_j \text{ for } j < K,\, Y_K = y_K\right] = \iota' h(c y_K \iota)$. Hence,

$$\begin{aligned}
\frac{\partial}{\partial y_K}\mathbb{E}\left[\iota' Y_{-K} \,|\, x_j^t \leq Y_j \leq b_j \text{ for } j < K,\, Y_K = y_K\right] &= \iota' \left(D_\mu h|_{\mu = c y_K \iota}\right) \iota \cdot c \\
&= \iota' \mathbb{V}\text{ar}\left[Y_{-K} \,|\, Y_{-K} \in B_{-K}, Y_K = y_K\right] \tilde{\Sigma}^{-1} \iota c \\
&= \iota' \mathbb{V}\text{ar}\left[Y_{-K} \,|\, Y_{-K} \in B_{-K}, Y_K = y_K\right] \iota c_1 c \\
&\geq 0
\end{aligned}$$

where the second line follows from Lemma A.3; the third line uses Lemma A.4 to obtain that $\tilde{\Sigma}^{-1}\iota = \iota c_1$ for $c_1 > 0$ (if $K = 2$, this holds trivially); and the inequality follows from the fact that $\mathbb{V}\text{ar}\left[Y_{-K} \,|\, Y_{-K} \in B_{-K}, Y_K = y_K\right]$ is positive semi-definite and c_1 and c are positive by construction. Thus, for all $y_K \in [x_k^t, b_k]$,

$$\begin{aligned}
&\mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j \text{ for } j < K,\, Y_K = y_K\right] \geq \\
&\mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j \text{ for } j < K,\, Y_K = x_k^t\right].
\end{aligned}$$

By the law of iterated expectations, we have

$$\begin{aligned}
&\mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j, \forall j\right] = \\
&\mathbb{E}\left[\mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j \text{ for } j < K, Y_K\right] \,|\, x_j^t \leq Y_j \leq b_j, \forall j\right] \geq \\
&\mathbb{E}\left[\mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j \text{ for } j < K, Y_K = x_K^t\right] \,|\, x_j^t \leq Y_j \leq b_j, \forall j\right] = \\
&\mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j \text{ for } j < K, Y_K = x_K^t\right],
\end{aligned}$$

as we wished to show.

□

Proof of Proposition 3.2 From Proposition 3.1, the desired result is equivalent to showing that

$$\Sigma_{12}\Sigma_{22}^{-1}\,\mathbb{E}\left[\hat{\beta}_{pre} - \beta_{pre} \,|\, \hat{\beta}_{pre} \in B\right] > 0.$$

By Lemma A.4, $\Sigma_{12}\Sigma_{22}^{-1} = c_1\iota'$ for $c_1 > 0$, so it suffices to show that $\iota'\mathbb{E}\left[\hat{\beta}_{pre} - \beta_{pre} \,|\, \hat{\beta}_{pre} \in B\right]$ is elementwise greater than zero. Note that by assumption $(\hat{\beta}_{pre} - \beta_{pre}) \sim \mathcal{N}(0, \Sigma_{22})$. Additionally, observe that $\hat{\beta}_{pre} \in B_{NIS} = \{\hat{\beta}_{pre} : |\hat{\beta}_{pre,j}|/\sqrt{\Sigma_{jj}} \leq c_\alpha \text{ for all } j\}$ iff $(\hat{\beta}_{pre} - \beta_{pre}) \in \tilde{B}_{NIS} = \{\beta : a_j \leq \beta_j \leq b_j\}$ for $a_j = -c_\alpha\sqrt{\Sigma_{jj}} - \beta_{pre,j}$ and $b_j = c_\alpha\sqrt{\Sigma_{jj}} - \beta_{pre,j}$. Since $\beta_{pre,j} < 0$ for all j, we have that $-b_j < a_j < b_j$ for all j. The result then follows immediately from Lemma A.5.

Proof of Proposition 3.3 Note that since $\hat{\beta}_{post} = \tilde{\beta}_{post} + \Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre}$, for any set S,

$$\begin{aligned}\mathbb{V}\text{ar}\left[\hat{\beta}_{post} \,|\, \hat{\beta}_{pre} \in S\right] &= \mathbb{V}\text{ar}\left[\tilde{\beta}_{post} \,|\, \hat{\beta}_{pre} \in S\right] + \mathbb{V}\text{ar}\left[\Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in S\right] \\ &\quad + 2\,\text{Cov}\left(\tilde{\beta}_{post}\,,\, \Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in S\right) \\ &= \mathbb{V}\text{ar}\left[\tilde{\beta}_{post}\right] + \mathbb{V}\text{ar}\left[\Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in S\right], \qquad (12)\end{aligned}$$

where we use the independence of $\tilde{\beta}_{post}$ and $\hat{\beta}_{pre}$ from Lemma A.1 to obtain that $\mathbb{V}\text{ar}\left[\tilde{\beta}_{post} \,|\, \hat{\beta}_{pre} \in B\right] = \mathbb{V}\text{ar}\left[\tilde{\beta}_{post}\right]$ and that the covariance term equals 0. Applying equation (12) for $S = B$ and for $S = \mathbb{R}^K$, and then taking the difference between the two, we have

$$\begin{aligned}\mathbb{V}\text{ar}\left[\hat{\beta}_{post}|\hat{\beta}_{pre} \in B\right] - \mathbb{V}\text{ar}\left[\hat{\beta}_{post}\right] &= \mathbb{V}\text{ar}\left[\Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] - \mathbb{V}\text{ar}\left[\Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre}\right] \\ &= (\Sigma_{12}\Sigma_{22}^{-1})\left(\mathbb{V}\text{ar}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] - \mathbb{V}\text{ar}\left[\hat{\beta}_{pre}\right]\right)(\Sigma_{12}\Sigma_{22}^{-1})',\end{aligned}$$

which gives the desired result.

Proof of Proposition 3.4 By Proposition 3.3, it suffices to show that

$$(\Sigma_{12}\Sigma_{22}^{-1})\left(\mathbb{V}\text{ar}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] - \mathbb{V}\text{ar}\left[\hat{\beta}_{pre}\right]\right)(\Sigma_{12}\Sigma_{22}^{-1})' \leq 0.$$

The result then follows immediately from the fact that $\mathbb{V}\text{ar}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] - \mathbb{V}\text{ar}\left[\hat{\beta}_{pre}\right]$ is negative semi-definite by Lemma A.3. □

B Uniform Asymptotic Results

In the main text of the paper, I consider a finite sample normal model for the event-study coefficients, which I use to evaluate the distribution of the event-study estimates conditional on passing a pre-test for the pre-period coefficients. In this section, I show that these finite-sample results translate to uniform asymptotic results over a large class of data-generating processes in which the probability of passing the pre-test does not go to zero asymptotically, i.e. when the pre-trend is $O(n^{-\frac{1}{2}})$.

B.1 Assumptions

We consider a class of data-generating processes $\mathcal{P}$. Let $\hat{\beta}_n = \sqrt{n}\hat{\beta}$ be the event-study estimates $\hat{\beta} = \begin{pmatrix} \hat{\beta}_{post} \\ \hat{\beta}_{pre} \end{pmatrix}$ scaled by $\sqrt{n}$. Likewise, let $\tau_{P,n} = \sqrt{n} \begin{pmatrix} \tau_{post}(P) \\ 0 \end{pmatrix}$ be the scaled vector of treatment effects under data-generating process $P \in \mathcal{P}$, where we assume there is no true effect of treatment in the pre-periods.

Assumption 2 (Unconditional uniform convergence). *Let BL_1 denote the set of Lipschitz functions which are bounded by 1 in absolute value and have Lipschitz constant bounded by 1. We assume*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{f\in BL_1} \left|\left| \mathbb{E}_P \left[f(\hat{\beta}_n - \tau_{P,n}) \right] - \mathbb{E}\left[f(\xi_{P,n}) \right] \right|\right| = 0,$$

where $\xi_{P,n} \sim \mathcal{N}(\delta_{P,n}, \Sigma_P)$.

Convergence in distribution is equivalent to convergence in bounded Lipschitz metric (see Theorem 1.12.4 in van der Vaart and Wellner (1996)), so Assumption 2 formalizes the notion of uniform convergence in distribution of $\hat{\beta}_n - \tau_{P,n}$ to a $\mathcal{N}(\delta_{P,n}, \Sigma_P)$ variable under P. Note that we allow δ to depend both on P and the sample size n.

We next assume that we have a uniformly consistent estimator of the variance Σ_P, and that the eigenvalues of Σ_P are bounded above and away from singularity.

Assumption 3 (Consistent estimation of Σ_P). *Our estimator $\hat{\Sigma}$ is uniformly consistent for Σ_P,*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(||\hat{\Sigma}_n - \Sigma_P|| > \epsilon \right) = 0,$$

for all $\epsilon > 0$.

Assumption 4 (Assumptions on Σ_P). *We assume that there exists $\bar{\lambda} > 0$ such that for all $P \in \mathcal{P}$, $\Sigma_P \in \mathcal{S} := \{\Sigma \,|\, 1/\bar{\lambda} \leq \lambda_{min}(\Sigma) \leq \lambda_{max}(\Sigma) \leq \bar{\lambda}\}$, where $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the minimal and maximal eigenvalues of a matrix A.*

Next, we assume that the pre-test takes the form of a polyhedral restriction on the vector of pre-period coefficients. Note that the test that no pre-period coefficient be individually significant can be written in this form.

