A More Credible Approach to Parallel Trends

Online Appendix

Ashesh Rambachan   Jonathan Roth

December 29, 2022

This online appendix contains proofs and additional results for the paper "A More Credible Approach to Parallel Trends" by Ashesh Rambachan and Jonathan Roth. Section A contains proofs and auxilliary lemmas for results stated in the main text. Section B contains additional details and results from our simulations. Section C compares our confidence sets to the sample analog to the identified set in our empirical applications.

A Proofs of Results in Main Text

Proof of Lemma 2.2

Proof. By equation (7), we can write the coverage requirement as

$$\inf_{\delta \in \Delta, \tau} \inf_{k} \inf_{\theta \in \mathcal{S}(\delta+\tau, \Delta_k)} \mathbb{P}_{\hat{\beta}_n \sim \mathcal{N}(\delta+\tau, \Sigma_n)} \left(\theta \in \bigcup_{k'} \mathcal{C}_{n,k'}(\hat{\beta}_n, \Sigma_n) \right) \geqslant 1-\alpha.$$

The left-hand side is bounded below by

$$\inf_{\delta \in \Delta, \tau} \inf_{k} \inf_{\theta \in \mathcal{S}(\delta+\tau, \Delta_k)} \mathbb{P}_{\hat{\beta}_n \sim \mathcal{N}(\delta+\tau, \Sigma_n)} \left(\theta \in \mathcal{C}_{n,k}(\hat{\beta}_n, \Sigma_n) \right),$$

which is at least $1-\alpha$ since $\mathcal{C}_{n,k}(\hat{\beta}_n, \Sigma_n)$ satisfies (10) for $\Delta = \Delta_k$ for all k. □

Proof of Proposition 3.1

Proof. We verify that the conditions of the proposition are sufficient for the conditions for size control for the conditional and hybrid tests given in Proposition 2 of ARP. Note that in our setting, the non-stochastic variable $\tilde{X}$ plays the role of the instruments Z in ARP, so all statements in ARP conditional on Z can be interpreted as unconditional in our context.

First, suppose that Assumption 5(A) holds. Then we can write $\tilde{Y}_n(\theta) = A\hat{\beta}_n - d - \tilde{A}_{(\cdot,1)}\theta = TU_n(\theta) - \zeta(\theta)$, where $U_n(\theta) = Q\hat{\beta}_n$ and $\zeta(\theta) = d + \tilde{A}_{(\cdot,1)}\theta$ is non-stochastic, which is the structure required by the first part of Assumption 1 of ARP.[39] Note that $\Omega_P := Var_P(U_n(\theta)) = Q\Sigma_P Q'$. Since Q is full-rank by assumption and Σ_P has eigenvalues bounded away from zero by Assumption 3, so too does $\Omega_P = Q\Sigma_P Q'$, as required by the latter part of Assumption 1 in ARP. Next, note that our estimate of the variance of $\tilde{Y}_n(\theta)$, $A\hat{\Sigma}_n A'$, can be expressed as $T\hat{\Omega}_n T$, for $\hat{\Omega}_n = Q\hat{\Sigma}_n Q'$. It is immediate from Assumption 4 that $\hat{\Omega}_n$ is uniformly consistent for Ω_P, as required in Assumption 2 in ARP. Next, note that if $f \in BL_1$, then $g(x) = ||G||_{op}^{-1} f(Gx)$ is also in BL_1, where $||\cdot||_{op}$ is the operator norm. This implies that

$$\sup_{f\in BL_1}\left|\mathbb{E}_P\left[f(\sqrt{n}Q(\hat{\beta}-\beta_P)\right]-\mathbb{E}\left[f(Q\xi_P)\right]\right| \leqslant ||Q||_{op}\sup_{f\in BL_1}\left|\mathbb{E}_P\left[f(\sqrt{n}(\hat{\beta}-\beta_P)\right]-\mathbb{E}\left[f(\xi_P)\right]\right|.$$

Since $U_n(\theta) = Q\hat{\beta}_n$, Assumption 2 together with the previous argument implies that

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{f\in BL_1}\left|\mathbb{E}_P\left[f(\sqrt{n}(U_n(\theta)-Q\beta_P)\right]-\mathbb{E}\left[f(\tilde{\xi}_P)\right]\right| = 0,$$

where $\tilde{\xi}_P \sim \mathcal{N}(0, \Omega_P)$. This verifies Assumption 3 in ARP. Note that Assumption 5(A) implies that Assumption D.1 in ARP is satisfied, and Assumption D.2 in ARP is trivially satisfied for $\mathcal{X} = \{\tilde{X}\}$. Hence, Proposition D.1 in ARP implies that Assumption 4 in ARP is satisfied. We have thus verified the conditions for size control in Proposition 2 of ARP.

Second, consider the case where Assumption 5(B) holds. In this case, we can write $\tilde{Y}_n(\theta) = TU_n(\theta) - \zeta(\theta)$, where now $T = A$, $U_n(\theta) = \hat{\beta}_n$, and $\zeta(\theta) = d + \tilde{A}_{(\cdot,1)}\theta$. Assumptions 1-3 in ARP can be verified analogously to the arguments above for the case where T is as given in Assumption 5(A). To verify Assumption 4 in ARP, we must show that

$$\sup_{\Sigma_P\in\mathbf{S}}\min_{\gamma,\tilde{\gamma}\in V_\dagger(\Sigma_P),\gamma\neq\tilde{\gamma},a\geqslant 0}(\gamma-a\tilde{\gamma})'A\Sigma_P A'(\gamma-c\tilde{\gamma}) > 0,$$

where $V_\dagger(\Sigma)$ is the subset of vertices in $V(\Sigma)$ that can be optimal when $\hat{\eta} > 0$ (see Lemma 4 in ARP). By Lemma A.1 below, each $\gamma \in V(\Sigma_P)$ can be written as $c_j(\Sigma_P)\bar{\gamma}_j$ for some element $\bar{\gamma}_j \in V(I)$. Moreover, $c_j(\Sigma_P) = (\bar{\gamma}_j'\tilde{\sigma}(\Sigma_P))^{-1}$, where $\tilde{\sigma}(\Sigma_P)$ is the square root of the diagonal elements of $\Omega_P = A\Sigma_P A'$. However, the jth diagonal element of Ω_P is $A_{(j,\cdot)}\Sigma_P A_{(j,\cdot)}'$, where $A_{(j,\cdot)}$ is the jth row of A. Since the eigenvalues of Σ_P are bounded above by $\bar{\lambda}$, it follows that $A_{(j,\cdot)}\Sigma_P A_{(j,\cdot)}'$ is bounded above by $\bar{\lambda}||A_{(j,\cdot)}||^2$. The elements of $\tilde{\sigma}(\Sigma_P)$ are thus

[39] Assumption 1 of ARP imposes the structure $Y_i = TU_i + \zeta_i$, where the index i corresponds with individual observations and the sample moments are formed by averaging across i. However, this structure is only used in the proofs of size control to show that the scaled sample moments, denoted $Y_{n,0}$ in ARP, have the structure $Y_{n,0} = TU_{n,0} + \zeta_{n,0}(\theta)$, where $U_{n,0}$ and $\zeta_{n,0}$ are sample averages of U_i and ζ_i. In our setting $\tilde{Y}_n$ is analogous to $\frac{1}{\sqrt{n}}Y_{n,0}$ in ARP, and we thus verify this structure directly.

bounded above, and hence $c_j(\Sigma_P)$ is bounded away from zero. Thus, there exists a $\underline{c}$ such that $c_j(\Sigma_P) \geqslant \underline{c}$ for all $\Sigma_P \in \mathcal{S}$. Hence,

$$\sup_{\Sigma_P \in \mathbf{S}} \min_{\gamma, \tilde{\gamma} \in V_\dagger(\Sigma_P), \gamma \neq \tilde{\gamma}, a \geqslant 0} (\gamma - a\tilde{\gamma})' A \Sigma_P A' (\gamma - a\tilde{\gamma}) \geqslant \underline{c}^2 \left(\sup_{\Sigma_P \in \mathbf{S}} \min_{\gamma, \tilde{\gamma} \in V(I), \gamma \neq \tilde{\gamma}, a \geqslant 0} (\gamma - a\tilde{\gamma})' A \Sigma_P A' (\gamma - a\tilde{\gamma}) \right)$$

$$\geqslant \underline{c}^2 \left(\min_{\gamma, \tilde{\gamma} \in V_\dagger(I), \gamma \neq \tilde{\gamma}, a \geqslant 0} ||(\gamma - a\tilde{\gamma})' A||^2 \underline{\lambda} \right),$$

where the second inequality uses the fact that the minimal eigenvalue of Σ_P is at least $\underline{\lambda}$. To complete the proof, it thus suffices to show that $V_\dagger(I)$ contains only vertices such that $\bar{\gamma}_j' A \neq 0$, so that the lower bound obtained in the previous display is strictly positive by Assumption 5(B). To show this, note that if $\bar{\gamma}_j' A = 0$, then $\bar{\gamma}_j' \tilde{Y}_n(\bar{\theta}) = \bar{\gamma}_j'(A\hat{\beta}_n - d - \tilde{A}_{(\cdot,1)}\bar{\theta}) = -\bar{\gamma}_j' d$. Since Δ is non-empty, there exists some δ such that $A\delta - d \leqslant 0$, which implies that $-\bar{\gamma}_j' d = \bar{\gamma}_j'(A\delta - d) \leqslant 0$ since $\bar{\gamma}_j \geqslant 0$ by construction. We have thus established that $\bar{\gamma}_j' \tilde{Y}(\bar{\theta}) \leqslant 0$, and hence $\bar{\gamma}_j$ can never be optimal when $\hat{\eta} > 0$, so $\bar{\gamma}_j \notin V_\dagger(I)$. We have thus verified that Assumption 4 in ARP holds, as needed.

□

A.1 Proof and auxiliary lemmas for uniform consistency

Proof of Proposition 3.2

Proof. Towards contradiction, suppose that the conditional test is not consistent. Then there exists an increasing sequence of sample sizes and distributions (n_m, P_m), $x > 0$, and $\omega > 0$ such that

$$\limsup_{m \to \infty} \mathbb{E}_{P_m} \left[\psi_\alpha^C(\hat{\beta}_{n_m}, A, d, \theta_{P_m}^{ub} + x, \frac{1}{n} \hat{\Sigma}_{n_m}) \right] \leqslant 1 - \omega.$$

It is straightforward to verify that the conditional test is invariant to a re-scaling of the units of $\hat{\beta}$, so that $\psi_\alpha^C(\hat{\beta}_{n_m}, A, d, \theta_{P_m}^{ub} + x, \frac{1}{n_m} \hat{\Sigma}_{n_m}) = \psi_\alpha^C(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m} d, \sqrt{n_m}(\theta_{P_m}^{ub} + x), \hat{\Sigma}_{n_m})$. Thus, along this sequence,

$$\limsup_{m \to \infty} \mathbb{E}_{P_m} \left[\psi_\alpha^C(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m} d, \sqrt{n_m}(\theta_{P_m}^{ub} + x), \hat{\Sigma}_{n_m}) \right] \leqslant 1 - \omega.$$

Since $\mathbf{V}$ is compact, we can extract a further subsequence m_1 under which $V_{P_{m_1}} \to V^*$ for $V^* \in \mathbf{V}$. Denote the top left block of V^* by Σ^*.

To obtain a contradiction, we will construct a further subsequence such that the conditions of Lemma A.2 hold asymptotically with probability at least 1-$\omega/2$. From Lemma A.1, each element $\gamma_{P_{m_1}} \in V(\Sigma_{P_{m_1}})$ can be written as $c_j(\Sigma_{P_{m_1}})\bar{\gamma}_j$, where $\bar{\gamma}_1, ..., \bar{\gamma}_J$ are the elements of $V(I)$. We argued in the proof to Proposition 3.1 that there exists a constant $\underline{c}$ such that $c_j(\Sigma_P) \geqslant \underline{c}$ for all j whenever Σ_P has eigenvalues bounded above by $\bar{\lambda}$. By an analogous argument, we can show that there exists a constant $\bar{c}$ such that $c_j(\Sigma_P) \leqslant \bar{c}$ whenever Σ_P has eigenvalues bounded below by $\underline{\lambda}$. Thus, $\underline{c} \leqslant c_j(\Sigma_P) \leqslant \bar{c}$ for $\Sigma_P \in \mathcal{S}$. For $\gamma \in V(\Sigma_P)$, $\gamma' A\Sigma_P A'\gamma = c_j(\Sigma_P)^2 \bar{\gamma}_j' A\Sigma_P A'\bar{\gamma}_j$ for some j, and thus for $\Sigma_P \in \mathbf{S}$, we have that

$$\underline{c}^2 ||\bar{\gamma}_j' A||^2 \underline{\lambda} \leqslant \gamma' A\Sigma_P A'\gamma \leqslant \bar{c}^2 ||\bar{\gamma}_j' A||^2 \bar{\lambda}.$$

Thus, either $\gamma' A\Sigma_P A'\gamma = 0$ (if $\bar{\gamma}_j' A = 0$), or

$$\underline{c}^2 \min_{j:\bar{\gamma}_j' A \neq 0} ||\bar{\gamma}_j' A||^2 \underline{\lambda} \leqslant \gamma' A\Sigma_P A'\gamma \leqslant \bar{c}^2 \max_{j:\bar{\gamma}_j' A \neq 0} ||\bar{\gamma}_j' A||^2 \bar{\lambda},$$

where the upper and lower bounds are finite and positive since $V(I)$ is finite. Now consider the vertex $\hat{\gamma}_{m_1,j} = c_j(\hat{\Sigma}_{n_{m_1}})\bar{\gamma}_j$. By the continuous mapping theorem, $\hat{\gamma}_{m_1,j}' A\hat{\Sigma}_P A'\hat{\gamma}_{m_1,j} \to_p c_j(\Sigma^*)^2 \bar{\gamma}_j' A\Sigma^* A'\bar{\gamma}_j$. From this convergence and the inequalities in the previous display, it follows that there exist constants $\underline{\sigma}^2$ and $\bar{\sigma}^2$ such that condition (i) of Lemma A.2 is satisfied w.p.a. 1.

