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

Jonathan Roth*

July 30, 2019

Abstract

Researchers using an event-study design often test for pre-event trends ("pre-trends"), yet typical estimation and inference does not account for this test. This paper analyzes the properties of conventional event-study estimates conditional on having survived a test for pre-trends. Pre-testing for pre-trends changes the bias of conventional estimates when parallel trends is violated. I show that in settings with homoskedastic errors and a monotone trend, the bias conditional on surviving a pre-test is larger than the unconditional bias. Hence, pre-trends tests meant to mitigate bias can actually exacerbate it. Pre-testing also distorts the coverage rates of conventional confidence intervals, which can be above or below their nominal level conditional on surviving the pre-test. Simulations based on a systematic review of recent papers in leading economics journals suggest that conventional pre-tests are often underpowered and substantial distortions from pre-testing are possible in practice. To address these issues, I develop a method to correct event-study plots for the distortions from pre-testing. I recommend that researchers who rely on pre-trends testing report these corrected event-studies along with calculations of the power of the pre-test against economically relevant alternatives.

*Department of Economics, Harvard University, jonathanroth@g.harvard.edu. I am grateful to Isaiah Andrews, Kirill Borusyak, Gary Chamberlain, Peter Ganong, Edward Glaeser, Nathan Hendren, Ariella Kahn-Lang, Max Kasy, Larry Katz, Sendhil Mullainathan, Andreas Pick, Jesse Shapiro, Jann Spiess, Jim Stock, and Elie Tamer for helpful comments, as well as to seminar participants at Harvard and Erasmus University. This material is based upon work supported by the NSF Graduate Research Fellowship under Grant DGE1144152.

1 Introduction

Difference-in-differences and related event-study designs are highly popular tools for quasi-experimental empirical research. The key identifying assumption in these research designs is the parallel trends assumption, which requires that had the treatment of interest not occurred the mean of the outcome would have evolved in parallel for the treatment and control groups. Researchers often assess the plausibility of the parallel trends assumption by testing for "pre-trends," i.e. differences in trends between the treatment and control group prior to the date of treatment assignment. Pre-trends tests are remarkably common: since 2014, over 70 papers in the journals of the American Economic Association have employed a so-called "event-study plot" to visually test for pre-trends.

A handful of recent papers have warned, however, that the common approach of testing for pre-trends will be imperfect in finite samples (Freyaldenhoven et al., 2018; Kahn-Lang and Lang, 2018; Bilinski and Hatfield, 2018). Owing to noise in the data, we may spuriously detect a violation of parallel trends when none exists, and perhaps more concerningly, we may fail to detect a violation of parallel trends when there is one. Given the imperfect performance of pre-trends tests in finite samples, it is important to understand the statistical properties of the estimates that survive a test for pre-trends.

This paper analyzes the properties of event-study estimates and confidence intervals conditional on having survived a pre-test for pre-trends. I show both theoretically and in simulations based on a systematic review of recent papers that many of the properties we expect of conventional estimates no longer hold conditional on passing a pre-test for parallel trends. Perhaps most concerningly, I illustrate that pre-trends tests meant to mitigate bias in the treatment effects estimates can actually amplify bias in published work. I therefore propose a method for producing a corrected event-study plot that corrects for the distortions from pre-testing.

I begin in Section 2 with a stylized example that highlights how pre-trends testing can amplify bias in point estimates and distort the coverage rate of confidence intervals. I consider a difference-in-differences setting with three periods in which there are potentially linear violations of parallel trends. A key insight of the example is that noise in the period before treatment impacts both the estimated treatment effect and the estimated pre-period coefficient. As it turns out, noise that leads us to fail to detect a violation of parallel trends also tends to amplify the bias in the treatment effect estimate created by the underlying trend. Thus, when the parallel trends assumption is violated in population, the bias conditional on surviving the pre-test is worse than the unconditional bias. Selection on noise in the pre-period can similarly lead confidence intervals (CIs) to have coverage rates that differ from their nominal level conditional on passing the pre-test.

Section 3 provides theoretical results that extend the intuition from the stylized model to a more general setting with an arbitrary number of pre-periods and non-linear violations of parallel trends. I first derive formulas for the bias and variance of event-study estimates after passing a pre-test for parallel trends. I then prove that under homoskedasticity, pre-testing necessarily amplifies bias in the treatment effects estimates when there is a monotone violation of parallel trends. Finally, I show that pre-testing reduces the variance of event-study estimates under very general conditions. The bias and reduction in variance induced by pre-testing have opposite effects on the coverage rates of conventional CIs; as a result, conventional CIs will tend to overcover when the underlying trend is close to zero (when bias is small), and will tend to undercover when the underlying trend is sufficiently large.

Section 4 evaluates the practical relevance of these distortions in simulations based on a systematic review of recent papers in three leading economics journals (the *American Economic Review*, *AEJ: Applied Economics*, and *AEJ: Economic Policy*). Although other recent papers have cautioned that pre-trends tests may have low power (Freyaldenhoven et al., 2018; Kahn-Lang and Lang, 2018; Bilinski and Hatfield, 2018), I provide the first systematic evaluation of the power of pre-trends tests in published papers. I find that, indeed, conventional pre-trends tests often have low power against meaningful violations of parallel trends. In many cases, linear trends against which conventional tests have power of only 50 percent would produce bias of a magnitude similar to the estimated treatment effect. Additionally, the bias conditional on failing to detect an underlying trend is worse than the unconditional bias in the large majority of cases, in line with the theoretical prediction for the homoskedastic case. This bias amplification can be substantial in magnitude: in some cases, the bias conditional on passing the pre-test is more than twice as large as the unconditional bias.

How can we evaluate event-study plots given the distortions from pre-testing? In Section 5, I develop a method for constructing a "corrected event-study plot" that recovers many of the properties we expect of conventional event-study estimates conditional on passing a pre-test for parallel trends. In particular, my method, which builds on results by Andrews and Kasy (forthcoming) and Lee et al. (2016), provides median-unbiased point estimates and valid confidence intervals for the population event-study coefficients conditional on passing a pre-trends test. My method can also be extended to correct for selection among multiple specifications on the basis of the observed pre-trends – e.g. using different subgroups or different sets of control variables. The corrected event-study plot thus gives the reader an unbiased way of evaluating the chosen event-study in light of the process by which the research design has been screened and/or selected. The proposed corrections perform well in simulations based on my survey of recent papers, although they typically produce somewhat wider confidence intervals than conventional methods.[1]

In practice, I recommend that researchers who rely on pre-trends tests report corrected event-study plots along with calculations of the power of their pre-test to detect meaningful violations of parallel trends.[2] The power calculations are important because while the corrections I propose eliminate the distortions from pre-testing and provide unbiased estimates and valid confidence intervals for the true event-study coefficients, as usual the post-period event-study coefficients will not correspond with the treatment effect of interest if parallel trends is violated. The power calculations give the reader a sense of the probability that a meaningful violation of parallel trends would be detected via the pre-trends test, whereas the corrected event study estimates allow the reader to evaluate the results in light of the screening process that has occurred.

Related Literature. This paper contributes to a large body of work on the econometrics of difference-in-differences and related research designs (e.g. Bertrand et al. (2004); Abadie (2005); Donald and Lang (2007); Borusyak and Jaravel (2016); Abraham and Sun (2018); Athey and Imbens (2018); de Chaisemartin and D'Haultfœuille (2018a,b); Goodman-Bacon (2018); Callaway and Sant'Anna (2019)). Most closely related, recent papers by Freyaldenhoven et al. (2018), Kahn-Lang and Lang (2018), and Bilinski and Hatfield (2018) have warned that traditional pre-tests may have low power to detect meaningful violations of parallel trends. I contribute to this literature in three ways. First, I characterize the distribution of treatment effects estimates conditional on having survived a pre-test for parallel trends, and I show that pre-testing can exacerbate the bias from an underlying trend.[3] Second, I provide the first systematic evaluation of the power of pre-trends tests and the distortions from pre-trends testing in published work. Third, I develop a method for producing a corrected event-study plot that eliminates the additional bias and coverage distortions from pre-testing.

More broadly, 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 (see, e.g., Giles and Giles (1993), Leeb and Pötscher (2005),

[1]The main contribution of the paper is to identify issues with the common approach of pre-testing for pre-trends and provide methods to correct for these distortions. For space constraints, I therefore do not showcase my corrections in a fully worked out empirical application.

[2]An interesting question for future research is the extent to which the current practice of pre-trends testing can be improved upon, either by changing the pre-testing criteria or by adopting a more continuous approach to accounting for potential confounding trends.

[3]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.

Lee et al. (2016), and references therein). 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), or using data-driven tuning parameters (Armstrong and Kolesár, 2018).

Finally, this paper relates to the literature on selective publication of scientific results (Rothstein et al. (2005) and Christensen and Miguel (2016) provide reviews). My procedure for producing corrected event-study plots builds on results developed by Andrews and Kasy (forthcoming) to correct for publication bias, as well as earlier results from Lee et al. (2016) and Pfanzagl (1994). 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 estimation and inference procedures that correct for the distortions created by selecting research designs based on the result of this sniff test.

2 Intuition: The effect of pre-trends testing in a stylized three-period model

This section develops intuition for how testing for pre-trends affects the distribution of event-study estimates in a simple model with three-periods, homoskedastic errors, and (potentially) linear violations of parallel trends. A key insight is that estimation error in the treatment-control difference in the reference period ($t = 0$) enters both the pre-period and post-period event-study coefficients, so pre-testing using the pre-period coefficients affects the distribution of the post-period coefficients via its effect on the distribution of this error.

2.1 Set-up of stylized model

Suppose that we observe an outcome y_{it} for individuals i in period t for three periods $t = -1, 0, 1$. Individuals in the treatment group ($D_i = 1$) receive a treatment of interest between periods 0 and 1, whereas individuals in the control group do not receive the treatment. We denote by $y_{it}(1)$ and $y_{it}(0)$ the potential outcomes for individual i in period t that would have occurred if they respectively did or did not receive treatment. The observed outcome can then be written as $y_{it} = D_i y_{it}(1) + (1 - D_i) y_{it}(0)$. For simplicity, we consider the case where there is no causal effect of treatment, i.e. $y_{it}(1) \equiv y_{it}(0)$, and the true data-generating process for $y_{it}(0)$ is given by:

$$y_{it}(0) = \alpha_i + \phi_t + D_i \times g(t) + \epsilon_{it} \tag{1}$$

for $\epsilon_{it} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$. The term $D_i \times g(t)$ represents a potential difference in trends between the treatment and control group. For instance, if $g(t) = t$, then the average outcome for the treatment group is increasing linearly relative to the control group, whereas if $g(t) = 0$ then the parallel trends assumption holds.

We suppose that the researcher estimates the canonical "event-study" difference-in-differences regression specification:

$$y_{it} = \alpha_i + \phi_t + \sum_{s \neq 0} \beta_s \times 1[s = t] \times D_i + \epsilon_{it}. \tag{2}$$

The estimate $\hat{\beta}_1$ is the canonical difference-in-differences treatment effect estimate,

$$\hat{\beta}_1 = \Delta\bar{y}_{t=1} - \Delta\bar{y}_{t=0},$$

where $\Delta\bar{y}_t$ is the difference in sample means between the treatment and control group in period t. Likewise, the estimate $\hat{\beta}_{-1}$ is the canonical pre-period event-study coefficient,

$$\hat{\beta}_{-1} = \Delta\bar{y}_{t=-1} - \Delta\bar{y}_{t=0}.$$

An important observation is that the term $\Delta\bar{y}_{t=0}$, the estimated difference in means between treatment and control in the reference period ($t = 0$), enters the expression for both the pre-period and post-period coefficients. As a result, if we select on the observed pre-period coefficient $\hat{\beta}_{-1}$ being close to zero, this will affect the distribution of $\Delta\bar{y}_{t=0}$, which in turn will impact the distribution of $\hat{\beta}_1$. The next two sections illustrate how this selection plays out, first in a setting where parallel trends is violated and next in the case where it holds.

2.2 When there is an underlying trend, pre-trends testing exacerbates bias

Figure 1 provides intuition for how pre-trends testing affects the distribution of estimated event-study estimates when parallel trends is violated. The top panel of the figure shows simulations from a data-generating process where in population there is an underlying upward linear trend, i.e. $g(t) = \gamma \cdot t$ for $\gamma > 0$.[4] The y-axis shows the difference in means between treatment and control in each period, which in population is a straight line.

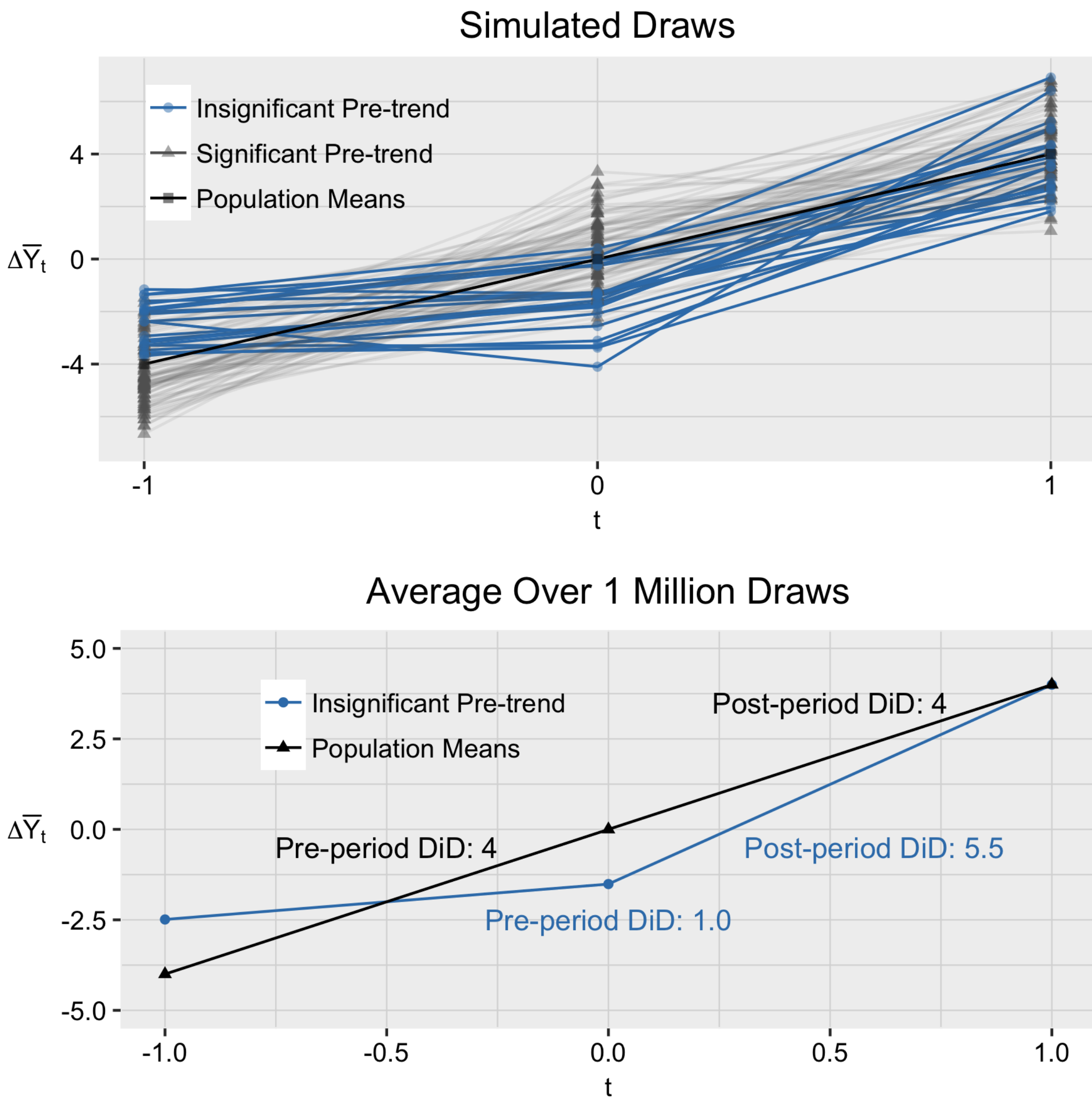
Although in population there is an underlying pre-trend, in finite samples there will sometimes be noise that prevents us from detecting this trend. The top panel highlights such draws in blue – in particular, the highlighted draws are those for which the t-statistic on $\hat{\beta}_{-1}$ is less than 1 in absolute value, so that at conventional levels we would not detect a significant pre-trend. I use a threshold t-stat of 1 (rather than 1.96) for this stylized example because it allows for an easier visual comparison of the cases with and without significant pre-trends, although results using the standard threshold of 1.96 are qualitatively similar.[5] By definition, the slope of the line between $t = -1$ and $t = 0$ corresponds with $-\hat{\beta}_{-1}$, whereas the slope between $t = 0$ and $t = 1$ corresponds with $\hat{\beta}_1$. Thus, the highlighted draws are those for which the slope between $t = -1$ and $t = 0$ is close to 0.

Figure 1 makes apparent that the draws where we fail to detect a significant pre-trend also have a particularly large slope between period $t = 0$ and $t = 1$, corresponding with a large value of $\hat{\beta}_1$. The reason for this is that the draws of the data in which we fail to detect the underlying trend tend to have below-average values of $\Delta \bar{y}_{t=0}$. Intuitively, this is because negative shocks to $\Delta \bar{y}_{t=0}$ help to mask the underlying trend and "flatten" out the observed slope in the pre-period. However, when we underestimate the difference in means between treatment and control in period 0, we tend to overestimate the growth in this difference between period $t = 0$ and $t = 1$ owing to mean reversion, and so the cases where we fail to detect the underlying trend tend to produce particularly large treatment estimates for period 1. Thus, the expected bias conditional on passing the pre-test is worse than the unconditional bias in OLS.

[4] Specifically, I choose σ^2 and N so that the sample mean for each treatment group is standard normal in each period. This leads $\hat{\beta}_1$ and $\hat{\beta}_{-1}$ to each be normally distributed with variance 4 and correlation 0.5. I set $\gamma = 4$, so that the pre-period mean is 2 standard errors below 0.

[5] In Section 4, I conduct simulations based on a review of recent published papers using the more traditional threshold of 1.96.

Figure 1: Intuition for how bias is worse conditional on not detecting a significant pre-trend



Note: The top panel of the figure shows simulated draws from a DGP in which in population the outcome of interest for the treatment group is increasing linearly relative to the control group. The y-axis shows the difference in sample means between the treatment and control group in each period ($\Delta\bar{y}_t$). I highlight in blue the draws of the data in which the t-stat on the pre-period coefficient $\hat{\beta}_{-1}$ is less than one in absolute value. The bottom panel shows the average of the blue lines over 1 million draws. The figure illustrates that the cases in which we fail to detect an upward underlying trend exhibit a small observable pre-trend, but produce particularly large treatment effects estimates.