Assumption 5 (Assumptions on B)**.** *We assume that the conditioning set $B(\Sigma)$ is of the form $B(\Sigma) = \{(\beta_{post}, \beta_{pre}) \mid A_{pre}(\Sigma)\beta_{pre} \leq b(\Sigma)\}$ for continuous functions A_{pre} and b. We further assume that for all Σ on an open set containing $\mathcal{S}$, $B(\Sigma)$ is bounded and has non-empty interior, and $A_{pre}(\Sigma)$ has no all-zero rows.*

For ease of notation, it will be useful to define $A(\Sigma) = [0, A_{pre}(\Sigma)]$, so that $\beta \in B(\Sigma)$ iff $A(\Sigma)\beta \leq b(\Sigma)$.

B.2 Main uniform asymptotic results

Our first result concerns the asymptotic distribution of the event-study coefficients *conditional* on passing the pre-test.

Proposition B.1 (Uniform conditional convergence in distribution)**.** *Under Assumptions 2-5,*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{f\in BL_1} \left|\left|\mathbb{E}_P\left[f(\hat{\beta}_n - \tau_{P,n}) \mid \hat{\beta}_n \in B(\hat{\Sigma}_n)\right] - \mathbb{E}\left[f(\xi_{P,n}) | \xi_{P,n} \in B(\Sigma_P)\right]\right|\right| \mathbb{P}_P\left(\hat{\beta}_n \in B(\hat{\Sigma}_n)\right) = 0,$$

where $\xi_{P,n} \sim \mathcal{N}\left(\delta_{P,n}, \Sigma_P\right)$.

Note that if we removed the $\mathbb{P}_P\left(\hat{\beta}_n \in B(\hat{\Sigma}_n)\right)$ term from the statement of Proposition B.1, then the proposition would imply uniform convergence in distribution of $(\hat{\beta}_n - \tau_{P,n}) | \hat{\beta}_n \in B(\hat{\Sigma}_n)$ to $\xi_{P,n} | \xi_{P,n} \in B(\Sigma_P)$. The Proposition thus guarantees such convergence in distribution along any sequence of distributions for which the probability of passing the pre-test is not going to zero.

Although Proposition B.1 gives uniform convergence of the treatment effect estimates conditional on passing the pre-test, it is well known that convergence in distribution need not imply convergence in expectations. Our next result shows that under the additional assumption of asymptotic uniform integrability, we also obtain uniform convergence in expectations, provided that the probability of passing the pre-test is not going to zero.

Proposition B.2 (Uniform convergence of expectations)**.** *Suppose Assumptions 2-5 hold. Let $\beta_{P,n} = \tau_{P,n} + \delta_{P,n}$. Assume that $\hat{\beta}_n - \beta_{P,n}$ is asymptotically uniformly integrable over the class $\mathcal{P}$,*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P \left[||\hat{\beta}_n - \beta_{P,n}|| \cdot 1[||\hat{\beta}_n - \beta_{P,n}|| > M] \right] = 0.$$

Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} 1 \left[\left|\left| \mathbb{E}_P \left[\hat{\beta}_n - \tau_{P,n} \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n) \right] - \mathbb{E}\left[\xi_{P,n} \,|\, \xi_{P,n} \in B(\Sigma_P) \right] \right|\right| > \epsilon \right] \mathbb{P}_P \left(\hat{\beta}_n \in B(\hat{\Sigma}_n) \right) = 0,$$

where $\xi_{P,n} \sim \mathcal{N}\left(\delta_{P,n}, \Sigma_P\right)$.

B.3 Proofs of main asymptotic results

Proof of Proposition B.1 Towards contradiction, suppose that the proposition is false. Then there exists an increasing sequence of sample sizes n_m and data-generating processes P_{n_m} such that

$$\liminf_{m\to\infty} \sup_{f\in BL_1} \left|\left| \mathbb{E}_{P_n} \left[f(\hat{\beta}_n - \tau_{P_{n_m},n_m}) \,|\, \hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m}) \right] - \mathbb{E}\left[f(\xi_{P_{n_m},n_m}) | \xi \in B(\Sigma_{P_{n_m}}) \right] \right|\right| \times$$
$$\mathbb{P}_{P_{n_m}} \left(\hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m}) \right) > 0. \tag{13}$$

Since the interval $[0,1]$ is compact, there exists a subsequence of increasing sample sizes, n_q, such that

$$\lim_{q\to\infty} \mathbb{P}_{P_{n_q}} \left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q}) \right) = p^*,$$

for $p^* \in [0,1]$.

Suppose first that $p^* = 0$. Note that by definition, a function $f \in BL_1$ is bounded in absolute value by 1. It then follows from the triangle inequality that for all $f \in BL_1$,

$$\left|\left| \mathbb{E}_{P_{n_q}} \left[f(\hat{\beta}_{n_q} - \tau_{P_{n_q},n_q}) \,|\, \hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q}) \right] - \mathbb{E}\left[f(\xi_{P_{n_q},n_q}) | \xi_{P_{n_q},n_q} \in B(\Sigma_{P_{n_q}}) \right] \right|\right| \le 2$$

for all q. But this implies that

$$\liminf_{q\to\infty} \sup_{f\in BL_1} \left|\left| \mathbb{E}_{P_{n_q}} \left[f(\hat{\beta}_{n_q} - \tau_{P_{n_q},n_q}) \,|\, \hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q}) \right] - \mathbb{E}\left[f(\xi_{P_{n_q}}) | \xi_{P_{n_q}} \in B(\Sigma_{P_{n_q}}) \right] \right|\right| \mathbb{P}_{P_{n_q}} \left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q}) \right)$$
$$\le 2p^* = 0,$$

which contradicts (13).

Now, suppose $p^* > 0$. Note that by Assumption 4, Σ_P falls in the set $\mathcal{S} = \{\Sigma | 1/\bar{\lambda} \le \lambda_{min}(\Sigma) \le \lambda_{max}(\Sigma) \le \bar{\lambda}\}$, which is compact (e.g., in the Frobenius norm). Thus, we can extract a further subsequence of increasing sample sizes, n_r, such that

$$\lim_{r\to\infty} \Sigma_{P_{n_r}} = \Sigma^*,$$

for some $\Sigma^* \in \mathcal{S}$.

Additionally, since $p^* > 0$, Lemma B.4 implies that $\delta^{pre}_{P_{n_r},n_r}$ is bounded, and thus we can extract a further subsequence n_s along which

$$\lim_{s\to\infty} \delta^{pre}_{P_{n_s},n_s} = \delta^{pre,*}.$$

By Lemma B.3, for $\delta^+_{n_s} = \begin{pmatrix} \delta^{post}_{P_{n_s},n_s} \\ 0 \end{pmatrix}$, $\delta^* = \begin{pmatrix} 0 \\ \delta^{pre,*} \end{pmatrix}$, and $\xi^* \sim \mathcal{N}\left(\delta^*, \Sigma^*\right)$, we have

$$(\hat{\beta}_{n_s} - \tau_{P,n_s} - \delta^+_{n_s})|\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s}) \xrightarrow{d} \xi^*|\xi^* \in B(\Sigma^*),$$

and

$$(\xi_{P_{n_s}} - \delta^+_{n_s})|\xi_{P_{n_s}} \in B(\Sigma_{P_{n_s}}) \xrightarrow{d} \xi^*|\xi^* \in B(\Sigma^*).$$

Recalling the convergence in distribution is equivalent to convergence in bounded Lipschitz metric, we see that

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s}) \,|\, \hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right] - \mathbb{E}\left[f(\xi^*)|\xi^* \in B(\Sigma^*)\right]\right|\right| = 0 \quad (14)$$

and

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^+_{n_s})|\xi_{P_{n_s}} \in B(\Sigma_{P_{n_s}})\right] - \mathbb{E}\left[f(\xi^*)|\xi^* \in B(\Sigma^*)\right]\right|\right| = 0. \quad (15)$$

Equations (14) and (15) together with the triangle inequality then imply that

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s}) \,|\, \hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right] - \mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^+_{n_s})|\xi_{P_{n_s}} \in B(\Sigma_{P_{n_s}})\right]\right|\right| = 0.$$

However, BL_1 is closed under horizontal transformation (i.e. $f(x) \in BL_1$ implies $f(x-c) \in BL_1$), and so this implies that

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s}) \,|\, \hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right] - \mathbb{E}\left[f(\xi_{P_{n_s}})|\xi_{P_{n_s}} \in B(\Sigma_{P_{n_s}})\right]\right|\right| = 0,$$

which contradicts (13). □

Proof of Proposition B.2 Towards contradiction, suppose the proposition is false. Then there exists an increasing sequence of sample sizes n_m and data-generating processes P_{n_m} such that for some $\epsilon > 0$,

$$\liminf_{m\to\infty} 1\left[\left|\left|\mathbb{E}\left[\hat{\beta}_{n_m} - \tau_{P_{n_m},n_m} \,|\, \hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m})\right] - \mathbb{E}\left[\xi_{P_{n_m}} \,|\, \xi_{P_{n_m}} \in B(\Sigma_{P_{n_m}})\right]\right|\right| > \epsilon\right] \times \\ \mathbb{P}_{P_{n_m}}\left(\hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m})\right) > 0. \tag{16}$$

Since the interval $[0,1]$ is compact, we can extract a subsequence of increasing sample sizes, n_q, along which

$$\lim_{q\to\infty} \mathbb{P}_{P_{n_q}}\left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q})\right) = p^*$$

for $p^* \in [0,1]$.