Next, define

$$\eta(\beta, A, d, \bar{\theta}, \Sigma) := \min_{\eta,\tilde{\tau}} \eta \text{ s.t. } A\beta - d - \tilde{A}_{(\cdot,1)}\bar{\theta} - \tilde{A}_{(\cdot,-1)}\tilde{\tau} \leqslant \eta\tilde{\sigma}, \tag{21}$$

where $\tilde{\sigma}$ is the square root of the diagonal elements of $A\Sigma A'$. Since $\theta_P^{ub} \in \mathcal{S}(\beta_P, \Delta)$, $\eta(\beta_P, A, d, \theta_P^{ub}, \Sigma_P) \leqslant 0$. By duality, we can write $\eta(\beta_P, A, d, \theta_P^{ub}, \Sigma_P) = \max_{\gamma \in V(\Sigma_P)} \gamma'(A\beta_P - d - \tilde{A}_{(\cdot,1)}\theta^{ub})$. It follows that there exists some $\tilde{\gamma}_P \in V(\Sigma_P)$ such that $\tilde{\gamma}_P'(A\beta_P - d - \tilde{A}_{(\cdot,1)}\theta^{ub}) = 0$ and $-\tilde{\gamma}_P'\tilde{A}_{(\cdot,1)} > 0$, since otherwise for $\epsilon > 0$ sufficiently small we would have that $\eta(\beta_P, A, d, \theta_P^{ub} + \epsilon, \Sigma_P) = \max_{\gamma \in V(\Sigma_P)} \gamma'(A\beta_P - d - \tilde{A}_{(\cdot,1)}(\theta^{ub} + \epsilon)) \leqslant 0$, which would imply that $\theta^{ub} + \epsilon \in \mathcal{S}(\beta_P, \Delta)$, which is a contradiction. From Lemma A.1, $\tilde{\gamma}_{P_{m_1}} \in V(\Sigma_{P_{m_1}})$ can be written as $c_j(\Sigma_{P_{m_1}})\bar{\gamma}_j$, where $c_j(\Sigma) \geqslant \underline{c} > 0$ for all $\Sigma \in \mathbf{S}$ and $\bar{\gamma}_1, ..., \bar{\gamma}_J$ are the elements of $V(I)$. Since $V(I)$ is finite, we can extract a further subsequence (n_l, P_l) such that $\tilde{\gamma}_{P_l} = c_{j*}(\Sigma_{P_l})\bar{\gamma}_{j*}$ for fixed j^*. For ease of notation, without loss of generality we assume $j^* = 1$. It follows that

$$\begin{aligned}
\eta(\sqrt{n_l}\hat{\beta}_{n_l}, A, \sqrt{n_l}d, \sqrt{n_l}(\theta_{P_l}^{ub} + x), \hat{\Sigma}_{n_l}) &= \max_{\gamma \in V(\hat{\Sigma}_{n_l})} \gamma' \sqrt{n_l}(A\hat{\beta}_{n_l} - d - \tilde{A}_{(\cdot,1)}(\theta_{P_l}^{ub} + x)) \\
&\geqslant \sqrt{n_l} c_1(\hat{\Sigma}_{n_l})\bar{\gamma}_1'(A\hat{\beta}_{n_l} - d - \tilde{A}_{(\cdot,1)}(\theta_{P_l}^{ub} + x)) \\
&= c_1(\hat{\Sigma}_{n_l})\sqrt{n_l}\bar{\gamma}_1' A(\hat{\beta}_{n_l} - \beta_{P_l}) + \sqrt{n_l} c_1(\hat{\Sigma}_{n_l})(-\bar{\gamma}_1'\tilde{A}_{(\cdot,1)})x.
\end{aligned}$$

By the continuous mapping theorem, $c_1(\hat{\Sigma}_{n_l}) \to_p c_1(\Sigma^*) > 0$. Assumption 6 and the continuous mapping theorem together imply that the first term in the previous display converges in distribution to a $\mathcal{N}(0,\, c_1(\Sigma^*)^2 \bar{\gamma}_1' A \Sigma^* A' \bar{\gamma}_1)$ distribution, while the second term converges in probability to ∞. It follows that $\eta(\sqrt{n_l}\hat{\beta}_l, A, \sqrt{n_l}d, \sqrt{n_l}(\theta^{ub}_{P_l} + x), \hat{\Sigma}_{n_l}) \to_p \infty$, and thus condition (ii) of Lemma A.2 holds w.p.a. 1 for any value of M.

To complete the proof, we construct a further subsequence such that condition (iii) of Lemma A.2 holds asymptotically with probability at least 1-$\omega/2$. Let $\tilde{Y}_l = A\hat{\beta}_{n_l} - d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_{P_l} + x)$ and $\tilde{\mu}_l = A\beta_{P_l} - d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_{P_l} + x)$. Recall that any element of $V(\hat{\Sigma}_{n_l})$, say $\gamma_{l,j}$, takes the form $\gamma_{l,j} = c_j(\hat{\Sigma}_{n_l})\bar{\gamma}_j$, and our argument above implies that $\bar{\gamma}_j'\tilde{\mu}_l \leqslant -\bar{\gamma}_j'\tilde{A}_{(\cdot,1)}x$. Since $c_j(\hat{\Sigma}_{n_l}) \to_p c_j(\Sigma^*) > 0$ by the continuous mapping theorem, and $\bar{\gamma}_j'\tilde{\mu}_l$ is bounded from above, we can extract a subsequence l_1 along which $\gamma_{l_1,j}'\tilde{\mu}_{l_1} \to_p \nu_j \in \mathbb{R} \cup \{-\infty\}$. The vertex set is finite, and so passing to further subsequences we obtain a subsequence indexed by k such that $\gamma_{k,j}'\tilde{\mu}_k \to_p \nu_j \in \mathbb{R} \cup \{-\infty\}$ for all j. Observe that for distinct vertices i and j with $\bar{\gamma}_i'A \neq 0$,

$$\begin{aligned}(c_i(\hat{\Sigma}_{n_k})\bar{\gamma}_i - c_j(\hat{\Sigma}_{n_k})\bar{\gamma}_j)'\sqrt{n_k}\tilde{Y}_k =& (c_i(\hat{\Sigma}_{n_k})\bar{\gamma}_i - c_j(\hat{\Sigma}_{n_k})\bar{\gamma}_j)'\sqrt{n_k}(\tilde{Y}_k - \tilde{\mu}_k) + \\ & \sqrt{n_k}(c_i(\hat{\Sigma}_{n_k}) - c_i(\Sigma^*))\bar{\gamma}_i'\tilde{\mu}_k - \sqrt{n_k}(c_j(\hat{\Sigma}_{n_k}) - c_j(\Sigma^*))\bar{\gamma}_j'\tilde{\mu}_k + \\ & \sqrt{n_k}(c_i(\Sigma^*)\bar{\gamma}_i' - c_j(\Sigma^*)\bar{\gamma}_j')\tilde{\mu}_k\end{aligned}$$

Consider first the case where $\gamma_{k,i}'\tilde{\mu}_k$ and $\gamma_{k,j}'\tilde{\mu}_k$ both have finite limits ν_i and ν_j. Since $\sqrt{n_k}(c_i(\Sigma^*)\bar{\gamma}_i' - c_j(\Sigma^*)\bar{\gamma}_j')\tilde{\mu}_k$ is non-stochastic, we can extra a further subsequence k_1 such that $\sqrt{n_{k_1}}(c_i(\Sigma^*)\bar{\gamma}_i' - c_j(\Sigma^*)\bar{\gamma}_j')\tilde{\mu}_{k_1} \to \nu^* \in \mathbb{R} \cup \{\pm\infty\}$. Assumption 6 and the continuous mapping theorem imply that $(c_i(\hat{\Sigma}_{n_{k_1}})\bar{\gamma}_i - c_j(\hat{\Sigma}_{n_{k_1}})\bar{\gamma}_j)'\sqrt{n_{k_1}}\tilde{Y}_{k_1}$ converges in distribution to

$$\zeta_{ij} = (c_i(\Sigma^*)\bar{\gamma}_i - c_j(\Sigma^*)\bar{\gamma}_j)'A\xi_\beta + \frac{\nu_i}{c_i(\Sigma^*)}Dc_i'\xi_\Sigma - \frac{\nu_j}{c_j(\Sigma^*)}Dc_j'\xi_\Sigma + \nu^*,$$

where $(\xi_\beta', \xi_\Sigma')' \sim \mathcal{N}(0, V^*)$ and Dc_i is the gradient of $c_i(\Sigma^*)$ with respect to $vec(\Sigma^*)$. The limiting distribution is normal, and limiting variance must be positive since Assumptions 5 and 7 imply that $(c_i(\Sigma^*)\bar{\gamma}_i - c_j(\Sigma^*)\bar{\gamma}_j)'A\xi_\beta$ has positive variance[40] and is not perfectly colinear with ξ_Σ. It follows that for any ϑ, there exists some $\epsilon > 0$ such that the probability that $\zeta_{ij} \in (-\epsilon, \epsilon)$ is less than ϑ. On the other hand, if $\bar{\gamma}_i'\tilde{\mu}_k \to -\infty$, then $c_i(\hat{\Sigma}_{n_k})\bar{\gamma}_i\sqrt{n_k}\tilde{Y}_k \to_p -\infty$, so $c_i(\hat{\Sigma}_{n_k})\bar{\gamma}_i$ is optimal for $\hat{\eta}(\sqrt{n_k}\hat{\beta}_{n_k}, \sqrt{n_k}d, \sqrt{n_k}(\theta^{ub}_{P_k} + x), \hat{\Sigma}_{n_k})$ w.p.a. 0, whereas if $\bar{\gamma}_j'\tilde{\mu}_k \to -\infty$, then $\hat{\eta}(\sqrt{n_k}\hat{\beta}_{n_k}, \sqrt{n_k}d, \sqrt{n_k}(\theta^{ub}_{P_k} + x), \hat{\Sigma}_{n_k}) - c_j(\hat{\Sigma}_k)\gamma_j'\sqrt{n_k}\tilde{Y}_k \to_p \infty$. Since there

[40] This is immediate under Assumption 5(B). Under Assumption 5(A), the proof of Proposition D.1 in ARP shows that if there is a positive constant c such $(\bar{\gamma}_i - c\bar{\gamma}_j)'A = 0$, then $c_i(\hat{\Sigma}_{n_k})\bar{\gamma}_i$ and $c_j(\hat{\Sigma}_{n_k})\bar{\gamma}_j$ can only be optimal vertices if $\hat{\eta} \leqslant 0$. Since we've shown $\hat{\eta} \to_p \infty$, such vertices will be optimal w.p.a. 0, and thus can be ignored when establishing part (iii) of Lemma A.2.

are a finite number of pairs of vertices, we can choose ϑ such that the probability that $\zeta_{ij} \in (-\epsilon, \epsilon)$ for any (i, j) is bounded above by $\omega/2$, and thus condition (iii) of Lemma A.2 is satisfied with probability at least $\omega/2$, as we wished to show. The result for the hybrid test is immediate from the fact that the hybrid test rejects whenever the size-$\frac{\alpha-\kappa}{1-\kappa}$ conditional test rejects. □

Lemma A.1. *Let $F(\Sigma) := \{\gamma : \tilde{A}'_{(\cdot,-1)}\gamma = 0, \tilde{\sigma}(\Sigma)'\gamma = 1, \gamma \geqslant 0\}$ be the feasible set of the dual problem, where $\tilde{\sigma}(\Sigma)$ is the vector containing the square-roots of the diagonal elements of $A\Sigma A'$. Let $V(\Sigma)$ denote the set of vertices of $F(\Sigma)$. Then, for any Σ positive definite,*

$$V(\Sigma) = \{c_1(\Sigma)\bar{\gamma}_1, ..., c_J(\Sigma)\bar{\gamma}_J\},$$

where $\bar{\gamma}_1, ..., \bar{\gamma}_J$ are the elements of $V(I)$ and $c_j(\Sigma) = (\bar{\gamma}_j'\tilde{\sigma}(\Sigma))^{-1}$.

Proof of Lemma A.1

Proof. Immediate from Lemma A.2 in ARP. □

Lemma A.2. *For any positive constants $\epsilon, \underline{\sigma}^2, \bar{\sigma}^2$, there exists a finite constant $\bar{C}$ such that the conditional test $\psi_\alpha^C(\hat{\beta}, A, d, \theta, \Sigma)$ rejects whenever the following conditions are satisfied*

(i) For all $\gamma \in V(\Sigma)$, either $\gamma' A\Sigma A'\gamma = 0$ or $\underline{\sigma}^2 \leqslant \gamma' A\Sigma A'\gamma \leqslant \bar{\sigma}^2$.

(ii) $\hat{\eta} = \max_{\gamma \in V(\Sigma)} \gamma'\tilde{Y} > \bar{C}$, where $\tilde{Y} = A\hat{\beta} - d - \tilde{A}_{(\cdot,1)}\theta$.

(iii) If the optimal vertex γ_* satisfies, $\gamma_*' A\Sigma A'\gamma_* > 0$, then for all $\tilde{\gamma} \in V(\Sigma)$ with $\tilde{\gamma} \neq \gamma_*$, we have that $|\gamma_*'\tilde{Y} - \tilde{\gamma}'\tilde{Y}| > \epsilon$.

Proof. Let $\tilde{\Sigma} = A\Sigma A'$. If the optimal vertex γ_* satisfies $\gamma_*'\tilde{\Sigma}\gamma_* = 0$, then the conditional test rejects whenever $\hat{\eta} > 0$, so condition (ii) with any $C > 0$ suffices. For the remainder of the proof, we show that conditions (i)-(iii) are sufficient when $\gamma_*'\Sigma\gamma_* \neq 0$. Observe that the conditional test rejects if and only if $\hat{\eta} > 0$ and

$$\frac{\Phi(t) - \Phi(z^{lo})}{\Phi(z^{up}) - \Phi(z^{lo})} > 1 - \alpha,$$

where $t = \frac{\hat{\eta}}{\sigma^*}$, $z^{lo} = \frac{v^{lo}}{\sigma^*}$, $z^{up} = \frac{v^{up}}{\sigma^*}$, and $\sigma^* = \sqrt{\gamma_*'\tilde{\Sigma}\gamma_*}$. It is clear that the left-hand side of the previous display is increasing in t and decreasing in z^{up}. It is also decreasing in z^{lo}, since the derivative with respect to z^{lo} is

$$-\frac{\phi(z_{lo})(\Phi(z^{up}) - \Phi(z^{lo}))}{(\Phi(z^{up}) - \Phi(z^{lo}))^2} < 0.$$

From Lemma A.3 below, condition (iii) implies that $\hat{\eta} - v^{lo} \geqslant \epsilon$, and thus $z^{lo} \leqslant t - \tilde{\epsilon}$, for $\tilde{\epsilon} = \epsilon/\bar{\sigma}$. This, combined with the previous discussion, implies that the conditional test rejects whenever $\hat{\eta} > 0$ and

$$\frac{\Phi(t) - \Phi(t-\tilde{\epsilon})}{1 - \Phi(t-\tilde{\epsilon})} > 1 - \alpha.$$

By L'Hopitale's rule, we have that

$$\lim_{t\to\infty} \frac{\Phi(t) - \Phi(t-\tilde{\epsilon})}{1 - \Phi(t-\tilde{\epsilon})} = \lim_{t\to\infty} \frac{\phi(t-\tilde{\epsilon}) - \phi(t)}{\phi(t-\tilde{\epsilon})} = \lim_{t\to\infty} 1 - \frac{\phi(t)}{\phi(t-\tilde{\epsilon})} = 1.$$

Hence, there exists $\tilde{C} > 0$ such that the conditional test rejects whenever $t \geqslant \tilde{C}$. But $t = \frac{\hat{\eta}}{\sigma_*}$ and thus $t > \tilde{C}$ whenever $\hat{\eta} > \bar{C}$ for $\bar{C} = \tilde{C}\bar{\sigma}$.

