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

Jonathan Roth*

December 21, 2021

This supplement contains additional results for the paper "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends." Section A provides proofs for Proposition 2 and Proposition 4, generalizing the proofs given in the main text for the case of $K = 1$ to arbitrary K. Section B states and proves asymptotic results. Section C provides additional simulation results in which the treatment and control group receive stochastic common shocks. Finally, Section D contains additional tables and figures.

A Generalized Proofs for Results in the Main Text

This section provides multivariate extensions to some of the proofs in the main text, which considered only the case $K = 1$. For ease of notation, I leave the dependence of B on Σ implicit unless needed for clarity.

We begin with a series of lemmas leading up to proofs of Proposition 2 and Proposition 4 for the more general case of $K > 1$. These results extend the argument in Alecos Papadopoulos (2013) for univariate truncated normals to the multivariate case.

Lemma A.1. *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.*

*Brown University. jonathanroth@brown.edu.

Proof.

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 **?**), from which it follows that H is log-concave in μ.

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

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

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

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

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

$$D_\mu \mathbb{E}[Y \mid Y \in B, \mu] = I + \Sigma D_\mu (D_\mu \log H)'. \tag{1}$$

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,

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

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 (2) thus becomes

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

Applying the chain-rule, the second term on the right-hand side of (2) becomes

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

Substituting the expressions in the previous two displays back into (2), we have

$$\begin{aligned} D_\mu \mathbb{E}\left[Y \,|\, Y \in B, \mu\right] &= \left(\mathbb{E}\left[YY' \,|\, Y \in B, \mu\right] - \mathbb{E}\left[Y \,|\, Y \in B, \mu\right] \mathbb{E}\left[Y \,|\, Y \in B, \mu\right]'\right)\Sigma^{-1} \\ &= \mathbb{V}\mathrm{ar}\left[Y \,|\, Y \in B, \mu\right]\Sigma^{-1}, \end{aligned} \tag{3}$$

which establishes the first result. Additionally, combining (1) and (3), we have that

$$\mathbb{V}\mathrm{ar}\left[Y \,|\, Y \in B, \mu\right]\Sigma^{-1} = I + \Sigma D_\mu (D_\mu \log H)', \tag{4}$$

which implies that

$$\mathbb{V}\mathrm{ar}\left[Y \,|\, Y \in B, \mu\right] - \Sigma = \Sigma\, D_\mu (D_\mu \log H)'\, \Sigma. \tag{5}$$

However, log-concavity implies that $D_\mu (D_\mu \log H)'$ is negative semi-definite, and thus $\mathbb{V}\mathrm{ar}\left[Y \,|\, Y \in B, \mu\right] - \Sigma$ is negative semi-definite as well, as we desired to show. □

Lemma A.2. *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.3. *Suppose* $Y \sim N(0, \Sigma)$ *is* K*-dimensional normal, with* Σ *satisfying Assumption 1. Let* $B = \{y \in \mathbb{R}^K \,|\, a_j \le y \le b_j \text{ for all } j\}$, *where* $-b_j < a_j < b_j$ *for all* j. *Then for* ι *the vector of ones,* $\mathbb{E}[\iota'Y \,|\, Y \in B] = \mathbb{E}[Y_1 + \ldots + Y_K \,|\, Y \in B] > 0$.

Proof. For any $x \in \mathbb{R}^K$ such that $x_j \le b_j$ for all j, define $B^X(x) = \{y \in \mathbb{R}^K \,|\, x_j \le y \le b_j \text{ for all } j\}$. Let $b = (b_1, \ldots, b_K)$. Note that the distribution of Y is symmetric around zero, and $B^X(-b)$ is likewise symmetric around 0, from which it follows that $\mathbb{E}\left[Y \,|\, Y \in B^X(-b)\right] = \mathbb{E}\left[-Y \,|\, Y \in B^X(-b)\right] = 0$. Now, define