First, suppose $p^* = 0$. Since the indicator function is bounded by 1,

$$\liminf_{s\to\infty} 1\left[\left|\left|\mathbb{E}\left[\hat{\beta}_{n_q} - \tau_{P_{n_q},n_q} \,|\, \hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q})\right] - \mathbb{E}\left[\xi_{P_{n_q}} \,|\, \xi_{P_{n_q}} \in B(\Sigma_{P_{n_q}})\right]\right|\right| > \epsilon\right] \mathbb{P}_{P_{n_q}}\left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q})\right) \leq \\ \liminf_{s\to\infty} \mathbb{P}_{P_{n_q}}\left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q})\right) = p^* = 0,$$

which contradicts (16).

Now, suppose $p^* > 0$. As argued in the proof to Proposition B.1, we can iteratively extract subsequences to obtain a subsequence, n_s, along which

$$\begin{aligned}
&\lim_{s\to\infty} \Sigma_{P_{n_s}} = \Sigma^*, \\
&\lim_{s\to\infty} \delta^{pre}_{P_{n_s},n_s} = \delta^{pre,*}, \\
&\lim_{s\to\infty} \mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) = p^* > 0,
\end{aligned}$$

where $\Sigma^* \in \mathcal{S}$.

Let $\delta^-_{n_s} = \begin{pmatrix} 0 \\ \delta^{pre}_{P_{n_s},n_s} \end{pmatrix}$ and $\delta^* = \begin{pmatrix} 0 \\ \delta^{pre,*} \end{pmatrix}$ be the vectors with zeros for the post-period coefficients and $\delta^{pre}_{P_{n_s},n_s}$ and $\delta^{pre,*}$, respectively, for the pre-period coefficients. Similarly, let $\delta^+_{n_s} = \begin{pmatrix} \delta^{post}_{P_{n_s},n_s} \\ 0 \end{pmatrix}$ be the vector with zeros for the pre-period coefficients and $\delta^{post}_{P_{n_s},n_s}$ for the post-period coefficients. From Lemma B.3, $(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s})|\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s}) \xrightarrow{d} \xi^*|\xi^* \in B(\Sigma^*)$, for $\xi^* \sim \mathcal{N}(\delta^*, \Sigma^*)$.

Additionally, from uniform integrability, we have

$$\lim_{M\to\infty}\limsup_{s\to\infty}\mathbb{E}_{P_{n_s}}\left[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||\cdot 1[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||>M]\right]=0.$$

Observe that

$$\begin{aligned}
&\mathbb{E}_{P_{n_s}}\left[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||\cdot 1[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||>M]\right]= \\
&\mathbb{E}_{P_{n_s}}\left[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||\cdot 1[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||>M]\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]\cdot\mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right)+ \\
&\mathbb{E}_{P_{n_s}}\left[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||\cdot 1[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||>M]\,|\,\hat{\beta}_{n_s}\notin B(\hat{\Sigma}_{n_s})\right]\cdot\mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s}\notin B(\hat{\Sigma}_{n_s})\right)\geq \\
&\mathbb{E}_{P_{n_s}}\left[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||\cdot 1[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||>M]\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]\cdot\mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right),
\end{aligned}$$

and hence

$$\lim_{M\to\infty}\limsup_{s\to\infty}\mathbb{E}_{P_{n_s}}\left[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||\cdot 1[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||>M]\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]\cdot\mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right)=0.$$

Further, since $\mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right)\to p^*>0$, it follows that

$$\lim_{M\to\infty}\limsup_{s\to\infty}\mathbb{E}_{P_{n_s}}\left[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||\cdot 1[||\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}||>M]\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]=0,$$

so $\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}$ is uniformly asymptotically integrable conditional on $\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})$. Note that $\hat{\beta}_{n_s}-\tau_{P_{n_s},n_s}-\delta^+_{n_s}=\hat{\beta}_{n_s}-\beta_{P_{n_s},n_s}+\delta^-_{n_s}$, and $\delta^-_{n_s}\to\delta^*$ as $s\to\infty$. It then follows from Lemma B.6 that $\hat{\beta}_{n_s}-\tau_{P_{n_s},n_s}-\delta^+_{n_s}$ is uniformly asymptotically integrable conditional on $\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})$.

Convergence in distribution along with uniform asymptotic integrability implies convergence in expectation (see Theorem 2.20 in van der Vaart (2000)), and thus

$$\lim_{s\to\infty}\left|\left|\mathbb{E}_{P_{n_s}}\left[\hat{\beta}_{n_s}-\tau_{P_{n_s},n_s}-\delta^+_{n_s}\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]-\mathbb{E}\left[\xi^*\,|\,\xi^*\in B(\Sigma^*)\right]\right|\right|=0.$$

Likewise, Lemma B.5 gives that

$$\lim_{s\to\infty}\left|\left|\mathbb{E}\left[\xi_{P_{n_s}}-\delta^+_{n_s}\,|\,\xi_{P_{n_s}}\in B(\Sigma_{P_{n_s}})\right]-\mathbb{E}\left[\xi^*\,|\,\xi^*\in B(\Sigma^*)\right]\right|\right|=0.$$

It then follows from the triangle inequality that

$$\lim_{s\to\infty}\left|\left|\mathbb{E}_{P_{n_s}}\left[\hat{\beta}_{n_s}-\tau_{P_{n_s},n_s}-\delta^+_{n_s}\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]-\mathbb{E}\left[\xi_{P_{n_s}}-\delta^+_{n_s}\,|\,\xi_{P_{n_s}}\in B(\Sigma_{P_{n_s}})\right]\right|\right|=0.$$

Cancelling the $\delta^+_{n_s}$ terms gives

$$\lim_{s\to\infty}\left|\left|\mathbb{E}_{P_{n_s}}\left[\hat{\beta}_{n_s}-\tau_{n_s,P_{n_s}}\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]-\mathbb{E}\left[\xi_{P_{n_s}}\,|\,\xi_{P_{n_s}}\in B(\Sigma_{P_{n_s}})\right]\right|\right|=0,$$

which contradicts (16). □

B.4 Auxiliary lemmas and proofs

Lemma B.1. *Suppose* $(\xi_n,\Sigma_n)\xrightarrow{d}(\xi^*,\Sigma^*)$, *for* $\xi^*\sim\mathcal{N}(\delta^*,\,\Sigma^*)$ *and* $\Sigma^*\in\mathcal{S}$. *Then, if* B *satisfies Assumption 5,*

$$\mathbb{P}_{P_n}\left(\xi_n\in B(\Sigma_n)\right)\longrightarrow\mathbb{P}\left(\xi^*\in B(\Sigma^*)\right).$$

Proof. By definition, $\xi_n\in B(\Sigma_n)$ iff $A(\Sigma_n)\xi_n\leq b(\Sigma_n)$. Now, consider the function

$$h(\xi,\Sigma)=1[A(\Sigma)\xi\leq b(\Sigma)].$$

Note that since $A(\cdot)$ and $b(\cdot)$ are continuous by Assumption 5, h is continuous at all (ξ,Σ) such that for all j, $(A(\Sigma)\xi)_j\neq b(\Sigma)_j$. However, the jth element of $A(\Sigma^*)\xi^*$ is normally distributed with variance $A(\Sigma^*)_{(j,\cdot)}\Sigma^*A(\Sigma^*)'_{(j,\cdot)}$, where $X_{(j,\cdot)}$ denotes the jth row of a matrix X. Since $A(\Sigma^*)$ has no non-zero rows by Assumption 5, and $\Sigma^*\in\mathcal{S}$ implies that Σ^* is positive definite, $A(\Sigma^*)_{(j,\cdot)}\Sigma^*A(\Sigma^*)'_{(j,\cdot)}>0$. This implies that for each j, $(A(\Sigma^*)\xi^*)_j=b(\Sigma^*)_j$ with probability zero, and hence $(A(\Sigma^*)\xi^*)_j\neq b(\Sigma^*)_j$ for all j with probability 1. Thus, h is continuous at (ξ^*,Σ^*) for almost every ξ.

Since $(\xi_n,\Sigma_n)\xrightarrow{d}(\xi^*,\Sigma^*)$, the Continuous Mapping Theorem gives that $1[A(\Sigma_n)\xi_n\leq b(\Sigma_n)]\xrightarrow{d}1[A(\Sigma^*)\xi^*\leq b(\Sigma^*)]$. Since the indicator functions are bounded, it follows that

$$\mathbb{P}\left(\xi_n\in B(\Sigma_n)\right)=\mathbb{E}\left[1[A(\Sigma_n)\xi_n\leq b(\Sigma_n)]\right]\longrightarrow\mathbb{E}\left[1[A(\Sigma^*)\xi^*\leq b(\Sigma^*)]\right]=\mathbb{P}\left(\xi^*\in B(\Sigma^*)\right),$$

which completes the proof. □

Lemma B.2. *Suppose that* $(\xi_n,\Sigma_n)\xrightarrow{d}(\xi^*,\Sigma^*)$, *for* $\xi^*\sim\mathcal{N}(\delta^*,\,\Sigma^*)$ *and* $\Sigma^*\in\mathcal{S}$. *Suppose further that* $\mathbb{P}\left(\xi^*\in B(\Sigma^*)\right)=p^*>0$ *for* $B(\Sigma)$ *satisfying Assumption 5. Then*

$$\xi_n\,|\,\xi_n\in B(\Sigma_n)\xrightarrow{d}\xi^*\,|\,\xi^*\in B(\Sigma^*).$$

Proof. By the Portmanteau Lemma (see Lemma 2.2. in van der Vaart (2000)),

$$\xi_n\,|\,\xi_n\in B(\Sigma_n)\xrightarrow{d}\xi^*\,|\,\xi^*\in B(\Sigma^*)$$

iff $\mathbb{E}\left[f(\xi_n) \,|\, \xi_n \in B(\Sigma_n)\right] \longrightarrow \mathbb{E}\left[f(\xi^*) \,|\, \xi^* \in B(\Sigma^*)\right]$ for all bounded, continuous functions f.