□

Lemma A.3. *Consider the conditional test* $\psi^C_\alpha(\hat{\beta}, A, d, \theta, \Sigma)$*. If the optimal vertex* γ_* *is such that* $\gamma_*' A\Sigma A'\gamma_* > 0$*, then* $\hat{\eta} - v^{lo} \geqslant \min_{\gamma\in V(\Sigma),\gamma\neq\gamma_*} |\gamma_*'\tilde{Y} - \gamma'\tilde{Y}|$ *where* $\tilde{Y} = A\hat{\beta} - d - \tilde{A}_{(\cdot,1)}\theta$*. Similarly,* $v^{up} - \hat{\eta} \geqslant \frac{\gamma_*' A\Sigma A'\gamma_*}{\max_{\gamma\in V(\Sigma)} \gamma' A\Sigma A'\gamma} \min_{\gamma\in V(\Sigma),\gamma\neq\gamma_*} |\gamma_*'\tilde{Y} - \gamma'\tilde{Y}|$.

Proof. Since $\hat{\eta}$ is finite, the results hold trivially when v^{lo} and v^{up} are infinite. For the remainder of the proof, we assume that they are finite. Let $\tilde{\Sigma} = A\Sigma A'$. Lemma 1 in ARP implies that

$$v^{lo} = \min_{\gamma\in V(\Sigma):\gamma_*'\tilde{\Sigma}\gamma_* - \gamma_*'\tilde{\Sigma}\gamma > 0} \frac{\gamma_*'\tilde{\Sigma}\gamma_*\gamma' S}{\gamma_*'\tilde{\Sigma}\gamma_* - \gamma_*'\tilde{\Sigma}\gamma},$$

where $S = (I - \frac{\tilde{\Sigma}\gamma_*}{\gamma_*'\tilde{\Sigma}\gamma_*}\gamma_*')\tilde{Y}$. Let $\tilde{\gamma}$ denote the vertex at which the minimum is obtained. Substituting in the definition of S and re-arranging terms, we obtain that

$$\hat{\eta} - v^{lo} = \frac{\gamma_*'\tilde{\Sigma}\gamma_*}{\gamma_*'\tilde{\Sigma}\gamma_* - \gamma_*'\tilde{\Sigma}\tilde{\gamma}}(\gamma_*'\tilde{Y} - \tilde{\gamma}'\tilde{Y}) \geqslant (\gamma_*'\tilde{Y} - \tilde{\gamma}'\tilde{Y}),$$

from which the result for v^{lo} is immediate. We can analogously show that

$$v^{up} - \hat{\eta} = \frac{\gamma_*'\tilde{\Sigma}\gamma_*}{\gamma_*'\tilde{\Sigma}\tilde{\gamma} - \gamma_*'\tilde{\Sigma}\gamma_*}(\gamma_*'\tilde{Y} - \tilde{\gamma}'\tilde{Y}),$$

for a vertex $\tilde{\gamma}$ such that $\gamma_*'\tilde{\Sigma}\tilde{\gamma} - \gamma_*'\tilde{\Sigma}\gamma_* > 0$. The result then follows from noting that

$$\frac{\gamma_*'\tilde{\Sigma}\gamma_*}{\gamma_*'\tilde{\Sigma}\tilde{\gamma} - \gamma_*'\tilde{\Sigma}\gamma_*} \geqslant \frac{\gamma_*'\tilde{\Sigma}\gamma_*}{\gamma_*'\tilde{\Sigma}\tilde{\gamma}} \geqslant \frac{\gamma_*'\tilde{\Sigma}\gamma_*}{\max_{\gamma\in V(\Sigma)} \gamma'\tilde{\Sigma}\gamma}.$$

□

A.2 Proof and auxiliary lemmas for uniform local asymptotic power

Proof of Proposition 3.3

Proof. By an invariance to scale argument as in Proposition 3.2, it is sufficient to show that

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_\epsilon} \left| \mathbb{E}_P \left[\psi_\alpha^C(\sqrt{n}\hat{\beta}_n, A, \sqrt{n}d, \sqrt{n}\theta_P^{ub} + x, \hat{\Sigma}_n) \right] - \rho_\alpha^*(P,x) \right| = 0.$$

To show this, it suffices to establish that for every subsequence (n_m, P_m) with $n_m \to \infty$, there exists a further subsequence l such that

$$\lim_{l\to\infty} \left| \mathbb{E}_{P_l} \left[\psi_\alpha^C(\sqrt{n_l}\hat{\beta}_{n_l}, A, \sqrt{n_l}d, \sqrt{n}\theta_{P_l}^{ub} + x, \hat{\Sigma}_{n_l}) \right] - \rho_\alpha^*(P_l,x) \right| = 0.$$

Since $P_m \in \mathcal{P}_\epsilon$, for each m there exists a B_m^* and a value $\tilde{\tau}_m^*$ such that

$$A_{(B_m^*,\cdot)}\beta_{P_m} - d_{B_m^*} - \tilde{A}_{(B_m^*,1)}\theta_{P_m}^{ub} - \tilde{A}_{(B_m^*,-1)}\tilde{\tau}_m^* = 0 \tag{22}$$

$$A_{(-B_m^*,\cdot)}\beta_{P_m} - d_{-B_m^*} - \tilde{A}_{(-B_m^*,1)}\theta_{P_m}^{ub} - \tilde{A}_{(-B_m^*,-1)}\tilde{\tau}_m^* < -\epsilon. \tag{23}$$

Since there are a finite number of possible values of B_m^*, we can extract a subsequence m_1 along which $B_{m_1}^*$ is constant. For simplicity of notation, we'll denote the constant value $B_{m_1}^*$ by B^*. Similarly, Lemma A.4 implies that there is a unique element $\gamma_{m_1}^* \in V(\Sigma_{P_{m_1}})$ such that the elements of $\gamma_{m_1}^*$ in positions $-B^*$ are all 0. By Lemma A.1, we can write $\gamma_{m_1}^* = c_j(\Sigma_{P_{m_1}})\bar{\gamma}_j$ for $c_j(\cdot)$ a continuous function and $\bar{\gamma}_j \in V(I)$. Since $V(I)$ is finite, we can extract a subsequence m_2 along which $\gamma_{m_2}^* = c_{j^*}(\Sigma_{P_{m_2}})\bar{\gamma}_{j^*}$ for a fixed j^*, which without loss of generality we normalize to $j^* = 1$. Moreover, since $\mathbf{S}$ is compact, we can extract a further subsequence l along which $\Sigma_{P_l} \to \Sigma^*$. By Assumption 4, $\hat{\Sigma}_{n_l} \to_p \Sigma^*$. The continuous mapping theorem then implies that $\gamma_l^* = c_1(\Sigma_{P_l})\bar{\gamma}_1 \to c_1(\Sigma^*)\bar{\gamma}_1$, and likewise, $\hat{\gamma}_l^* := c_1(\hat{\Sigma}_{n_l})\bar{\gamma}_1 \to_p c_1(\Sigma^*)\bar{\gamma}_1$. From Lemma A.9, we have that

$$\rho_\alpha^*(P_l,x) = \Phi\left(\frac{-\gamma_l^{*\prime}\tilde{A}_{(\cdot,1)}x}{\sqrt{\gamma_l^{*\prime}A\Sigma_{P_l}A'\gamma_l^*}} - z_{1-\alpha} \right),$$

which combined with the convergences shown above implies that

$$\rho_\alpha^*(P_l,x) \to \Phi\left(\frac{-\bar{\gamma}_1'\tilde{A}_{(\cdot,1)}x}{\sqrt{\bar{\gamma}_1'A\Sigma^*A'\bar{\gamma}_1}} - z_{1-\alpha} \right). \tag{24}$$

Now, for the function $\eta(\cdot)$ defined in (21), let

$$\hat{\eta}_l = \eta(\sqrt{n_l}\hat{\beta}_{n_l}, A, \sqrt{n_l}d, \sqrt{n_l}\theta_{P_l}^{ub} + x, \hat{\Sigma}_{n_l}).$$

By duality, we have that

$$\hat{\eta}_l = \max_{\gamma \in V(\hat{\Sigma}_{n_l})} \gamma' \left(\sqrt{n} A \hat{\beta}_{n_l} - \sqrt{n_l} d - \sqrt{n_l} \tilde{A}_{(\cdot,1)} \theta^{ub}_{P_l} - \tilde{A}_{(\cdot,1)} x \right)$$
$$\geqslant \hat{\gamma}^{*\prime}_l \left(\sqrt{n} A \hat{\beta}_{n_l} - \sqrt{n_l} d - \sqrt{n_l} \tilde{A}_{(\cdot,1)} \theta^{ub}_{P_l} - \tilde{A}_{(\cdot,1)} x \right).$$

By construction, $\hat{\gamma}^*_l$ has zero elements in positions $-B^*$ and satisfies $\hat{\gamma}^{*\prime}_l \tilde{A}_{(\cdot,-1)} = 0$. This, combined with equation (22) implies that

$$\hat{\gamma}^{*\prime}_l \left(\sqrt{n_l} A \hat{\beta}_{n_l} - \sqrt{n_l} d - \sqrt{n_l} \tilde{A}_{(\cdot,1)} \theta^{ub}_{P_l} - \tilde{A}_{(\cdot,1)} x \right) = \hat{\gamma}^{*\prime}_l A \sqrt{n_l} (\hat{\beta}_{n_l} - \beta_{P_l}) - \hat{\gamma}^{*\prime}_l \tilde{A}_{(\cdot,1)} x.$$

From Assumption 2 combined with Slutsky's lemma, we have that

$$\hat{\gamma}^{*\prime}_l A \sqrt{n_l} (\hat{\beta}_{n_l} - \beta_{P_l}) - \hat{\gamma}^{*\prime}_l \tilde{A}_{(\cdot,1)} x \to_d \mathcal{N} \left(-c_1(\Sigma^*) \bar{\gamma}_1' \tilde{A}_{(\cdot,1)} x,\, c_1(\Sigma^*)^2 \bar{\gamma}_1' A \Sigma^* A' \bar{\gamma}_1 \right).$$

Now, consider $\hat{\gamma}_{l,j} = c_j(\hat{\Sigma}_{n_l}) \bar{\gamma}_j$ for $j \neq 1$. By construction $\bar{\gamma}_j \geqslant 0$, and Lemma A.4 implies that $\bar{\gamma}_j$ has a non-zero element in at least one component in B^*. But this, combined with equations (22) and (23) and the fact that $c_j(\hat{\Sigma}_{n_l}) \to_p c_j(\Sigma^*) > 0$, implies that

$$\hat{\gamma}_{l,j}' \left(\sqrt{n_l} A \beta_{P_l} - \sqrt{n_l} d - \sqrt{n_l} \tilde{A}_{(\cdot,1)} \theta^{ub}_{P_l} - \tilde{A}_{(\cdot,1)} x \right) \to_p -\infty,$$

and thus

$$\hat{\gamma}_{l,j}' \left(\sqrt{n_l} A \hat{\beta}_{n_l} - \sqrt{n_l} d - \sqrt{n_l} \tilde{A}_{(\cdot,1)} \theta^{ub}_{P_l} - \tilde{A}_{(\cdot,1)} x \right) \to_p -\infty,$$

as well, since as before $\hat{\gamma}_{l,j}' A \sqrt{n} (\hat{\beta}_{n_l} - \beta_{P_l})$ converges in distribution to a normal distribution with finite variance. This implies that $\hat{\gamma}^*_l$ is the optimizer of the problem for $\hat{\eta}_l$ with probability approaching 1, and thus

$$\hat{\eta}_l \to_d \mathcal{N} \left(-c_1(\Sigma^*) \bar{\gamma}_1' \tilde{A}_{(\cdot,1)} x,\, c_1(\Sigma^*)^2 \bar{\gamma}_1' A \Sigma^* A' \bar{\gamma}_1 \right).$$

This also implies that for any $j \neq 1$, $|\hat{\eta}_l - \hat{\gamma}_{l,j}' \tilde{Y}_l| \to_p \infty$, where $\tilde{Y}_l = \sqrt{n_l} A \hat{\beta}_{n_l} - \sqrt{n_l} d - \sqrt{n_l} \tilde{A}_{(\cdot,1)} \theta^{ub}_{P_l} - \tilde{A}_{(\cdot,1)} x$. Since there are a finite number of vertices, it follows that $\min_{j \neq 1} |\hat{\eta}_l - \hat{\gamma}_{l,j}' \tilde{Y}_l| \to -\infty$. This together with the result of Lemma A.3 implies that $|\hat{\eta}_l - v^{lo}_l| \to_p \infty$ and $|\hat{\eta}_l - v^{up}_l| \to_p \infty$, where v^{lo}_l, v^{up}_l are the values of v^{lo}, v^{up} associated with the $\psi^C_\alpha(\sqrt{n_l} \hat{\beta}_{n_l}, A, \sqrt{n_l} d, \sqrt{n_l} \theta^{ub}_{P_l} + x, \hat{\Sigma}_{n_l})$ test. Since $\hat{\eta}_l$ is stochastically bounded, and by construction $v^{lo}_l \leqslant \hat{\eta}_l \leqslant v^{up}_l$, it follows that $v^{lo}_l \to_p -\infty$ and $v^{up}_l \to_p \infty$. Let $\hat{\sigma}^2_l = \gamma_{*,l}' A \hat{\Sigma}_{n_l} A' \gamma_{*,l}$ denote the variance at the optimal vertex used by the $\psi^C_\alpha(\sqrt{n_l} \hat{\beta}_{n_l}, A, \sqrt{n_l} d, \sqrt{n_l} \theta^{ub}_{P_l} + x, \hat{\Sigma}_{n_l})$ test. Since, we've shown that $\hat{\gamma}^*_l$ is optimal w.p.a. 1, we have that $\hat{\sigma}^2_l \to_p c_1(\Sigma^*)^2 \bar{\gamma}_1' A \Sigma^* A' \bar{\gamma}_1$. From another application of the continuous mapping theorem, we have that

$$\frac{\Phi(\hat{\eta}_l/\hat{\sigma}_l) - \Phi(v_l^{lo}/\hat{\sigma}_l)}{\Phi(v_l^{up}/\hat{\sigma}_l) - \Phi(v_l^{lo}/\hat{\sigma}_l)} \to_d \frac{\Phi(\xi) - \Phi(-\infty)}{\Phi(\infty) - \Phi(-\infty)} = \Phi(\xi),$$

where $\xi \sim \mathcal{N}\left(-\bar{\gamma}_1' \tilde{A}_{(\cdot,1)} x / \sqrt{\bar{\gamma}_1' A \Sigma^* A \bar{\gamma}_1},\, 1\right)$. The limiting distribution is continuous, and thus

$$\mathbb{P}_{P_l}\left(\frac{\Phi(\hat{\eta}_l/\hat{\sigma}_l) - \Phi(v_l^{lo}/\hat{\sigma}_l)}{\Phi(v_l^{up}/\hat{\sigma}_l) - \Phi(v_l^{lo}/\hat{\sigma}_l)} > 1-\alpha\right) \to \mathbb{P}\left(\Phi(\xi) > 1-\alpha\right) = \Phi\left(\frac{-\bar{\gamma}_1 \tilde{A}_{(\cdot,1)} x}{\sqrt{\bar{\gamma}_1 A \Sigma^* A' \bar{\gamma}_1}} - z_{1-\alpha}\right).$$