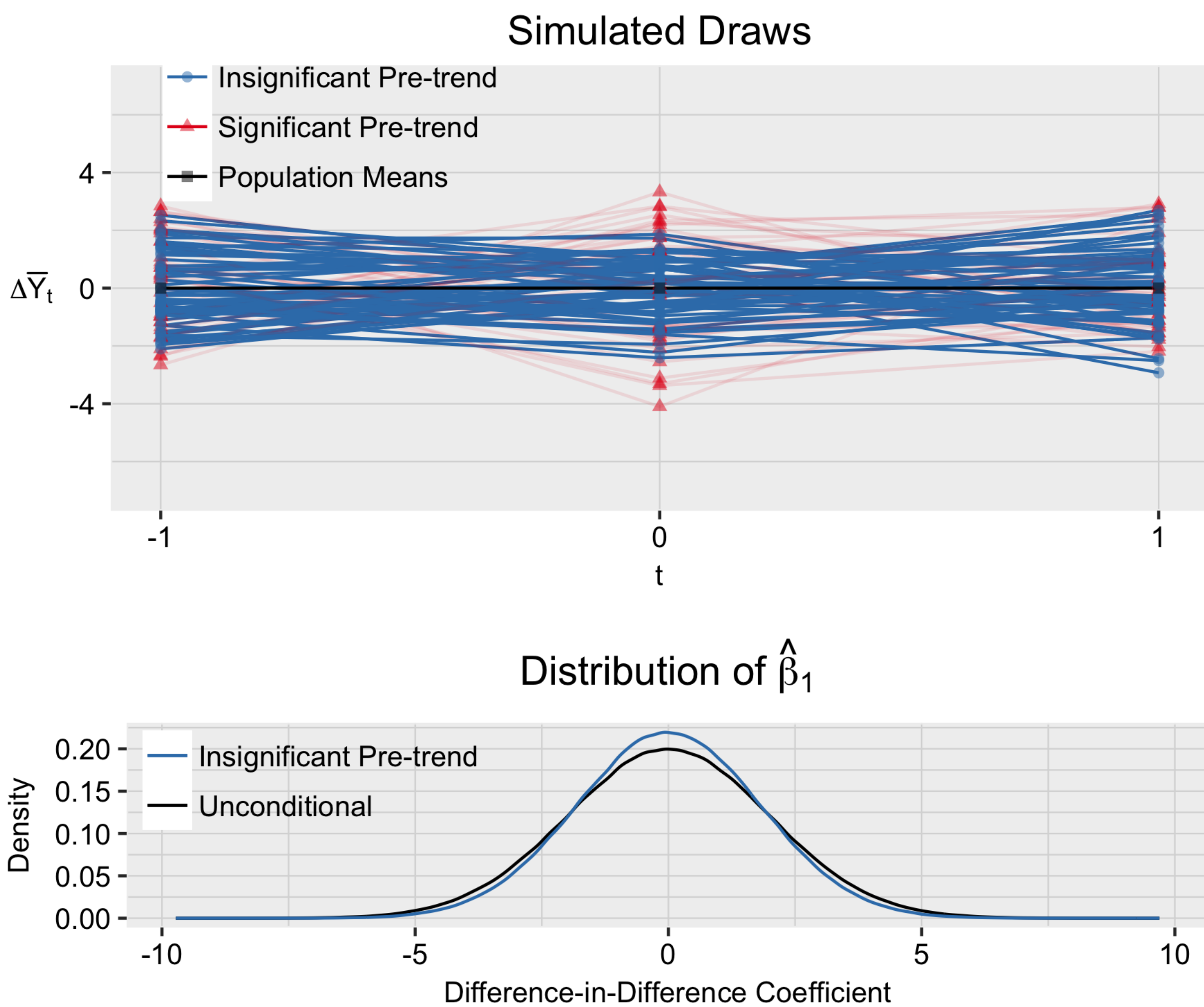
2.3 When parallel trends holds, there is still no bias after pre-trends testing

What happens if parallel trends holds in population? The top panel of Figure 2 shows draws from our DGP when parallel trends holds ($g(t) \equiv 0$). We again highlight in blue the draws in which the t-stat on $\hat{\beta}_{-1}$ is less than 1 in absolute value, and we now highlight in red the draws in which the t-stat is greater than 1 in absolute value. We see from the figure that the draws where we detect a significant pre-trend (in red) disproportionately have extreme values for $\Delta \bar{y}_{t=0}$. This is because when in population there is no underlying trend, we tend to incorrectly detect one when the noise in the reference period is large. As a result, by conditioning on having not found a significant pre-trend, we are excluding cases that on average have a high degree of noise in period 0, and consequently have noisy treatment effects estimates. This leads to a reduction in the variance of the treatment effects estimates, but since we are equally likely to throw out cases where the noise in the reference period is positive or negative, we do not induce any bias in the treatment effects estimates. This can be seen in the bottom panel of Figure 2, which shows that the distribution of estimated treatment effects conditional on finding an insignificant pre-trend is centered around zero but has lower variance than the unconditional distribution.

2.4 After pre-testing, coverage rates can be too high or too low

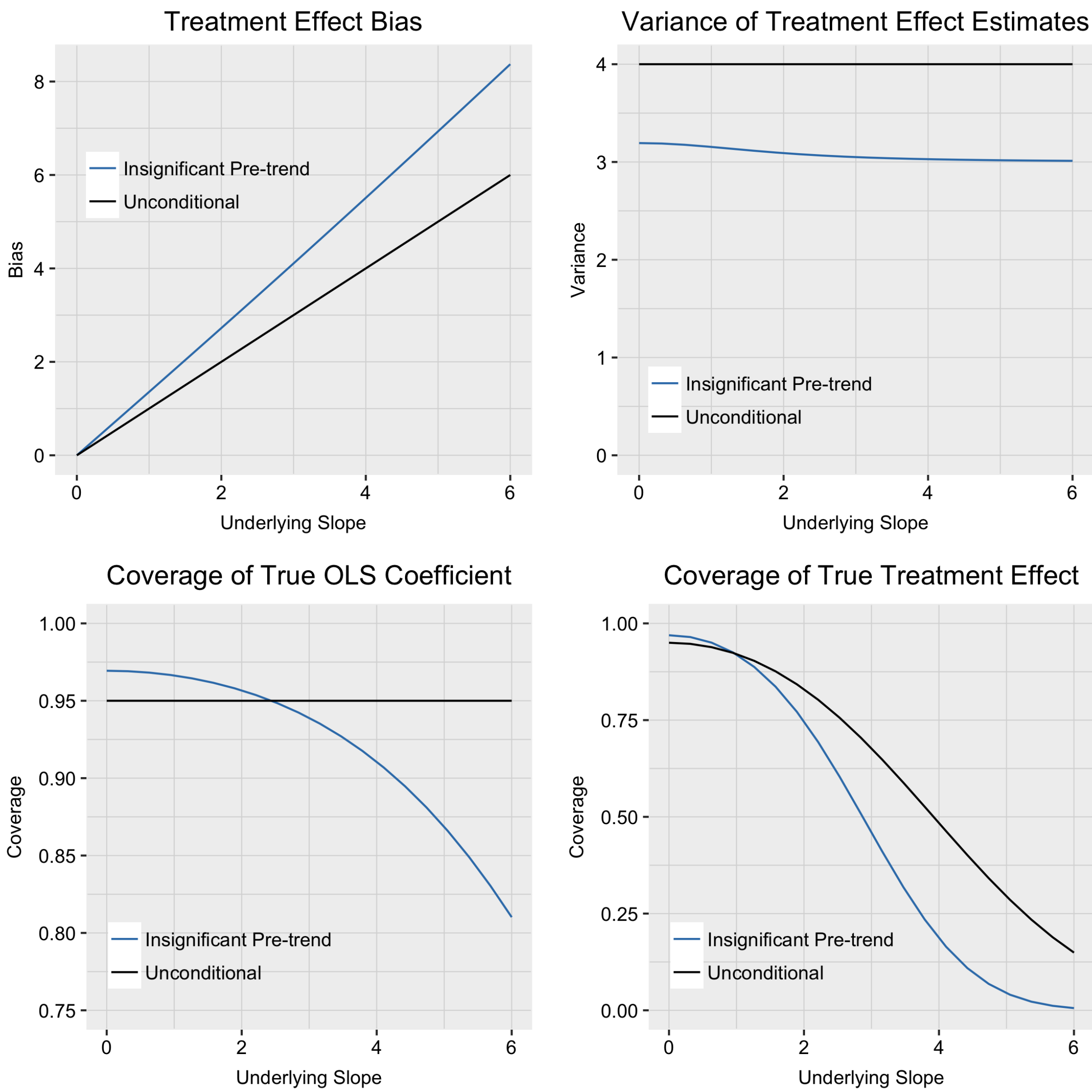
What do these two examples imply for the performance of traditional confidence intervals conditional on finding an insignificant pre-trend? Intuitively, when parallel trends (approximately) holds, the conditional treatment effects estimates are (approximately) unbiased but have lower variance, so traditional confidence intervals will tend to overcover. For larger violations of parallel trends, the bias in the OLS estimates is worse conditional on finding an insignificant pre-trend, leading to undercoverage. These dynamics are captured in Figure 3, which summarizes the performance of the OLS treatment effects estimates and CIs under linear violations of parallel trends as a function of the underlying slope. We see that for underlying slopes close to zero, traditional CIs overcover both the true OLS coefficient and true treatment effect conditional on finding an insignificant pre-trend. However, as the bias gets larger, the traditional CIs undercover the true OLS coefficient conditional on finding an insignificant pre-trend. Additionally, for larger values of the underlying trend, undercoverage of the true treatment effect is substantially worse conditional on not finding a significant pre-trend.

Figure 2: Intuition for why pre-testing reduces variance but preserves unbiasedness when parallel trends holds



Note: The top panel of the figure shows simulated draws from a DGP in which parallel trends holds. The y-axis shows the difference in sample means between the treatment and control group in each period ($\Delta\bar{y}_t$). I highlight in blue the draws of the data in which the t-stat on the pre-period coefficient $\hat{\beta}_{-1}$ is less than one in absolute value, and I highlight in red the draws where the t-stat exceeds 1 in absolute value. The bottom panel shows the distribution of the treatment effects estimates unconditionally and conditional on having found an insignificant pre-trend. The figure illustrates that conditional on finding an insignificant pre-trend, the treatment effects remain unbiased but have lower variance.

Figure 3: Bias, Variance, and Coverage of OLS Treatment Effect Estimates Under Linear Violations of Parallel Trends



Note: This figure shows the performance of the OLS treatment effect estimate under linear violations of parallel trends, both unconditionally and conditional on not detecting a pre-trend ($|t| < 1$). The top two panels show the bias and variance of the treatment effects estimates. The bottom left panel shows the coverage of the true OLS coefficient for a nominal 95% interval; the bottom right panel shows the coverage of the true treatment effect.

2.5 Implications of publication rules that require pre-testing

So far, the analysis has conditioned on whether or not there is an underlying trend in population. A natural follow-up question is what happens when researchers try many different studies, and parallel trends is satisfied in some of these but not others.

This section illustrates via a simple extension of the base model that requiring insignificant pre-trends to publish can either reduce or increase bias in published work in this setting. Intuitively, when we require an insignificant pre-trend to publish, there is a tradeoff between two effects: the parallel trends assumption holds for a higher fraction of published studies, but for any given violation of parallel trends, the expected bias is worse conditional on not finding a significant pre-trend. As a result, whether pre-trends testing reduces or increases bias in published work will depend on both the power of the pre-test to detect meaningful violations of parallel trends, and the fraction of latent research designs in which parallel trends holds.

To clarify the tradeoffs of requiring insignificant pre-trends to publish, we consider a simple extension to the stylized model in which parallel trends holds in fraction $1-\theta$ of latent studies, and in fraction θ of latent studies there is a linear violation of parallel trends with slope $\bar{\gamma} > 0$. We will denote the underlying slope in population by γ, so that $\gamma = \bar{\gamma}$ when parallel trends is violated, and $\gamma = 0$ when parallel trends holds. If we didn't test for parallel trends and published everything, the expected bias in published studies would be:

$$Bias^{\text{No test}} = P(\gamma = \bar{\gamma})\bar{\gamma} = \theta\bar{\gamma}.$$

Likewise, if we only publish cases where we accept the pre-trend, the bias in published studies is:

$$Bias^{\text{Test}} = P(\gamma = \bar{\gamma} \,|\, Accept)\mathbb{E}\left[bias \,|\, \gamma = \bar{\gamma}, Accept\right].$$

The ratio of biases across the two regimes is then:

$$\frac{Bias^{Test}}{Bias^{Notest}} = \underbrace{\frac{P(\gamma = \bar{\gamma} \,|\, Accept)}{P(\gamma = \bar{\gamma})}}_{\substack{\text{Relative fraction of}\\ \text{studies with biased}\\ \text{design } (\leq 1)}} \cdot \underbrace{\frac{\mathbb{E}\left[bias \,|\, \gamma = \bar{\gamma}, Accept\right]}{\bar{\gamma}}}_{\substack{\text{Ratio of bias when accept}\\ \text{biased design } (\geq 1)}}. \tag{3}$$

Equation (3) makes clear the tradeoffs involved in requiring an insignificant pre-trend to publish. The first term represents the relative fraction of published studies with a biased design ($\gamma = \bar{\gamma}$) across the two regimes. Pre-testing makes us relatively more likely to accept a study where parallel trends holds, so this term will tend to be less than 1. However, the second term represents the ratio of biases in the published studies where parallel trends does not hold in population. As demonstrated in Section 2.2, this bias is worse conditional on the pre-test, so the second term will be greater than 1.

The bias under the pre-testing regime will tend to be worse when either the fraction of latent studies with a biased design (θ) is high, or if the pre-test has low power. To see why this is the case, using Bayes' rule we can re-write the first term in (3) as:

$$\frac{1}{\theta + (1-\theta)BF} \tag{4}$$

where

$$BF := \frac{P(Accept|\gamma = 0)}{P(Accept|\gamma = \bar{\gamma})}$$

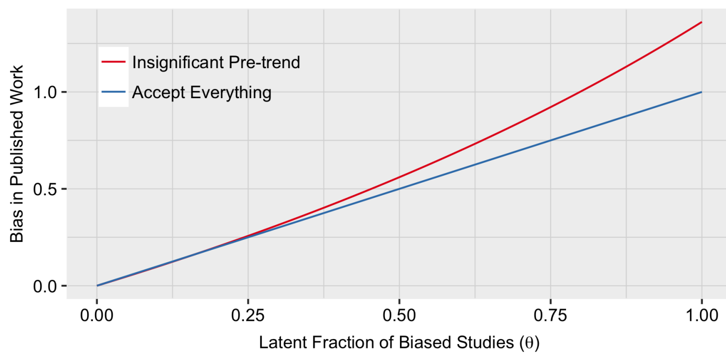
is the Bayes factor, i.e. the ratio of the likelihood of finding an insignificant pre-trend when parallel trends holds relative to when it is violated. The pre-testing regime will tend to have larger bias when the expression in (4) is close to 1. This will occur if θ is close to 1, meaning that a high fraction of latent research designs are biased, or if the Bayes Factor is close to 1, meaning that the pre-test has low power.

These dynamics are captured in Figure 4, which shows the (mean) bias in published studies as a function of θ for three values of $\bar{\gamma}$. These three values of $\bar{\gamma}$ correspond with β_{pre} being 1, 2, and 3 standard errors away from zero, and lead to Bayes factors of 1.4, 4.3, and 30. The top panel shows that when the Bayes factor is small, so that the pre-test is poorly powered, requiring an insignificant pre-test to publish leads to weakly larger bias in published work for all values of θ, with larger differences when θ is large. In the second and third panels, where the power of the pre-test is larger, we see that requiring an insignificant pre-test to publish can substantially reduce bias for lower values of θ, but will nonetheless exacerbate bias if θ is sufficiently large.

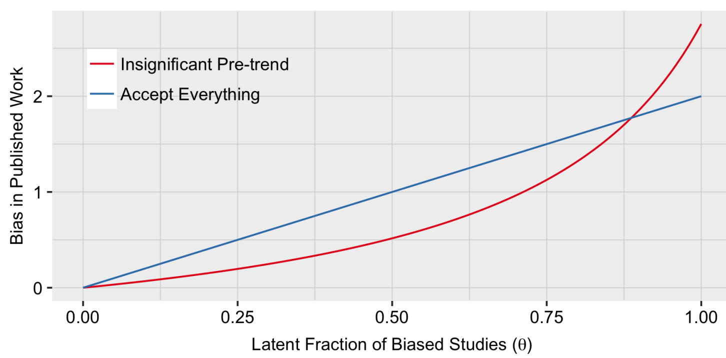
An implication of this section is that researchers should consider both the power of their pre-test to reject meaningful violations of parallel trends as well as the ex ante plausibility of the research design (as proxied by $1-\theta$). If either of these is low, then pre-testing for an insignificant pre-trend in the usual way will likely be ineffective, and can even increase bias in published work.

Figure 4: Comparing bias in published studies when requiring an insignificant pre-trend to publish versus publishing everything

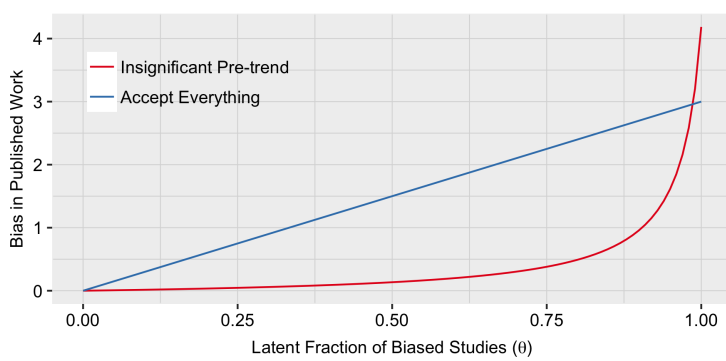
(a) $\bar{\gamma} = \sigma_{pre}$, Bayes factor $= 1.4$



(b) $\bar{\gamma} = 2\sigma_{pre}$, Bayes factor $= 4.3$



(c) $\bar{\gamma} = 3\sigma_{pre}$, Bayes factor $= 30$



Note: Each figure shows the (mean) bias in published work in the setting described in Section 2.5 as a function of the fraction of latent studies in which parallel trends is violated (θ). The Insignificant Pre-trend regime only publishes studies in which the t-stat on the pre-period coefficient is less than 1. The three panels show results for three different values of the slope of the underlying trend when parallel trends fails. See Section 2.5 for further detail.

3 Theory: The effect of pre-trends testing in a more general setting

Section 2 considered the performance of conventional treatment effects estimates after pre-testing in a stylized setting with 3 periods, i.i.d. shocks to the outcome across periods, and linear violations of parallel trends. This section formalizes the intuition from Section 2 and extends the analysis to allow for additional periods, more complicated covariance structures, and non-linear violations of parallel trends.

3.1 The generalized set-up

I consider a setting where the researcher observes a vector of pre-period and post-period coefficients that is jointly normally distributed with known variance:

$$\begin{pmatrix} \hat{\beta}_{post} \\ \hat{\beta}_{pre} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta_{post} \\ \beta_{pre} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right).$$

I denote by K the dimension of the pre-period coefficient vector $\hat{\beta}_{pre}$, and by M the dimension of the post-period coefficients $\hat{\beta}_{post}$. For ease of notation, I will consider the case where $M = 1$ unless noted otherwise; all of the results for $M = 1$ will then apply to each individual post-period coefficient in the case when $M > 1$.

I will analyze the properties of the distribution of $\hat{\beta}_{post}$ conditional on a pre-test using the pre-period coefficients – i.e. conditional on the event that $\hat{\beta}_{pre} \in B$ for some set B. For instance, researchers often test to see whether any of the pre-period coefficients is individually statistically significant at the 5% level, which is captured by the event $\hat{\beta}_{pre} \in B_{NIS} := \{\hat{\beta}_{pre} : |\hat{\beta}_{pre,j}|/\sqrt{\Sigma_{jj}} \leq 1.96 \text{ for all } j\}$.

We decompose the population mean as

$$\begin{pmatrix} \beta_{post} \\ \beta_{pre} \end{pmatrix} = \begin{pmatrix} \tau_{post} \\ 0 \end{pmatrix} + \begin{pmatrix} \delta_{post} \\ \delta_{pre} \end{pmatrix},$$

where τ_{post} is the true causal parameter of interest, and δ represents the (unconditional) bias in conventional estimates from an underlying trend. For instance, in the example in Section 2, the true treatment effect was $\tau_{post} = 0$, but the researcher estimating regression (2) would have bias from the underlying trend given by $\begin{pmatrix} \delta_{post} \\ \delta_{pre} \end{pmatrix} = \begin{pmatrix} g(1) - g(0) \\ g(-1) - g(0) \end{pmatrix}$. If parallel trends holds, then $\delta = 0$. We assume that the treatment of interest has no causal effect prior to its implementation, so that $\beta_{pre} = \delta_{pre}$.

The finite-sample normal model specified above will hold exactly if we assume normal errors, as in the example in Section 2, but can more reasonably be thought of as an asymptotic approximation, since a wide variety of estimation procedures will yield asymptotically normal coefficients via the central limit theorem. For instance, the traditional two-way fixed effects model (2) will lead to asymptotically normal coefficients as N grows large under mild regularity conditions. Other procedures, such as the GMM estimator proposed by Freyaldenhoven et al. (2018) or "instrumental variable" event-studies, will also have an asymptotically normal distribution under suitable regularity conditions, and thus the results here can also be used to analyze the distribution of treatment effects estimates conditional on not finding significant pre-period placebo coefficients in these models as well.

Appendix C 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.[6]

3.2 A formula for how pre-testing impacts treatment effect bias

I first provide a formula for the bias in the treatment effect estimate conditional on passing a pre-test for parallel trends.