Let f be a bounded, continuous function. Since $(\xi_n, \Sigma_n) \xrightarrow{d} (\xi^*, \Sigma^*)$, the Continuous Mapping Theorem together with the Dominated Convergence Theorem imply that $\mathbb{E}\left[g(\xi_n, \Sigma_n)\right] \xrightarrow{p} \mathbb{E}\left[g(\xi^*, \Sigma^*)\right]$ for any bounded function g that is continuous for almost every (ξ^*, Σ^*). It follows that

$$\mathbb{E}\left[f(\xi_n) \cdot 1\left[\xi_n \in B(\Sigma_n)\right]\right] \longrightarrow \mathbb{E}\left[f\left(\xi^*\right) \cdot 1\left[\xi^* \in B(\Sigma^*)\right]\right],$$

where we use the fact that the function $1[\xi \in B(\Sigma)]$ is continuous at (ξ^*, Σ^*) for almost every ξ^*, as shown in the proof to Lemma B.1, and that the product of bounded and continuous functions is bounded and continuous. Additionally, by Lemma B.1, we have that

$$\mathbb{P}\left(\xi_n \in B(\xi_n)\right) \longrightarrow \mathbb{P}\left(\xi^* \in B(\Sigma^*)\right) = p^* > 0.$$

We can thus apply the Continuous Mapping Theorem to obtain

$$\frac{\mathbb{E}\left[f(\xi_n) \cdot 1\left[\xi_n \in B(\Sigma_n)\right]\right]}{\mathbb{P}\left(\xi_n \in B(\Sigma_n)\right)} \longrightarrow \frac{\mathbb{E}\left[f\left(\xi^*\right) \cdot 1\left[\xi^* \in B(\Sigma^*)\right]\right]}{\mathbb{P}\left(\xi^* \in B(\Sigma^*)\right)},$$

which by the definition of the conditional expectation, implies

$$\mathbb{E}\left[f(\xi_n) \,|\, \xi_n \in B(\Sigma_n)\right] \longrightarrow \mathbb{E}\left[f(\xi^*) \,|\, \xi^* \in B(\Sigma^*)\right],$$

as needed. □

Lemma B.3. *Suppose Assumptions 2-5 hold, and n_s is an increasing sequence of sample sizes such that*

$$\lim_{s\to\infty} \Sigma_{P_{n_s}} = \Sigma^*,$$
$$\lim_{s\to\infty} \delta^{pre}_{P_{n_s},n_s} = \delta^{pre,*},$$
$$\lim_{s\to\infty} \mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) = p^* > 0$$

for $\Sigma^* \in \mathcal{S}$. Let $\delta^+_{n_s} = \begin{pmatrix} \delta^{post}_{P_{n_s},n_s} \\ 0 \end{pmatrix}$ be the vector with elements corresponding with $\delta_{P_{n_s},n_s}$ for the post-period coefficients, and zeros for the pre-period coefficients. Likewise, let $\delta^* = \begin{pmatrix} 0 \\ \delta^{pre,*} \end{pmatrix}$ be the vector with zeros for the post-period coefficients and $\delta^{pre,*}$ for the pre-period *coefficients. Then*

$$(\hat{\beta}_{n_s} - \tau_{P,n_s} - \delta^{+}_{n_s}) | \hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s}) \xrightarrow{d} \xi^* | \xi^* \in B(\Sigma^*)$$

and

$$(\xi_{P_{n_s},n_s} - \delta^{+}_{n_s}) \,|\, \xi_{P_{n_s},n_s} \in B(\Sigma_{P_{n_s}}) \xrightarrow{d} \xi^* | \xi^* \in B(\Sigma^*),$$

for $\xi^* \sim \mathcal{N}\left(\delta^*,\, \Sigma^*\right)$.

Proof. By assumption, $\xi_{P_{n_s}} \sim \mathcal{N}\left(\delta_{P_{n_s}},\, \Sigma_{P_{n_s}}\right)$, and thus $\xi_{P_{n_s}} - \delta^{+}_{n_s} \sim \mathcal{N}\left(\delta^{-}_{n_s},\, \Sigma_{P_{n_s}}\right)$. Since by construction $\delta^{-}_{n_s} \longrightarrow \delta^*$ and $\Sigma_{P_{n_s}} \longrightarrow \Sigma^*$, it follows that $\xi_{P_{n_s}} - \delta^{+}_{n_s} \xrightarrow{d} \xi^*$, for $\xi^* \sim \mathcal{N}\left(\delta^*,\, \Sigma^*\right)$. Convergence in distribution is equivalent to convergence in bounded Lipschitz metric, so

$$\lim_{s \to \infty} \sup_{f \in BL_1} \left|\left| \mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^{+}_{n_s})\right] - \mathbb{E}\left[f(\xi^*)\right]\right|\right| = 0. \tag{17}$$

Additionally, Assumption 2 gives that

$$\lim_{s \to \infty} \sup_{f \in BL_1} \left|\left| \mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s})\right] - \mathbb{E}\left[f(\xi_{P_{n_s}})\right]\right|\right| = 0.$$

Since the class of BL_1 functions is closed under horizontal transformations, it follows that

$$\lim_{s \to \infty} \sup_{f \in BL_1} \left|\left| \mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^{+}_{n_s})\right] - \mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^{+}_{n_s})\right]\right|\right| = 0. \tag{18}$$

Equations (17) and (18), together with the triangle inequality, imply that

$$\lim_{s \to \infty} \sup_{f \in BL_1} \left|\left| \mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^{+}_{n_s})\right] - \mathbb{E}\left[f(\xi^*)\right]\right|\right| = 0, \tag{19}$$

or equivalently, $(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^{+}_{n_s}) \xrightarrow{d} \xi^*$. By Assumption 5, the pre-test is invariant to shifts that only affect the post-period coefficients, and so $\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})$ iff $(\hat{\beta}_{n_s} - \tau_{n_s,P_{n_s}} - \delta^{+}_{n_s}) \in B(\hat{\Sigma}_{n_s})$. Lemma B.1 thus implies that $\lim_{s \to \infty} \mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) = \mathbb{P}\left(\xi^* \in B(\Sigma^*)\right)$, and hence $\mathbb{P}\left(\xi^* \in B(\Sigma^*)\right) = p^* > 0$. We have thus shown that $(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^{+}_{n_s}, \hat{\Sigma}_{n_s}) \xrightarrow{d} (\xi^*, \Sigma^*)$, $(\xi_{P_{n_s}} - \delta^{+}_{n_s}, \Sigma_{P_{n_s}}) \xrightarrow{d} (\xi^*, \Sigma^*)$, and $\mathbb{P}\left(\xi^* \in B(\Sigma^*)\right) > 0$. The result then follows immediately from Lemma B.2.

□

Lemma B.4. *Suppose that Assumptions 2-5 hold. Then for any increasing sequence of* *sample sizes n_q and corresponding data-generation processes P_{n_q} such that*

$$\lim_{q\to\infty} ||\delta^{pre}_{P_{n_q},n_q}|| = \infty,$$

we have

$$\lim_{q\to\infty} \mathbb{P}_{P_{n_q}}\left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q})\right) = 0.$$

Proof. Towards contradiction, suppose that there exists a sequence n_q such that

$$\lim_{q\to\infty} ||\delta^{pre}_{P_{n_q},n_q}|| = \infty,$$

and

$$\liminf_{q\to\infty} \mathbb{P}_{P_{n_q}}\left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q})\right) > 0. \tag{20}$$

Since $\mathcal{S}$ is compact, we can extract a subsequence n_r along which $\Sigma_{P_{n_r}} \to \Sigma^*$ for some $\Sigma^* \in \mathcal{S}$. Assumption 3 then implies that $\hat{\Sigma}_{n_r} \xrightarrow{p} \Sigma^*$.