Moreover, for $\alpha < 0.5$, z^{lo} sufficiently small, and z^{up} sufficiently large, $(\Phi(\hat{\eta}_l) - \Phi(z^{lo}))/(\Phi(z^{up}) - \Phi(z^{lo})) > 1-\alpha$ only if $\hat{\eta}_l > 0$. It follows that

$$\mathbb{P}_{P_l}\left(\frac{\Phi(\hat{\eta}_l/\hat{\sigma}_l) - \Phi(v_l^{lo}/\hat{\sigma}_l)}{\Phi(v_l^{up}/\hat{\sigma}_l) - \Phi(v_l^{lo}/\hat{\sigma}_l)} > 1-\alpha, \hat{\eta}_l > 0\right) \to \Phi\left(\frac{-\bar{\gamma}_1 \tilde{A}_{(\cdot,1)} x}{\sqrt{\bar{\gamma}_1 A \Sigma^* A' \bar{\gamma}_1}} - z_{1-\alpha}\right).$$

However, the event in the previous display is precisely the event that $\psi_\alpha^C(\sqrt{n_l}\hat{\beta}_l, A, \sqrt{n_l} d, \sqrt{n_l}\theta_P^{ub} + x, \hat{\Sigma}_{n_l}) = 1$, and thus

$$\mathbb{E}_{P_l}\left[\psi_\alpha^C(\sqrt{n_l}\hat{\beta}_l, A, \sqrt{n_l} d, \sqrt{n_l}\theta_P^{ub} + x, \hat{\Sigma}_{n_l})\right] \to \Phi\left(\frac{-\bar{\gamma}_1 \tilde{A}_{(\cdot,1)} x}{\sqrt{\bar{\gamma}_1 A \Sigma^* A' \bar{\gamma}_1}} - z_{1-\alpha}\right).$$

The result is then immediate from the previous display combined with (24).

□

Lemma A.4. *If LICQ holds in direction l at β_P, then there exists a unique $\bar{\gamma} \in V(\Sigma_P)$ such that $\bar{\gamma}_{-B*} = 0$, where B^* is the set of binding moments at the optimum to (16).*

Proof. We first show that there is at most one such $\bar{\gamma}$. By definition, any vertex $\gamma \in V(\Sigma_P)$ satisfies $\gamma' \tilde{A}_{(\cdot,-1)} = 0$. Recall that $\tilde{A} = A_{(\cdot,post)}\Gamma^{-1}$, where Γ is full rank. LICQ implies that $A_{(B*,post)}$ has full row rank, and thus so does $\tilde{A}_{(B*,\cdot)}$. It follows that $\tilde{A}_{(B*,-1)}$ has rank at least $|B^*|-1$. If the rank is $|B^*|$, then there are no non-zero solutions to $\gamma_{B*}' \tilde{A}_{(B*,-1)} = 0$, and thus there are no vertices with $\gamma_{-B*} = 0$. If the rank is $|B^*|-1$, then any solution to $\gamma' \tilde{A}_{(\cdot,-1)} = 0$ with $\gamma_{-B*} = 0$ takes the form $\gamma_{B*} = c \cdot \nu$ for some constant c and ν a vector the generates the one-dimensional nullspace of $\tilde{A}_{(B*,-1)}$. However, any $\gamma \in V(\Sigma_P)$ also must satisfy $\gamma' \tilde{\sigma} = 1$, which uniquely pins down the constant c. Thus, there is at most one element of the feasible set with $\gamma_{-B*} = 0$.

We next show that there exists such a $\bar{\gamma}$. Consider the optimization $\eta(\beta_P, A, d, \theta_P^{ub}, \Sigma_P)$ for $\eta(\cdot)$ defined in (21). As argued in the proof to Proposition 3.2, since θ_P^{ub} is on the boundary of the identified set, we must have $\eta(\beta_P, A, d, \theta_P^{ub}, \Sigma_P) = 0$. However, LICQ implies that there exists a value $\tilde{\tau}^*$ such that

$$A_{(B*,\cdot)}\beta_P - d_{B*} - \tilde{A}_{(B*,1)}\theta_P^{ub} - \tilde{A}_{(B*,-1)}\tilde{\tau}^* = 0$$
$$A_{(-B*,\cdot)}\beta_P - d_{-B*} - \tilde{A}_{(-B*,1)}\theta_P^{ub} - \tilde{A}_{(-B*,1)}\tilde{\tau}^* < 0.$$

In particular, this holds for $\tilde{\tau}^* = \Gamma_{(-1,\cdot)}\tau^*$. It follows that $(\eta, \tilde{\tau}) = (0, \tilde{\tau}^*)$ is a solution to $\eta(\beta_P, A, d, \theta_P^{ub}, \Sigma_P)$. By duality, there is some $\bar{\gamma} \in V(\Sigma_P)$ that is a Lagrange multiplier for this optimization problem. The complementary slackness conditions imply, however, that $\bar{\gamma}_{-B*} = 0$, as needed.

□

Lemma A.5. *Suppose $\hat{\beta} \sim \mathcal{N}(\beta, \Sigma)$ for Σ known. Let B_0 be a closed, convex set. Then the most-powerful size α test of $H_0 : \beta \in B_0$ against the point alternative $H_A : \beta = \beta_A$ is equivalent to the most powerful test of $H_0 : \beta = \tilde{\beta}$ against $H_A : \beta = \beta_A$, where $\tilde{\beta} = \arg\min_{\beta \in B_0} ||\beta - \beta_A||_\Sigma$ and $||\cdot||_\Sigma$ is the Mahalanobis norm in Σ, $||x||_\Sigma = \sqrt{x'\Sigma^{-1}x}$. The most powerful test rejects for values of $(\beta_A - \tilde{\beta})'\Sigma^{-1}\hat{\beta}$ greater than $(\beta_A - \tilde{\beta})'\Sigma^{-1}\tilde{\beta} + z_{1-\alpha}||\beta_A - \tilde{\beta}||_\Sigma$, and has power against the alternative of $\Phi(||\beta_A - \tilde{\beta}||_\Sigma - z_{1-\alpha})$, for $z_{1-\alpha}$ the $1-\alpha$ quantile of the standard normal.*

Proof. Define $< \cdot, \cdot >_\Sigma$ by $< x, y >_\Sigma = x'\Sigma^{-1}y$, and observe that $< \cdot, \cdot >_\Sigma$ is an inner product. The result then follows immediately from the discussion in Section 2.4.3 of Ingster and Suslina (2003), replacing all instances of the usual euclidean inner product with $< \cdot, \cdot >_\Sigma$. □

Lemma A.6. *Let $\mathcal{B}$ be a closed, convex subset of $\mathbb{R}^K$, and $\beta_A \notin \mathcal{B}$. Let $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} ||\beta - \beta_A||_\Sigma$, where $||x||_\Sigma^2 = x'\Sigma^{-1}x$ for some positive definite matrix Σ. Then for any $\beta \in \mathcal{B}$, $(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta - \tilde{\beta}) \geqslant 0$.*

Proof. Consider any $\beta \in \mathcal{B}$. Define $\beta_\theta = \theta(\beta - \tilde{\beta}) + \tilde{\beta}$, and note that since $\mathcal{B}$ is convex $\beta_\theta \in \mathcal{B}$ for any $\theta \in [0, 1]$. Further,

$$||\beta_\theta - \beta_A||_\Sigma^2 = \theta^2 ||\beta - \tilde{\beta}||_\Sigma^2 + 2\theta(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta - \tilde{\beta}) + ||\tilde{\beta} - \beta_A||_\Sigma^2.$$

Differentiating with respect to θ, we have

$$\frac{\partial}{\partial\theta}||\beta_\theta - \beta_A||_\Sigma^2 = 2\theta||\beta - \tilde{\beta}||_\Sigma^2 + 2(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta - \tilde{\beta}),$$

from which we see that the derivative evaluated at $\theta = 0$ is $2(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta_A - \tilde{\beta})$. Since $\tilde{\beta}$ minimizes the norm, it follows that we must have $2(\tilde{\beta} - \beta_A)'\Sigma^{-1}(\beta_A - \tilde{\beta}) \geqslant 0$, else we could achieve a lower value of the norm at β_θ by choosing $\theta > 0$ sufficiently small. □

Lemma A.7. *Let $\mathcal{B} = \{\beta \in \mathbb{R}^K : v'\beta \leqslant d\}$ for some $v \in \mathbb{R}^K \backslash \{0\}$ and $d \in \mathbb{R}$. Let $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} ||\beta - \beta_A||_\Sigma$ for some $\beta_A \notin \mathcal{B}$, where $||x||_\Sigma^2 = x'\Sigma^{-1}x$ and Σ is positive definite. Then $(\beta_A - \tilde{\beta})'\Sigma^{-1} = c \cdot v'$ for the positive constant $c = \frac{v'\beta_A - d}{v'\Sigma v}$.*

Proof. Note that we can form a basis $v, \tilde{v}_2, ..., \tilde{v}_K$ such that $v'\tilde{v}_j = 0$ for $j = 2, ..., K$. It follows by construction that for any $j = 2, .., K$ and any $t \in \mathbb{R}$, $\tilde{\beta} + t \cdot \tilde{v}_j \in \mathcal{B}$. Hence, from Lemma A.6, $-(\beta_A - \tilde{\beta})'\Sigma^{-1}(t\tilde{v}_j) \geqslant 0$. Since we can choose t both positive and negative, it follows that $(\beta_A - \tilde{\beta})'\Sigma^{-1}\tilde{v}_j = 0$ for all j. Since $(\beta_A - \tilde{\beta})'\Sigma^{-1}$ is orthogonal to $\{\tilde{v}_2, ..., \tilde{v}_K\}$, and $\{v, \tilde{v}_2, ..., \tilde{v}_K\}$ form a basis, we have that $(\beta_A - \tilde{\beta})'\Sigma^{-1} = c \cdot v'$, for some $c \in \mathbb{R}$. Multiplying both sides of the equation on the right by Σv, we obtain that $(\beta_A - \tilde{\beta})'v = c \cdot v'\Sigma v$. However, since $\tilde{\beta}$ is the closest point to β_A in Mahalanobis distance, it must be on the boundary of $\mathcal{B}$, and so $v'\tilde{\beta} = d$. It follows that $c = (v'\beta_A - d)/(v'\Sigma v)$, which is clearly positive since $\beta_A \notin \mathcal{B}$ and thus $v'\beta_A > d$. □

Lemma A.8 (Power of optimal test for linear subspace)**.** *Let $\mathcal{B} = \{\beta \in \mathbb{R}^K : v'\beta \leqslant d\}$ for some $v \in \mathbb{R}^K \backslash \{0\}$ and $d \in \mathbb{R}$. Suppose $\hat{\beta} \sim \mathcal{N}(\beta, \Sigma)$ for Σ positive definite known, and consider the problem of testing $H_0 : \beta \in \mathcal{B}$ against $H_A : \beta = \beta_A$ for some $\beta_A \notin \mathcal{B}$. Then the most powerful size-α test of H_0 against H_A is a one-sided t-test that rejects for large values of $v'\hat{\beta}$, and has power equal to $\Phi((v'\beta_A - d)/\sqrt{v'\Sigma v} - z_{1-\alpha})$.*

Proof. From Lemma A.5, the most powerful test rejects for large values of $(\beta_A - \tilde{\beta})'\Sigma^{-1}\hat{\beta}$, where $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} ||\beta - \beta_A||_\Sigma$, and has power $\Phi(||\beta_A - \tilde{\beta}||_\Sigma - z_{1-\alpha})$. By Lemma A.7, $(\beta_A - \tilde{\beta})'\Sigma^{-1} = cv'$, for $c = (v'\beta_A - d)/(v'\Sigma v)$. It follows that

$$\begin{aligned} ||\beta_A - \tilde{\beta}||_\Sigma^2 &= (\beta_A - \tilde{\beta})'\Sigma^{-1}(\beta_A - \tilde{\beta}) \\ &= cv'(\beta_A - \tilde{\beta}) \\ &= c(v'\beta_A - d) = (v'\beta_A - d)^2/(v'\Sigma v), \end{aligned}$$

where we use the fact that $v'\tilde{\beta} = d$, since $\tilde{\beta}$ must be on the boundary of $\mathcal{B}$, as argued in the proof to Lemma A.7. The result then follows immediately. □

Lemma A.9. *If $P \in \mathcal{P}_\epsilon$, then $\rho_\alpha^*(P, x) = \Phi\left(\frac{-\bar{\gamma}'\tilde{A}_{(\cdot,1)}x}{\sqrt{\bar{\gamma}'A\Sigma_P A'\bar{\gamma}}} - z_{1-\alpha}\right)$, where $\bar{\gamma} \in V(\Sigma_P)$ is the unique element of $V(\Sigma_P)$ with $\bar{\gamma}_{-B^*} = 0$ (see Lemma A.4).*