Proposition 3.1. *For any conditioning set B,*

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

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-period coefficients from pre-testing, as well as on the normalized covariance between the pre-period and post-period coefficients.

3.3 Unbiasedness after pre-testing when parallel trends holds

In the simple example in Section 2, we saw that under parallel trends, $\hat{\beta}_{post}$ was unbiased conditional on not finding a significant pre-trend. From Proposition 3.1, we see that $\hat{\beta}_{post}$ is conditionally unbiased for τ_{post} if $\delta_{post} = 0$ and $\mathbb{E}\left[\hat{\beta}_{pre} \mid \hat{\beta}_{pre} \in B\right] = \beta_{pre}$. One can show that if $\delta_{pre} = 0$, then $\mathbb{E}\left[\hat{\beta}_{pre} \mid \hat{\beta}_{pre} \in B\right] = \beta_{pre}$ provided that the pre-test B is symmetric in

[6] The condition that the probability of passing the pre-test does not disappear asymptotically requires that the pre-period 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; 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.

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 any two-sided test of significance. It follows that, as in the simple example, $\hat{\beta}$ is conditionally unbiased for the treatment effect of interest when parallel trends holds, so long as the pre-test is symmetric.

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

$$\mathbb{E}\left[\hat{\beta}_{post} \mid \hat{\beta}_{pre} \in B\right] = \tau_{post}.$$

3.4 Conditions under which pre-testing amplifies bias

In the stylized example in Section 2, we saw that when in population there was a linear underlying trend, the bias in the treatment effect estimate was worse conditional on having not detected a significant pre-period coefficient – i.e. the pre-test bias and the bias from trend went in the same direction. In this section, I show that under certain conditions on the covariance structure, this feature holds regardless of the number of pre-periods or the functional form of the underlying trend. I first state the formal result, and then discuss the assumptions on the covariance structure, which are implied by homoskedasticity in the general multi-period case.

Assumption 1. *Let K denote the dimension of $\hat{\beta}_{pre}$.*

1. *If $K = 1$, then we assume that $\Sigma_{12} = \mathrm{Cov}\left(\hat{\beta}_{pre}, \hat{\beta}_{post}\right) > 0$.*
2. *If $K > 1$, we assume that Σ has a common term σ^2 on the diagonal and a common term $\rho > 0$ off of the diagonal, with $\sigma^2 > \rho$.*

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

$$\mathbb{E}\left[\hat{\beta}_{post} \mid \hat{\beta}_{pre} \in B_{NIS}\right] > \beta_{post} > \tau_{post}.$$

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

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

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

If the researcher estimates regression (2), 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, i.e. $\frac{1}{\#\{i|D_i=1\}} \sum_{i \in \{i|D_i=1\}} \epsilon_{it} - \frac{1}{\#\{i|D_i=0\}} \sum_{i \in \{i|D_i=0\}} \epsilon_{it}$. It follows that $\text{Cov}\left(\hat{\beta}_j, \hat{\beta}_k\right) = \text{Cov}\left(\Delta\bar{\epsilon}_j - \Delta\bar{\epsilon}_0, \Delta\bar{\epsilon}_k - \Delta\bar{\epsilon}_0\right)$. Hence, Assumption 1 will hold if $\Delta\bar{\epsilon}_t$ is *iid* across time, since we will have $\mathbb{V}\text{ar}\left[\hat{\beta}_k\right] = 2\sigma^2$ and $\text{Cov}(\hat{\beta}_k, \hat{\beta}_j) = \sigma^2$ for $\sigma^2 := \mathbb{V}\text{ar}\left[\Delta\bar{\epsilon}_t\right]$. A sufficient condition for $\Delta\bar{\epsilon}_t$ to be *iid* across time is for the individual-level errors ϵ_{it} to be *iid*.

With a one-dimensional pre-period coefficient, the requirements of Assumption 1 are less restrictive, as we only require the pre-period and post-period coefficients to be positively correlated. One can show, for instance, that if ϵ_{it} follows an AR(1) process, then Assumption 1 will hold so long as the AR coefficient is strictly less than 1. Although in practice having only one pre-period may be rare, when there are multiple pre-periods researchers may test for a pre-trend using a parametric linear trend, such as

$$y_{it} = \alpha_i + \phi_t + \beta_{trend} \times t \times D_i + \sum_{s>0} \beta_s \times 1[s = t] \times D_i + \epsilon_{it}. \tag{5}$$

In this case, testing the significance of β_{trend} amounts to testing a one-dimensional pre-period coefficient.

What happens if Assumption 1 is not satisfied? One can construct examples using a covariance matrix that violates Assumption 1 in which the conditional bias is less than the unconditional bias, so there is no universal guarantee that the bias is exacerbated with arbitrary covariance structures. However, it is straightforward to calculate whether pre-testing will exacerbate bias for any particular underlying trend and covariance matrix using Proposition 3.1. More specifically, Proposition 3.1 implies that the conditional bias in the treatment effect estimate will be worse than the unconditional bias if and only if the bias from trend δ_{post} and the pre-test bias $\Sigma_{12}\Sigma_{22}^{-1}\left(\mathbb{E}\left[\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in B\right] - \beta_{pre}\right)$ have the same sign. The conditional expectation can be easily calculated via simulation, and for the pre-test of no individually significant coefficient can even be calculated analytically using the formulas of Manjunath and Wilhelm (2012). In Section 4, I apply this approach to calculate the pre-test bias under linear violations of parallel trends in a sample of recently published papers. I show that although in practice homoskedasticity typically does not hold, in most published papers the pre-test bias nonetheless goes in the same direction as the underlying trend.

A second limitation of the result in Proposition 3.2 is that the result applies only to the pre-test that no individual coefficient is statistically significant, as opposed to an arbitrary pre-test. It seems likely that similar results may be available for tests of joint significance using the results on elliptically-truncated normal distributions from Tallis (1963) and Arismendi Zambrano and Broda (2016), but I leave this to future work. However, as in the previous paragraph, I note that the researcher interested in the pre-testing bias from a particular violation of parallel trends can calculate it using Proposition 3.1.

3.5 Pre-testing reduces the variance of estimates

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.

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

In the model in Section 2, we found that the variance of the treatment effect estimate conditional on passing the pre-test for parallel trends was smaller than the unconditional variance. We now show that that this feature holds more broadly under very mild conditions. In particular, we only require that the pre-test is convex, meaning that if we don't reject parallel trends for $\hat{\beta}_{pre,1}$ and $\hat{\beta}_{pre,2}$, then for $\theta \in (0, 1)$, we also will not reject parallel trends for $\theta\hat{\beta}_{pre,1} + (1-\theta)\hat{\beta}_{pre,2}$. This property holds for most common pre-tests – including tests of individual statistical significance, joint tests for significance, and tests for significant linear slopes.

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

4 The practical relevance of pre-testing distortions: evidence from a review of recent papers

Sections 2 and 3 illustrate that pre-trends testing can lead to undesirable distortions in the distribution of treatment effects estimates, particularly in the case when parallel trends is violated. The extent to which these distortions are consequential in practice, however, will depend on the power of pre-tests against meaningful violations of parallel trends. The practical relevance of theses concerns will also depend on whether the covariance matrices, which may not be homoskedastic in practice, are typically such that bias is exacerbated under monotone violations of parallel trends.

This section provides evidence that the theoretical concerns raised in the previous sections are relevant in practice. First, in a systematic review of recent papers in three leading economics journals, I illustrate that conventional pre-tests for parallel trends often have low power even against substantial linear violations of parallel trends. This suggests that the possibility of failing to detect a meaningful violation of parallel trends is not merely a theoretical concern. Second, I show that while homoskedasticity typically does not hold in practice, the bias from pre-testing nonetheless typically amplifies the bias from an underlying trend, and can be of a substantial magnitude.

4.1 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. I chose the phrase "event study" since papers that evaluate pre-trends often do so in a so-called "event study plot." The search returned 70 total papers that include a figure displaying the results from what the authors describe as an event-study.

For my analysis, I further restricted to papers meeting the following criteria:

1. The data to replicate the event-study plot was publicly available.
2. The event-study plot shows point estimates and confidence intervals for dynamic treatment effects relative to some reference period, which is normalized to zero.
3. The authors do not explicitly reject a causal interpretation of the event-study.

Table 1 shows the number of papers that were eliminated by each of the criteria. Unfortunately, the constraint that the data be publicly available eliminated roughly two-thirds

Meets criteria:	Number of Papers
Contains event study plot	70
& Replication data available	27
& Provides standard errors	18
& Normalizes a period to 0	15
& Doesn't reject causal interpretation	12

Table 1: Number of papers meeting criteria for inclusion in review of papers

of the original sample of papers.[7] The second constraint eliminated two groups of papers. First, some papers portray the time-series of the outcome of interest for the treatment group and control group, typically without standard errors. I omit these papers, since I would like to rely on the author's determination of what the appropriate clustering scheme is for standard errors. Second, the restriction that a pre-period be normalized to zero primarily rules out a handful of papers employing a more traditional finance event-study, which examines the time-series of cumulative abnormal return around some event of interest. The final constraint eliminated a handful of papers in which the authors recognize that the pre-trends do not appear to be flat, and either subsequently add time-varying controls or suggest a non-causal interpretation.

Twelve papers contained event-study plots that matched all of the above criteria. Some of these papers present multiple event-study plots, many of which show robustness checks or heterogeneity analyses. I focus here on the first event-study plot presented in each paper, which I view as a reasonable proxy for the main specification in the paper.

4.2 What pre-tests are researchers using?

It is not entirely clear in practice what criteria researchers are using to evaluate pre-trends. By far the most commonly mentioned criterion is that none of the pre-period coefficients is individually statistically significant – e.g. "the estimated coefficients of 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. Moreover, it is clear that a statistically significant pre-period coefficient does not necessarily preclude publication. As shown in Table 2, 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.[8]

[7] I also omit one paper in which the replication code produced different results from the actual paper.

[8] 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.

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 2: 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 4 for more detail on the sample of papers reviewed.

4.3 Evaluating power and pre-test bias in practice

In this section, I evaluate two questions: i) to what extent might pre-trends tests fail to detect meaningful violations of parallel trends?, and ii) to what extent will pre-testing exacerbate bias in practice?

In light of the emphasis in published work on the individual statistical significance of the pre-period coefficients, I base my calculations on this criterion. For each study in my sample, I evaluate the power of the pre-trends test to detect linear violations of parallel trends. Of course, we might also be concerned about non-linear violations of parallel trends – otherwise, we would just control parametrically for a linear time trend. However, it seems reasonable in most settings to consider the possibility of a linear violation of parallel trends, potentially alongside other functional forms for the violation. In this case, my calculations can be viewed as an upper bound on the ability of pre-trends tests to detect meaningful violations of parallel trends.[9]

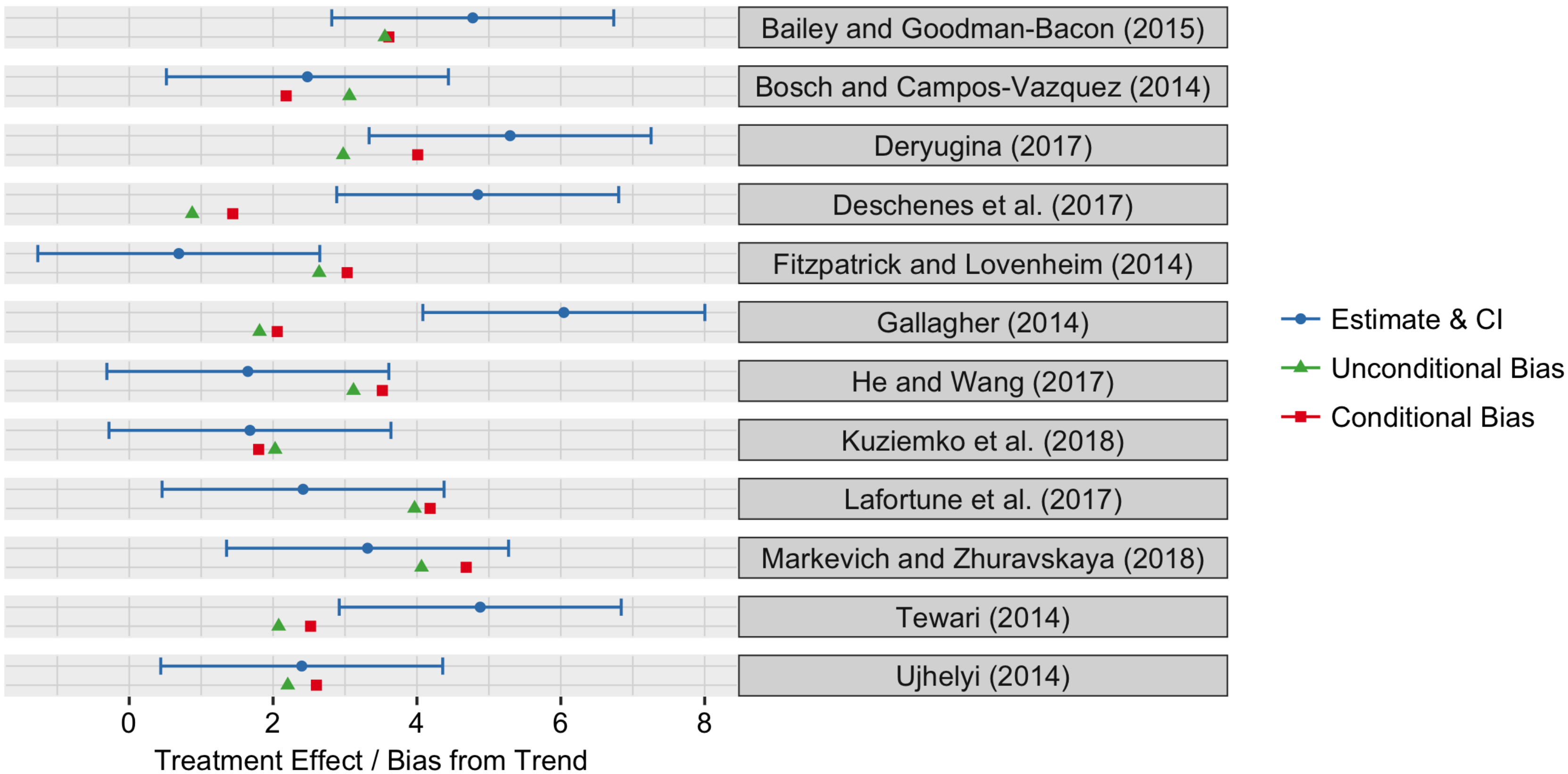
To evaluate the extent to which pre-testing will detect meaningful violations of parallel trends, I compute the linear slope for which the pre-test will achieve 50 or 80 percent power. I

[9]Formally, the worst-case bias over any class of potential violations that includes the class of linear violations is weakly worse than the worst-case bias over the class of linear violations.

choose 80 percent since this is often used as a benchmark for an acceptable degree of power in power analyses (Cohen, 1988). To do this, I obtain the estimated variance-covariance matrix Σ for the pre- and post-period coefficients from each of the event-study regressions studied. Let Σ_{pre} denote the component of the covariance matrix corresponding with the pre-trends coefficients, and let $\beta_{pre}(\gamma) = \gamma \cdot (-K, ..., -1)$ denote a vector of means with an upward slope of γ. I then ask the following question: if the pre-trends coefficients $\hat{\beta}_{pre} \sim \mathcal{N}(\beta_{pre}(\gamma), \Sigma)$, for what value of γ would we reject the research design 50 (80) percent of the time? I refer to this slope as the slope against which we have 50 (80) percent power.

I find that the magnitude of the violations of parallel trends against which we have 50 and 80 percent power can be sizable relative to the magnitude of the estimated treatment effect. Figure 5 plots in green the magnitude of the unconditional bias in the average post-period treatment effect from the linear trends against which we have 80 percent power. The bias from such trends is often of a magnitude comparable to, and in some cases larger than, the estimated treatment effect. Appendix Figure D1 shows the equivalent results for the trends against which we have 50 percent power; even at the lower power threshold, the biases are often of a comparable magnitude to the estimated treatment effects. Appendix Figures D2 and D3 present the analogous results when we restrict attention to the first period only, and show similar patterns.

Figure 5: OLS Estimates and Bias from Linear Trends for Which We Have 80 Percent Power – Average Treatment Effect



Note: I calculate the linear trend against which we would have a rejection probability of 80 percent if we rejected the research design whenever any of the pre-period event-study coefficients was statistically significant at the 5% level. I plot in red the bias that would result from such a trend conditional on not rejecting the research design; I plot in green the unconditional bias from such a trend. In blue, I plot the original OLS estimates and 95% CIs. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The parameter of interest is the average of the treatment effects in all periods after treatment began.

	Treatment Effect:			
	1st Period		All Periods	
	Power Against Slope:			
Paper	0.5	0.8	0.5	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. $100 \cdot \frac{ConditionalBias - UnconditionalBias}{UnconditionalBias}$. See notes to Figure 5 for additional detail on the simulation exercise.

I also find that the bias conditional on passing the pre-test can be substantially worse than the unconditional bias. In Figure 5, I plot in red the average bias that would be induced from the trends against which we have 80 percent power, *conditional* on not finding any significant pre-period coefficient. I calculate these values using the formula in Proposition 3.1. Appendix Figures D1 to D3 show the analogous results for the other specifications. I also summarize the additional bias from pre-testing as a percentage of the unconditional bias from the underlying trend in Table 3. For the trend against which we have 50 percent power, the pre-test bias can be as much as 103 percent of the bias from the trend for the first period after treatment, and as much as 48 percent for the average of the post-periods.[10] The analogous values are even larger when looking at the trend against which we have 80 percent power. Moreover, the pre-test bias and the bias from trend go in the same direction for the average treatment effect in all but two of the studies in the sample, and all but three of the studies for the first period. Thus, although not always true, the prediction of the direction of the bias from the homoskedastic case holds in most cases I consider.

[10]We expect the bias from pre-testing to be a larger fraction of the bias from the trend in periods closer to treatment, since the bias from the 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).

5 Pre-test Corrected Event Studies

The results developed so far suggest that many of the properties we expect of conventional event-study estimates and confidence intervals do not hold in the cases when they're actually reported – i.e. conditional on not detecting a significant pre-trend. In this section, I derive alternative estimators, along with associated confidence intervals, that correct for the bias and coverage distortions induced by conditioning on having not observed a significant pre-trend.

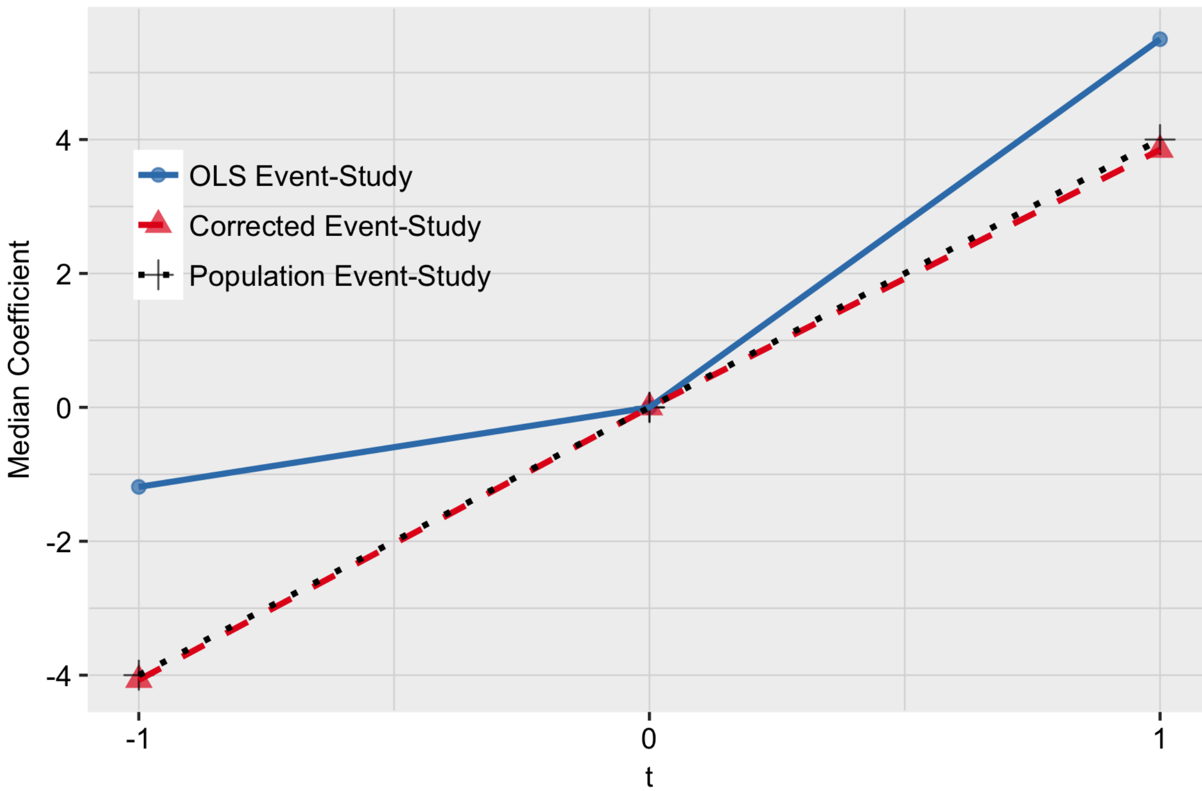
The estimator I develop provides median-unbiased estimates and correct confidence intervals for the true event-study coefficients, conditional on passing the pre-test. To make this concrete, consider the example from Section 2 in which in population there was a linear violation of parallel trends. Figure 6 shows in blue the median OLS event-study coefficients from estimating specification (2), conditional on not finding a significant pre-period coefficient. We see from the figure that conditional on passing the pre-test, we tend to observe a relatively flat pre-trend, and a large kink after treatment – despite the fact that in population the trend is a straight line. In other words, the distortions from pre-testing "mask" the underlying trend in the event-study. By contrast, the estimator developed in this section, plotted in red in Figure 6, on-median shows a straight line conditional on passing the pre-test. The corrections in this section thus undo the distortions to the event-study plot induced by pre-testing. It is important to note, however, that while these corrections remove the distortions from pre-testing, as usual the population post-period coefficients in the event-study will not correspond with the true treatment effect if parallel trends is violated.