By Assumption 5, $B_{pre}(\Sigma)$ is bounded for every Σ. Let $\tilde{M}(\Sigma) = \sup_{\beta_{pre}\in B_{pre}(\Sigma)} ||\beta_{pre}||$. Assumption 5 implies that $B_{pre}(\Sigma)$ is a compact-valued continuous correspondence, and so $\tilde{M}(\Sigma)$ is a continuous function by the theorem of the maximum. It follows that for any Σ in a sufficiently small neighborhood of Σ^*, $\tilde{M}(\Sigma) \leq \tilde{M}(\Sigma^*) + 1 =: \bar{M}$. Since $\hat{\Sigma}_{n_r} \xrightarrow{p} \Sigma^*$, it follows that $\tilde{M}(\hat{\Sigma}_{n_r}) \to_p \tilde{M}(\Sigma^*)$, and thus for r sufficiently large, $\tilde{M}(\hat{\Sigma}_{n_r}) \leq \bar{M}$ with probability 1. Thus, for r sufficiently large, $\mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B(\hat{\Sigma}_{n_r})\right) \leq \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B_{\bar{M}}\right)$, where $B_{\bar{M}} = \{(\beta_{post}, \beta_{pre}) \,|\, ||\beta_{pre}|| \leq \bar{M}\}$. It follows that

$$\begin{aligned}\liminf_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B(\hat{\Sigma}_{n_r})\right) &\leq \liminf_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B_{\bar{M}}\right) \\ &= 1 - \limsup_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B^c_{\bar{M}}\right).\end{aligned}$$

We now show that $\limsup_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B^c_{\bar{M}}\right) = 1$, which along with the display above implies that $\liminf_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B(\hat{\Sigma}_{n_r})\right) = 0$, contradicting (20).

Consider the function $h(\beta) = \min(d(\beta, B_{\bar{M}}), 1)$, where for a set S we define $d(\beta, S) = \inf_{\tilde{\beta}\in S} ||\beta - \tilde{\beta}||$. It is easily verified that $h \in BL_1$, and that $h(\beta) \leq 1[\beta \in B^c_{\bar{M}}]$ for all β. Thus,

$$\limsup_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B^c_{\bar{M}}\right) \geq \limsup_{r\to\infty} \mathbb{E}_{P_{n_r}}\left[h(\hat{\beta}_{n_r})\right]. \tag{21}$$

Note that $d(\hat{\beta}, B_{\bar{M}})$ depends only on the components of $\hat{\beta}$ corresponding with the pre-period, and thus $h(\hat{\beta}) = h(\hat{\beta} - \tau)$ for any value $\tau = \begin{pmatrix} \tau_{post} \\ 0 \end{pmatrix}$ that has zeros in the positions corresponding with β_{pre}. This, along with Assumption 2, implies that

$$\lim_{r \to \infty} \left|\left| \mathbb{E}_{P_{n_r}} \left[h(\hat{\beta}_{n_r}) \right] - \mathbb{E} \left[h(\xi_{P_{n_r}, n_r}) \right] \right|\right| = 0.$$

Using the triangle inequality and the fact that h is a non-negative function, we have

$$\mathbb{E}_{P_{n_r}} \left[h(\hat{\beta}_{n_r}) \right] \geq \mathbb{E} \left[h(\xi_{P_{n_r}, n_r}) \right] - \left|\left| \mathbb{E}_{P_{n_r}} \left[h(\hat{\beta}_{n_r}) \right] - \mathbb{E} \left[h(\xi_{P_{n_r}, n_r}) \right] \right|\right|.$$

It then follows that

$$\limsup_{r \to \infty} \mathbb{E}_{P_{n_r}} \left[h(\hat{\beta}_{n_r}) \right] \geq \limsup_{r \to \infty} \mathbb{E} \left[h(\xi_{P_{n_r}, n_r}) \right]. \tag{22}$$

Now, since $\lim_{r \to \infty} ||\delta^{pre}_{P_{n_r}, n_r}|| = \infty$, there exists at least one component j of $\delta^{pre}_{P_{n_r}, n_r}$ that diverges. Let $\delta^{pre}_{j,r}$ denote the jth element of $\delta^{pre}_{P_{n_r}, n_r}$, and suppose WLOG that $\delta^{pre}_{j,r} \to \infty$. Likewise, let $\xi^{pre}_{j,r}$ denote the jth element of $\xi^{pre}_{P_{n_r}, n_r}$. Note that $h(\xi_{P_{n_r}, n_r}) = 1$ whenever $\xi^{pre}_{j,r} > \bar{M} + 1$, and thus $\mathbb{E} \left[h(\xi_{P_{n_r}, n_r}) \right] \geq \mathbb{E} \left[1[\xi^{pre}_{j,r} > \bar{M} + 1] \right]$. Hence,

$$\limsup_{r \to \infty} \mathbb{E} \left[h(\xi_{P_{n_r}, n_r}) \right] \geq \limsup_{r \to \infty} \mathbb{E} \left[1[\xi^{pre}_{j,r} > \bar{M} + 1] \right]. \tag{23}$$

Since $\xi^{pre}_{j,r} \sim \mathcal{N} \left(\delta^{pre}_{j,r}, \sigma^2_{j,r} \right)$, for $\sigma^2_{j,r}$ the jth diagonal element of $\Sigma_{P_{n_r}}$, we have

$$\mathbb{E} \left[1[\xi^{pre}_{j,r} > \bar{M} + 1] \right] = 1 - \Phi \left(\frac{\bar{M} + 1 - \delta^{pre}_{j,r}}{\sigma_{j,r}} \right).$$

However, by construction $\sigma_{j,r} \to \sigma^*_j$ as $r \to \infty$, where σ^{*2}_j is the jth diagonal element of Σ^*. Additionally, $\sigma^*_j > 0$ by Assumption 4. Thus, since $\delta^{pre}_{j,r} \to \infty$, we have that $\Phi \left(\frac{\bar{M}+1-\delta^{pre}_{j,r}}{\sigma_{j,r}} \right) \to 0$, and hence $\mathbb{E} \left[1[\xi^{pre}_{j,r} > \bar{M} + 1] \right] \to 1$. This, combined with the inequalities (21), (22), (23), gives the desired result.

□

Lemma B.5. *Suppose Assumptions 2-5 hold. Consider a subsequence of increasing sample sizes, n_s, such that*

$$\lim_{s \to \infty} \Sigma_{P_{n_s}} = \Sigma^*, \tag{24}$$

$$\lim_{s \to \infty} \delta^{pre}_{P_{n_s}, n_s} = \delta^{pre,*}, \tag{25}$$

$$\lim_{s \to \infty} \mathbb{P}_{P_{n_s}} \left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s}) \right) = p^* > 0 \tag{26}$$

for $\Sigma^* \in \mathcal{S}$. *Then*

$$\lim_{s\to\infty} \left|\left|\mathbb{E}\left[\xi_{P_{n_s},n_s} - \delta^+_{n_s} \,|\, \xi_{P_{n_s},n_s} \in B(\Sigma_{P_{n_s}})\right] - \mathbb{E}\left[\xi^* \,|\, \xi^* \in B(\Sigma^*)\right]\right|\right| = 0,$$

for $\xi^* \sim \mathcal{N}\left(\delta^*, \Sigma^*\right)$, *where* $\delta^* = \begin{pmatrix} 0 \\ \delta^{pre,*} \end{pmatrix}$ *and* $\delta^+_{n_s} = \begin{pmatrix} \delta^{post}_{P_{n_s},n_s} \\ 0 \end{pmatrix}$

Proof. Let $\xi_{j,s}$ denote the jth element of $\xi_{P_{n_s},n_s} - \delta^+_{n_s}$. We show that $\mathbb{E}\left[\xi_{j,s} \,|\, \xi_{P_{n_s},n_s} \in B(\hat{\Sigma}_{P_{n_s}})\right] \longrightarrow \mathbb{E}\left[\xi^*_j \,|\, \xi^* \in B(\Sigma^*)\right]$ for each element j, which implies the desired result.

Note that $\xi_{P_{n_s},n_s} \sim \mathcal{N}\left(\delta_{P_{n_s},n_s}, \Sigma_{P_{n_s}}\right)$, so $\xi_{P_{n_s},n_s} - \delta^+_{n_s} \sim \mathcal{N}\left(\delta^-_{n_s}, \Sigma_{P_{n_s}}\right)$, where $\delta^-_{n_s} = \begin{pmatrix} 0 \\ \delta^{pre}_{P_{n_s},n_s} \end{pmatrix}$. Since by construction $\delta^-_{n_s} \longrightarrow \delta^*$ and $\Sigma_{P_{n_s}} \longrightarrow \Sigma^*$, it follows that $\xi_{P_{n_s},n_s} - \delta^+_{n_s} \xrightarrow{d} \xi^*$. The continuous mapping theorem then gives that $(\xi_{P_{n_s},n_s} - \delta^+_{n_s}) \cdot 1[\xi_{P_{n_s},n_s} \in B(\hat{\Sigma}_{P_{n_s}})] \xrightarrow{d} \xi^* 1[\xi^* \in B(\Sigma^*)]$, where the function is continuous for almost every ξ^* as shown in the proof to Lemma B.1, and we use the fact that $\xi_{P_{n_s},n_s} \in B(\hat{\Sigma}_{P_{n_s}})$ iff $\xi_{P_{n_s},n_s} - \delta^+_{n_s} \in B(\hat{\Sigma}_{P_{n_s}})$ by Assumption 5. Next, observe that

$$|\xi_{j,s} \cdot 1[\xi_{P_{n_s},n_s} \in B(\hat{\Sigma}_{P_{n_s}})]| \leq |\xi_{j,s}|.$$