Proof. Suppose $\hat{\beta}_n \sim \mathcal{N}(\beta_P, \Sigma_P/n)$. Let $\mathcal{B}_n = \{\beta : \theta_P^{ub} + x/\sqrt{n} \in \mathcal{S}(\beta, \Delta)\}$ be the set of values for β consistent with the null that $\theta = \theta_P^{ub} + x/\sqrt{n}$. Observe that $\mathcal{B}_n = \{\beta : \eta(\beta, A, d, \theta^{ub} + x/\sqrt{n}, \Sigma_P) \leqslant 0\}$, where $\eta(\cdot)$ is defined in (21). From Lemma A.5, the most powerful test of $H_0 : \beta \in \mathcal{B}_n$ against $H_1 : \beta = \beta_P$ rejects for large values of $(\beta_P - \tilde{\beta})'\Sigma_P^{-1}\hat{\beta}_n$. To derive the optimal test, it is instructive to first consider a simpler testing problem. From Lemma A.4, there exists a unique $\bar{\gamma} \in V(\Sigma_P)$ such that $\bar{\gamma}_{-B*} = 0$, where B^* are the binding rows at the solution to (16) satisfying LICQ. Define $\mathcal{B}_n^{\bar{\gamma}} = \{\beta : \bar{\gamma}'(A\beta - d - \tilde{A}_{(\cdot,1)}(\theta_P^{ub} + x/\sqrt{n})) \leqslant 0\}$. We first consider testing $\tilde{H}_0 : \beta \in \mathcal{B}_n^{\bar{\gamma}}$ against $H_1 : \beta = \beta_P$. From Lemma A.7, the optimal test rejects for large values of $\bar{\gamma}' A\hat{\beta}_n$ and has power $\Phi(\frac{h}{\sqrt{\bar{\gamma}' A\Sigma_P A'\bar{\gamma}/n}} - z_{1-\alpha})$, where

$$h = \bar{\gamma}'(A\beta_P - d - \tilde{A}_{(\cdot,1)}(\theta_P^{ub} + x/\sqrt{n})). \tag{25}$$

From the definition of LICQ in direction l, however, there exists a value $\tilde{\tau}^*$ such that

$$A_{(B*,\cdot)}\beta_P - d_{B*} - \tilde{A}_{(B,1)}\theta_P^{ub} - \tilde{A}_{(B,-1)}\tilde{\tau}^* = 0 \tag{26}$$

$$A_{(-B*,\cdot)}\beta_P - d_{-B*} - \tilde{A}_{(-B,1)}\theta_P^{ub} - \tilde{A}_{(-B,1)}\tilde{\tau}^* < 0 \tag{27}$$

By construction, $\bar{\gamma}'\tilde{A}_{(\cdot,-1)} = 0$ and $\bar{\gamma}_{-B*} = 0$, which combined with the previous two displays implies that $h = -\bar{\gamma}'\tilde{A}_{(\cdot,1)}x/\sqrt{n}$, and hence the power of the optimal test of $\tilde{H}_0$ is $\Phi\left(\frac{-\bar{\gamma}'\tilde{A}_{(\cdot,1)}x}{\sqrt{\bar{\gamma}' A\Sigma_P A'\bar{\gamma}}} - z_{1-\alpha}\right)$.

To complete the proof, it thus suffices to show that the optimal test of $\tilde{H}_0$ against H_1 is the same as the optimal test of H_0 against H_1 for n sufficiently large. To this end, note that $\mathcal{B}_n \subseteq \mathcal{B}_n^{\bar{\gamma}}$, since by duality,

$$\eta(\beta, A, d, \theta^{ub}+x/\sqrt{n}, \Sigma_P) = \max_{\gamma \in V(\Sigma_P)} \gamma'(A\beta - d - \tilde{A}_{(\cdot,1)}(\theta_P^{ub}+x/\sqrt{n})) \geqslant \bar{\gamma}'(A\beta - d - \tilde{A}_{(\cdot,1)}(\theta_P^{ub}+x/\sqrt{n})).$$

Thus, Lemma A.5 implies that the optimal test under H_0 coincides with the optimal test under H_1 whenever $\tilde{\beta}_n = \arg\min_{\beta \in \mathcal{B}_n^{\bar{\gamma}}} ||\beta - \beta_P||_{\Sigma_P/n}$ is in $\mathcal{B}_n$. From Lemma A.7, however, $\tilde{\beta}_n' = \beta_P' + \frac{h}{\sqrt{\bar{\gamma}' A\Sigma_P A'\bar{\gamma}/n}} v'(\Sigma_P/n)$, for h defined in (25). Using the equality $h = -\bar{\gamma}'\tilde{A}_{(\cdot,1)}x/\sqrt{n}$ derived above, we see that

$$\tilde{\beta}_n' = \beta_P' - \frac{1}{\sqrt{n}} \frac{\bar{\gamma}'\tilde{A}_{(\cdot,1)}x}{\sqrt{\bar{\gamma}' A\Sigma_P A'\bar{\gamma}}} \bar{\gamma}' A\Sigma_P,$$

and thus we can write $\tilde{\beta} = \beta_P - \nu/\sqrt{n}$ for a finite vector ν. From Lemma A.4, every $\gamma \in V(\Sigma_P)$ with $\gamma \neq \bar{\gamma}$ has $\gamma_{-B*} \neq 0$. Since $\gamma \geqslant 0$ by construction, equations (26) and (27) imply that

$$\gamma'(A\beta_P - d - \tilde{A}_{(\cdot,1)}\theta_P^{ub}) < 0$$

for all $\gamma \neq \bar{\gamma}$, where we use the fact that $\gamma'\tilde{A}_{(\cdot,-1)} = 0$ by construction. We've shown, however, that $\bar{\gamma}'(A\beta_P - d - \tilde{A}_{(\cdot,1)}\theta_P^{ub}) = 0$. By continuity arguments, it follows that for n

sufficiently large,

$$\eta(\tilde{\beta}, A, d, \theta_P^{ub} + x/\sqrt{n}, \Sigma_P) = \max_{\gamma \in V(\Sigma_P)} \gamma'(A(\beta_P - \nu/\sqrt{n}) - d - \tilde{A}_{(\cdot,1)}(\theta_P^{ub} + x/\sqrt{n}))$$

is equal to

$$\bar{\gamma}'(A(\beta_P - \nu/\sqrt{n} - d - \tilde{A}_{(\cdot,1)}(\theta_P^{ub} + x/\sqrt{n})),$$

and thus $\tilde{\beta}_n \in \mathcal{B}_n$, as we wished to show.

□

A.3 Proofs and auxiliary lemmas for FLCIs

Proof of Proposition 4.2

Proof. First, suppose Assumption 9 holds. Without loss of generality, we show $\mathbb{P}\left((\theta^{ub} + x) \in \mathcal{C}_{\alpha,n}^{FLCI}\right) \to 0$ for any $x > 0$. By Lemma A.11 there exists $(\bar{a}, \bar{v})$ such that $\bar{b}(\bar{a}, \bar{v}) = \frac{1}{2}LID(\delta_{pre}, \Delta) =: \bar{b}_{min}$ and $\mathbb{E}_{\hat{\beta}_n \sim \mathcal{N}(\delta+\tau, \Sigma_n)}\left[\bar{a} + \bar{v}'\hat{\beta}_n\right] = \frac{1}{2}(\theta^{ub} + \theta^{lb}) =: \theta^{mid}$. Let $\bar{\mathcal{C}}_n := \bar{a} + \bar{v}'\hat{\beta}_n \pm \chi_n(\bar{a}, \bar{v})$ denote the fixed length confidence interval based on $(\bar{a}, \bar{v})$.

By construction, $\bar{\chi}_n := \chi_n(\bar{a}, \bar{v})$ is the $1 - \alpha$ quantile of the $|\mathcal{N}\left(\bar{b}_{min}, \sigma_{\bar{v},n}^2\right)|$ distribution. Since $\sigma_{\bar{v},n}^2 = \frac{1}{n}\sigma_{\bar{v},1}^2 \to 0$, the $|\mathcal{N}\left(\bar{b}_{min}, \sigma_{\bar{v},n}^2\right)|$ distribution collapses to a point mass at $\bar{b}_{min}$, and thus $\bar{\chi}_n \to \bar{b}_{min}$. By construction, the half-length of the shortest FLCI $\chi_n := \chi_n(a_n, v_n)$ must be less than or equal to $\bar{\chi}_n$, and so $\limsup_{n\to\infty} \chi_n \leqslant \bar{b}_{min}$. Let $b_n := \bar{b}(a_n, v_n)$ be the worst-case bias of the optimal FLCI. Since $\alpha \in (0, 0.5]$, Lemma A.12 implies that $\chi_n \geqslant b_n$. Additionally, Lemma A.10 implies that $b_n \geqslant \frac{1}{2}LID(\delta_{pre}, \Delta) = \bar{b}_{min}$, and thus $\chi_n \geqslant \bar{b}_{min}$. Hence, $\chi_n \to \bar{b}_{min}$ implies $b_n \to \bar{b}_{min}$. Additionally, note that for $\alpha \in (0, 0.5]$, $\chi_n(a, v)$ is increasing in both $\bar{b}(a, v)$ and $\sigma_{v,n}$. Since $\bar{b}_{min} \leqslant b_n$ and $\chi_n \leqslant \bar{\chi}_n$, it must be that $\sigma_{v_n,n} \leqslant \sigma_{\bar{v},n}$, from which it follows that $\sigma_{v_n,n} \to 0$.

Now, we claim that $\mu_n := \mathbb{E}_{\hat{\beta}_n \sim \mathcal{N}(\delta+\tau, \Sigma_n)}\left[a_n + v_n'\hat{\beta}_n\right]$ converges to $\theta^{mid} := \frac{1}{2}(\theta^{ub} + \theta^{lb})$. To show this, note that $\mu_n = a_n + v_n'\beta$ for $\beta = \delta + \tau$. Since $\theta^{ub}, \theta^{lb} \in \mathcal{S}(\beta, \Delta)$, by the definition of the identified set there exist $\delta^{ub}, \delta^{lb} \in \Delta$ and τ^{ub}, τ^{lb} such that $\beta = \delta^{ub} + \tau^{ub} = \delta^{lb} + \tau^{lb}$, $\theta^{ub} = l'\tau_{post}^{ub}$, and $\theta^{lb} = l'\tau_{post}^{lb}$. Thus, $\theta^{ub} - \mathbb{E}_{\hat{\beta}_n \sim \mathcal{N}(\beta, \Sigma_n)}\left[a_n + v_n'\hat{\beta}_n\right] = \theta^{ub} - \mu_n$ and $\mathbb{E}_{\hat{\beta}_n \sim \mathcal{N}(\beta, \Sigma_n)}\left[a_n + v_n'\hat{\beta}_n\right] - \theta^{lb} = \mu_n - \theta^{lb}$. This implies that $b_n \geqslant \max\{\theta^{ub} - \mu_n, \mu_n - \theta^{lb}\} = \bar{b}_{min} + |\mu_n - \theta^{mid}|$, where the equality uses the fact that $\theta^{ub} - \theta^{lb} = LID(\delta_{A,pre}, \Delta) = 2\bar{b}_{min}$. Since we've shown that $b_n \to \bar{b}_{min}$, it follows that $\mu_n \to \theta^{mid}$, as desired.

Next, note that if $\hat{\beta}_n \sim \mathcal{N}(\delta + \tau, \Sigma_n)$, then $a_n + v_n'\hat{\beta}_n \sim \mathcal{N}\left(\mu_n, \sigma_{v_n,n}^2\right)$. Observe that $\bar{\theta} \in \mathcal{C}_{\alpha,n}^{FLCI}$ if and only if $a_n + v_n'\hat{\beta}_n \in [\bar{\theta} - \chi_n, \bar{\theta} + \chi_n]$. Thus,

$$\mathbb{P}_{\hat{\beta}_n \sim \mathcal{N}(\beta, \Sigma_n)}\left(\bar{\theta} \in \mathcal{C}^{FLCI}_{\alpha,n}\right) = \Phi\left(\frac{\bar{\theta} + \chi_n - \mu_n}{\sigma_{v_n,n}}\right) - \Phi\left(\frac{\bar{\theta} - \chi_n - \mu_n}{\sigma_{v_n,n}}\right).$$

Now, recalling that $\theta^{ub} = \theta^{mid} + \bar{b}_{min}$ by construction, we have $\mathbb{P}_{\hat{\beta}_n \sim \mathcal{N}(\beta, \Sigma_n)}\left((\theta^{ub} + x) \in \mathcal{C}^{FLCI}_{\alpha,n}\right)$ equals

$$\Phi\left(\frac{\theta^{mid} + \bar{b}_{min} + x + \chi_n - \mu_n}{\sigma_{v_n,n}}\right) - \Phi\left(\frac{\theta^{mid} + \bar{b}_{min} + x - \chi_n - \mu_n}{\sigma_{v_n,n}}\right). \tag{28}$$

Note that the term inside the second normal CDF in the previous display equals

$$-\frac{\chi_n - b_n}{\sigma_{v_n,n}} + \frac{x + \theta^{mid} - \mu_n + \bar{b}_{min} - b_n}{\sigma_{v_n,n}}.$$

However, the first summand above is bounded between $-z_{1-\alpha/2}$ and $-z_{1-\alpha}$ by Lemma A.12. Additionally, we've shown that $\theta^{mid} - \mu_n \to 0$ and $\bar{b}_{min} - b_n \to 0$, so the numerator of the second summand converges to $x > 0$. Since the denominator $\sigma_{v_n,n} \to 0$, the expression in the previous display diverges to ∞, and hence the second normal CDF term in (28) converges to 1, which implies that $\mathbb{P}\left((\theta^{ub} + x) \in \mathcal{C}^{FLCI}_{\alpha,n}\right) \to 0$, as needed.

In order to prove the other direction, we proceed via the contrapositive. Towards this, suppose Assumption 9 fails. Let $L := LID(\delta_{pre}, \Delta)$ and $\bar{L} := \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\tilde{\delta}_{pre}, \Delta)$. By Lemma A.10, $b_n := \bar{b}(a_n, v_n) \geqslant \frac{1}{2}\bar{L} =: \bar{b}_{min}$. As argued earlier in the proof, since $\alpha \in (0, .5]$, $\chi_n \geqslant b_n \geqslant \frac{1}{2}\bar{L}$. If $\bar{L} = \infty$, then $\mathcal{C}^{FLCI}_{\alpha,n}$ is the real line, and thus never rejects, so $\mathcal{C}^{FLCI}_{\alpha,n}$ is trivially inconsistent under the assumption that $\mathcal{S}(\delta + \tau, \Delta) \neq \mathbb{R}$. For the remainder of the proof, we assume $L < \bar{L} < \infty$. From Lemma 2.1, $\mathcal{S}(\delta + \tau, \Delta) = [\theta^{lb}, \theta^{ub}]$, where $\theta^{ub} - \theta^{lb} = LID(\delta_{pre}, \Delta) = L$. Let $\epsilon = \frac{1}{4}(\bar{L} - L)$, and set $\theta^{out}_1 := \theta^{ub} + \epsilon$ and $\theta^{out}_2 := \theta^{lb} - \epsilon$. Let $\theta^{mid} = \frac{1}{2}(\theta^{ub} + \theta^{lb})$ be the midpoint of the identified set. By construction, $\theta^{out}_1 - \theta^{mid} = \theta^{mid} - \theta^{out}_2 = \frac{1}{2}L + \epsilon < \frac{1}{2}\bar{L}$. Since $\mathcal{C}^{FLCI}_{\alpha,n}$ is an interval with half-length at least $\frac{1}{2}\bar{L}$, it follows that if $\theta^{mid} \in \mathcal{C}^{FLCI}_{\alpha,n}$ then at least one of $\theta^{out}_1, \theta^{out}_2$ is also in $\mathcal{C}^{FLCI}_{\alpha,n}$. Hence, $\mathbb{P}\left(\theta^{out}_1 \in \mathcal{C}^{FLCI}_{\alpha,n}\right) + \mathbb{P}\left(\theta^{out}_2 \in \mathcal{C}^{FLCI}_{\alpha,n}\right) \geqslant \mathbb{P}\left(\theta^{mid} \in \mathcal{C}^{FLCI}_{\alpha,n}\right) \geqslant 1 - \alpha$, where the final bound follows since $\mathcal{C}^{FLCI}_{\alpha,n}$ satisfies the coverage requirement (10). It follows that $\limsup_{n\to\infty} \mathbb{P}\left(\theta^{out}_j \in \mathcal{C}^{FLCI}_{\alpha,n}\right) \geqslant \frac{1}{2}(1 - \alpha) > 0$ for at least one $j \in \{1, 2\}$. □

Lemma A.10 (Bounds for worst-case bias)**.** *For any* (a, v), $\bar{b}(a, v) \geqslant \frac{1}{2}\sup_{\delta_{pre} \in \Delta_{pre}} LID(\delta_{pre}, \Delta)$.