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

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

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

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

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

$$\begin{aligned} &\frac{\partial g(x^t)}{\partial x_K} = \\ &\frac{\partial}{\partial x_K}\left[\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}(y_1+\ldots+y_K)\,\phi_\Sigma(y)\,dy_1\ldots dy_K\right)\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}\ \phi_\Sigma(y)\,dy_1\ldots dy_K\right)^{-1}\right] = \\ &\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}(y_1+\ldots+y_K)\,\phi_\Sigma(y)\,dy_1\ldots dy_K \times \int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}\ \phi_\Sigma\left(\begin{pmatrix} y_{-K} \\ x_K^t \end{pmatrix}\right)\,dy_1\ldots dy_{K-1}\right. \\ &\left. - \int_{x_1^t}^{b_1}\cdots\int_{x_{K-1}^t}^{b_{K-1}}(y_1+\ldots+y_{K-1}+x_K^t)\,\phi_\Sigma\left(\begin{pmatrix} y_{-K} \\ x_K^t \end{pmatrix}\right)\,dy_1\ldots dy_{K-1}\ \times\ \int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}\ \phi_\Sigma(y)\,dy_1\ldots dy_K\right) \\ &\times\left(\int_{x_1^t}^{b_1}\cdots\int_{x_K^t}^{b_K}\ \phi_\Sigma(y)\,dy_1\ldots dy_K\right)^{-2} \end{aligned} \tag{6}$$

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

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

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 $\mathbb{E}\left[Y_K \,|\, x_j^t \leq Y_j \leq b_j, \forall j\right] > x_K^t$, 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 Jack Cartinhour (1990)). This completes the proof for the case where $K = 1$. For $K > 1$, it suffices to show that

$$\mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j, \forall j\right] \geq \mathbb{E}\left[Y_1 + \ldots + Y_{K-1} \,|\, x_j^t \leq Y_j \leq b_j, \text{ for } j < K, \, Y_K = x_K^t\right]. \tag{7}$$

To see why (7) 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. It is straightforward to verify that $\tilde{Y}_{-K}$ is uncorrelated, and hence (by normality) 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 $\tilde{\Sigma}$ satisfies Assumption 1, and then apply Lemma A.2 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. Hence, $\Sigma_{-K,K}\Sigma_{K,K}^{-1}y_K = c\, y_K\, \iota$ by Lemma A.2. We can therefore write

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

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

$\iota' h(c y_K \iota)$. Hence,

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

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

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

By the law of iterated expectations, we have

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

as we wished to show.

□

Proof of Proposition 2 From Proposition 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.2, $\Sigma_{12}\Sigma_{22}^{-1} = c_1\iota'$ for $c_1 > 0$, so it suffices to show that $\iota'\mathbb{E}\left[\hat{\beta}_{pre} - \beta_{pre}\,|\,\hat{\beta}_{pre}\in B\right] > 0$. 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.3.

Proof of Proposition 4 By Proposition 3, it suffices to show that

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

The result then follows immediately from the fact that $\mathbb{V}\text{ar}\left[\hat{\beta}_{pre}\,|\,\hat{\beta}_{pre}\in B\right]-\mathbb{V}\text{ar}\left[\hat{\beta}_{pre}\right]$ is negative semi-definite by Lemma A.1. □

B Uniform Asymptotic Results

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

B.1 Assumptions

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

Assumption 1 (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 $\xi_{P,n}\sim\mathcal{N}\left(\delta_{P,n},\,\Sigma_P\right)$.

Convergence in distribution is equivalent to convergence in bounded Lipschitz metric (see Theorem 1.12.4 in **?**), so Assumption 1 formalizes the notion of uniform convergence in distribution of $\hat{\beta}_n-\tau_{P,n}$ to a $\mathcal{N}\left(\delta_{P,n},\,\Sigma_P\right)$ 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 2 (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 3 (Assumptions on Σ_P)**.** *We assume that there exists* $\bar{\lambda} > 0$ *such that for all* $P \in \mathcal{P}$*,* $\Sigma_P \in \mathcal{S} := \{\Sigma \,|\, 1/\bar{\lambda} \leq \lambda_{min}(\Sigma) \leq \lambda_{max}(\Sigma) \leq \bar{\lambda}\}$*, where* $\lambda_{min}(A)$ *and* $\lambda_{max}(A)$ *denote the minimal and maximal eigenvalues of a matrix* A*.*

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

Assumption 4 (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 $A(\Sigma)\beta \leq b(\Sigma)$.