Since the absolute value function is continuous and $\xi_{j,s} \xrightarrow{d} \xi^*_j$, $|\xi_{j,s}| \xrightarrow{d} |\xi^*_j|$ by the continuous mapping theorem. Further, each $|\xi_{j,s}|$ has a folded-normal distribution, as does $|\xi^*_j|$, and since the mean of a folded-normal distribution is finite and continuous in the mean and variance parameters, we have $\mathbb{E}\left[|\xi_{j,s}|\right] \to \mathbb{E}\left[|\xi^*_j|\right] < \infty$. Thus, by the generalized dominated convergence theorem,

$$\mathbb{E}\left[\xi_{j,s} \cdot 1[\xi_{P_{n_s},n_s} \in B(\hat{\Sigma}_{P_{n_s}})]\right] \xrightarrow{d} \mathbb{E}\left[\xi^*_j \cdot 1[\xi^* \in B(\Sigma^*)]\right].$$

However, by Lemma B.1 we have that

$$\mathbb{P}\left(\xi_{P_{n_s}} \in B(\hat{\Sigma}_{P_{n_s},n_s})\right) \longrightarrow \mathbb{P}\left(\xi^* \in B(\Sigma^*)\right) = p^* > 0.$$

Thus, by the continuous mapping theorem,

$$\frac{\mathbb{E}\left[\xi_{j,s} \cdot 1[\xi_{P_{n_s}} \in B(\hat{\Sigma}_{P_{n_s}})]\right]}{\mathbb{P}\left(\xi_{P_{n_s},n_s} \in B(\hat{\Sigma}_{P_{n_s}})\right)} \longrightarrow \frac{\mathbb{E}\left[\xi^*_j \cdot 1[\xi^* \in B(\Sigma^*)]\right]}{\mathbb{P}\left(\xi^* \in B(\Sigma^*)\right)},$$

as we wished to show. □

Lemma B.6. *Suppose that a sequence of random variables Y_n is asymptotically uniformly integrable,*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n|| \cdot 1[||Y_n|| > M]\right] = 0.$$

If c_n is a sequence of constants with $c_n \to c$ and $Y_n - c_n$ converges in distribution, then $Y_n - c_n$ is also asymptotically uniformly integrable.

Proof. Note that $||Y_n - c_n|| \leq ||Y_n|| + ||c_n||$. Thus,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n - c_n|| \cdot 1[||Y_n - c_n|| > M]\right] \leq$$
$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n|| \cdot 1[||Y_n - c_n|| > M]\right] + \lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||c_n|| \cdot 1[||Y_n - c_n|| > M]\right]. \tag{27}$$

We now show that each of the two terms on the right hand side of (27) is zero. To see why the first term is zero, note that since $c_n \to c$, for n sufficiently large, $||c_n|| \leq ||c+1||$. By the triangle inequality, $||Y_n - c_n|| \leq ||Y_n|| + ||c_n||$ and so for n sufficiently large, $1[||Y_n - c_n|| > M] \leq 1[||Y_n|| > M - ||c+1||]$. Thus,

$$\begin{aligned}\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n|| \cdot 1[||Y_n - c_n|| > M]\right] &\leq \lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n|| \cdot 1[||Y_n|| > M - ||c+1||]\right] \\ &= \lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n|| \cdot 1[||Y_n|| > M]\right],\end{aligned}$$

and $\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n|| \cdot 1[||Y_n|| > M]\right] = 0$ by assumption.

To show that the second term in (27) is zero, note again that since $c_n \longrightarrow c$, for n sufficiently large, $||c_n|| \leq ||c+1||$, and thus

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||c_n|| \cdot 1[||Y_n - c_n|| > M]\right] \leq ||c+1|| \lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[1[||Y_n - c_n|| > M]\right].$$

However, since $Y_n - c_n$ converges in distribution, Prohorov's theorem gives that $Y_n - c_n$ is uniformly tight, so

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[1[||Y_n - c_n|| > M]\right] = 0.$$

□

C Power Calculations Under Stochastic Differential Trends

This section considers data-generating processes in which there are stochastic differential trends between the treated and control groups. In particular, we consider the following hierarchical model:

$$\delta \sim \mathcal{N}\left(0,\, V\right) \tag{28}$$

$$\hat{\beta} \,|\, \delta \sim \mathcal{N}\left(\delta + \tau,\, \Sigma\right). \tag{29}$$

The distribution for $\hat{\beta}|\delta$ in (29) is identical to the model considered in Section 3. However, we now treat δ as stochastic, rather than as a fixed parameter (e.g. linear in event-time). Treating δ as stochastic is sensible in situations in which we think that there may be common shocks to the treated and control groups (e.g. if each of these is a state, and there are macro-level shocks).

I now evaluate the power of pre-tests against such stochastic shocks in data-generating processes calibrated to the sample of papers reviewed in Section 2. For a given value of (V, Σ), we define the power of the pre-test to be the probability, $\mathbb{P}_{\delta,\hat{\beta}}\left(\hat{\beta}_{pre} \in B(\Sigma)\right)$, where $\mathbb{P}_{\delta,\hat{\beta}}\left(\cdot\right)$ denotes the probability taken over the realization of the joint distribution of $(\delta, \hat{\beta})$. We explicitly write the pre-test acceptance region as $B(\Sigma)$ to denote that the pre-test region depends on Σ (but not V). We again set Σ to be the estimated variance-covariance matrix from each of the papers in the sample. Calibrating the covariance matrix V for the common stochastic shocks is more difficult, as it cannot be consistently estimated from the data. For simplicity, I set $V = c \cdot \Sigma$ for a constant $c > 0$. Under this specification, the marginal distribution of $\hat{\beta}$ under the hierarchical model defined above is $\mathcal{N}\left(0,\, (1+c)\Sigma\right)$. The parameter c can thus be interpreted as the factor by which we have underestimated the variance matrix by treating δ as fixed and ignoring common stochastic shocks.

I then calculate the values of c for which the pre-test rejects 50 or 80% percent of the time, which I denote $c_{0.5}$ and $c_{0.8}$. As in Section 2, I use the pre-test criterion that no pre-period coefficient is significant at the 95% level. I compute the null rejection probabilities of conventional confidence intervals for the average post-treatment effect $\bar{\tau}$ and the first-period treatment effect τ_1 under the DGPs with $c_{0.5}$ and $c_{0.9}$. The null rejection probabilities are computed over the joint distribution of $(\hat{\beta}, \delta)$.[17] As in Section 2, I report these probabilities both unconditionally, and conditional on surviving the pre-test. Tables C1 and C2 show the results for τ_1 and $\bar{\tau}$, respectively. Across all specifications, the null rejection probabilities

[17] Recall that $\hat{\beta} \sim \mathcal{N}\left(0,\, (1+c)\Sigma\right)$. Thus, this is the probability that τ falls inside a confidence interval based on the assumption that $\hat{\beta} \sim \mathcal{N}\left(\tau,\, \Sigma\right)$ when in fact $\hat{\beta} \sim \mathcal{N}\left(\tau,\, (1+c)\Sigma\right)$.

substantially exceed the nominal level of 5% for most of the papers. Conditioning on passing the pre-test generally reduces the null rejection probability, but only moderately so in most cases. Conditional on passing the pre-test, null rejection probabilities are often many multiples of the nominal size. The results thus suggest that conventional pre-tests may be underpowered against detecting common stochastic shocks, in addition to the linear secular trends considered in the main text. Concurrent work by Ferman (2020) reaches a similar conclusion in a related model with stochastic violations of parallel trends.

I do not report results for bias as in the main text, since δ is mean-zero and so $\hat{\beta}$ is unbiased when the expectation is taken over the joint distribution of $(\hat{\beta}, \delta)$. It would be straightforward to combine this simulation design with one such as in the main test so that there are both stochastic shocks and a non-zero average difference in trends.

	Conditional on passing pre-test?					
	No			Yes		
	Scaling factor for stochastic variance					
	0	$c_{0.5}$	$c_{0.8}$	0	$c_{0.5}$	$c_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.17	0.34	0.05	0.16	0.33
Bosch and Campos-Vazquez (2014)	0.05	0.19	0.38	0.03	0.12	0.26
Deryugina (2017)	0.05	0.19	0.38	0.01	0.04	0.09
Deschenes et al. (2017)	0.05	0.17	0.35	0.03	0.10	0.19
Fitzpatrick and Lovenheim (2014)	0.05	0.23	0.45	0.05	0.21	0.43
Gallagher (2014)	0.05	0.14	0.30	0.04	0.12	0.26
He and Wang (2017)	0.05	0.26	0.48	0.05	0.23	0.46
Kuziemko et al. (2018)	0.05	0.29	0.55	0.04	0.20	0.42
Lafortune et al. (2017)	0.05	0.19	0.38	0.05	0.18	0.37
Markevich and Zhuravskaya (2018)	0.05	0.22	0.44	0.04	0.18	0.38
Tewari (2014)	0.05	0.10	0.22	0.04	0.08	0.18
Ujhelyi (2014)	0.05	0.22	0.43	0.04	0.18	0.36

Table C1: Null Rejection Probabilities for Nominal 5% Test of Average Treatment Effect Under Stochastic Trends Against Which Pre-tests Have 50 or 80% Power

Note: This table shows null rejection probabilities, i.e. the probability that the true parameter falls outside a nominal 95% confidence interval, using data-generating processes in which parallel trends holds (scaling factor = 0) or in which there are stochastic violations of parallel trends that conventional pre-tests would detect 50 or 80% of the time ($c_{0.5}$ and $c_{0.8}$). The first three columns show unconditional null rejection probabilities, whereas the latter three columns condition on passing the pre-test. The estimand is the average of the post-treatment causal effects, $\bar{\tau}$. See Section C for details on the data-generating process.