Proof. Since $\beta = \delta + \tau$, we can write the bias of the affine estimator $a + v'\hat{\beta}$ as $b = a + v'\delta + (v_{post} - l)'\tau_{post}$. Since τ_{post} is unrestricted in the maximization in (17), we see that the worst-case bias will be infinite if $v_{post} \neq l$ and the lemma holds trivially. We can thus restrict attention to affine estimators with $v_{post} = l$, in which case the worst-case bias reduces to

$$\bar{b}(a, v) = \sup_{\delta \in \Delta} |a + v'\delta| = \sup_{\delta \in \Delta} |a + v'_{pre}\delta_{pre} + l'\delta_{post}|. \tag{29}$$

Now, pick any $\delta^*_{pre} \in \Delta_{pre}$. First, suppose that the minimum $\left(\min_\delta l'\delta_{post},\ \text{s.t. } \delta \in \Delta, \delta_{pre} = \delta^*_{pre}\right)$ and the maximum $\left(\max_\delta l'\delta_{post},\ \text{s.t. } \delta \in \Delta, \delta_{pre} = \delta^*_{pre}\right)$ are finite. Let δ^{min} and δ^{max} be the associated solutions. By construction, $\delta^{max}_{pre} = \delta^{min}_{pre} = \delta^*_{pre}$. For any v_{pre}, we can apply the triangle inequality to show that

$$\begin{aligned}\left|a + v'_{pre}\delta^{max}_{pre} + l'\delta^{max}_{post}\right| + \left|a + v'_{pre}\delta^{min}_{pre} + l'\delta^{min}_{post}\right| &\geqslant \left|\left(a + v'_{pre}\delta^{max}_{pre} + l'\delta^{max}_{post}\right) - \left(a + v'_{pre}\delta^{min}_{pre} + l'\delta^{min}_{post}\right)\right| \\ &= |l'\delta^{max}_{post} - l'\delta^{min}_{post}| = LID(\delta^*_{pre}, \Delta).\end{aligned}$$

Note that for any $x_1, x_2 \geqslant 0$, $\max\{x_1, x_2\} \geqslant \frac{1}{2}(x_1 + x_2)$. It then follows from the previous display that

$$\max\{\left|a + v'_{pre}\delta^{max}_{pre} + l'\delta^{max}_{post}\right|, \left|a + v'_{pre}\delta^{min}_{pre} + l'\delta^{min}_{post}\right|\} \geqslant \frac{1}{2}LID(\delta^*_{pre}, \Delta).$$

Since δ^{max}_{pre} and δ^{min}_{pre} are feasible in the maximization (29), we see that $\bar{b} \geqslant \frac{1}{2}LID(\delta^*_{pre}, \Delta)$, as needed. To complete the proof, now suppose without loss of generality that

$$\left(\max_\delta l'\delta_{post},\ \text{s.t. } \delta \in \Delta, \delta_{pre} = \delta^*_{pre}\right) = \infty.$$

Then, we can replay the argument above replacing δ^{max} with a sequence of values $\{\delta_j\}$ such that $l'\delta_j$ diverges, which gives that $\bar{b}$ is infinite and the result follows. □

Lemma A.11. *Suppose Δ is convex, and there exists $\delta \in \Delta$ such that $LID(\delta_{pre}, \Delta) = \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\tilde{\delta}_{pre}, \Delta) < \infty$. Then there exists (a, v) such that $\bar{b}(a, v) = \frac{1}{2}\sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\tilde{\delta}_{pre}, \Delta)$. Additionally, for any τ and Σ_n, $\mathbb{E}_{\hat{\beta}_n \sim \mathcal{N}(\delta+\tau, \Sigma_n)}\left[a + v'\hat{\beta}_n\right] = \frac{1}{2}(\theta^{ub} + \theta^{lb})$, where θ^{ub} and θ^{lb} are the upper and lower bounds of the identified set $\mathcal{S}(\delta + \tau, \Delta)$.*

Proof. Let $b^{max}(\delta^*_{pre}) := \left(\max_{\tilde{\delta}} l'\tilde{\delta}_{post},\ \text{s.t. } \tilde{\delta} \in \Delta, \tilde{\delta}_{pre} = \delta^*_{pre}\right)$, where we define $b^{max} = -\infty$ if $\delta^*_{pre} \notin \Delta_{pre}$. Likewise, define $b^{min}(\delta^*_{pre}) := \left(\min_{\tilde{\delta}} l'\tilde{\delta}_{post},\ \text{s.t. } \tilde{\delta} \in \Delta, \tilde{\delta}_{pre} = \delta^*_{pre}\right)$, where we define $b^{min} = \infty$ if $\delta^*_{pre} \notin \Delta_{pre}$. Note that Δ convex implies that b^{max} is concave and b^{min} is convex. Thus, $-LID(\delta^*_{pre}) = b^{min}(\delta^*_{pre}) - b^{max}(\delta^*_{pre})$ is convex (where we define $LID(\delta^*_{pre}) = -\infty$ if $\delta^*_{pre} \notin \Delta_{pre}$). The domain of $-LID(\delta^*_{pre})$ (i.e. the set of values for which it is finite) is Δ_{pre}, since it is infinite for $\delta^*_{pre} \notin \Delta_{pre}$ by construction, and by assumption, $LID(\delta^*_{pre})$ is finite for all $\delta^*_{pre} \in \Delta_{pre}$. Since Δ is assumed to be convex, Δ_{pre} is a non-empty convex set, and thus has non-empty relative interior, so the relative interior of the domain of $-LID$ is non-empty.[41] It follows from Theorem 8.2 in Mau Nam (2019) that $\partial(-LID) = \partial(-b^{max}) + \partial(b^{min})$ where for a convex function f, ∂f is the subdifferential

[41]The relative interior of a set is the interior of the set relative to its affine hull. See, e.g., Mau Nam (2019), Chapter 5.

$\partial f(\bar{x}) := \{v \,:\, f(\bar{x}) + v'(x - \bar{x}) \leqslant f(x), \forall x\}$ and $\partial(-b^{max}) + \partial(b^{min})$ is the Minkowski sum of the two subdifferentials.

Additionally, if $LID(\delta_{pre}) = \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\tilde{\delta}_{pre})$, then $-LID(\delta_{pre}) = \inf_{\tilde{\delta}_{pre} \in \Delta_{pre}} -LID(\tilde{\delta}_{pre})$. Thus, standard results in convex analysis (see, e.g., Theorem 16.2 in Mau Nam (2019)) give that $0 \in \partial(-LID)(\delta_{pre}) + N(\Delta; \delta_{pre})$, where $N(\Delta; \delta_{pre}) = \{v_{pre} \,:\, v'_{pre}(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant 0, \forall \tilde{\delta}_{pre} \in \Delta_{pre}\}$ is the normal cone to Δ_{pre} at δ_{pre}. Hence, there exist vectors $\bar{v}_{min}, \bar{v}_{max}$ such that for all $\tilde{\delta}_{pre} \in \Delta_{pre}$,

$$b^{min}(\delta_{pre}) + \bar{v}'_{min}(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant b^{min}(\tilde{\delta}_{pre}) \tag{30}$$

$$-b^{max}(\delta_{pre}) + \bar{v}'_{max}(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant -b^{max}(\tilde{\delta}_{pre}) \tag{31}$$

$$-(\bar{v}_{min} + \bar{v}_{max})'(\tilde{\delta}_{pre} - \delta_{pre}) \leqslant 0. \tag{32}$$

The inequalities (31) and (32) together imply that for all $\tilde{\delta}_{pre} \in \Delta_{pre}$,

$$b^{max}(\delta_{pre}) + \bar{v}'_{min}(\tilde{\delta}_{pre} - \delta_{pre}) \geqslant b^{max}(\tilde{\delta}_{pre}). \tag{33}$$

Now, let v be the vector such that $v_{post} = l$ and $v_{pre} = -\bar{v}_{min}$. Observe that

$$\begin{aligned} \max_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} &= \max_{\tilde{\delta}_{pre} \in \Delta_{pre}} \left(a + v'_{pre}\tilde{\delta}_{pre} + \max_{\bar{\delta} \in \Delta, \bar{\delta}_{pre} = \tilde{\delta}_{pre}} l'\bar{\delta}_{post} \right) \\ &= \max_{\tilde{\delta}_{pre} \in \Delta_{pre}} a + v'_{pre}\tilde{\delta}_{pre} + b^{max}(\tilde{\delta}_{pre}) \\ &\leqslant a + v'_{pre}\delta_{pre} + b^{max}(\delta_{pre}), \end{aligned} \tag{34}$$

where the first equality nests the maximization, the second equality uses the definition of b^{max}, and the inequality follows from (33). An analogous argument using (30) yields that

$$\begin{aligned} \min_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} &= \min_{\tilde{\delta}_{pre} \in \Delta_{pre}} a + v'_{pre}\tilde{\delta}_{pre} + b^{min}(\tilde{\delta}_{pre}) \\ &\geqslant a + v'_{pre}\delta_{pre} + b^{min}(\delta_{pre}). \end{aligned} \tag{35}$$

Now, it is apparent from equation (29) that

$$\bar{b}(a, v) = \max \left\{ \left| \max_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} \right|, \left| \min_{\tilde{\delta} \in \Delta} a + v'_{pre}\tilde{\delta}_{pre} + l'\tilde{\delta}_{post} \right| \right\}, \tag{36}$$

which is bounded above by $\max\left\{a + v'_{pre}\delta_{pre} + b^{max}(\delta_{pre}), -(a + v'_{pre}\delta_{pre} + b^{min}(\delta_{pre}))\right\}$ from the results above. Setting $a = -v'_{pre}\delta_{pre} - \frac{1}{2}(b^{max}(\delta_{pre}) + b^{min}(\delta_{pre}))$, this upper bound reduces to $\frac{1}{2}(b^{max}(\delta_{pre}) - b^{min}(\delta_{pre}))$. Since $LID(\delta_{pre}, \Delta) = b^{max}(\delta_{pre}) - b^{min}(\delta_{pre})$ and

$LID(\delta_{pre}, \Delta) = \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\tilde{\delta}_{pre}, \Delta)$ by assumption, it is then immediate that $\bar{b} \leqslant \frac{1}{2} \sup_{\tilde{\delta}_{pre} \in \Delta_{pre}} LID(\tilde{\delta}_{pre}, \Delta)$. The inequality in the opposite direction follows from Lemma A.10. Finally, substituting in the definition of a and v above and simplifying, we see that $\mathbb{E}_{\hat{\beta}_n \sim \mathcal{N}(\delta+\tau, \Sigma_n)} \left[a + v' \hat{\beta}_n \right] = l' \beta_{post} - \frac{1}{2}(b^{max}(\delta_{pre}) + b^{min}(\delta_{pre}))$, which from (5) and (6) we see is the midpoint of the identified set. □

Lemma A.12. *Let χ_α be the $1-\alpha$ quantile of the $|\mathcal{N}(b, \sigma^2)|$ distribution for $b \geqslant 0$. Then $b + \sigma z_{1-\alpha} \leqslant \chi_\alpha \leqslant b + \sigma z_{1-\alpha/2}$.*

Proof. Since $|\xi| \geqslant \xi$, we have that $q_{1-\alpha}(|\xi| \,|\, \xi \sim \mathcal{N}(b, \sigma^2)) \geqslant q_{1-\alpha}(\xi \,|\, \xi \sim \mathcal{N}(b, \sigma^2)) = b + \sigma z_{1-\alpha}$, which yields the first inequality. For the second inequality, observe that

$$\begin{aligned} q_{1-\alpha}(|\xi| \,|\, \xi \sim \mathcal{N}\left(b, \sigma^2\right)) &= q_{1-\alpha}(|\xi + b| \,|\, \xi \sim \mathcal{N}\left(0, \sigma^2\right)) \\ &\leqslant b + q_{1-\alpha}(|\xi| \,|\, \xi \sim \mathcal{N}\left(0, \sigma^2\right)) = b + \sigma z_{1-\alpha/2} \end{aligned}$$

where the first inequality uses the triangle inequality, and the final equality uses the fact that a mean-zero normal distribution is symmetric about 0. □

B Additional Simulation Results

This section contains additional simulation results that complement the simulations presented in the main text. Section B.1 describes the computation of the optimal bound for expected excess length. Section B.2 contains additional results from the normal data-generating process considered in the main text. Section B.3 presents results from a non-normal data-generating process in which the covariance matrix is estimated from the data, which show that our proposed procedures have (approximate) size control, with similar power curves to those in the normal simulations.

B.1 Optimal bounds on excess length

We now discuss the computation of optimal bounds on the excess length of confidence intervals that satisfy the uniform coverage requirement (10). In Section 5, we benchmark the performance of our proposed procedures in Monte Carlo simulations relative to these bounds.

The following result restates Theorem 3.2 of Armstrong and Kolesár (2018) in the notation of our paper, which provides a formula for the optimal expected length of a confidence set that satisfies the uniform coverage requirement.