B.2 Main uniform asymptotic results

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

Proposition B.1 (Uniform conditional convergence in distribution)**.** *Under Assumptions 1-4,*

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

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

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

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

Proposition B.2 (Uniform convergence of expectations)**.** *Suppose Assumptions 1-4 hold. Let $\beta_{P,n} = \tau_{P,n} + \delta_{P,n}$. Assume that $\hat{\beta}_n - \beta_{P,n}$ is asymptotically uniformly integrable over the class $\mathcal{P}$,*

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

Then, for any $\epsilon > 0$,

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

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

B.3 Proofs of main asymptotic results

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

$$\liminf_{m\to\infty} \sup_{f\in BL_1} \left|\left| \mathbb{E}_{P_n} \left[f(\hat{\beta}_n - \tau_{P_{n_m},n_m}) \,|\, \hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m}) \right] - \mathbb{E}\left[f(\xi_{P_{n_m},n_m}) | \xi \in B(\Sigma_{P_{n_m}}) \right] \right|\right| \times \mathbb{P}_{P_{n_m}} \left(\hat{\beta}_{n_m} \in B(\hat{\Sigma}_{n_m}) \right) > 0. \quad (8)$$

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

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

which contradicts (8).

Now, suppose $p^* > 0$. Note that by Assumption 3, Σ_P falls in the set $\mathcal{S} = \{\Sigma | 1/\bar{\lambda} \le \lambda_{min}(\Sigma) \le \lambda_{max}(\Sigma) \le \bar{\lambda}\}$, which is compact (e.g., in the Frobenius norm). Thus, we can extract a further subsequence of increasing sample sizes, n_r, such that

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

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

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

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

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

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s}-\tau_{P_{n_s},n_s}-\delta^+_{n_s})\,|\,\hat{\beta}_{n_s}\in B(\hat{\Sigma}_{n_s})\right]-\mathbb{E}\left[f(\xi^*)|\xi^*\in B(\Sigma^*)\right]\right|\right| = 0 \quad (9)$$

and

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|\left|\mathbb{E}\left[f(\xi_{P_{n_s}}-\delta^+_{n_s})|\xi_{P_{n_s}}\in B(\Sigma_{P_{n_s}})\right]-\mathbb{E}\left[f(\xi^*)|\xi^*\in B(\Sigma^*)\right]\right|\right| = 0. \quad (10)$$

Equations (9) and (10) 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 (8). $\square$

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

$$\liminf_{m\to\infty} 1\left[\left|\left|\mathbb{E}\left[\hat{\beta}_{n_m}-\tau_{P_{n_m},n_m}\,|\,\hat{\beta}_{n_m}\in B(\hat{\Sigma}_{n_m})\right]-\mathbb{E}\left[\xi_{P_{n_m}}\,|\,\xi_{P_{n_m}}\in B(\Sigma_{P_{n_m}})\right]\right|\right|>\epsilon\right]\times \mathbb{P}_{P_{n_m}}\left(\hat{\beta}_{n_m}\in B(\hat{\Sigma}_{n_m})\right)>0. \tag{11}$$

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 (11).

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

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

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

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

Additionally, from uniform integrability, we have

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

Observe that

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

and hence

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

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

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

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

Lemma B.6 that $\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s}$ is uniformly asymptotically integrable conditional on $\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})$.

Convergence in distribution along with uniform asymptotic integrability implies convergence in expectation (see Theorem 2.20 in **?**), and thus

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

Likewise, Lemma B.5 gives that

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

It then follows from the triangle inequality that

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

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

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

which contradicts (11). $\square$

B.4 Auxiliary lemmas and proofs

Lemma B.1. *Suppose $(\xi_n, \Sigma_n) \xrightarrow{d} (\xi^*, \Sigma^*)$, for $\xi^* \sim \mathcal{N}(\delta^*, \Sigma^*)$ and $\Sigma^* \in \mathcal{S}$. Then, if B satisfies Assumption 4,*