	Conditional on passing pre-test?					
	No			Yes		
	Scaling factor for stochastic variance					
	0	$c_{0.5}$	$c_{0.8}$	0	$c_{0.5}$	$c_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.17	0.34	0.04	0.14	0.30
Bosch and Campos-Vazquez (2014)	0.05	0.19	0.38	0.05	0.17	0.35
Deryugina (2017)	0.05	0.19	0.38	0.04	0.13	0.29
Deschenes et al. (2017)	0.05	0.17	0.35	0.04	0.11	0.22
Fitzpatrick and Lovenheim (2014)	0.05	0.23	0.45	0.05	0.22	0.44
Gallagher (2014)	0.05	0.14	0.30	0.03	0.09	0.19
He and Wang (2017)	0.05	0.26	0.48	0.04	0.23	0.45
Kuziemko et al. (2018)	0.05	0.29	0.55	0.04	0.21	0.45
Lafortune et al. (2017)	0.05	0.19	0.38	0.05	0.18	0.37
Markevich and Zhuravskaya (2018)	0.05	0.22	0.44	0.04	0.17	0.36
Tewari (2014)	0.05	0.10	0.22	0.04	0.08	0.19
Ujhelyi (2014)	0.05	0.22	0.43	0.04	0.18	0.35

Table C2: Null Rejection Probabilities for Nominal 5% Test of First Period Treatment Effect Under Stochastic Trends Against Which Pre-tests Have 50 or 80% Power

Note: This table shows null rejection probabilities, i.e. the probability that the true parameter falls outside a nominal 95% confidence interval, using data-generating processes in which parallel trends holds (scaling factor = 0) or in which there are stochastic violations of parallel trends that conventional pre-tests would detect 50 or 80% of the time ($c_{0.5}$ and $c_{0.8}$). The first three columns show unconditional null rejection probabilities, whereas the latter three columns condition on passing the pre-test. The estimand is the causal effect for the first period after treatment, τ_1. See Section C for details on the data-generating process.

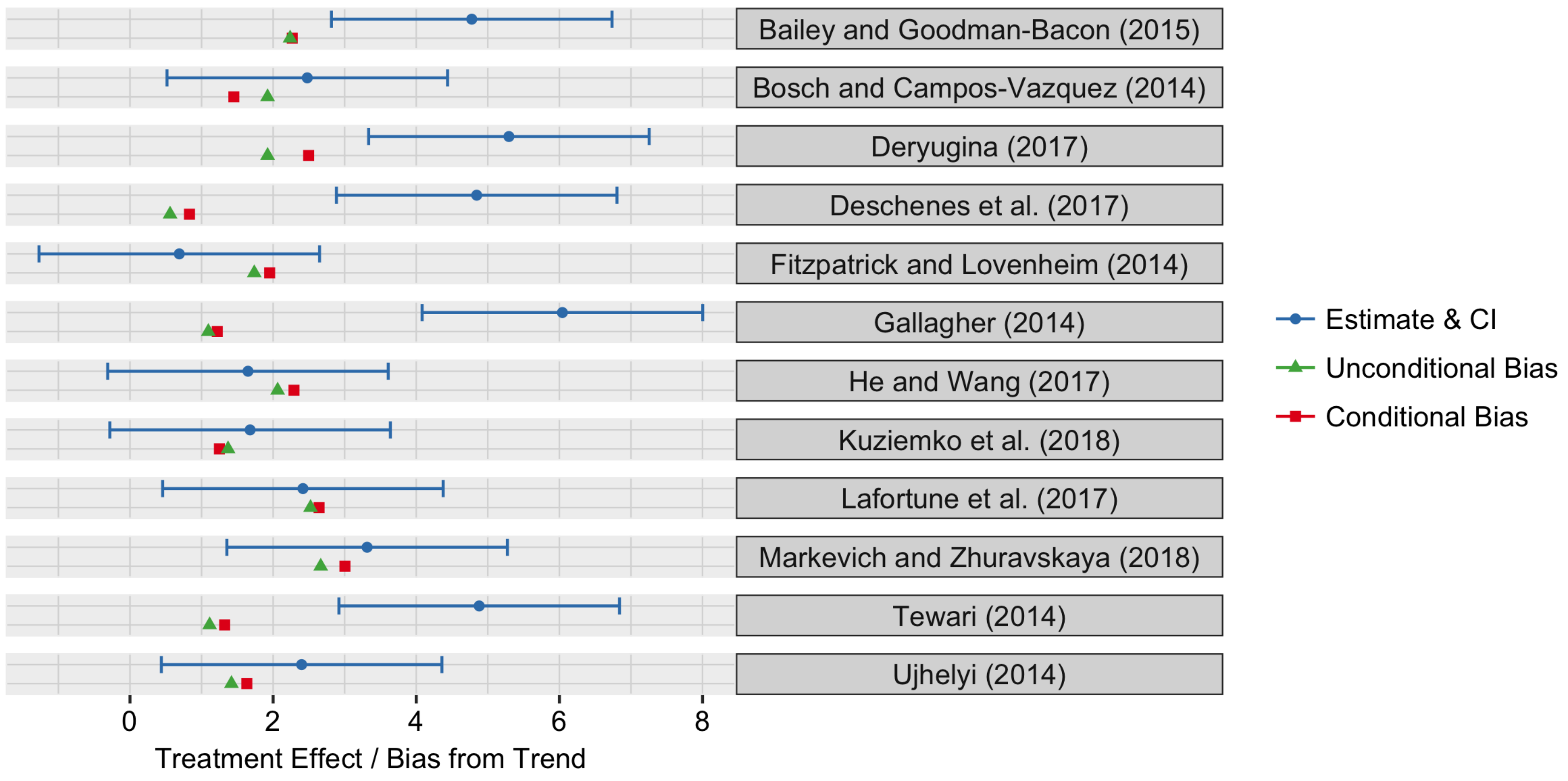
D Additional tables and figures

	Conditional on passing pre-test?					
	No			Yes		
	Slope of differential trend:					
	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.06	0.09	0.04	0.07	0.13
Bosch and Campos-Vazquez (2014)	0.05	0.12	0.22	0.05	0.08	0.11
Deryugina (2017)	0.05	0.07	0.09	0.04	0.09	0.21
Deschenes et al. (2017)	0.05	0.06	0.06	0.04	0.05	0.08
Fitzpatrick and Lovenheim (2014)	0.05	0.10	0.18	0.05	0.13	0.26
Gallagher (2014)	0.05	0.05	0.06	0.03	0.04	0.05
He and Wang (2017)	0.05	0.15	0.29	0.04	0.21	0.47
Kuziemko et al. (2018)	0.05	0.13	0.22	0.04	0.07	0.11
Lafortune et al. (2017)	0.05	0.19	0.41	0.05	0.17	0.34
Markevich and Zhuravskaya (2018)	0.05	0.11	0.19	0.04	0.17	0.42
Tewari (2014)	0.05	0.06	0.07	0.04	0.06	0.11
Ujhelyi (2014)	0.05	0.09	0.15	0.04	0.12	0.28

Table D1: Null Rejection Probabilities for Nominal 5% Test of First Period Treatment Effect Under Linear Trends Against Which Pre-tests Have 50 or 80% Power