Lemma B.1. *Suppose that Δ is convex. Let $\mathcal{I}_\alpha$ denote the set of confidence sets that satisfy the coverage requirement (10). Then, for any $\delta^* \in \Delta$, $\tau^*_{post} \in \mathbb{R}^{\bar{T}}$, and Σ_n positive definite,*

$$\inf_{\mathcal{C}\in\mathcal{I}_\alpha} \mathbb{E}_{\hat{\beta}_n\sim\mathcal{N}(\delta^*+L_{post}\tau^*,\Sigma_n)}\left[\lambda(\mathcal{C})\right] = (1-\alpha)\mathbb{E}\left[\bar{\omega}(z_{1-\alpha}-Z) - \underline{\omega}(z_{1-\alpha}-Z)\,|\,Z < z_{1-\alpha}\right],$$

where $Z \sim \mathcal{N}(0,1)$, $z_{1-\alpha}$ is the $1-\alpha$ quantile of Z, and

$$\bar{\omega}(b) := \sup\{l'\tau \,|\, \tau \in \mathbb{R}^{\bar{T}}, \exists \delta \in \Delta \text{ s.t. } \|\delta + L_{post}\tau - \beta^*\|^2_{\Sigma_n} \leqslant b^2\}$$
$$\underline{\omega}(b) := \inf\{l'\tau \,|\, \tau \in \mathbb{R}^{\bar{T}}, \exists \delta \in \Delta \text{ s.t. } \|\delta + L_{post}\tau - \beta^*\|^2_{\Sigma_n} \leqslant b^2\},$$

for $\beta^* := \delta^* + L_{post}\tau^*_{post}$, and $||x||_\Sigma = x'\Sigma^{-1}x$.

The proof of this result follows from observing that the confidence set that optimally directs power against $(\delta^*, \tau^*_{post})$ inverts Neyman-Pearson tests of $H_0 : \delta \in \Delta, \theta = \bar{\theta}$ against $H_A : (\delta, \tau_{post}) = (\delta^*, \tau^*_{post})$ for each value $\bar{\theta}$. The formulas above are then obtained by integrating one minus the power function of these tests over $\bar{\theta}$. By the same argument, the optimal excess length for confidence sets that control size is the integral of one minus the power function over all points $\bar{\theta}$ outside of the identified set. Additionally, for any value $\bar{\theta} \in \mathcal{S}(\beta, \Delta)$, the null and alternative hypotheses are observationally equivalent, and so the most powerful test trivially has size α. It follows that the lowest achievable expected excess length is $(1-\alpha)\cdot LID(\delta^*_{pre}, \Delta)$ shorter than the lowest achievable expected length, where as in Section 4, LID denotes the length of the identified set.

Corollary B.1. *Under the conditions of Lemma B.1,*

$$\inf_{\mathcal{C}\in\mathcal{I}_\alpha} \mathbb{E}_{\hat{\beta}_n\sim\mathcal{N}(\beta^*,\Sigma_n)}\left[EL(\mathcal{C};\beta^*)\right] = \inf_{\mathcal{C}\in\mathcal{I}_\alpha} \mathbb{E}_{\hat{\beta}_n\sim\mathcal{N}(\beta^*,\Sigma_n)}\left[\lambda(\mathcal{C})\right] - (1-\alpha)LID(\beta^*, \Delta),$$

where $EL(\mathcal{C};\beta) = \lambda(\mathcal{C}\backslash\mathcal{S}(\beta,\Delta))$ is the excess length of the confidence set $\mathcal{C}$, i.e. the length of the part of the confidence set that falls outside of the identified set.

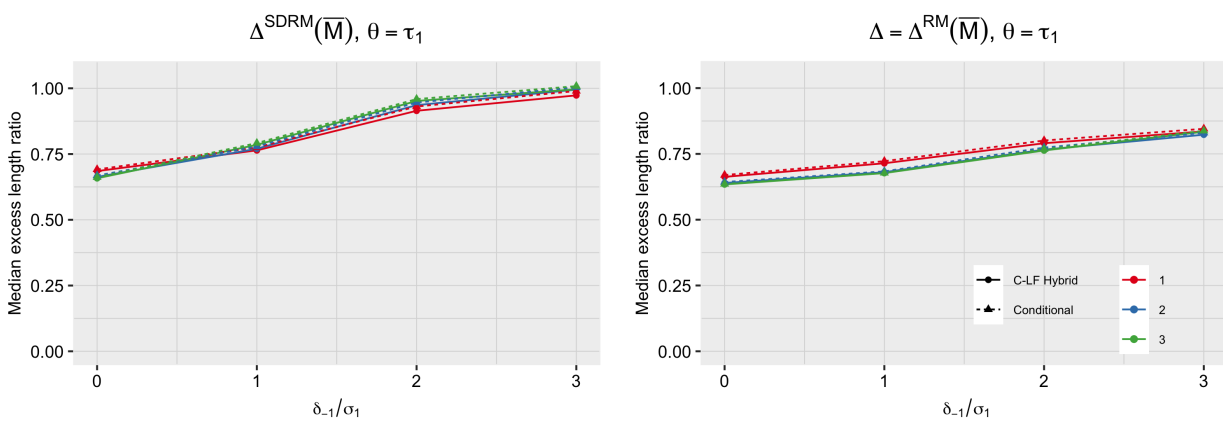
Recall that when Δ is the union of polyhedra ($\Delta = \bigcup_{k=1}^K \Delta_k$), the identified set is the union of the identified sets for each of the Δ_k. Thus, any $\mathcal{C}_\alpha$ that satisfies (10) for Δ must also satisfy (10) for each Δ_k. It follows that the expected excess length for $\mathcal{C}$ is bounded below by the optimal excess length for confidence sets satisfying (10) for Δ_k for each k. For Δs that are unions of polyhedra, we therefore use the largest lower bound implied by the individual Δ_k, which is a potentially non-sharp lower bound on the excess length of a procedure that satisfies (10) for Δ.

B.2 Additional Results for Normal Simulations

In the main text, we report efficiency in terms of excess length for the parameter $\theta = \tau_1$ for $\Delta^{SD}(M)$, $\Delta^{SDPB}(M)$, $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$. In this section, we provide additional simulation results.

Alternative choices of $\bar{M}$ for $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$. The main text reports efficiency in terms of excess length over $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$ for $\bar{M} = 1$. We now report additional results for $\bar{M} = 1, 2, 3$. The results are qualitatively similarly, suggesting that the choice of $\bar{M}$ does not appear to have a large effect on the performance of our proposed procedures.

Figure I1: $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$: Median efficiency ratios for proposed procedures when $\theta = \tau_1$ as $\bar{M}$ varies.

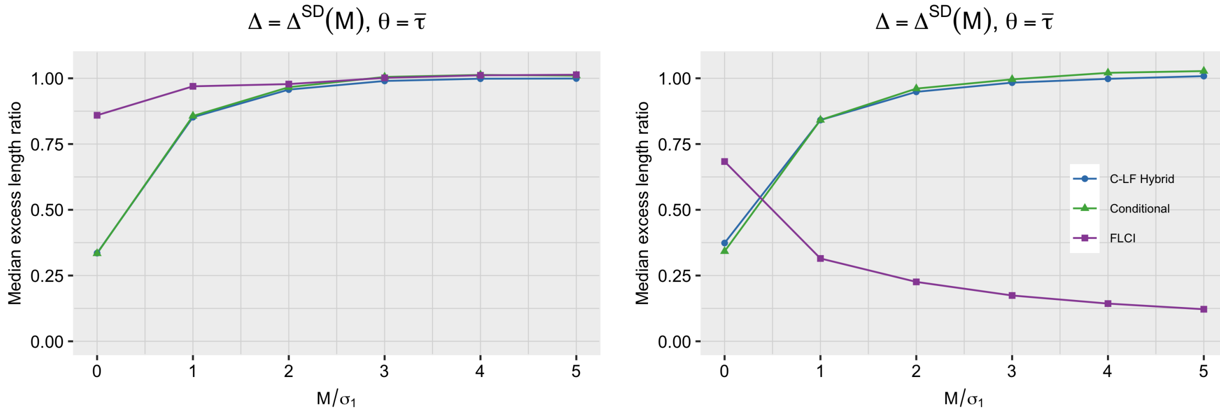


Note: This figure shows the median efficiency ratio for our proposed confidence sets for $\theta = \tau_1$ over $\Delta^{SDRM}(\bar{M})$, $\Delta^{RM}(\bar{M})$ and $\bar{M} = 1, 2, 3$. The efficiency ratio for a procedure is defined as the excess length bound divided by the procedure's expected excess length. The results for $\bar{M} = 1$ are plotted in red, $\bar{M} = 2$ are plotted in blue, and $\bar{M} = 3$ are plotted in green. The results for the conditional-least favorable hybrid confidence set ("C-LF Hybrid") are plotted in the solid line with circles. The results for the conditional confidence set are plotted in the dashed line with triangles. Results are averaged over 1000 simulations for each of the 12 papers surveyed, and the median across papers is reported here.

Alternative choice of target parameter. The main text reports efficiency in terms of excess length for the parameter $\theta = \tau_1$. We now report additional results using the average of post-period treatment effects, $\theta = \bar{\tau}_{post}$, as the target parameter.

Figure I2 plots the efficiency results for $\theta = \bar{\tau}_{post}$ over $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$. As in the main text, we conduct these simulations under the assumption of parallel trends and zero treatment effects (i.e., $\beta = 0$), reporting results as M/σ_1 varies.

Figure I2: Median efficiency ratios for $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$ when $\theta = \bar{\tau}_{post}$.

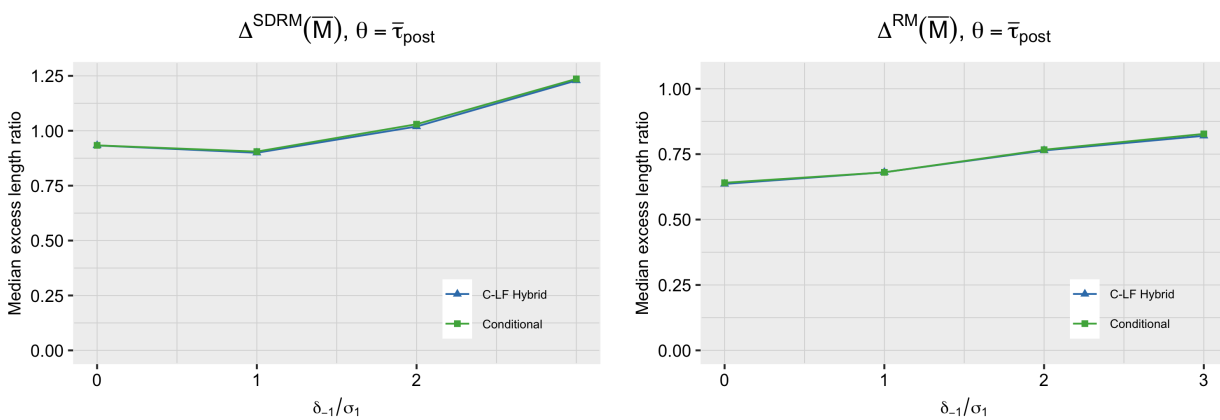


Note: This figure shows the median efficiency ratios for our proposed confidence sets for $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$ when $\theta = \bar{\tau}_{post}$. The efficiency ratio for a procedure is defined as the optimal bound divided by the procedure's expected excess length. The results for the FLCI are plotted in purple, the results for the conditional-LF ("C-LF Hybrid") in blue, and the results for the conditional confidence set are in green. Results are averaged over 1000 simulations for each of the 12 papers surveyed, and the median across papers is reported here.

Figure I3 plots the efficiency results for $\theta = \bar{\tau}_{post}$ over $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$. As in the main text, we conduct these simulations under the assumption of zero treatment effects and a "pulse" pre-trend (i.e., $\beta_{-1} = \delta_{-1}$ and $\beta_t = 0$ for all $t \neq -1$), reporting results for $\bar{M} = 1$ over $\delta_{-1}/\sigma_1 = 0, 1, 2, 3$.[42]

[42] We note that over $\Delta^{SDRM}(\bar{M})$ the median efficiency ratio for our proposed confidence sets is larger than one for $\bar{M} = 3$. For $\bar{M} = 3$, the length of the identified set for $\theta = \bar{\tau}_{post}$ can be quite large when there are many post-treatment periods (e.g., as mentioned in the main text, 5 papers in the survey have $\bar{T} > 10$), and so this behavior occurs due to computational constraints on the grid size for the underlying test inversion.

Figure I3: Median efficiency ratios for $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$ when $\theta = \bar{\tau}_{post}$.



Note: This figure shows the median efficiency ratios for our proposed confidence sets for $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$ when $\theta = \bar{\tau}_{post}$ and $\bar{M} = 1$. The efficiency ratio for a procedure is defined as the optimal bound divided by the procedure's expected excess length. The results for the conditional-least favorable ("C-LF") hybrid in blue and the results for the conditional confidence set in green. Results are averaged over 1000 simulations for each of the 12 papers surveyed, and the median across papers is reported here.

B.3 Non-normal simulation results with estimated covariance matrix

In the main text, we presented simulations results where $\hat{\beta}$ is normally distributed and its covariance matrix is treated as known. In this section, we present Monte Carlo results using a data-generating process in which $\hat{\beta}$ is not normally distributed and the covariance matrix is estimated from the data. Specifically, we consider simulations based on the empirical distribution in Bailey and Goodman-Bacon (2015). We find that all of our procedures achieve (approximate) size control, and our results on the relative power of the various procedures are quite similar to those presented in the main text.

B.3.1 Simulation design

The simulations are calibrated using the empirical distribution of the data in Bailey and Goodman-Bacon (2015).[43] Let $\hat{\beta}$, $\hat{\Sigma}$ denote the original, estimated event-study coefficients and variance-covariance matrix from the event-study regression in the paper. We simulate data using a clustered bootstrap sampling scheme at the county level (i.e. the level of clustering used by the authors in their event-study regression). For each bootstrap sample b, we re-estimate the event-study coefficients $\hat{\beta}_b$ and the variance-covariance matrix $\hat{\Sigma}_b$ also using the clustering scheme specified by the authors. We then re-center the bootstrapped coefficient so that under our simulated data-generating process either parallel trends holds (i.e., $\hat{\beta}_b^{centered} = \hat{\beta}_b - \hat{\beta}$) or the "pulse" pre-trend holds (i.e., $\hat{\beta}_b^{centered} = \hat{\beta}_b - \hat{\beta} + \delta_{-1} * e_{-1}$ where e_{-1} is the $(\underline{T} + \bar{T})$-dimensional vector with one in $t = -1$ entry and zeroes everywhere else). We construct our proposed confidence sets for bootstrap draw b using the pair $(\hat{\beta}_b^{centered}, \hat{\Sigma}_b)$.

As in the main text, we focus on the performance of our proposed confidence sets for $\Delta^{SD}(M)$, $\Delta^{SDPB}(M)$ under parallel trends and $\Delta^{SDRM}(\bar{M})$, $\Delta^{RM}(\bar{M})$ under the "pulse" pre-trend. The parameter of interest in these simulations is the causal effect in the first post-period ($\theta = \tau_1$). For $\Delta^{SD}(M)$ and $\Delta^{SDPB}(M)$, we report the performance of the FLCI, conditional confidence set, and conditional-least favorable confidence set. For $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$, we report the performance of the conditional confidence set and the conditional-least favorable confidence set. All results are averaged over 1000 bootstrap samples.