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

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

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

Note that since $A(\cdot)$ and $b(\cdot)$ are continuous by Assumption 4, 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 4, and $\Sigma^* \in \mathcal{S}$ implies that Σ^* is positive definite, $A(\Sigma^*)_{(j,\cdot)}\Sigma^* A(\Sigma^*)'_{(j,\cdot)} > 0$. This implies that for each j, $(A(\Sigma^*)\xi^*)_j = b(\Sigma^*)_j$ with probability zero, and hence $(A(\Sigma^*)\xi^*)_j \neq b(\Sigma^*)_j$ for all j with probability 1. Thus, h is continuous at (ξ^*, Σ^*) for almost every ξ.

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

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

which completes the proof. □

Lemma B.2. *Suppose that* $(\xi_n, \Sigma_n) \xrightarrow{d} (\xi^*, \Sigma^*)$, *for* $\xi^* \sim \mathcal{N}(\delta^*, \Sigma^*)$ *and* $\Sigma^* \in \mathcal{S}$. *Suppose further that* $\mathbb{P}\left(\xi^* \in B(\Sigma^*)\right) = p^* > 0$ *for* $B(\Sigma)$ *satisfying Assumption 4. 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 **?**),

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

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

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

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

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

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

We can thus apply the Continuous Mapping Theorem to obtain

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

which by the definition of the conditional expectation, implies

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

as needed. □

Lemma B.3. *Suppose Assumptions 1-4 hold, and n_s is an increasing sequence of sample sizes such that*

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

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

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

and

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

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

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

$$\lim_{s\to\infty} \sup_{f \in BL_1} \left|\left| \mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^+_{n_s})\right] - \mathbb{E}\left[f(\xi^*)\right] \right|\right| = 0. \tag{12}$$

Additionally, Assumption 1 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

$$\lim_{s\to\infty} \sup_{f \in BL_1} \left|\left| \mathbb{E}_{P_{n_s}}\left[f(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s})\right] - \mathbb{E}\left[f(\xi_{P_{n_s}} - \delta^+_{n_s})\right] \right|\right| = 0. \tag{13}$$

Equations (12) and (13), together with the triangle inequality, imply that

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

or equivalently, $(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s}) \xrightarrow{d} \xi^*$. By Assumption 4, the pre-test is invariant to shifts that only affect the post-period coefficients, and so $\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})$ iff $(\hat{\beta}_{n_s} - \tau_{n_s,P_{n_s}} - \delta^+_{n_s}) \in B(\hat{\Sigma}_{n_s})$. Lemma B.1 thus implies that $\lim_{s\to\infty} \mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) = \mathbb{P}\left(\xi^* \in B(\Sigma^*)\right)$, and hence $\mathbb{P}\left(\xi^* \in B(\Sigma^*)\right) = p^* > 0$. We have thus shown that $(\hat{\beta}_{n_s} - \tau_{P_{n_s},n_s} - \delta^+_{n_s}, \hat{\Sigma}_{n_s}) \xrightarrow{d} (\xi^*, \Sigma^*)$, $(\xi_{P_{n_s}} - \delta^+_{n_s}, \Sigma_{P_{n_s}}) \xrightarrow{d} (\xi^*, \Sigma^*)$, and $\mathbb{P}\left(\xi^* \in B(\Sigma^*)\right) > 0$. The result then follows immediately from Lemma B.2.

□

Lemma B.4. *Suppose that Assumptions 1-4 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

$$\liminf_{q\to\infty} \mathbb{P}_{P_{n_q}}\left(\hat{\beta}_{n_q} \in B(\hat{\Sigma}_{n_q})\right) > 0. \tag{15}$$

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 2 then implies that $\hat{\Sigma}_{n_r} \xrightarrow{p} \Sigma^*$.

By Assumption 4, $B_{pre}(\Sigma)$ is bounded for every Σ. Let $\tilde{M}(\Sigma) = \sup_{\beta_{pre}\in B_{pre}(\Sigma)} ||\beta_{pre}||$. Assumption 4 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

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

We now show that $\limsup_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B_{\bar{M}}^c\right) = 1$, 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 (15).