5.1 Construction of the corrected estimator and CIs

In this section, I discuss the construction of a median-unbiased estimator and confidence intervals for a parameter of the form $\eta'\beta$, where $\beta = (\beta_{pre}, \beta_{post})$ and η is a vector of the appropriate length. By setting η to be the appropriate basis vector, we can therefore do estimation and inference for each individual event-study coefficient, and we can then collect the results to form a corrected event-study. If we have M periods after treatment, we could also for instance do inference on the average post-period treatment effect by setting η to put weight 0 on the pre-period coefficients and weight $1/M$ on each of the post-period coefficients.

I begin with a derivation for the case where the researcher does a pre-test for a fixed specification. I then show that the results can be extended to cases where the researcher searches over multiple specifications – e.g. with different sets of control variables or focusing on different subpopulations – and selects the final specification on the basis of the observed pre-trends.

Figure 6: Median OLS and corrected event-study estimates, conditional on not finding a significant pre-trend, in the example from Section 2



Note: This figure shows the median OLS and corrected-event study estimates, conditional on not finding a significant pre-trend, in the example considered in Figure 1. The results are based on 1,000 simulations from the DGP considered in Figure 1. See Section 2 for additional details on the DGP considered, and Section 5 for details on the construction of the corrected event-study.

5.1.1 Correcting for a pre-test with a fixed specification

We begin by deriving the distribution of $\eta'\hat{\beta}$ conditional on the event $\hat{\beta} \in B$.[11] In general, the conditional distribution of $\eta'\hat{\beta}$ will depend on the full parameter vector β, and we will therefore condition also on a minimal sufficient statistic for the other components of β. The following result extends Theorem 5.2 in Lee et al. (2016), who show the result for the particular case where B is a polyhedron.

Lemma 5.1 (Conditional distribution of $\eta'\hat{\beta}$)**.** *Let $\hat{\beta} = (\hat{\beta}_{post}, \hat{\beta}_{pre})$ and $\eta \neq 0$ be in $\mathbb{R}^{K+M}$. Define $c = \Sigma\eta/(\eta'\Sigma\eta)$ and $Z = (I - c\eta')\hat{\beta}$. Then*

$$\eta'\hat{\beta} \,|\, \hat{\beta} \in B, Z = z \sim \xi \,|\, \xi \in \Xi(z),$$

for $\xi \sim \mathcal{N}(\eta'\beta,\, \eta'\Sigma\eta)$, and $\Xi(z) := \{x : \exists \hat{\beta} \in B \text{ s.t. } x = \eta'\hat{\beta} \text{ and } z = (I - c\eta')\hat{\beta}\}$.

Having derived the conditional distribution $\eta'\hat{\beta}$, we can then make use of results on optimal quantile-unbiased estimators and inference for exponential family distributions, which were originally developed by Pfanzagl (1994). Similar techniques have been used recently in papers by Andrews and Kasy (forthcoming) on publication bias, Lee et al. (2016) on inference for the LASSO, and Andrews et al. (2018) on inference for "winners".

Proposition 5.1 (Optimal quantile-unbiased estimation)**.** *Let $\eta \neq 0$ be in $\mathbb{R}^{K+M}$. Assume that $\hat{\beta} \in B$ with positive probability, and that Σ is full rank. Let F^{Ξ}_{μ,σ^2} denote the CDF of the normal distribution with mean μ and variance σ^2 truncated to the set Ξ. Define $\hat{b}_\alpha(\eta'\hat{\beta}, z)$ to be the value of x that solves $F^{\Xi(z)}_{x,\eta'\Sigma\eta}(\eta'\hat{\beta}) = 1 - \alpha$, for $\Xi(z)$ as defined in Lemma 5.1. Then for any $\alpha \in (0,1)$, $\hat{b}_\alpha$ is α-quantile unbiased conditional on $\hat{\beta} \in B$,*

$$P\left(\hat{b}_\alpha(\eta'\hat{\beta}, Z) \leq \eta'\beta \,|\, \hat{\beta} \in B\right) = 1 - \alpha.$$

Further, suppose that the parameter space for β is an open set, and that the distribution of $\eta'\hat{\beta} \,|\, Z, \hat{\beta} \in B$ is continuous for almost every Z. Then $\hat{b}_\alpha$ is uniformly most concentrated in the class of α-quantile-unbiased estimators, in the sense that for any other α-quantile unbiased estimator $\tilde{b}_\alpha$, and any loss function $L(x, \eta'\beta)$ that attains its minimum at $x = \eta'\beta$ and is increasing as x moves away from $\eta'\beta$,

$$\mathbb{E}\left[L\left(\hat{b}_\alpha(\eta'\hat{\beta}, Z)\,, \eta'\beta\right) \,|\, \hat{\beta} \in B\right] \leq \mathbb{E}\left[L\left(\tilde{b}_\alpha\,, \eta'\beta\right) \,|\, \hat{\beta} \in B\right].$$

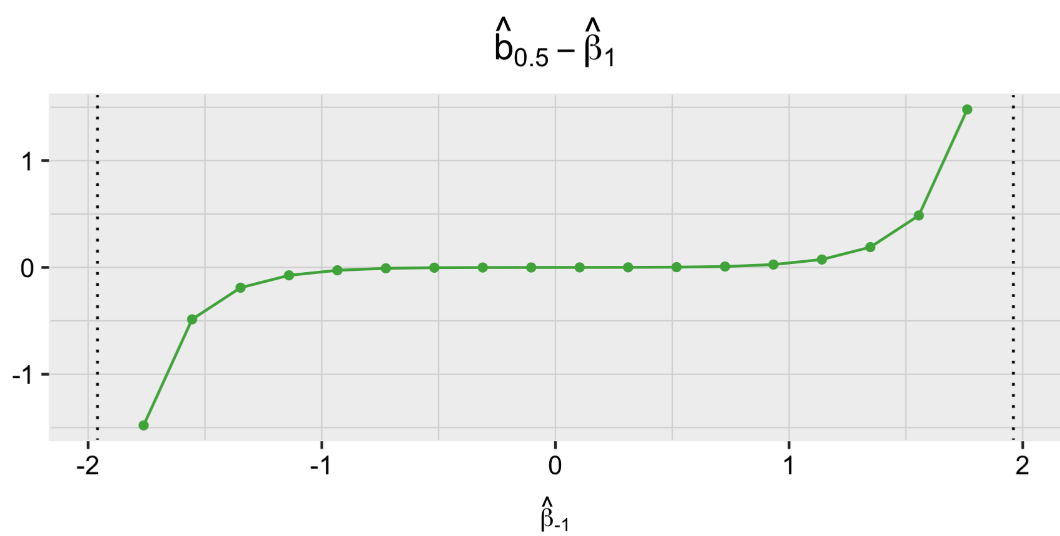
[11] In a slight change of notation, I will now refer to B as the conditioning set for the full parameter vector $\hat{\beta} = (\hat{\beta}_{pre}, \hat{\beta}_{post})$ rather than for $\hat{\beta}_{pre}$ only. Note that we can write the event $\hat{\beta}_{pre} \in B_{pre} \subset \mathbb{R}^K$ as $\hat{\beta} \in B = \{(\beta_{pre}, \beta_{post}) \in \mathbb{R}^{K+M} \,|\, \beta_{pre} \in B_{pre}\}$.

Corollary 5.1. *Under the conditions of Proposition 5.1, conditional on $\hat{\beta} \in B$, $\hat{b}_{0.5}(\eta'\hat{\beta}, Z)$ is a uniformly most-concentrated median-unbiased estimate of $\eta'\beta$, and the interval $\mathcal{C}_{1-\alpha} := [\hat{b}_{\alpha/2}(\eta'\hat{\beta}, Z), \hat{b}_{1-\alpha/2}(\eta'\hat{\beta}, Z)]$ is a $1-\alpha$ level confidence interval for $\eta'\beta$.*

Applying these results in practice requires calculation of the set $\Xi(z)$. In Appendix B, I derive easy-to-calculate formulas for $\Xi(z)$ for the cases where B is based on linear or quadratic restrictions on $\hat{\beta}$, which covers the common cases of tests based on individual or joint significance.

Intuitively, the median-unbiased estimator proposed above chooses the value $\hat{b}_{0.5}$ so that if the parameter of interest $\eta'\beta$ were equal to $\hat{b}_{0.5}$, then the observed value $\eta'\hat{\beta}$ would be at the median of the distribution conditional on passing the pre-test. Figure 7 illustrates how the median-unbiased estimator for β_1 differs from the traditional estimate $\hat{\beta}_1$ in a simple case where we have one pre-period and one post-period coefficient, $\Sigma = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$, and we condition on the event $|\hat{\beta}_{pre}| \leq 1.96$. As can be seen, the adjustment to the traditional estimate is small for values of $\hat{\beta}_1$ close to 0, but becomes (arbitrarily) large as $\hat{\beta}_1$ approaches the boundary. This is because the median of $\hat{\beta}_1$ conditional on $\hat{\beta}_{-1} \in [-1.96, 1.96]$ is close to the boundary only if β_1 is very large in magnitude.

Figure 7: Difference between median-unbiased and traditional estimator in 2-period example



Note: This figure shows the difference between the median unbiased estimator $\hat{b}_{0.5}$ and the traditional estimate $\hat{\beta}_1$ in a simple example with one pre-period and one post-period, where $\hat{\beta}_1$ and $\hat{\beta}_{-1}$ each have variance 1 and correlation 0.5, and we condition on $|\hat{\beta}_{-1}| \leq 1.96$

5.1.2 Correcting for specification search using pre-trends

So far, I have considered the case where the researcher accepts or rejects a fixed research design on the basis of pre-trends. In practice, however, researchers may choose among multiple specifications on the basis of pre-trends. For instance, a researcher may first evaluate tests for pre-trends in a large sample, and then upon finding a significant pre-trend, restrict to a subsample in which the pre-trends appear to be better. Likewise, a researcher may evaluate the pre-trends both with and without certain controls in their regression, and choose the specification with the flattest observable pre-trends.

The machinery developed so far can easily be adapted to handle selection among a finite number of specifications on the basis of the pre-trends. Suppose that we have M models, each with estimated event-study coefficients $\hat{\beta}^m = (\hat{\beta}^m_{pre}, \hat{\beta}^m_{post})$. Let $\hat{\beta}^{stacked} = (\hat{\beta}^1, ..., \hat{\beta}^M)$ denote the stacked vector of coefficients across the M models. For OLS, the stacked vector of coefficients can be estimated using Seemingly Unrelated Regressions (SUR), and so will typically be asymptotically normal. Under a normal approximation for $\hat{\beta}^{stacked}$, we can immediately apply the results from the previous section to obtain median-unbiased estimates and valid confidence intervals for each of the elements of β^m, conditional on model m being chosen. That is, letting B^*_m denote the set of values for $\hat{\beta}^{stacked}$ such that model m is chosen, we can obtain adjusted estimates with the property that $\mathbb{P}\left(\hat{b}^m_j \leq \beta^m_j \mid \hat{\beta}^{stacked} \in B^*_m\right) = 0.5$. Conditional coverage of the corrected confidence intervals can be defined analogously.

Implementing these corrected estimates and confidence intervals in practice requires calculation of the set $\Xi(z)$ accounting for the model selection rule. In Appendix B, I show that $\Xi(z)$ can be easily calculated for a variety of model selection rules, including rules where the researcher tries a series of models and stops when she finds one without a significant pre-trend, or where she chooses the model with the smallest pre-trend.

5.2 Practical Recommendations and Caveats

In practice, I recommend that researchers who rely on pre-tests for parallel trends report corrected event-studies along with power calculations for what they perceive to be economically relevant violations of parallel trends. For a given pre-trend of interest δ_{pre}, the power of the pre-test B is the probability that a $\mathcal{N}(\delta_{pre}, \Sigma_{pre})$ variable falls outside the region B, which can be easily calculated by simulation or, for instance, using the R package mvtnorm (Genz et al., 2018). The power calculations are important because although the corrections proposed in this section correct the bias from pre-testing, they do not solve the issue that the power of the pre-test against violations of parallel trends may be low. In other words, while the corrected point estimates for the pre-period coefficients will be median-unbiased for the population event-study coefficients, as with an event-study plot that is not distorted by pre-testing, the post-period event-study coefficients will be biased for the treatment effect of interest if $\delta_{post} \neq 0$. The power calculations thus give the reader a sense of the probability that a violation of parallel trends of a given magnitude would be detected via the pre-trends test, whereas the corrected event study estimates allow us to evaluate the results in light of the distortions from the pre-test.

The relevant functional form for the potential violation of parallel trends depends on the economic context.[12] In many contexts, we may be worried that the outcome of interest was growing steadily at a different rate relative to the control group, in which case considering linear violations of parallel trends would be a sensible approach. In other contexts, we may worry that the policy of interest was implemented in response to shocks that occurred in close proximity to the timing of treatment, in which case the researcher should consider violations of parallel trends where the treatment and control group begin to diverge only in the pre-periods close to treatment.

Finally, an important caveat to these corrections is that journal editors may be inclined to only publish papers whose conclusions are "robust" to correcting for a pre-trends test. However, the guarantees of unbiasedness and proper coverage are valid conditional on the original pre-test for parallel trends, *not* conditional on passing the original pre-test *and* having the event-study "survive" the correction. Thus, if all authors apply these corrections to their working papers, these desirable properties will hold over the sample of working papers, but likely will not hold conditional on the paper being published. In a sense, this is analogous to the application of Bonferroni adjustments to studies with multiple hypothesis tests. An author who Bonferroni-adjusts all of her studies will falsely reject a null in no more than 5% of the studies she runs (assuming a 5% significance threshold). However, if journals are more likely to accept the papers where the results are significant after Bonferroni-adjustment, then she may falsely reject a null in more than 5% of her published papers. Likewise, if a researcher applies my corrections for pre-testing to each project where she does not reject the research design on the basis of pre-trends, median-unbiasedness and proper coverage will hold over the event-study plots in her studies where she accepts the research design. However, these properties may not hold in the author's published papers if the probability of publication depends on the results of the correction. I refer the reader interested in correcting published

[12]Kahn-Lang and Lang (2018) rightly observe that the parallel trends assumption necessary for identification differs depending on the functional form used for the outcome (e.g. levels versus logs), and argue that the choice of functional form should be motivated by context-specific economic knowledge. I argue here that when authors test for parallel trends, they are trying to differentiate between different models of the world, in some of which the parallel trends assumption holds and in some of which it is violated. As in Kahn-Lang and Lang (2018), the relevant models to consider should likewise be determined by context-specific economic knowledge.

estimates to the supplement of Andrews and Kasy (forthcoming), who show how to correct for publication bias provided that one knows the probability of publishing as a function of the both the estimated pre-trend coefficients and the estimated treatment effects.

5.3 Simulations based on the survey of papers

I evaluate the performance of the corrected event-studies in simulations based on the survey of papers discussed in Section 4. As in Section 4, for each paper I consider a setting where the unconditional distribution of the event-study coefficients $\hat{\beta}$ is $\mathcal{N}(\beta(\gamma), \Sigma)$, where Σ is the estimated variance covariance matrix in the regression in the paper and $\beta(\gamma)$ represents a linear upward trend with slope γ. I consider the case where in population parallel trends holds, so that $\gamma = 0$, and where γ is chosen such that the power of the pre-test is 0.5 or 0.8 (see Section 4.3 for details on this calculation). For each paper and each value of the underlying trend, I conduct 5,000 Monte Carlo simulations, and I evaluate the performance of conventional estimates and of the corrected event-studies in the cases that pass the pre-test – i.e. where no significant pre-trend coefficient is detected.

I calculate the median bias for each estimator relative to the population event-study coefficients β, with the results shown in Table 4. For comparability between studies with different units, I measure the median bias in the estimate for the coefficient β_i as the difference between the exceedance probability $\mathbb{P}\left(\hat{\beta}_i > \beta_i\right)$ and 0.5, as in Andrews et al. (2018). A median bias of 0 therefore represents unbiasedness, whereas the maximum value of 0.5 represents the case where $\hat{\beta}_i$ always exceeds β_i. I compute these probabilities for each individual event-study coefficient, and then report the average across all pre-periods and all post-periods for each study.

The first three columns of Table 4 show the median bias of conventional estimates. When parallel trends holds, conventional estimates are median-unbiased in both the pre- and post-periods (first column). When there is an upward trend, however, both the pre-period and post-period coefficients tend to be upward biased (columns 2 and 3). That is, the pre-period coefficients are biased towards zero (since their means are negative) whereas the post-period coefficients are biased away from zero (since their means are positive). Thus, as in the simple example in Figure 6, conditional on passing the pre-test conventional estimates tend to mask the underlying linear trend by showing a kink rather than a straight line. By contrast, the final three columns of the table indicate that the corrected event-study estimates are median-unbiased (up to simulation error) for the true event-study coefficients conditional on passing the pre-test, regardless of the underlying trend. The corrected event-studies thus undo the kink introduced by pre-testing (in the median case).

Table 5 shows that the confidence intervals for the corrected event-study estimates also have (approximately) correct size for the population event-study coefficients, whereas the size of traditional CIs can be above or below the nominal rate. In particular, when parallel trends holds, rejection rates conditional on passing the pre-test are too low for conventional CIs, since point estimates are unbiased but – as discussed in Section 3.5 – the variance of the conventional estimator conditional on passing the pre-test is lower than the unconditional variance, on which the standard errors are based. When the underlying trend is non-zero, the conditional variance again remains lower than the unconditional variance, but the CIs are non-centered, so rejection rates can be either too high or too low.

Table 6 illustrates that the improved bias and coverage properties of the corrected event-study estimates do come at a cost in terms of the width of the confidence intervals. The table shows the median ratio of the width of the 95% CI for the corrected event-study estimates relative to the conventional CI.[13] A notable feature of the corrected event-study CIs, however, is that the inflation in width relative to conventional CIs tends to be smaller in the case when parallel trends holds in population. When parallel trends holds, the inflation in the post-period CIs is between 0 and 26 percent, and the inflation in the pre-period CIs is between 33 and 62 percent. By contrast, the inflation in the CIs can be above 100 percent in both the pre- and post-periods when we consider the data-generating process with a linear slope against which we have 80 percent power. This disparity occurs because the adjusted CIs become large as the vector of pre-period coefficients approaches the rejection boundary, and this occurs more frequently the larger is the underlying trend in population.

[13] I focus on medians since the results of Kivaranovic and Leeb (2018) imply that the corrected CIs will have infinite average length in this context. Although this is a somewhat undesirable feature, I note that the CIs are typically informative, and the expectation diverges only because the adjusted CIs become arbitrarily wide as the pre-period coefficients approach the boundary of the pre-test.