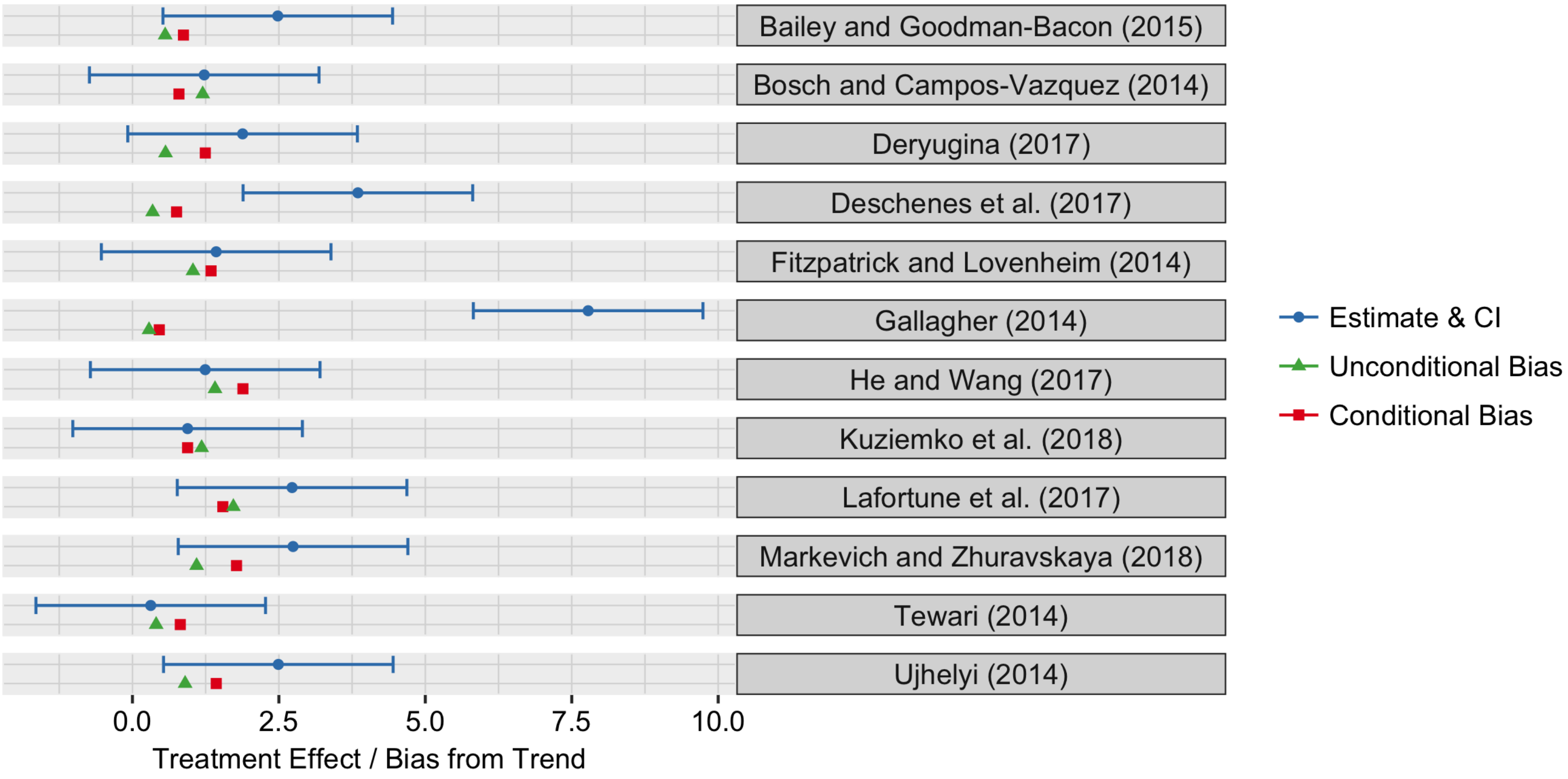
Note: This table shows null rejection probabilities, i.e. the probability that the true parameter falls outside a nominal 95% confidence interval, using data-generating processes in which parallel trends holds (slope of differential trend = 0) and in which there are linear violations of parallel trends that conventional pre-tests would detect 50 or 80% of the time ($\gamma_{0.5}$ and $\gamma_{0.8}$). The first three columns show unconditional null rejection probabilities, whereas the latter three columns condition on passing the pre-test. The estimand is the treatment effect in the first period after treatment, τ_1.

Figure D1: OLS Estimates and Bias from Linear Trends for Which Pre-tests Have 50 Percent Power – Average Treatment Effect



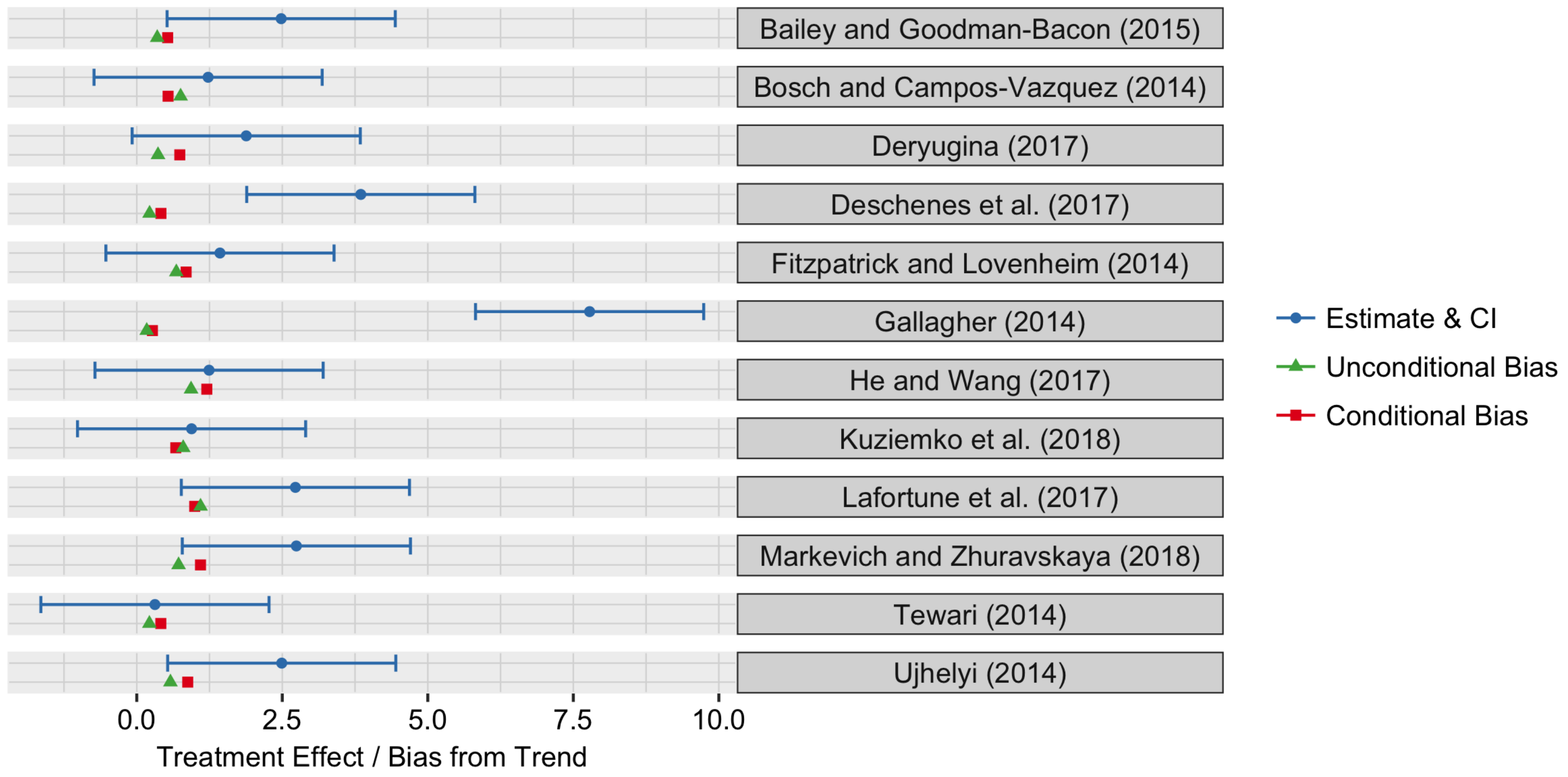
Note: I calculate the linear trend against which conventional pre-tests would reject 50 percent of the time ($\gamma_{0.5}$). The red squares show the bias that would result from such a trend conditional on passing the pre-test ($\mathbb{E}\left[\hat{\tau} - \tau_* \,|\, \hat{\beta}_{pre} \in B_{NIS}(\Sigma)\right]$); the green triangles show the unconditional bias from such a trend ($\mathbb{E}\left[\hat{\tau} - \tau_*\right]$). As a benchmark, I plot in blue the OLS estimates and 95% CIs from the original paper. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The estimand is the average of the treatment effects in all periods after treatment began, $\tau_* = \bar{\tau}$.

Figure D2: OLS Estimates and Bias from Linear Trends for Which Pre-tests Have 80 Percent Power – First Period Treatment Effect



Note: I calculate the linear trend against which conventional pre-tests would reject 80 percent of the time ($\gamma_{0.8}$). The red squares show the bias that would result from such a trend conditional on passing the pre-test ($\mathbb{E}\left[\hat{\tau} - \tau_* \mid \hat{\beta}_{pre} \in B_{NIS}(\Sigma)\right]$); the green triangles show the unconditional bias from such a trend ($\mathbb{E}\left[\hat{\tau} - \tau_*\right]$). As a benchmark, I plot in blue the OLS estimates and 95% CIs from the original paper. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The estimand is the treatment effect in the first period after treatment began, $\tau_* = \tau_1$.

Figure D3: OLS Estimates and Bias from Linear Trends for Which Pre-tests Have 50 Percent Power – First Period Treatment Effect



Note: I calculate the linear trend against which conventional pre-tests would reject 50 percent of the time ($\gamma_{0.5}$). The red squares show the bias that would result from such a trend conditional on passing the pre-test ($\mathbb{E}\left[\hat{\tau} - \tau_* \,|\, \hat{\beta}_{pre} \in B_{NIS}(\Sigma)\right]$); the green triangles show the unconditional bias from such a trend ($\mathbb{E}\left[\hat{\tau} - \tau_*\right]$). As a benchmark, I plot in blue the OLS estimates and 95% CIs from the original paper. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The estimand is the treatment effect in the first period after treatment began, $\tau_* = \tau_1$.

Supplement References

Cartinhour, J. (1990). One-dimensional marginal density functions of a truncated multivariate normal density function. *Communications in Statistics-Theory and Methods*, 19:197–203.

Ferman, B. (2020). Inference in Differences-in-Differences: How Much Should We Trust in Independent Clusters? *arXiv:1909.01782 [econ]*. arXiv: 1909.01782.

Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: A review. *arXiv:1404.5886 [math, stat]*.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media. Google-Books-ID: seH8dMrEgggC.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press. Google-Books-ID: UEuQEM5RjWgC.