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 $h(\beta) \leq 1[\beta \in B_{\bar{M}}^c]$ for all β. Thus,

$$\limsup_{r\to\infty} \mathbb{P}_{P_{n_r}}\left(\hat{\beta}_{n_r} \in B_{\bar{M}}^c\right) \geq \limsup_{r\to\infty} \mathbb{E}_{P_{n_r}}\left[h(\hat{\beta}_{n_r})\right]. \tag{16}$$

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 1, 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

$$\limsup_{r\to\infty} \mathbb{E}_{P_{n_r}}\left[h(\hat{\beta}_{n_r})\right] \geq \limsup_{r\to\infty} \mathbb{E}\left[h(\xi_{P_{n_r},n_r})\right]. \tag{17}$$

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,

$$\limsup_{r\to\infty} \mathbb{E}\left[h(\xi_{P_{n_r},n_r})\right] \geq \limsup_{r\to\infty} \mathbb{E}\left[1[\xi^{pre}_{j,r} > \bar{M} + 1]\right]. \tag{18}$$

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 3. 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 (16), (17), (18), gives the desired result.

□

Lemma B.5. *Suppose Assumptions 1-4 hold. Consider a subsequence of increasing sample sizes, n_s, such that*

$$\lim_{s\to\infty} \Sigma_{P_{n_s}} = \Sigma^*, \tag{19}$$

$$\lim_{s\to\infty} \delta^{pre}_{P_{n_s},n_s} = \delta^{pre,*}, \tag{20}$$

$$\lim_{s\to\infty} \mathbb{P}_{P_{n_s}}\left(\hat{\beta}_{n_s} \in B(\hat{\Sigma}_{n_s})\right) = p^* > 0 \tag{21}$$

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 B.1, and we use the fact that $\xi_{P_{n_s},n_s} \in B(\hat{\Sigma}_{P_{n_s}})$ iff $\xi_{P_{n_s},n_s} - \delta^+_{n_s} \in B(\hat{\Sigma}_{P_{n_s}})$ by Assumption 4. 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 $\mathbb{E}[|\xi_{j,s}|] \to \mathbb{E}\left[|\xi_j^*|\right] < \infty$. Thus, by the generalized dominated convergence theorem,

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

However, by Lemma B.1 we have that

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

Thus, by the continuous mapping theorem,

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

as we wished to show. □

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

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

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

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

$$\begin{aligned}&\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n - c_n|| \cdot 1[||Y_n - c_n|| > M]\right] \leq \\ &\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||Y_n|| \cdot 1[||Y_n - c_n|| > M]\right] + \lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[||c_n|| \cdot 1[||Y_n - c_n|| > M]\right].\end{aligned} \tag{22}$$

We now show that each of the two terms on the right hand side of (22) 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 (22) is zero, note again that since $c_n \longrightarrow c$, for n sufficiently large, $||c_n|| \leq ||c+1||$, and thus

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

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

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

□

C Power Calculations Under Stochastic Differential Trends

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

$$\delta \sim \mathcal{N}(0,\, V) \tag{23}$$

$$\hat{\beta} \,|\, \delta \sim \mathcal{N}(\delta + \tau,\, \Sigma)\,. \tag{24}$$

The distribution for $\hat{\beta}|\delta$ in (24) is identical to the model considered in Section II. However, we now treat δ as stochastic, rather than as a fixed parameter (e.g. linear in event-time). Treating δ as stochastic is sensible in situations in which we think that there may be common shocks to the treated and control groups (e.g. if each of these is a state, and there are macro-level shocks).