	Estimator:					
	$\hat{\beta}^{Conventional}$			$\hat{b}_{0.5}$		
	Slope of Pre-trend:					
Paper	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	-0.00	0.22	0.40	0.00	-0.00	-0.00
Bosch and Campos-Vazquez (2014)	-0.00	0.33	0.48	-0.00	-0.01	-0.02
Deryugina (2017)	-0.01	0.20	0.40	-0.00	-0.01	-0.00
Deschenes et al. (2017)	-0.01	0.20	0.37	-0.01	-0.02	-0.02
Fitzpatrick and Lovenheim (2014)	-0.00	0.23	0.39	-0.00	-0.01	-0.02
Gallagher (2014)	0.00	0.10	0.24	0.00	0.00	0.01
He and Wang (2017)	-0.00	0.28	0.45	-0.01	-0.01	0.02
Kuziemko et al. (2018)	0.01	0.22	0.48	0.01	0.01	-0.01
Lafortune et al. (2017)	0.01	0.27	0.45	0.01	-0.01	-0.01
Markevich and Zhuravskaya (2018)	0.00	0.22	0.42	0.00	-0.00	-0.02
Tewari (2014)	-0.00	0.15	0.33	-0.00	-0.01	-0.00
Ujhelyi (2014)	0.01	0.28	0.48	0.01	-0.01	-0.00

(a) Median Bias for Pre-period Event-Study Coefficients

	Estimator:					
	$\hat{\beta}^{Conventional}$			$\hat{b}_{0.5}$		
	Slope of Pre-trend:					
Paper	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	-0.00	0.01	0.03	0.00	-0.00	0.01
Bosch and Campos-Vazquez (2014)	-0.00	-0.19	-0.34	-0.00	0.01	0.02
Deryugina (2017)	0.00	0.18	0.31	0.00	-0.00	-0.00
Deschenes et al. (2017)	-0.00	0.11	0.24	0.00	-0.01	-0.01
Fitzpatrick and Lovenheim (2014)	-0.00	0.08	0.13	0.00	0.01	-0.01
Gallagher (2014)	-0.00	0.03	0.06	-0.00	-0.00	-0.00
He and Wang (2017)	0.00	0.08	0.14	0.00	-0.00	0.01
Kuziemko et al. (2018)	-0.00	-0.05	-0.09	-0.00	-0.00	0.01
Lafortune et al. (2017)	0.00	0.01	0.02	0.00	-0.00	-0.01
Markevich and Zhuravskaya (2018)	0.00	0.12	0.20	0.00	0.00	-0.02
Tewari (2014)	-0.00	0.06	0.14	-0.01	-0.01	0.00
Ujhelyi (2014)	0.00	0.08	0.16	0.00	-0.01	-0.00

(b) Median Bias for Post-period Event-Study Coefficients

Table 4: Median Bias for Event-Study Coefficients Conditional on Not Finding a Significant Pre-trend Coefficient

Note: This table shows the median bias of conventional and corrected event-study estimates in simulations based on a sample of recent papers. All results are shown conditional on not finding a significant pre-period event study coefficient. For comparability across studies, I measure median bias as the probability the parameter estimate is greater than the population event-study coefficient minus one half, as in Andrews et al. (2018). For each study, I calculate median bias for each individual pre-period and post-period coefficient, and then average over all of the pre periods (top panel) or post periods (bottom panel). I present results for the case where parallel trends holds in population, and when in population there is a linear violation of parallel trends with slope such that we have 0.5 or 0.8 power ($\gamma_{0.5}$ and $\gamma_{0.8}$). See Section 5.3 for details.

	Estimator:					
	$\hat{\beta}^{Conventional}$			$\hat{b}_{0.5}$		
	Slope of Pre-trend:					
Paper	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	0.00	0.04	0.09	0.05	0.06	0.05
Bosch and Campos-Vazquez (2014)	0.00	0.04	0.12	0.05	0.05	0.07
Deryugina (2017)	0.00	0.04	0.08	0.05	0.05	0.05
Deschenes et al. (2017)	0.00	0.03	0.07	0.05	0.05	0.05
Fitzpatrick and Lovenheim (2014)	0.00	0.04	0.09	0.05	0.05	0.06
Gallagher (2014)	0.00	0.02	0.03	0.05	0.05	0.05
He and Wang (2017)	0.00	0.05	0.11	0.05	0.05	0.06
Kuziemko et al. (2018)	0.00	0.03	0.05	0.04	0.04	0.04
Lafortune et al. (2017)	0.00	0.04	0.11	0.05	0.05	0.05
Markevich and Zhuravskaya (2018)	0.00	0.04	0.08	0.05	0.05	0.05
Tewari (2014)	0.00	0.03	0.07	0.05	0.05	0.05
Ujhelyi (2014)	0.00	0.04	0.10	0.05	0.05	0.05

(a) Rejection Probability for True Pre-period Event-Study Coefficients

	Estimator:					
	$\hat{\beta}^{Conventional}$			$\hat{b}_{0.5}$		
	Slope of Pre-trend:					
Paper	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.05	0.06	0.05	0.05	0.06
Bosch and Campos-Vazquez (2014)	0.03	0.04	0.09	0.05	0.05	0.06
Deryugina (2017)	0.03	0.04	0.08	0.05	0.05	0.05
Deschenes et al. (2017)	0.03	0.03	0.04	0.05	0.05	0.04
Fitzpatrick and Lovenheim (2014)	0.05	0.05	0.06	0.05	0.05	0.06
Gallagher (2014)	0.04	0.04	0.04	0.05	0.05	0.05
He and Wang (2017)	0.04	0.05	0.07	0.05	0.05	0.06
Kuziemko et al. (2018)	0.04	0.03	0.02	0.05	0.05	0.04
Lafortune et al. (2017)	0.05	0.04	0.05	0.05	0.05	0.05
Markevich and Zhuravskaya (2018)	0.04	0.05	0.07	0.05	0.05	0.06
Tewari (2014)	0.04	0.04	0.06	0.05	0.05	0.05
Ujhelyi (2014)	0.04	0.04	0.05	0.05	0.05	0.05

(b) Rejection Probability for True Post-period Event-Study Coefficients

Table 5: Rejection Probabilities for Population Event-Study Coefficients, Conditional on Not Finding a Significant Pre-trend Coefficient

Note: This table shows the probability that a 95% confidence interval rejects the population event-study coefficient, using simulations based on a sample of recent papers. The first three columns show results for conventional CIs, whereas the latter three columns show results for the corrected event-study estimates. The rejection rates are calculated for each period, and then the average is calculated pooling all pre-period coefficients (top panel) and post-period coefficients (bottom panel). I present results for the case where parallel trends holds in population, and when in population there is a linear violation of parallel trends with slope such that we have 0.5 or 0.8 power ($\gamma_{0.5}$ and $\gamma_{0.8}$). See Section 5.3 for details.

	Pre-periods			Post-periods		
	Slope of Pre-trend:					
Paper	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	1.49	1.77	2.18	1.00	1.00	1.00
Bosch and Campos-Vazquez (2014)	1.62	1.96	2.61	1.21	1.42	1.88
Deryugina (2017)	1.45	1.74	2.14	1.26	1.52	1.86
Deschenes et al. (2017)	1.50	1.81	2.31	1.14	1.40	1.83
Fitzpatrick and Lovenheim (2014)	1.43	1.71	2.11	1.01	1.04	1.13
Gallagher (2014)	1.55	1.77	2.17	1.04	1.07	1.16
He and Wang (2017)	1.45	1.84	2.23	1.01	1.05	1.13
Kuziemko et al. (2018)	1.33	1.60	2.06	1.09	1.42	2.01
Lafortune et al. (2017)	1.51	1.85	2.30	1.01	1.04	1.08
Markevich and Zhuravskaya (2018)	1.39	1.68	2.16	1.03	1.16	1.43
Tewari (2014)	1.58	1.71	2.07	1.08	1.12	1.27
Ujhelyi (2014)	1.49	1.79	2.29	1.04	1.12	1.32

Table 6: Median Ratio of CI Widths (Corrected Event Study CIs relative to Conventional CIs)

Note: This table shows the median ratio of the width of the 95% confidence interval for the corrected event-study estimates relative to conventional CIs, using simulations based on a sample of recent papers. The ratios are calculated for each period, and then the median is calculated pooling all pre-period coefficients (first 3 columns) and post-period coefficients (last 3 columns). I present results for the case where parallel trends holds in population, and when in population there is a linear violation of parallel trends with slope such that we have 0.5 or 0.8 power ($\gamma_{0.5}$ and $\gamma_{0.8}$). See Section 5.3 for details.

6 Conclusion

This paper illustrates the importance of accounting for tests for pre-trends, which are common in applied work. I show both theoretically and in simulations based on a survey of recent papers that pre-trends testing can lead to undesirable distortions in the distribution of conventional treatment effect estimates, and potentially even amplify bias in published work. I introduce a method for correcting event-study plots for pre-testing or model selection using pre-trends, and I encourage researchers relying on pre-testing to report these corrected estimates along with simple calculations of the power of their pre-test against economically relevant violations of parallel trends.

I conclude by highlighting two potential directions for future research. First, while I have analyzed the implications of researchers pre-testing or choosing among a finite number of models on the basis of pre-trends, the synthetic control method (Abadie et al., 2010) essentially automates the selection of a control group on the basis of pre-trends. Future research might investigate the extent to which the synthetic control method can cause bias amplification similar to that under simple pre-testing when parallel trends is violated. Second, while I have analyzed the current practice of testing for pre-trends, an interesting question for future research is the extent to which the current pre-testing regime could be improved upon, either by modifying the rules used for the pre-test or adopting a different approach to account for possible violations of the parallel trends assumption.

References

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

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505.

Abraham, S. and Sun, L. (2018). Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects. SSRN Scholarly Paper ID 3158747, Social Science Research Network, Rochester, NY.

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. ("forthcoming"). Identification of and Correction for Publication Bias. *American Economic Review.*

Andrews, I., Kitagawa, T., and McCloskey, A. (2018). Inference on winners. Technical Report CWP31/18, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Arismendi Zambrano, J. and Broda, S. A. (2016). Multivariate Elliptical Truncated Moments. SSRN Scholarly Paper ID 2841401, Social Science Research Network, Rochester, NY.

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

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.

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

Bilinski, A. and Hatfield, L. A. (2018). Seeking evidence of absence: Reconsidering tests of model assumptions. *arXiv:1805.03273 [stat].*

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. (2019). Difference-in-Differences with Multiple Time Periods. SSRN Scholarly Paper ID 3148250, Social Science Research Network, Rochester, NY.

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. (2018a). Fuzzy Differences-in-Differences. *The Review of Economic Studies*, 85(2):999–1028.

de Chaisemartin, C. and D'Haultfœuille, X. (2018b). Two-way fixed effects estimators with heterogeneous treatment effects. *arXiv:1803.08807 [econ]*.

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.

Donald, S. G. and Lang, K. (2007). Inference with Difference-in-Differences and Other Panel Data. *The Review of Economics and Statistics*, 89(2):221–233.

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. (2018). Pre-event Trends in the Panel Event-study Design. Working Paper 24565, National Bureau of Economic Research.

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.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2018). *Mvtnorm: Multivariate Normal and t Distributions*.

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). Difference-in-Differences with Variation in Treatment Timing. Working Paper 25018, National Bureau of Economic Research.

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.

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.

Kivaranovic, D. and Leeb, H. (2018). Expected length of post-model-selection confidence intervals conditional on polyhedral constraints. *arXiv:1803.01665 [math, stat]*. arXiv: 1803.01665.

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.

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.

Pfanzagl, J. (1994). *Parametric Statistical Theory*. W. de Gruyter. Google-Books-ID: 1S20QgAACAAJ.

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.

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.

Tallis, G. M. (1963). Elliptical and Radial Truncation in Normal Populations. *The Annals of Mathematical Statistics*, 34(3):940–944.

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.

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

Supplement to the paper

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

For online publication

Jonathan Roth

July 30, 2019

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 provides additional results for the corrections developed in Section 5. Section C states and proves asymptotic results. Finally, Section D contains additional 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. 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. $\square$

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

$\square$

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. $\square$

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.[14]
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

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

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

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

$$D_\mu \mathbb{E}[Y \mid Y \in B, \mu] = I + \Sigma D_\mu (D_\mu \log H)'. \tag{6}$$

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,

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

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

$$\begin{aligned} D_\mu \mathbb{E}[Y \mid Y \in B, \mu] = & \frac{\int_B y\, D_\mu \phi_\Sigma(y-\mu)\, dy}{\int_B \phi_\Sigma(y-\mu)\, dy} + \\ & \left[\int_B y\, \phi_\Sigma(y-\mu)\, dy\right] \cdot D_\mu \left[\int_B \phi_\Sigma(y-\mu)\, dy\right]^{-1}. \end{aligned} \tag{7}$$

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 (7) thus becomes

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

Applying the chain-rule, the second term on the right-hand side of (7) becomes

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

Substituting the expressions in the previous two displays back into (7), we have

$$
\begin{aligned}
D_\mu \mathbb{E}\left[Y \mid Y \in B, \mu\right] &= \left(\mathbb{E}\left[YY' \mid Y \in B, \mu\right] - \mathbb{E}\left[Y \mid Y \in B, \mu\right] \mathbb{E}\left[Y \mid Y \in B, \mu\right]'\right) \Sigma^{-1} \\
&= \mathbb{V}\mathrm{ar}\left[Y \mid Y \in B, \mu\right] \Sigma^{-1},
\end{aligned} \tag{8}
$$

which establishes the first result. Additionally, combining (6) and (8), we have that

$$
\mathbb{V}\mathrm{ar}\left[Y \mid Y \in B, \mu\right] \Sigma^{-1} = I + \Sigma D_\mu (D_\mu \log H)', \tag{9}
$$

which implies that

$$
\mathbb{V}\mathrm{ar}\left[Y \mid Y \in B, \mu\right] - \Sigma = \Sigma \, D_\mu (D_\mu \log H)' \, \Sigma. \tag{10}
$$

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

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

where the inequality follows from the fact that $D_\mu (D_\mu \log H)'$ is negative semi-definite. Since $\mathbb{V}\mathrm{ar}\left[Y \mid Y \in B, \mu\right] - \Sigma$ 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 $t \in (0, 1)$,

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

By assumption, $(a+b)$ is elementwise greater than 0. It thus suffices to show that all elements of $\nabla g\,(x^t)$ 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

$$\frac{\partial g(x^t)}{\partial x_K} =$$

$$\frac{\partial}{\partial x_K}\left[\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}(y_1+\ldots+y_K)\,\phi_\Sigma(y)\,dy_1\ldots dy_K\right)\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}\,\phi_\Sigma(y)\,dy_1\ldots dy_K\right)^{-1}\right] =$$

$$\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}(y_1+\ldots+y_K)\,\phi_\Sigma(y)\,dy_1\ldots dy_K \times \int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}\,\phi_\Sigma\left(\begin{pmatrix} y_{-K}\\ x_K^t\end{pmatrix}\right)\,dy_1\ldots dy_{K-1}\right.$$

$$\left. - \int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}(y_1+\ldots+y_{K-1}+x_K^t)\,\phi_\Sigma\left(\begin{pmatrix} y_{-K}\\ x_K^t\end{pmatrix}\right)\,dy_1\ldots dy_{K-1}\;\times\;\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}\,\phi_\Sigma(y)\,dy_1\ldots dy_K\right)$$

$$\times\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}\,\phi_\Sigma(y)\,dy_1\ldots dy_K\right)^{-2} \qquad (11)$$

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 (11) that $\dfrac{\partial g(x^t)}{\partial x_K} > 0$ if and only if

$$\frac{\int_{x_1^t}^{b_1}\cdots\int_{x_k^t}^{b_K}(y_1+\ldots+y_K)\,\phi_\Sigma(y)\,dy_1\ldots dy_K}{\int_{x_1^t}^{b_1}\cdots\int_{x_k^t}^{b_K}\,\phi_\Sigma(y)\,dy_1\ldots dy_K} >$$

$$\frac{\int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}(y_1+\cdots+y_{K-1}+x_K^t)\,\phi_\Sigma\left(\begin{pmatrix} y_{-K}\\ x_K^t\end{pmatrix}\right)\,dy_1\ldots dy_{K-1}}{\int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}\,\phi_\Sigma\left(\begin{pmatrix} y_{-K}\\ x_K^t\end{pmatrix}\right)\,dy_1\ldots dy_{K-1}}$$

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

$$\mathbb{E}\left[Y_1+\ldots+Y_{K-1}\,|\,x_j^t\leq Y_j\leq b_j,\forall j\right]\geq\mathbb{E}\left[Y_1+\ldots+Y_{K-1}\,|\,x_j^t\leq Y_j\leq b_j,\ \text{for }j<K,\,Y_K=x_K^t\right]. \tag{12}$$

To see why (12) 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 $\mathbb{E}\left[\iota'Y_{-K}\,|\,x_j^t\leq Y_j\leq b_j\ \text{for }j<K,\,Y_K=y_K\right]=\iota'h(cy_K\iota)$. Hence,

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

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 $\mathbb{V}\mathrm{ar}\left[Y_{-K}\,|\,Y_{-K}\in B_{-K}, Y_K = y_K\right]$ is positive semi-definite and c_1 and c are positive by construction. Thus, for all $y_K \in [x_k^t, b_k]$,

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

By the law of iterated expectations, we have

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

as we wished to show.

$\square$

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 $\hat{\beta}_{pre}\in B_{NIS} = \{\hat{\beta}_{pre} : |\hat{\beta}_{pre,j}|/\sqrt{\Sigma_{jj}} \le c_\alpha \text{ for all } j\}$ iff $(\hat{\beta}_{pre}-\beta_{pre}) \in \tilde{B}_{NIS} = \{\beta : a_j \le \beta_j \le b_j\}$ 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,

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\hat{\beta}_{post} \,|\, \hat{\beta}_{pre} \in S\right] &= \mathbb{V}\mathrm{ar}\left[\tilde{\beta}_{post} \,|\, \hat{\beta}_{pre} \in S\right] + \mathbb{V}\mathrm{ar}\left[\Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in S\right] \\ &\quad + 2\,\mathrm{Cov}\left(\tilde{\beta}_{post}\,,\, \Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in S\right) \\ &= \mathbb{V}\mathrm{ar}\left[\tilde{\beta}_{post}\right] + \mathbb{V}\mathrm{ar}\left[\Sigma_{12}\Sigma_{22}^{-1}\hat{\beta}_{pre} \,|\, \hat{\beta}_{pre} \in S\right], \qquad (13)\end{aligned}$$

where we use the independence of $\tilde{\beta}_{post}$ and $\hat{\beta}_{pre}$ from Lemma A.1 to obtain that $\mathbb{V}\mathrm{ar}\left[\tilde{\beta}_{post} \,|\, \hat{\beta}_{pre} \in B\right] = \mathbb{V}\mathrm{ar}\left[\tilde{\beta}_{post}\right]$ and that the covariance term equals 0. Applying equation (13) for $S = B$ and for $S = \mathbb{R}^K$, and then taking the difference between the two, we have

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

which gives the desired result.

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

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

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

Proof of Lemma 5.1

Proof. Note that by construction, $\eta'\hat{\beta}$ and Z are jointly normal and uncorrelated, hence independent. Thus, without conditioning on $\hat{\beta} \in B$, we have

$$\eta'\hat{\beta} \,|\, Z = z \sim \mathcal{N}\left(\eta'\beta,\, \eta'\Sigma\eta\right).$$

Conditioning further on $\hat{\beta} \in B$ implies that $\eta'\hat{\beta} \in \Xi(z)$, but owing to the (unconditional) independence of Z and $\eta'\hat{\beta}$, provides no additional information about $\eta'\hat{\beta}$. It follows that

$$\eta'\hat{\beta} \,|\, \hat{\beta} \in B, Z = z \sim \xi \,|\, \xi \in \Xi(z),$$

for $\xi \sim \mathcal{N}(\eta'\beta, \eta'\Sigma\eta)$, and $\Xi(z) := \{x : \exists \hat{\beta} \in B \text{ s.t. } x = \eta'\hat{\beta} \text{ and } z = (I - c\eta')\hat{\beta}\}$. □

Proof of Proposition 5.1

Proof. The result follows immediately from Proposition D.2 of the supplement to Andrews and Kasy (forthcoming). □

Proof of Corollary 5.1

Proof. Median-unbiasedness of $\hat{b}_{0.5}(\eta'\hat{\beta}, Z)$ follows immediately from Proposition 5.1. To show that $\mathcal{C}_{1-\alpha}$ controls size, note that $\eta'\beta \notin \mathcal{C}_{1-\alpha}$ only if either $\hat{b}_{\alpha/2}(\eta'\hat{\beta}, Z) > \eta'\beta$ or $\hat{b}_{1-\alpha/2}(\eta'\hat{\beta}, Z) < \eta'\beta$. However, Proposition 5.1 implies that each of these events occurs with probability bounded above by $\alpha/2$, and thus $\eta'\beta \notin \mathcal{C}_{1-\alpha}$ with probability bounded above by α. □

B Additional Results for Corrected Event-Studies

Applying the results in Section 5 to produce corrected event-studies in practice requires computation of the set $\Xi(z)$. In this appendix, I derive the form of $\Xi(z)$ for polyhedral and quadratic pre-tests, which respectively cover the cases of pre-tests based on individual and joint significance. I then derive the form of $\Xi(z)$ for a variety of model selection criteria.