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

I then calculate the values of c for which the pre-test rejects 50 or 80% percent of the time, which I denote $c_{0.5}$ and $c_{0.8}$. As in Section I, I use the pre-test criterion that no pre-period coefficient is significant at the 95% level. I compute the null rejection probabilities of conventional confidence intervals for the average post-treatment effect $\bar{\tau}$ and the first-period treatment effect τ_1 under the DGPs with $c_{0.5}$ and $c_{0.9}$. The null rejection probabilities are computed over the joint distribution of $(\hat{\beta}, \delta)$.[1] As in Section I, I report these probabilities both unconditionally, and conditional on surviving the pre-test. Tables C1 and C2 show the results for τ_1 and $\bar{\tau}$, respectively. Across all specifications, the null rejection probabilities substantially exceed the nominal level of 5% for most of the papers. Conditioning on passing the pre-test generally reduces the null rejection probability, but only moderately so in most cases. Conditional on passing the pre-test, null rejection probabilities are often many multiples of the nominal size. The results thus suggest that conventional pre-tests may be underpowered against detecting common stochastic shocks, in addition to the linear secular trends considered in the main text. Concurrent work by **?** reaches a similar conclusion in a related model with stochastic violations of parallel trends.

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

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

	Unconditional			Cond'l on Passing Pre-test		
	Scaling factor for stochastic variance					
	0	$c_{0.5}$	$c_{0.8}$	0	$c_{0.5}$	$c_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.17	0.34	0.05	0.16	0.33
Bosch and Campos-Vazquez (2014)	0.05	0.19	0.38	0.03	0.12	0.26
Deryugina (2017)	0.05	0.19	0.38	0.01	0.04	0.09
Deschenes et al. (2017)	0.05	0.17	0.35	0.03	0.10	0.19
Fitzpatrick and Lovenheim (2014)	0.05	0.23	0.45	0.05	0.21	0.43
Gallagher (2014)	0.05	0.14	0.30	0.04	0.12	0.26
He and Wang (2017)	0.05	0.26	0.48	0.05	0.23	0.46
Kuziemko et al. (2018)	0.05	0.29	0.55	0.04	0.20	0.42
Lafortune et al. (2017)	0.05	0.19	0.38	0.05	0.18	0.37
Markevich and Zhuravskaya (2018)	0.05	0.22	0.44	0.04	0.18	0.38
Tewari (2014)	0.05	0.10	0.22	0.04	0.08	0.18
Ujhelyi (2014)	0.05	0.22	0.43	0.04	0.18	0.36

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

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

	Unconditional			Cond'l on Passing Pre-test		
	Scaling factor for stochastic variance					
	0	$c_{0.5}$	$c_{0.8}$	0	$c_{0.5}$	$c_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.17	0.34	0.04	0.14	0.30
Bosch and Campos-Vazquez (2014)	0.05	0.19	0.38	0.05	0.17	0.35
Deryugina (2017)	0.05	0.19	0.38	0.04	0.13	0.29
Deschenes et al. (2017)	0.05	0.17	0.35	0.04	0.11	0.22
Fitzpatrick and Lovenheim (2014)	0.05	0.23	0.45	0.05	0.22	0.44
Gallagher (2014)	0.05	0.14	0.30	0.03	0.09	0.19
He and Wang (2017)	0.05	0.26	0.48	0.04	0.23	0.45
Kuziemko et al. (2018)	0.05	0.29	0.55	0.04	0.21	0.45
Lafortune et al. (2017)	0.05	0.19	0.38	0.05	0.18	0.37
Markevich and Zhuravskaya (2018)	0.05	0.22	0.44	0.04	0.17	0.36
Tewari (2014)	0.05	0.10	0.22	0.04	0.08	0.19
Ujhelyi (2014)	0.05	0.22	0.43	0.04	0.17	0.35

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

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

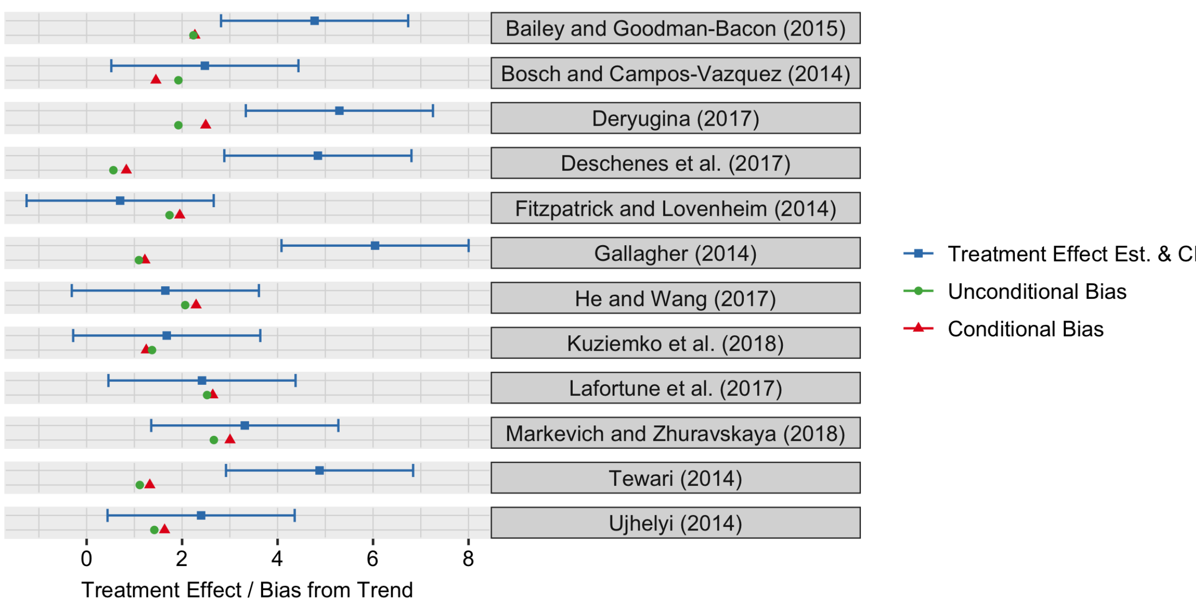
D Additional tables and figures

	Unconditional			Cond'l on Passing Pre-test		
	Slope of differential trend:					
	0	$\gamma_{0.5}$	$\gamma_{0.8}$	0	$\gamma_{0.5}$	$\gamma_{0.8}$
Bailey and Goodman-Bacon (2015)	0.05	0.06	0.09	0.04	0.07	0.13
Bosch and Campos-Vazquez (2014)	0.05	0.12	0.22	0.05	0.08	0.11
Deryugina (2017)	0.05	0.07	0.09	0.04	0.09	0.21
Deschenes et al. (2017)	0.05	0.06	0.06	0.04	0.05	0.08
Fitzpatrick and Lovenheim (2014)	0.05	0.10	0.18	0.05	0.13	0.26
Gallagher (2014)	0.05	0.05	0.06	0.03	0.04	0.05
He and Wang (2017)	0.05	0.15	0.29	0.04	0.21	0.47
Kuziemko et al. (2018)	0.05	0.13	0.22	0.04	0.07	0.11
Lafortune et al. (2017)	0.05	0.19	0.41	0.05	0.17	0.34
Markevich and Zhuravskaya (2018)	0.05	0.11	0.19	0.04	0.17	0.42
Tewari (2014)	0.05	0.06	0.07	0.04	0.06	0.11
Ujhelyi (2014)	0.05	0.09	0.15	0.04	0.12	0.28