B.1 Calculating $\Xi(z)$ for polyhedral pre-tests

We first consider the case where $B = \{\beta \,|\, A\beta \leq b\}$. Note that the test that no pre-period coefficient is significant can be written in this form, $B_{NIS} = \{\beta \,|\, A^{NIS}\beta \leq b^{NIS}\}$, for $A^{NIS} = \begin{pmatrix} I_{K\times K} & 0_{1\times M} \\ -I_{K\times K} & 0_{1\times M} \end{pmatrix}$ and $b^{NIS} = \begin{pmatrix} c_\alpha \times \sqrt{diag(\Sigma)} \\ c_\alpha \times \sqrt{diag(\Sigma)} \end{pmatrix}$. For the polyhedral case, the form of $\Xi(z)$ follows immediately from the results of Lee et al. (2016).

Lemma B.1 (Application to polyhedral conditioning sets)**.** *Suppose that the conditioning set $B = \{\beta \,|\, A\beta \leq b\}$ for A an $R \times K + M$ matrix and b an $R \times 1$ vector. Then $\Xi(z)$, as defined in Lemma 5.1, is an interval in $\mathbb{R}$, with endpoints $V^-(z)$ and $V^+(z)$ given by:*

$$V^-(z) = \underset{\{j:(Ac)_j<0\}}{max} \frac{b_j - (Az)_j}{(Ac)_j} \tag{14}$$

$$V^+(z) = \underset{\{j:(Ac)_j>0\}}{min} \frac{b_j - (Az)_j}{(Ac)_j}. \tag{15}$$

Additionally, if $\mathbb{P}\left(\hat{\beta} \in B\right) > 0$*, then the conditions for the optimality of the* α*-quantile-unbiased estimator in Proposition 5.1 are met.*

B.2 Calculating $\Xi(z)$ for quadratic pre-tests

I next derive the form of $\Xi(z)$ for tests based on a quadratic form of the parameters, such as tests based on the joint significance or the euclidean norm of the pre-period coefficients.

Lemma B.2. *Let* $B = \{\beta \,|\, \beta' A \beta \leq b\}$ *for* A *an* $(K+M) \times (K+M)$ *matrix and* b *a scalar. Let* $\mathcal{A} = c'Ac$, $\mathcal{B} = 2c'Az$, $\mathcal{C} = z'Az - b$, *and* $\mathcal{D} = \mathcal{B}^2 - 4\mathcal{A} \cdot \mathcal{C}$, *for* c *and* z *as defined in Lemma 5.1. Then:*

1. *If* $\mathcal{A} > 0, \mathcal{D} \geq 0$, $\Xi(z) = \left[\frac{-\mathcal{B}-\sqrt{\mathcal{D}}}{2\mathcal{A}}, \frac{-\mathcal{B}+\sqrt{\mathcal{D}}}{2\mathcal{A}}\right]$.
2. *If* $\mathcal{A} < 0, \mathcal{D} \geq 0$, $\Xi(z) = \left(-\infty, \frac{-\mathcal{B}+\sqrt{\mathcal{D}}}{2\mathcal{A}}\right] \cup \left[\frac{-\mathcal{B}-\sqrt{\mathcal{D}}}{2\mathcal{A}}, \infty\right)$.
3. *If* $\mathcal{A} < 0, \mathcal{D} < 0$, $\Xi(z) = \mathbb{R}$.
4. *If* $\mathcal{A} > 0, \mathcal{D} < 0$, *then* $\Xi(z) = \emptyset$.
5. *If* $\mathcal{A} = 0$, $\mathcal{B} > 0$ *then* $\Xi(z) = (-\infty, -\frac{\mathcal{C}}{\mathcal{B}}]$.
6. *If* $\mathcal{A} = 0$, $\mathcal{B} < 0$, $\Xi(z) = [-\frac{\mathcal{C}}{\mathcal{B}}, \infty)$.
7. *If* $\mathcal{A} = 0$, $\mathcal{B} = 0$, *then* $\Xi(z) = \mathbb{R}$ *if* $\mathcal{C} \leq 0$ *and* $\Xi(z) = \emptyset$ *if* $\mathcal{C} > 0$.

   Additionally, if $\mathbb{P}\left(\hat{\beta} \in B\right) > 0$*, then the conditions for the optimality of the* α*-quantile-unbiased estimator in Proposition 5.1 are met.*

B.3 Calculating $\Xi(z)$ after model selection

I next discuss the computation of $\Xi(z)$ after selection among a finite number of models, as discussed in Section 5.1.2.

The form of Ξ will of course depend on the criteria for the specification search, but I note that a wide variety of specification searches will generate a Ξ that is the union of intervals in $\mathbb{R}$. To see why this is the case, note first that from the definition of $\Xi(z)$, it follows easily that if $B = B_1 \cup B_2$, then $\Xi_B(z) = \Xi_{B_1}(z) \cup \Xi_{B_2}(z)$, and likewise, if $B = B_1 \cap B_2$, then $\Xi_B(z) = \Xi_{B_1}(z) \cap \Xi_{B_2}(z)$. $\Xi_B(z)$ will therefore take the form of a union of intervals if the conditioning set B can be written as the union and intersection of a sequence of conditioning sets that themselves produce unions of intervals for $\Xi(z)$.[15] Further, Propositions B.1 and B.2 show

[15] Note that the complement of a collection of intervals is also a collection of intervals, and the intersection of collections of intervals can therefore be re-cast as a union of intervals using DeMorgan's laws.

that when conditioning on linear or quadratic restrictions on $\hat{\beta}$, $\Xi(z)$ is the union of intervals. Note also that the norm of $\hat{\beta}^m_{pre}$ is less than that of $\hat{\beta}^{m'}_{pre}$ if and only if $\hat{\beta}'(A_m - A_{m'})\hat{\beta} \leq 0$ for A_m the matrix with 1s on the diagonal in the positions corresponding with the elements of $\hat{\beta}^m_{pre}$ and zero otherwise. Thus, any selection rule that depends on logical combinations of the (individual or joint) significance of the pre-trends coefficients from each model and/or the relative magnitudes of the models will generate a $\Xi(z)$ that is the union of intervals.

A few examples are of note. First, suppose the researcher considers models sequentially and stops at the first model that has an insignificant pre-trend (either jointly, or based on the significance of each of the individual coefficients). Then if the mth model is chosen, $\Xi(z)$ is the intersection of the sets on which models $1, ..., m-1$ have a significant pre-trend, intersected with the set on which model m does not have a significant pre-trend. Second, suppose the researcher chooses the model that minimizes the norm of the pre-period coefficients. Then $\Xi(z)$ is the intersection of the sets on which the chosen model m^* has a lower norm than model m' for each candidate m'. Finally, suppose that the researcher first tests model 1 on the full population, and then if it has a significant pre-trend, chooses whichever has the smaller pre-trend among models 2 and 3, which each restrict to different subsets of the population. Then $\Xi(z)$ for the event model 2 is selected will correspond with the union of intervals on which model 1 is significant intersected with the interval(s) corresponding with the event that the norm of model 2 is less than that of model 3.

B.4 Proofs for the results on $\Xi(z)$

Proof of Lemma B.1

Proof. The form for $\Xi(z)$ follows immediately from Lemma 5.1 in Lee et al. (2016).

We now verify that the distribution of $\eta'\hat{\beta} \mid Z, A\hat{\beta} \leq b$ is continuous for almost every Z. Note that by Lemma 5.1, $\eta'\hat{\beta} \mid Z = z, A\hat{\beta} \leq b$ is truncated normal with truncation points $V^-(z)$ and $V^+(z)$ and untruncated variance $\eta'\Sigma\eta$. The untruncated variance is strictly positive since Σ is positive definite and $\eta \neq 0$, and so the conditional distribution of $\eta'\hat{\beta}$ is continuous if $V^-(z) < V^+(z)$. Since conditional on $A\hat{\beta} \leq b$ and $Z = z$, $V^-(z) \leq \eta'\hat{\beta} \leq V^+(z)$, we have $V^-(z) = V^+(z)$ only if $V^-(z) = \eta'\hat{\beta}$.

It thus suffices to show that $\mathbb{P}\left(\eta'\hat{\beta} = V^-(Z) \mid A\hat{\beta} \leq b\right) = 0$. Note though that

$$\mathbb{P}\left(\eta'\hat{\beta} = V^-(Z)\right) = \mathbb{E}\left[\mathbb{P}\left(\eta'\hat{\beta} = V^-(z) \mid Z = z\right)\right].$$

Next, observe that for any fixed value z, $\mathbb{P}\left(\eta'\hat{\beta} = V^-(z) \mid Z = z\right) = 0$ since $\eta'\hat{\beta}$ and Z are independent by construction and the distribution of $\eta'\hat{\beta}$ is continuous since $\hat{\beta}$ is normally distributed, Σ is full rank, and $\eta \neq 0$. It follows that

$$\begin{aligned}
0 &= \mathbb{P}\left(\eta'\hat{\beta} = V^{-}(Z)\right) \\
&= \mathbb{P}\left(\eta'\hat{\beta} = V^{-}(Z) \,|\, A\hat{\beta} \leq b\right) \mathbb{P}\left(A\hat{\beta} \leq b\right) + \mathbb{P}\left(\eta'\hat{\beta} = V^{-}(Z) \,|\, A\hat{\beta} \not\leq b\right) \mathbb{P}\left(A\hat{\beta} \not\leq b\right) \\
&\geq \mathbb{P}\left(\eta'\hat{\beta} = V^{-}(Z) \,|\, A\hat{\beta} \leq b\right) \mathbb{P}\left(A\hat{\beta} \leq b\right).
\end{aligned}$$

Since $\mathbb{P}\left(A\hat{\beta} \leq b\right) > 0$ by assumption, it follows that $\mathbb{P}\left(\eta'\hat{\beta} = V^{-}(Z) \,|\, A\hat{\beta} \leq b\right) = 0$, as needed. □

Proof of Lemma B.2 Note that by $\hat{\beta} \in B$ iff $\hat{\beta}'A\hat{\beta} - b \leq 0$. Further, by construction $\hat{\beta} = z + c\eta'\hat{\beta}$, so

$$\begin{aligned}
\hat{\beta}'A\hat{\beta} - b &= \left(z + c\eta'\hat{\beta}\right)' A \left(z + c\eta'\hat{\beta}\right) - b \\
&= \underbrace{(c'Ac)}_{:=\mathcal{A}}(\eta'\hat{\beta})^2 + \underbrace{2c'Az}_{:=\mathcal{B}}(\eta'\hat{\beta}) + \underbrace{(z'Az - b)}_{:=\mathcal{C}},
\end{aligned}$$

which is a quadratic in $(\eta'\hat{\beta})$. The first part of the result then follows by solving for the region where the parabola $\mathcal{A}x^2 + \mathcal{B}x + \mathcal{C} \leq 0$ using the quadratic formula.

To verify the conditions for optimality, note that the first part of the result implies that $\Xi(Z)$ is the finite union of intervals on the real line. (We can safely ignore the situations in which $\Xi(Z) = \emptyset$, since conditional on $\hat{\beta} \in B$, $\Xi(Z)$ is non-empty with probability 1). Since $\eta'\hat{\beta} \,|\, \hat{\beta} \in B, Z = z$ is truncated normal, it will be continuous unless $\Xi(z)$ collapses to a set of measure 0. However, examining the possible cases, we see that this could only occur if $\mathcal{A} > 0, \mathcal{D} \geq 0$ and the interval collapses to a point. By the same argument as in the proof to Lemma B.1, this occurs with probability zero, which completes the proof. □

C Uniform Asymptotic Results

In the main text of the paper, I consider a finite sample normal model for the event-study coefficients. I evaluate the distribution of the event-study estimates conditional on passing a pre-test for the pre-period coefficients, and derive corrected event-study plot estimates and CIs in the context of this model. 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 $O(n^{-\frac{1}{2}})$. I focus here on results for polyhedral pre-tests, which include the common pre-test that no pre-period coefficient be individually statistically significant.

C.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. Recall 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 $B(\Sigma) = \{(\beta_{post}, \beta_{pre}) \,|\, A_{pre}(\Sigma)\beta_{pre} \leq b(\Sigma)\}$ 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 $A(\Sigma)\beta \leq b(\Sigma)$.

C.2 Main uniformity results

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

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

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

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

Finally, our last main result concerns the asymptotic validity of the corrected event-studies.

Proposition C.3 (Uniform asymptotic α-quantile unbiasedness)**.** *Let $\eta\neq 0$, and consider $\hat{b}_\alpha(\hat{\beta}_n,\hat{\Sigma}_n)$ the α-quantile-unbiased estimator of $\eta'\beta$ conditional on $\hat{\beta}_n\in B(\hat{\Sigma}_n)$. Define $\beta_{P,n}=\tau_{P,n}+\delta_{P,n}$. Then under Assumptions 2-5,*

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}}\left|\left|\mathbb{P}_P\left(\hat{b}_\alpha(\hat{\beta}_n,\hat{\Sigma}_n)\leq\eta'\beta_{P,n}\,|\,\hat{\beta}_n\in B(\hat{\Sigma}_n)\right)-(1-\alpha)\right|\right|\,\mathbb{P}_P\left(\hat{\beta}_n\in B(\hat{\Sigma}_n)\right)=0.$$

Proposition C.3 states that the corrected α-quantile-unbiased estimator $\hat{b}_\alpha$ is uniformly α-quantile unbiased along any sequence of distributions such that the limiting probability of passing the pre-test is not going to zero. It follows immediately that under any such sequence $\hat{b}_{0.5}$ is asymptotically median-unbiased and the interval $[\hat{b}_{\alpha/2},\hat{b}_{1-\alpha/2}]$ is a valid $1-\alpha$ level confidence interval.

Corollary C.1 (Median unbiasedness and coverage of equal-tailed CIs)**.** *Suppose the conditions of Proposition C.3 hold. Then*

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}}\left|\left|\mathbb{P}_P\left(\hat{b}_{0.5}(\hat{\beta}_n,\hat{\Sigma}_n)\leq\eta'\beta_{P,n}\,|\,\hat{\beta}_n\in B(\hat{\Sigma}_n)\right)-0.5\right|\right|\,\mathbb{P}_P\left(\hat{\beta}_n\in B(\hat{\Sigma}_n)\right)=0$$

and, for $\mathcal{C}_{1-\alpha}(\hat{\beta}_n,\hat{\Sigma}_n)=[\hat{b}_{\alpha/2}(\hat{\beta}_n,\hat{\Sigma}_n),\hat{b}_{1-\alpha/2}(\hat{\beta}_n,\hat{\Sigma}_n)]$,

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}}\left|\left|\mathbb{P}_P\left(\eta'\beta_{P,n}\in\mathcal{C}_{1-\alpha}(\hat{\beta}_n,\hat{\Sigma}_n)\,|\,\hat{\beta}_n\in B(\hat{\Sigma}_n)\right)-(1-\alpha)\right|\right|\,\mathbb{P}_P\left(\hat{\beta}_n\in B(\hat{\Sigma}_n)\right)=0.$$

C.3 Proofs for main uniformity results

Proof of Proposition C.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

$$\liminf_{m\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}_{P_n}\left[f(\hat{\beta}_n - \tau_{P_{n_m},n_m}) \,|\, \hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m})\right] - \mathbb{E}\left[f(\xi_{P_{n_m},n_m})|\xi \in B(\Sigma_{P_{n_m}})\right]\right|\right| \times$$
$$\mathbb{P}_{P_{n_m}}\left(\hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m})\right) > 0. \tag{16}$$

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 $f \in BL_1$,

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

for all q. But this implies that

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

which contradicts (16).

Now, suppose $p^* > 0$. Note that by Assumption 4, Σ_P falls in the set $\mathcal{S} = \{\Sigma | 1/\bar{\lambda} \leq \lambda_{min}(\Sigma) \leq \lambda_{max}(\Sigma) \leq \bar{\lambda}\}$, 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 C.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 C.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 $\xi^{*} \sim \mathcal{N}(\delta^{*}, \Sigma^{*})$, 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

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^{+}_{n_s}) \,|\, \hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right] - \mathbb{E}\left[f(\xi^{*})|\xi^{*} \in B(\Sigma^{*})\right]\right|\right| = 0 \quad (17)$$

and

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^{+}_{n_s})|\xi_{P_{n_s}} \in B(\Sigma_{P_{n_s}})\right] - \mathbb{E}\left[f(\xi^{*})|\xi^{*} \in B(\Sigma^{*})\right]\right|\right| = 0. \quad (18)$$

Equations (17) and (18) 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 (16). $\square$

Proof of Proposition C.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 $\epsilon > 0$,

$$\liminf_{m\to\infty} 1\left[\left|\left|\mathbb{E}\left[\hat{\beta}_{n_m} - \tau_{P_{n_m},n_m} \,|\, \hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m})\right] - \mathbb{E}\left[\xi_{P_{n_m}} \,|\, \xi_{P_{n_m}} \in B(\Sigma_{P_{n_m}})\right]\right|\right| > \epsilon\right] \times \mathbb{P}_{P_{n_m}}\left(\hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m})\right) > 0. \tag{19}$$

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 (19).

Now, suppose $p^* > 0$. As argued in the proof to Proposition C.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 C.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 C.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 C.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 (19). $\square$

Proof of Proposition C.3 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

$$\liminf_{m\to\infty}\left|\left|\mathbb{P}_{P_{n_m}}\left(\hat{b}_\alpha(\hat{\beta}_{n_m},\hat{\Sigma}_{n_m})\leq\eta'\beta_{P_{n_m},n_m}\,|\,\hat{\beta}_{n_m}\in B(\hat{\Sigma}_{n_m})\right)-(1-\alpha)\right|\right|\,\mathbb{P}_{P_{n_m}}\left(\hat{\beta}_{n_m}\in B(\hat{\Sigma}_{n_m})\right)>0. \tag{20}$$

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

First, suppose $p^*=0$. Note that

$$\left|\left|\mathbb{P}_{P_{n_q}}\left(\hat{b}_\alpha(\hat{\beta}_{n_q},\hat{\Sigma}_{n_q})\leq\eta'\beta_{n_q,P_{n_q}}\,|\,\hat{\beta}_{n_q}\in B(\hat{\Sigma}_{n_q})\right)-(1-\alpha)\right|\right|\leq 1,$$

and hence

$$\liminf_{q\to\infty}\left|\left|\mathbb{P}_{P_{n_q}}\left(\hat{b}_\alpha(\hat{\beta}_{n_q},\hat{\Sigma}_{n_q})\leq\eta'\beta_{P_{n_q},n_q}\,|\,\hat{\beta}_{n_q}\in B(\hat{\Sigma}_{n_q})\right)-(1-\alpha)\right|\right|\,\mathbb{P}_{P_{n_q}}\left(\hat{\beta}_{n_q}\in B(\hat{\Sigma}_{n_q})\right)\leq p^*=0,$$

which contradicts (20).

Next, suppose $p^*>0$. As argued in the proof to Proposition C.1, we can extract a subsequence of increasing sample sizes, n_s, such that

$$\lim_{s\to\infty} \Sigma_{P_{n_s}} = \Sigma^*, \tag{21}$$

$$\lim_{s\to\infty} \delta^{pre}_{P_{n_s},n_s} = \delta^{pre,*}, \tag{22}$$

$$\lim_{s\to\infty} \mathbb{P}_{P_{n_s}} \left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) = p^*, \tag{23}$$

where $\Sigma^* \in \mathcal{S}$ and $p^* \in [0,1]$.