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

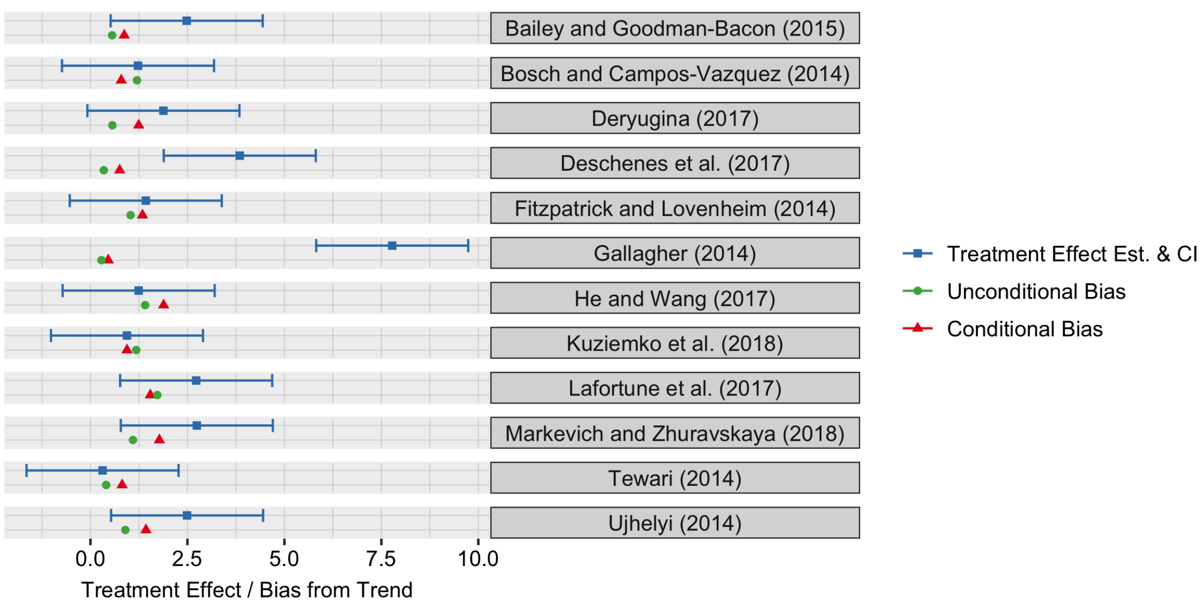
Note: This table shows null rejection probabilities, i.e. the probability that the true parameter falls outside a nominal 95% confidence interval, using data-generating processes in which parallel trends holds (slope of differential trend = 0) and in which there are linear violations of parallel trends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

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



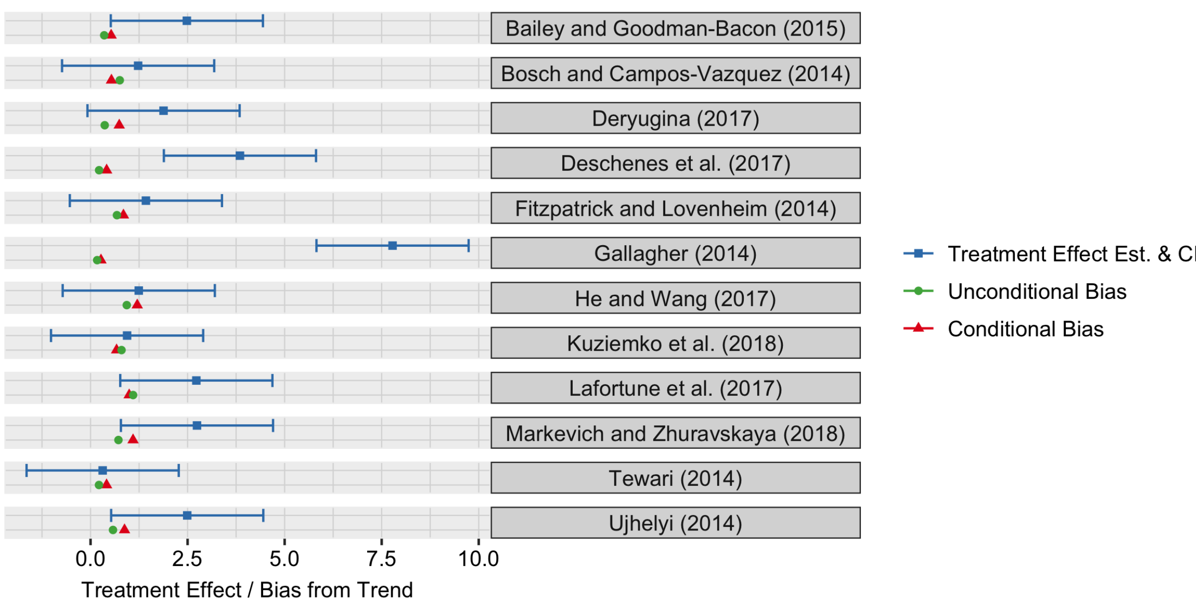
Note: I calculate the linear trend against which conventional pre-tests would reject 50 percent of the time $(\gamma_{0.5})$. The red triangles show the bias that would result from such a trend conditional on passing the pre-test $\left(\mathbb{E}\left[\hat{\tau} - \tau_* \,|\, \hat{\beta}_{pre} \in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

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



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

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



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