We wish to obtain a contradiction of (20) by showing that

$$\lim_{s\to\infty} \left|\left|\mathbb{P}_{P_{n_s}} \left(\hat{b}_\alpha(\hat{\beta}_{n_s}, \hat{\Sigma}_{n_s}) \le \eta' \beta_{P_{n_s},n_s} \,|\, \hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) - (1-\alpha)\right|\right| = 0.$$

Let $\delta^+_{n_s} = \begin{pmatrix} \delta^{post}_{P_{n_s},n_s} \\ 0 \end{pmatrix}$, $\delta^-_{n_s} = \begin{pmatrix} 0 \\ \delta^{pre}_{P_{n_s},n_s} \end{pmatrix}$, and $\delta^* = \begin{pmatrix} 0 \\ \delta^{pre,*} \end{pmatrix}$. From Lemma C.7, it suffices to show that, for $\hat{\beta}^*_{n_s} = \hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s}$,

$$\lim_{s\to\infty} \left|\left|\mathbb{P}_{P_{n_s}} \left(\hat{b}_\alpha(\hat{\beta}^*_{n_s}, \hat{\Sigma}_{n_s}) \le \eta' \delta^-_{n_s} \,|\, \hat{\beta}^*_{n_s} \in B(\hat{\Sigma}_{n_s})\right) - (1-\alpha)\right|\right| = 0.$$

Further, Lemma C.8 implies that this is equivalent to:

$$\lim_{s\to\infty} \left|\left|\mathbb{P}_{P_{n_s}} \left(g(\hat{\beta}^*_{n_s}, \hat{\Sigma}_{n_s}, \delta^-_{n_s}) \le 1-\alpha \,|\, \hat{\beta}^*_{n_s} \in B(\hat{\Sigma}_{n_s})\right) - (1-\alpha)\right|\right| = 0,$$

for g as defined in Lemma C.8.

Note that by construction, for $\xi^* \sim \mathcal{N}(\delta^*, \Sigma^*)$,

$$g(\xi^*, \Sigma^*, \delta^*) \,|\, \xi^* \in B(\Sigma^*) \sim U[0,1].$$

Thus,

$$\mathbb{P}\left(g(\xi^*, \Sigma^*, \delta^*) \le 1-\alpha \,|\, \xi^* \in B(\Sigma^*)\right) = 1-\alpha,$$

and the distribution of $g(\xi^*, \Sigma^*, \delta^*) \,|\, \xi^* \in B(\Sigma^*)$ is continuous at $1-\alpha$. Additionally, Lemma C.3, along with (21) and (22), imply that

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

Since Lemma C.12 gives that the function g is continuous for almost every $(\xi^*, \Sigma^*, \delta^*)$, conditional on $\xi^* \in B(\Sigma^*)$, the result then follows from the Continuous Mapping Theorem. $\square$

Proof of Corollary C.1

Proof. The result for $\hat{b}_{0.5}$ is immediate from Proposition C.3. To show the second result, note that

$$\mathbb{P}_P\left(\eta'\beta_{P,n} \notin \mathcal{C}_{1-\alpha}(\hat{\beta}_n, \hat{\Sigma}_n) \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n)\right) =$$
$$\mathbb{P}_P\left(\hat{b}_{\alpha/2}(\hat{\beta}_n, \hat{\Sigma}_n) > \eta'\beta_{P,n} \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n)\right) + \mathbb{P}_P\left(\hat{b}_{1-\alpha/2}(\hat{\beta}_n, \hat{\Sigma}_n) < \eta'\beta_{P,n} \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n)\right),$$

since $\eta'\beta_{P,n}$ falls outside of $\mathcal{C}_{1-\alpha}$ only if it is greater than the upper bound or less than the lower bound, and both of these events cannot occur simultaneously. Applying the result in the previous display along with the triangle inequality and the fact that for any event E, $\mathbb{P}(E) = 1 - \mathbb{P}(E^c)$, we obtain that

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \left|\left|\mathbb{P}_P\left(\eta'\beta_{P,n} \in \mathcal{C}_{1-\alpha}(\hat{\beta}_n, \hat{\Sigma}_n) \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n)\right) - (1-\alpha)\right|\right| \mathbb{P}_P\left(\hat{\beta}_n \in B(\hat{\Sigma}_n)\right) \leq$$
$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \left|\left|\mathbb{P}_P\left(\hat{b}_{1-\alpha/2}(\hat{\beta}_n, \hat{\Sigma}_n) \leq \eta'\beta_{P,n} \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n)\right) - \alpha/2\right|\right| \mathbb{P}_P\left(\hat{\beta}_n \in B(\hat{\Sigma}_n)\right) +$$
$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \left|\left|\mathbb{P}_P\left(\hat{b}_{\alpha/2}(\hat{\beta}_n, \hat{\Sigma}_n) \leq \eta'\beta_{P,n} \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n)\right) - (1-\alpha/2)\right|\right| \mathbb{P}_P\left(\hat{\beta}_n \in B(\hat{\Sigma}_n)\right) +$$
$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \left|\left|\mathbb{P}_P\left(\hat{b}_{1-\alpha/2}(\hat{\beta}_n, \hat{\Sigma}_n) = \eta'\beta_{P,n} \,|\, \hat{\beta}_n \in B(\hat{\Sigma}_n)\right)\right|\right| \mathbb{P}_P\left(\hat{\beta}_n \in B(\hat{\Sigma}_n)\right)$$

The first two terms on the right-hand side of the previous display converge to 0 by Proposition C.3. That the final term is 0 can be shown using an argument analogous to that in the proof of Proposition C.3. Specifically, using the notation from the proof of Proposition C.3, note that $\hat{b}_\alpha(\hat{\beta}_{n_s}, \hat{\Sigma}_{n_s}) = \eta'\beta_{P_{n_s},n_s}$ iff $g(\hat{\beta}^*_{n_s}, \hat{\Sigma}_{n_s}, \delta^-_{n_s}) = 1-\alpha$. However, we show in the proof to Proposition C.3 that for $\xi^* \sim \mathcal{N}(\delta^*, \Sigma^*)$, $g(\xi^*, \Sigma^*\delta^*)|\xi^* \in B(\Sigma^*)$ is uniformly distributed, and thus equal to $1-\alpha$ with probability 0. The desired result then follows from an application of the continuous mapping theorem as in the proof to Proposition C.3.

□

C.4 Auxiliary lemmas and proofs

Lemma C.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,*

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

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 C.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 $\mathbb{P}(\xi^* \in B(\Sigma^*)) = p^* > 0$ 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 C.1, and that the product of bounded and continuous functions is bounded and continuous. Additionally, by Lemma C.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 C.3. *Suppose Assumptions 2-5 hold, and n_s is an increasing sequence of sample sizes such that*

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

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 $\xi^* \sim \mathcal{N}(\delta^*, \Sigma^*)$. Convergence in distribution is equivalent to convergence in bounded Lipschitz metric, so

$$\lim_{s\to\infty} \sup_{f \in BL_1} \left|\left|\mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^+_{n_s})\right] - \mathbb{E}\left[f(\xi^*)\right]\right|\right| = 0. \tag{24}$$

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

$$\lim_{s\to\infty} \sup_{f \in BL_1} \left|\left|\mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s})\right] - \mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^+_{n_s})\right]\right|\right| = 0. \tag{25}$$

Equations (24) and (25), together with the triangle inequality, imply that

$$\lim_{s\to\infty} \sup_{f \in BL_1} \left|\left|\mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s})\right] - \mathbb{E}\left[f(\xi^*)\right]\right|\right| = 0, \tag{26}$$

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 C.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 C.2.

□

Lemma C.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

$$\liminf_{q\to\infty} \mathbb{P}_{P_{n_q}} \left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q}) \right) > 0. \tag{27}$$

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 (27).

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,

$$\limsup_{r\to\infty} \mathbb{P}_{P_{n_r}} \left(\hat{\beta}_{n_r} \in B^c_{\bar{M}} \right) \geq \limsup_{r\to\infty} \mathbb{E}_{P_{n_r}} \left[h(\hat{\beta}_{n_r}) \right]. \tag{28}$$

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

$$\limsup_{r\to\infty} \mathbb{E}_{P_{n_r}}\left[h(\hat{\beta}_{n_r})\right] \geq \limsup_{r\to\infty} \mathbb{E}\left[h(\xi_{P_{n_r},n_r})\right]. \tag{29}$$

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,

$$\limsup_{r\to\infty} \mathbb{E}\left[h(\xi_{P_{n_r},n_r})\right] \geq \limsup_{r\to\infty} \mathbb{E}\left[1[\xi^{pre}_{j,r} > \bar{M} + 1]\right]. \tag{30}$$

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 (28), (29), (30), gives the desired result.

□

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

$$\lim_{s\to\infty} \Sigma_{P_{n_s}} = \Sigma^*, \tag{31}$$

$$\lim_{s\to\infty} \delta^{pre}_{P_{n_s},n_s} = \delta^{pre,*}, \tag{32}$$

$$\lim_{s\to\infty} \mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) = p^* > 0 \tag{33}$$

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

$$\begin{aligned}&\lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[||Y_n-c_n||\cdot 1[||Y_n-c_n||>M]\right]\leq\\&\lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[||Y_n||\cdot 1[||Y_n-c_n||>M]\right]+\lim_{M\to\infty}\limsup_{n\to\infty}\mathbb{E}\left[||c_n||\cdot 1[||Y_n-c_n||>M]\right].\end{aligned}\tag{34}$$

We now show that each of the two terms on the right hand side of (34) 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 (34) 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

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

$\square$

Lemma C.7. *Suppose Assumption 5 holds. Suppose that* $\tau = \begin{pmatrix} \tau_{post} \\ 0 \end{pmatrix}$ *has zeros in the positions corresponding with* $\hat{\beta}_{pre}$*. Then for any* $\hat{\beta}$ *and* $\hat{\Sigma}$*,*

$$\hat{b}_\alpha(\hat{\beta} - \tau, \hat{\Sigma}) = \hat{b}_\alpha(\hat{\beta}, \hat{\Sigma}) - \eta'\tau.$$

Proof. Recall that for any values $(\tilde{\beta}, \tilde{\Sigma})$, $\hat{b}_\alpha(\tilde{\beta}, \tilde{\Sigma})$ solves

$$\frac{\Phi\left(\frac{\eta'\tilde{\beta} - \hat{b}_\alpha}{\tilde{\sigma}}\right) - \Phi\left(\frac{V^-(\tilde{Z}, \tilde{\Sigma}) - \hat{b}_\alpha}{\tilde{\sigma}}\right)}{\Phi\left(\frac{V^+(\tilde{Z}, \tilde{\Sigma}) - \hat{b}_\alpha}{\tilde{\sigma}}\right) - \Phi\left(\frac{V^-(\tilde{Z}, \tilde{\Sigma}) - \hat{b}_\alpha}{\tilde{\sigma}}\right)} = 1 - \alpha, \tag{35}$$

where $\tilde{Z}$ is shorthand for $Z(\tilde{\beta}, \tilde{\Sigma}) = (I - \tilde{c}\eta')\tilde{\beta}$, for $\tilde{c} = \tilde{\Sigma}\eta'/(\eta'\tilde{\Sigma}\eta)$, $\tilde{\sigma} = \sqrt{\eta'\tilde{\Sigma}\eta}$, and the functions V^+ and V^- are as defined in Lemma B.1 (replacing Σ with $\tilde{\Sigma}$). Let $\hat{\beta}^* = \hat{\beta} - \tau$, $\hat{Z} = Z(\hat{\beta}, \hat{\Sigma})$, and $\hat{Z}^* = Z(\hat{\beta}^*, \hat{\Sigma})$. We now show that

(i) $V^-(\hat{Z}^*, \hat{\Sigma}) = V^-(\hat{Z}, \hat{\Sigma}) - \eta'\tau$

(ii) $V^+(\hat{Z}^*, \hat{\Sigma}) = V^-(\hat{Z}, \hat{\Sigma}) - \eta'\tau$

(iii) $\eta'\hat{\beta}^* = \eta'\hat{\beta} - \eta'\tau$,

which together imply that $\hat{b}_\alpha$ solves (35) for $(\hat{\beta}, \hat{Z})$ iff $\hat{b}^*_\alpha = \hat{b}_\alpha - \eta'\tau$ solves (35) for $(\hat{\beta}^*, \hat{Z}^*)$, from which the claim follows.

To establish (i), note that $\hat{Z}^* - \hat{Z} = -(I - \hat{c}\eta')\begin{pmatrix} \tau_{post} \\ 0 \end{pmatrix}$, where $\hat{c}$ depends only on $\hat{\Sigma}$ and η, and not on $\hat{\beta}$. From this we see that

$$\begin{aligned}
A\hat{Z}^* &= A\hat{Z} + A(\hat{Z}^* - \hat{Z}) \\
&= A\hat{Z} - A(I - \hat{c}\eta')\begin{pmatrix} \tau_{post} \\ 0 \end{pmatrix} \\
&= A\hat{Z} + (A\hat{c})\eta'\begin{pmatrix} \tau_{post} \\ 0 \end{pmatrix}
\end{aligned}$$

where $A\begin{pmatrix}\tau_{post}\\ 0\end{pmatrix}=0$ since $A=[0,\, A_{post}]$. Additionally, from the definition of V^-,

$$V^-(\hat{Z},\hat{\Sigma}) = \underset{\{j:(A\hat{c})_j<0\}}{max}\frac{b_j-(A\hat{Z})_j}{(A\hat{c})_j}.$$

It follows from the previous two displays that $V^-(\hat{Z}^*)=V^-(\hat{Z})-\eta'\begin{pmatrix}\tau_{post}\\ 0\end{pmatrix}$. An analogous argument establishes (ii), and (iii) follows immediately from the definition of $\hat{\beta}^*$. $\square$

Lemma C.8. *Fix $\eta\neq 0$. For any $(\hat{\beta},\hat{\Sigma})$ and $x\in\mathbb{R}$, let $F^{\Xi(\hat{\beta},\hat{\Sigma})}_{x,\eta'\hat{\Sigma}\eta}(\cdot)$ denote the CDF of a $\mathcal{N}\left(x,\, \eta'\hat{\Sigma}\eta\right)$ variable truncated to the set $\Xi(\hat{\beta},\hat{\Sigma})=[V^-(Z(\hat{\beta},\hat{\Sigma}),\hat{\Sigma}),V^+(Z(\hat{\beta},\hat{\Sigma}),\hat{\Sigma})]$, where the functions V^-, V^+, and Z are as defined in Lemma C.10 below. Define $g(\hat{\beta},\hat{\Sigma},\delta)=F^{\Xi(\hat{\beta},\hat{\Sigma})}_{\eta'\delta,\eta'\hat{\Sigma}\eta}(\eta'\hat{\beta})$. Then $\hat{b}_\alpha(\hat{\beta},\hat{\Sigma})\leq\eta'\delta$ iff $g(\hat{\beta},\hat{\Sigma},\delta)\leq 1-\alpha$.*

Proof. By definition, $\hat{b}_\alpha$ solves:

$$F^{\Xi(\hat{\beta},\hat{\Sigma})}_{\hat{b}_\alpha,\eta'\hat{\Sigma}\eta}(\eta'\hat{\beta})=1-\alpha.$$

However, $F^{\Xi(\hat{\beta},\hat{\Sigma})}_{x,\eta'\hat{\Sigma}\eta}(\eta'\hat{\beta})$ is weakly decreasing in x (see, e.g. Lemma A.1 in Lee et al. (2016)), from which the result follows immediately.

$\square$

Lemma C.9. *Suppose Σ is a positive definite matrix such that for some j, $(Ac)_j=0$ for $c=\Sigma\eta/(\eta'\Sigma\eta)$. Let $\xi\sim\mathcal{N}(\delta,\,\Sigma)$ and $Z=(I-c\eta')\xi$. Let $B=\{\beta\,|\,A\beta\leq b\}$ such that $\mathbb{P}(\xi\in B)>0$. Assume further that none of the rows of A are zero. Then $\mathbb{P}(b_j-(AZ)_j>0\,|\,\xi\in B)=1$.*

Proof. By Lemma 5.1 in Lee et al. (2016), $\xi\in B$ only if $b_j-(AZ)_j\geq 0$. It thus suffices to show that $\mathbb{P}((Az)_j=b_j\,|\,\xi\in B)=0$. Note that

$$\begin{aligned}(AZ)_j &= (A_{(j,\cdot)}-(Ac)_j\eta')\xi\\ &= A_{(j,\cdot)}\xi\end{aligned}$$

where $A_{(j,\cdot)}$ denotes the jth row of A, and we use the fact that $(Ac)_j=0$. Since by assumption Σ is positive definite and $A_{(j,\cdot)}\neq 0$, it follows that $(Az)_j=A_{(j,\cdot)}\xi$ is normal with variance $A_{(j,\cdot)}\Sigma A'_{(j,\cdot)}>0$. Hence, $\mathbb{P}((Az)_j=b_j)=0$. Since $\mathbb{P}(\xi\in B)>0$, it follows that $\mathbb{P}((Az)_j=b_j\,|\,\xi\in B)=0$, as needed. $\square$

Lemma C.10. *Fix $\eta\neq 0$. Let $\xi^*\sim\mathcal{N}(\delta^*,\,\Sigma^*)$ for $\Sigma^*\in\mathcal{S}$ such that $\mathbb{P}(\xi^*\in B(\Sigma^*))=p^*>0$. Let $Z(\xi,\Sigma)=(I-c(\Sigma)\eta')\xi$ for $c(\Sigma)=\Sigma\eta/(\eta'\Sigma\eta)$, and let $V^-(Z,\Sigma)$ and $V^+(Z,\Sigma)$*

be as defined in Lemma B.1. Suppose $B(\Sigma)$ satisfies Assumption 5. Then for almost every $\xi^* \,|\, \xi^* \in B(\Sigma^*)$,

(i) $V^-(Z(\xi,\Sigma),\Sigma)$ is continuous at (ξ^*,Σ^*) as a function into $\mathbb{R}\cup\{-\infty\}$, where we define the max over the empty set to be $-\infty$.

(ii) $V^+(Z(\xi,\Sigma),\Sigma)$ is continuous at (ξ^*,Σ^*) as a function into $\mathbb{R}\cup\{\infty\}$ for almost every $\xi \,|\, \xi \in B(\Sigma)$, where we define the min over the empty set to be ∞.

(iii) $V^-(Z(\xi^*,\Sigma^*),\Sigma^*) < V^+(Z(\xi^*,\Sigma^*),\Sigma^*)$.

Proof. To prove (i), begin by fixing a value ξ^*. Consider a sequence (ξ_h,Σ_h) that converges to (ξ^*,Σ^*) as $h\to\infty$. Let $z_h = Z(\xi_h,\Sigma_h)$, and note that $z_h \longrightarrow z^* := Z(\xi^*,\Sigma^*)$, since the function Z is clearly continuous for values of Σ where the denominator in $c(\Sigma)$ is non-zero, i.e. when $\eta'\Sigma\eta > 0$, and this holds for Σ^* since $\Sigma^* \in \mathcal{S}$ and thus positive definite.

Suppose first that there is no j such that $(A(\Sigma^*)c(\Sigma^*))_j = 0$. The function $A(\Sigma)$ is continuous at Σ^* by Assumption 5, and we just argued that $c(\Sigma)$ is continuous at Σ^* as well. Thus, for h sufficiently large, $\{((A(\Sigma_h)c(\Sigma_h))_j > 0\} = \{((A(\Sigma^*)c(\Sigma^*))_j > 0\}$. Likewise, the function $b(\Sigma)$ is continuous at Σ^* by Assumption 5, and so $(b_j(\Sigma)-(AZ(\xi,\Sigma))_j)/(A(\Sigma)c(\Sigma))_j$ is continuous at (ξ^*,Σ^*), from which it follows that

$$\max_{\{j:(A(\Sigma_h)c(\Sigma_h))_j<0\}} \frac{b(\Sigma_h)_j - (A(\Sigma_h)z_h)_j}{(A(\Sigma_h)c(\Sigma_h))_j} \to \max_{\{j:(A(\Sigma^*)c(\Sigma^*))_j<0\}} \frac{b(\Sigma^*)_j - (A(\Sigma^*)z^*)_j}{(A(\Sigma^*)c(\Sigma^*))_j} \tag{36}$$

when $\{((A(\Sigma^*)c(\Sigma^*))_j > 0\}$ is non-empty. By an analogous argument, if $\{((A(\Sigma^*)c(\Sigma^*))_j > 0\}$ is empty, then for h sufficiently large, $\{((A(\Sigma_h)c(\Sigma_h))_j > 0\}$ is empty as well, and so (36) holds regardless of whether $\{((A(\Sigma^*)c(\Sigma^*))_j > 0\}$ is empty.

Now, let $\mathcal{J} = \{j \,|\, ((A(\Sigma^*)c(\Sigma^*))_j = 0\}$. Note that by the same argument as in the previous paragraph,

$$\max_{\{j\notin\mathcal{J}:(A(\Sigma_h)c(\Sigma_h))_j<0\}} \frac{b(\Sigma_h)_j - (A(\Sigma_h)z_h)_j}{(A(\Sigma_h)c(\Sigma_h))_j} \to \max_{\{j:(A(\Sigma^*)c(\Sigma^*))_j<0\}} \frac{b(\Sigma^*)_j - (A(\Sigma^*)z^*)_j}{(A(\Sigma^*)c(\Sigma^*))_j}. \tag{37}$$

Additionally, by Lemma C.9, for $j\in\mathcal{J}$, $(b(\Sigma^*)_j - A(\Sigma^*)z^*)_j > 0$ for almost every value of ξ^*. For such a ξ^*, it follows from the continuous mapping theorem that for h sufficiently large, $(b(\Sigma_h) - A(\Sigma_h)z_h)_j > 0$. Thus for any $j \in \mathcal{J}$ and any subsequence $\{h_1\} \subset \{h\}$ for which $(A(\Sigma_{h_1})c(\Sigma_{h_1}))_j < 0$, we have

$$\frac{b(\Sigma_{h_1})_j - (A(\Sigma_{h_1})z_{h_1})_j}{(A(\Sigma_{h_1})c(\Sigma_{h_1}))_j} \to -\infty.$$

This implies that

$$\underset{\{j\in\mathcal{J}:(A(\Sigma_h)c(\Sigma_h))_j<0\}}{max} \frac{b(\Sigma_h)_j - (A(\Sigma_h)z_h)_j}{(A(\Sigma_h)c(\Sigma_h))_j} \to -\infty,$$

and thus

$$\lim_{h\to\infty} \underset{\{j\notin\mathcal{J}:(A(\Sigma_h)c(\Sigma_h))_j<0\}}{max} \frac{b(\Sigma_h)_j - (A(\Sigma_h)z_h)_j}{(A(\Sigma_h)c(\Sigma_h))_j} = \lim_{h\to\infty} \underset{\{j:(A(\Sigma_h)c(\Sigma_h))_j<0\}}{max} \frac{b(\Sigma_h)_j - (A(\Sigma_h)z_h)_j}{(A(\Sigma_h)c(\Sigma_h))_j}.$$

Result (i) then follows from (37). The proof of result (ii) is analogous to that for proof (i), replacing *max* with *min* and $-\infty$ with ∞. Result (iii), that $V^- < V^+$, follows from the same argument as in the proof of Lemma B.1.

□

Lemma C.11. *Let* $\tilde{g}(\xi, \Sigma, V^-, V^+, \delta) = F^{[V^-,V^+]}_{\eta'\delta,\eta'\Sigma\eta}(\eta'\xi)$. *Then* $\tilde{g}$ *is continuous in* $(\xi, \Sigma, V^-, V^+, \delta)$ *on the set* $\{(\xi, \Sigma, V^-, V^+, \delta) \,|\, \xi \in \mathbb{R}^{K+M}, \Sigma \in B(\mathcal{S}), V^- \in \mathbb{R}\cup\{-\infty\}, V^+ \in \mathbb{R}\cup\{\infty\}, \delta \in \mathbb{R}^{K+M}\}$, *for* $B(\mathcal{S})$ *an open set of positive definite matrices containing* $\mathcal{S}$.

Proof. By definition,

$$\tilde{g}(\xi, \Sigma, V^-, V^+, \delta) = \frac{\Phi\left(\frac{\eta'\xi - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)}{\Phi\left(\frac{V^+ - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)},$$

for $\sigma = \sqrt{\eta'\Sigma\eta}$. Since $\Sigma \in B(\mathcal{S})$ implies that Σ is full rank, and hence $\sigma > 0$, it is immediate from the functional form that $\tilde{g}$ is continuous when all of the values are finite.

Moreover, for V^- finite and $(\xi, \Sigma, V^-, \delta)$ fixed,

$$\lim_{V^+\to\infty} \frac{\Phi\left(\frac{\eta'\xi - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)}{\Phi\left(\frac{V^+ - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)} = \frac{\Phi\left(\frac{\eta'\xi - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)}{\Phi\left(\frac{\infty}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)}.$$

Moreover, for V^+ finite and $(\xi, \Sigma, V^+, \delta)$ fixed,

$$\lim_{V^- \to -\infty} \frac{\Phi\left(\frac{\eta'\xi - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)}{\Phi\left(\frac{V^+ - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)} = \frac{\Phi\left(\frac{\eta'\xi - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{-\infty}{\sigma}\right)}{\Phi\left(\frac{V^+ - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{-\infty}{\sigma}\right)}.$$

Finally, for (ξ, Σ, δ) fixed,

$$\lim_{(V^-, V^+) \to (-\infty, \infty)} \frac{\Phi\left(\frac{\eta'\xi - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)}{\Phi\left(\frac{V^+ - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{V^- - \eta'\delta}{\sigma}\right)} = \frac{\Phi\left(\frac{\eta'\xi - \eta'\delta}{\sigma}\right) - \Phi\left(\frac{-\infty}{\sigma}\right)}{\Phi\left(\frac{\infty}{\sigma}\right) - \Phi\left(\frac{-\infty}{\sigma}\right)}.$$

$\square$

Lemma C.12. *Suppose the function $B(\Sigma)$ satisfies Assumption 5. Let $\xi^* \sim \mathcal{N}(\delta^*, \Sigma^*)$ for $\Sigma^* \in \mathcal{S}$ such that $\mathbb{P}(\xi^* \in B(\Sigma^*)) = p > 0$. Then $g(\xi^*, \Sigma^*, \delta^*)$ is continuous for almost every $\xi^* \mid \xi^* \in B(\Sigma^*)$ for the function g as defined in Lemma C.8.*

Proof. Observe that

$$g(\xi, \Sigma, \delta) = \tilde{g}(\xi, \Sigma, V^-(\xi, \Sigma), V^+(\xi, \Sigma), \delta),$$

for the function $\tilde{g}$ as defined in Lemma C.11. Lemma C.10 gives that for almost every value of $\xi^* \mid \xi^* \in B(\Sigma^*)$, the functions V^- and V^+ are continuous in (ξ, Σ) at (ξ^*, Σ^*) with $V^-(\xi^*, \Sigma^*) < V^+(\xi^*, \Sigma^*)$. Lemma C.11 gives that $\tilde{g}$ is continuous on $\{(\xi, \Sigma, V^-, V^+, \delta) \mid \xi \in \mathbb{R}^{K+M}, \Sigma \in B(\mathcal{S}), V^- \in \mathbb{R} \cup \{-\infty\}, V^+ \in \mathbb{R} \cup \{\infty\}, \delta \in \Delta\}$, for $B(\mathcal{S})$ an open set containing $\mathcal{S}$. The result then follows immediately from the fact that the composition of continuous functions is continuous. $\square$

References

Andrews, I. and Kasy, M. ("forthcoming"). Identification of and Correction for Publication Bias. *American Economic Review.*

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

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.

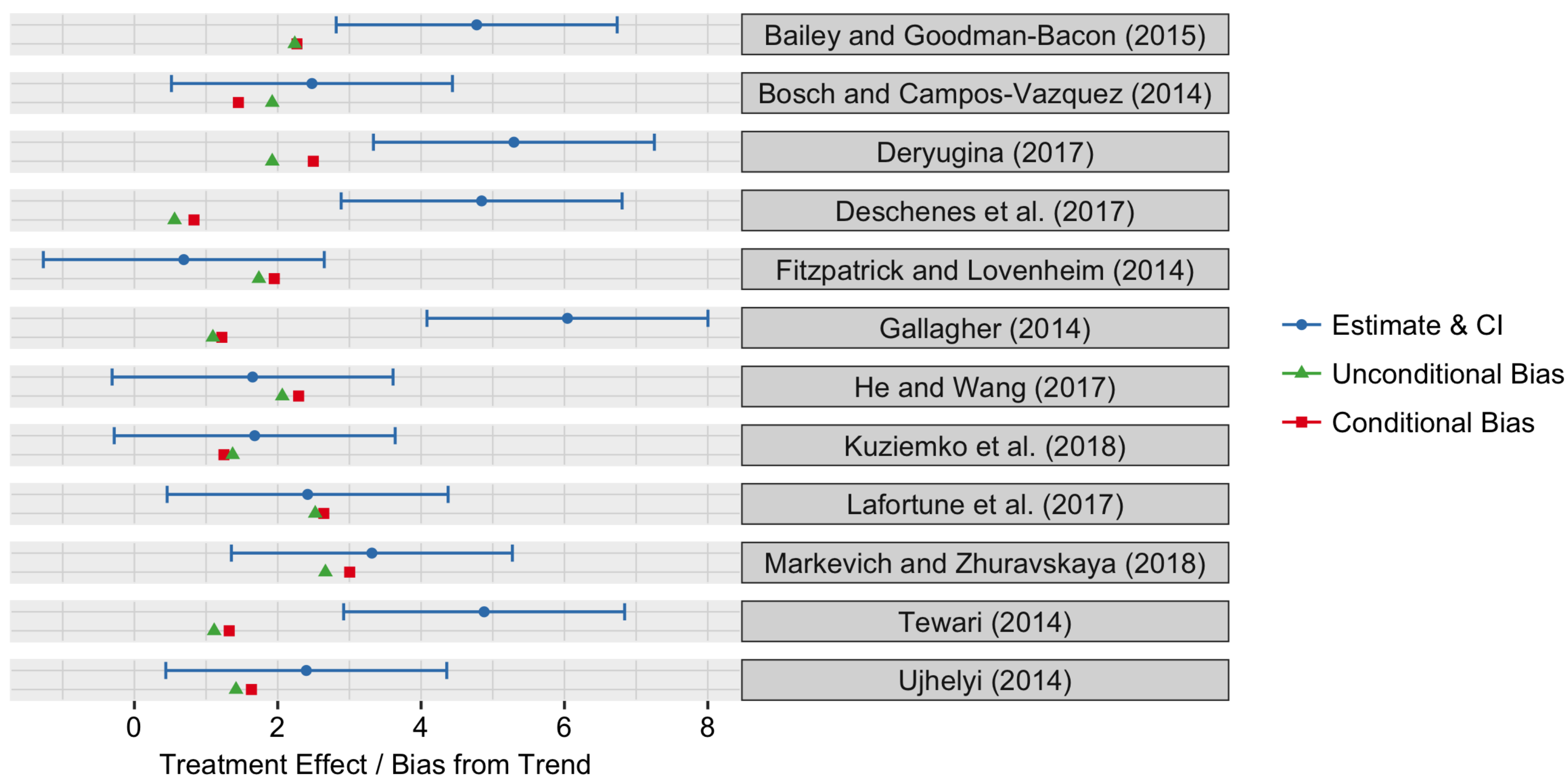
Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: A review. *arXiv:1404.5886 [math, stat]*.

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.

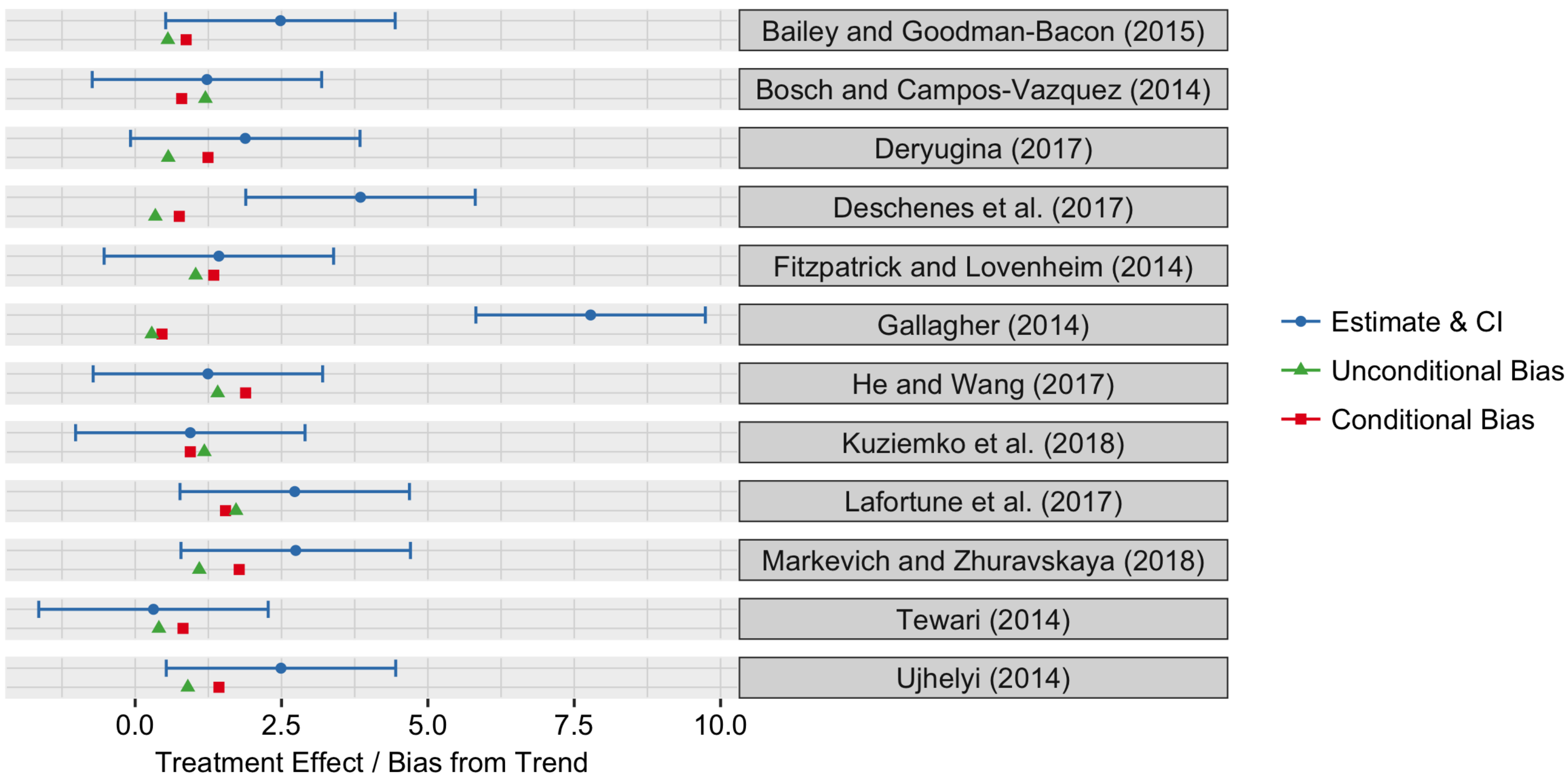
D Additional tables and figures

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



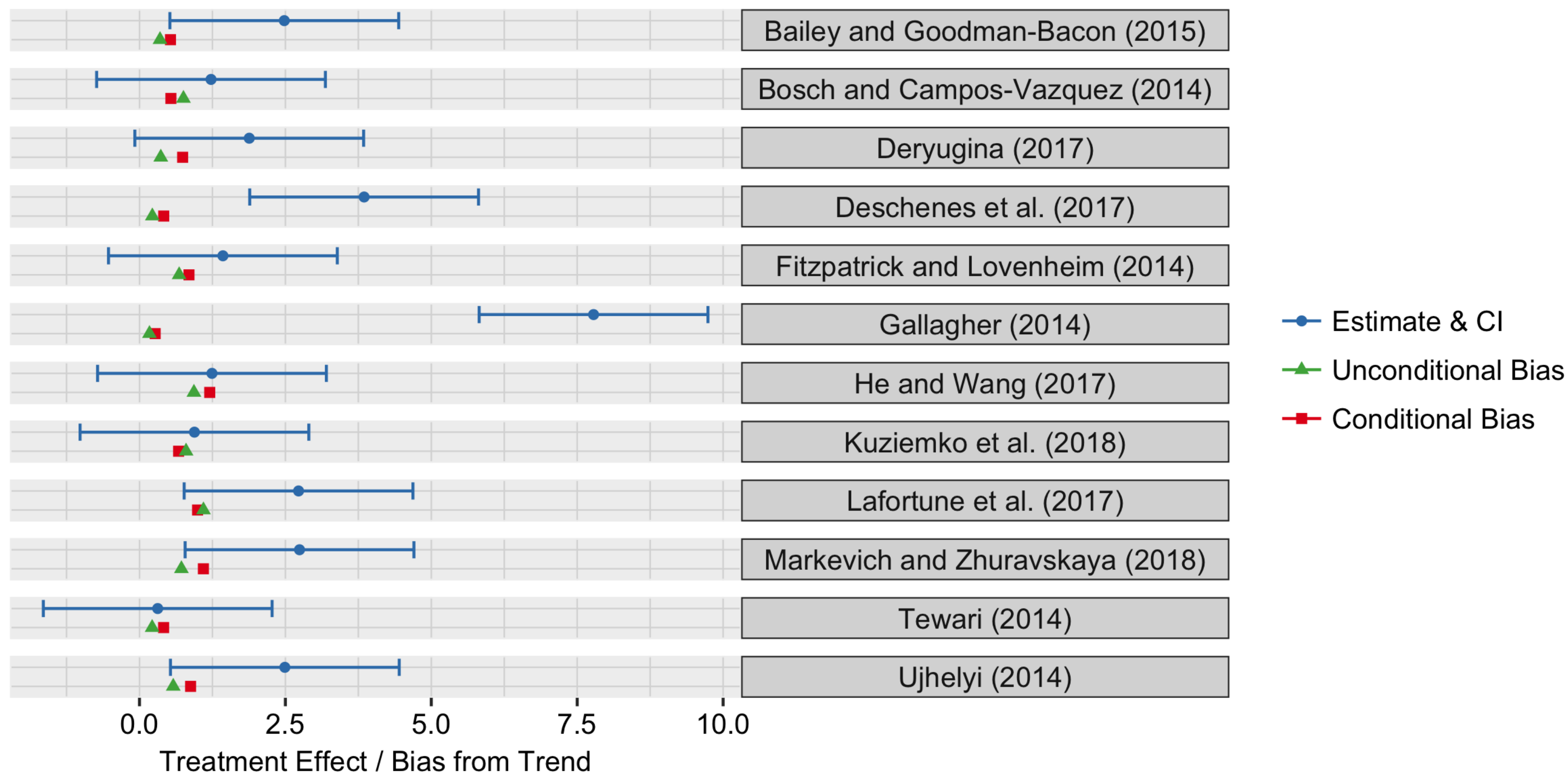
Note: I calculate the linear trend against which we would have a rejection probability of 50 percent if we rejected the research design whenever any of the pre-period event-study coefficients was statistically significant at the 5% level. I plot in red the bias that would result from such a trend conditional on not rejecting the research design; I plot in green the unconditional bias from such a trend. In blue, I plot the original OLS estimates and 95% CIs. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The parameter of interest is the average of the treatment effects in all periods after treatment began.

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



Note: I calculate the linear trend against which we would have a rejection probability of 80 percent if we rejected the research design whenever any of the pre-period event-study coefficients was statistically significant at the 5% level. I plot in red the bias that would result from such a trend conditional on not rejecting the research design; I plot in green the unconditional bias from such a trend. In blue, I plot the original OLS estimates and 95% CIs. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The parameter of interest is the treatment effect in the first period after treatment began.

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



Note: I calculate the linear trend against which we would have a rejection probability of 50 percent if we rejected the research design whenever any of the pre-period event-study coefficients was statistically significant at the 5% level. I plot in red the bias that would result from such a trend conditional on not rejecting the research design; I plot in green the unconditional bias from such a trend. In blue, I plot the original OLS estimates and 95% CIs. All values are normalized by the standard error of the estimated treatment effect and so the OLS treatment effect estimate is positive. The parameter of interest is the treatment effect in the first period after treatment began.