[43]Since implementing the bootstrap in practice is logistically challenging, we do so for one paper rather than the full 12 papers in the survey. We chose the first paper alphabetically to minimize concerns about cherry-picking.

B.3.2 Size control simulations

Table 2 reports the maximum rejection rate of each procedure over a grid of parameter values θ within the identified set $\mathcal{S}(\beta, \Delta)$ for $\Delta = \Delta^{SD}(M)$ and $\Delta = \Delta^{SDPB}(M)$ under parallel trends (i.e., $\beta = 0$). We report results for $M/\sigma_1 = 0, 1, 2, 3, 4, 5$. The table shows that all our procedures approximately control size, with null rejection rates not exceeding 0.08.

Δ	M/σ_1	Conditional	FLCI	C-LF Hybrid
$\Delta^{SD}(M)$				
	0	0.073	0.078	0.069
	1	0.046	0.061	0.044
	2	0.038	0.072	0.037
	3	0.040	0.072	0.038
	4	0.049	0.072	0.045
	5	0.059	0.072	0.051
$\Delta^{SDPB}(M)$				
	0	0.079	0.078	0.074
	1	0.052	0.047	0.048
	2	0.046	0.055	0.042
	3	0.051	0.058	0.046
	4	0.055	0.058	0.051
	5	0.059	0.058	0.057

Table 2: Maximum null rejection probability over the identified set $\mathcal{S}(\beta, \Delta)$ for $\Delta = \Delta^{SD}(M)$ and $\Delta = \Delta^{SDPB}(M)$ under parallel trends (i.e., $\beta = 0$) using the empirical distribution from Bailey and Goodman-Bacon (2015).

Table 3 reports the maximum rejection rate of the conditional test and the conditional-least favorable test over a grid of parameter values θ within the identified set $\mathcal{S}(\beta, \Delta)$ for $\Delta = \Delta^{SDRM}(\bar{M})$ and $\Delta = \Delta^{RM}(\bar{M})$ under the "pulse" pre-trend (i.e., $\beta_{-1} = \delta_{-1}$ and $\beta_t = 0$ for all $t \neq -1$). We report results for $\bar{M} = 1$ and $\delta_{-1}/\sigma_1 = 1, 2, 3$. The table shows that all our procedures approximately control size, with worst-case null rejection probability of 0.058.

B.3.3 Comparison with normal simulations

We next compare results from the non-normal simulations with estimated covariance discussed above to the normal model simulations the main text, in which $\hat{\beta}$ is normal and Σ is treated as known.

Figures I4-I5 shows the rejection probabilities at different values of the parameter θ using both simulation methods for $\Delta^{SD}(M)$, $\Delta^{SDPB}(M)$ at $M/\sigma_1 = 0, 5$ respectively. The results are quite similar for all values of M/σ_1 considered, and we thus omit the intermediate values.

Δ	δ_{-1}/σ_1	Conditional	C-LF Hybrid
$\Delta^{SDRM}(\bar{M})$			
	1	0.009	0.008
	2	0.037	0.035
	3	0.058	0.054
$\Delta^{RM}(\bar{M})$			
	1	0.005	0.005
	2	0.017	0.016
	3	0.024	0.023

Table 3: Maximum null rejection probability over the identified set $\mathcal{S}(\beta, \Delta)$ for $\Delta = \Delta^{SDRM}(\bar{M})$ and $\Delta = \Delta^{RM}(\bar{M})$ under the "pulse" pre-trend (i.e., $\beta_{-1} = \delta_{-1}$ and $\beta_t = 0$ for all $t \neq -1$) and $\bar{M} = 1$ using the empirical distribution from Bailey and Goodman-Bacon (2015). We report results for $\delta_{-1}/\sigma_1 = 1, 2, 3$.

The estimated average rejection rates of each procedure are quite similar in the non-normal simulations and the normal simulations across each choice of Δ. As a result, the relative rankings of the procedures in terms of power are the same in the non-normal simulations as in the normal simulations discussed in the main text. Similarly, Figures I6-I7 shows the rejection probabilities at different values of the parameter θ using both simulation methods for $\Delta^{SDRM}(\bar{M})$, $\Delta^{RM}(\bar{M})$ at $\delta_{-1}/\sigma_1 = 1, 2, 3$ respectively and $\bar{M} = 1$.

Figure I4: Comparison of rejection probabilities using bootstrap and normal simulations. Results are shown for $\theta = \tau_1$, and each choice of $\Delta = \Delta^{SD}(M), \Delta^{SDPB}(M)$, and $M/\sigma_1 = 0$. The average rejection rate for the non-normal simulations are in red and the average rejection rate for the normal simulations are in blue; the dashed black lines indicate the identified set bounds. Results are averaged over 1000 simulations.

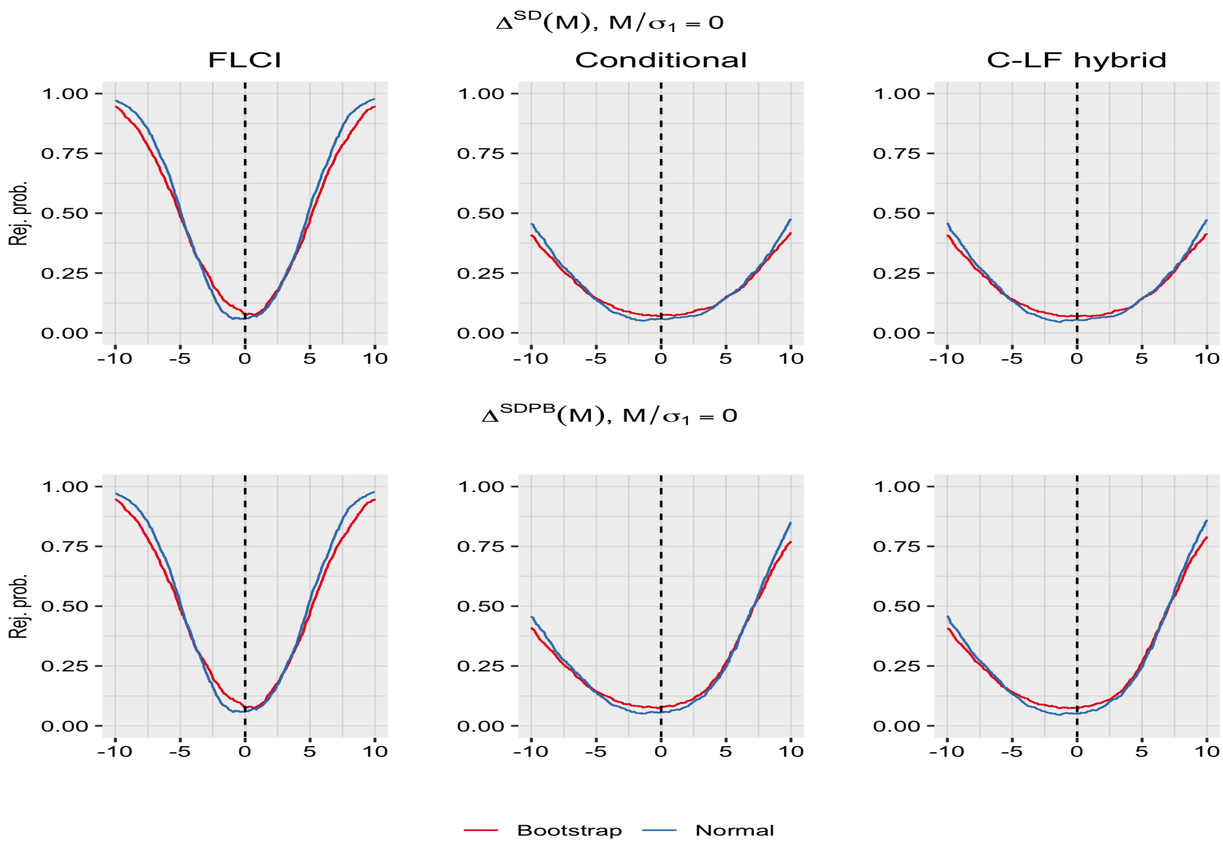


Figure I5: Comparison of rejection probabilities using bootstrap and normal simulations. Results are shown for $\theta = \tau_1$, and each choice of $\Delta = \Delta^{SD}(M), \Delta^{SDPB}(M)$, and $M/\sigma_1 = 5$. The average rejection rate for the non-normal simulations are in red and the average rejection rate for the normal simulations are in blue; the dashed black lines indicate the identified set bounds. Results are averaged over 1000 simulations.

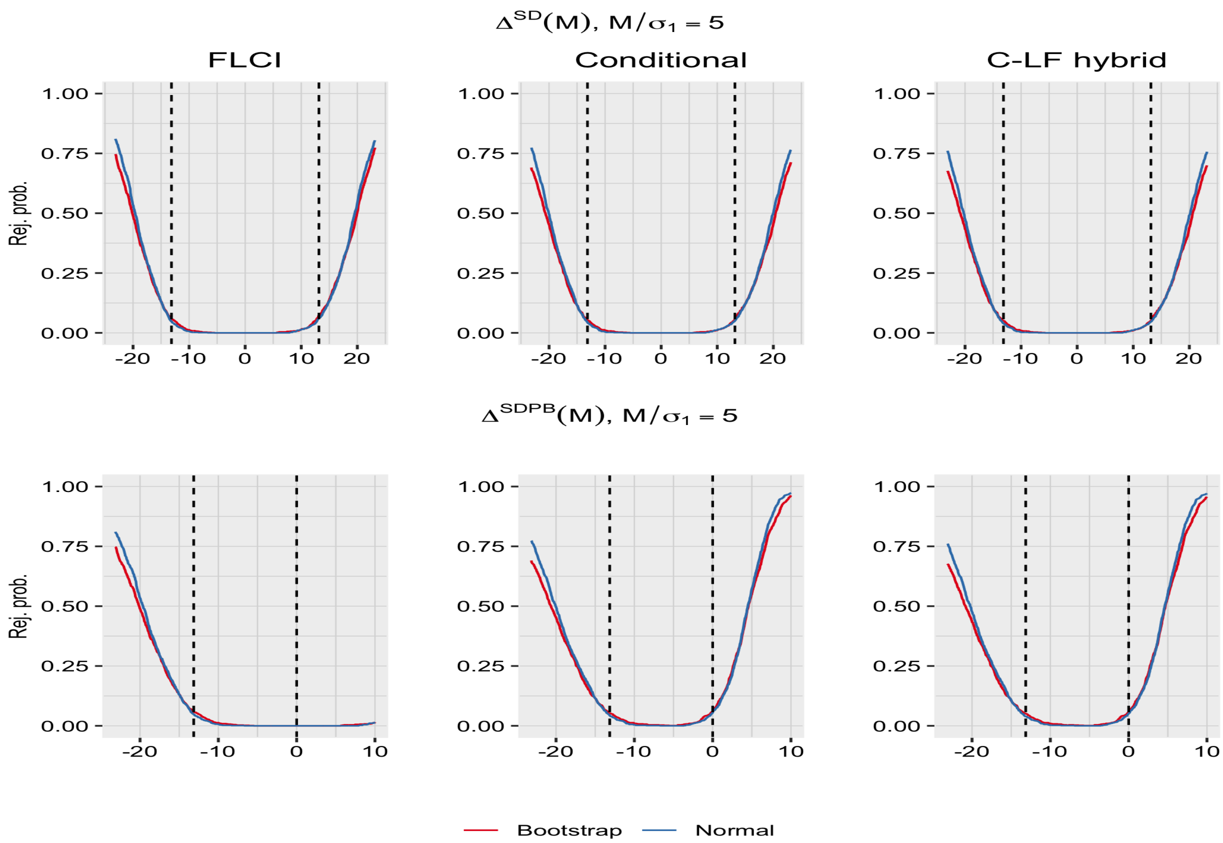


Figure I6: Comparison of rejection probabilities using bootstrap and normal simulations for $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$. Results are shown for $\theta = \tau_1$, $\bar{M} = 1$ and $\delta_{-1}/\sigma_1 = 1$. The average rejection rate for the non-normal simulations are in red and the average rejection rate for the normal simulations are in blue; the dashed black lines indicate the identified set bounds. Results are averaged over 1000 simulations.

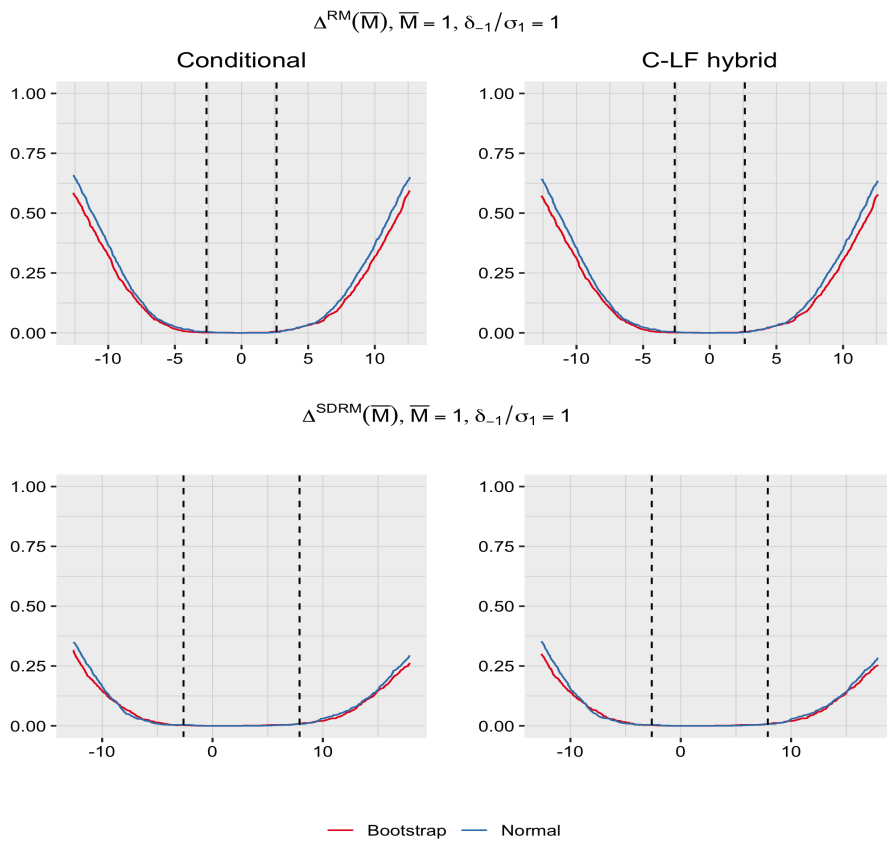
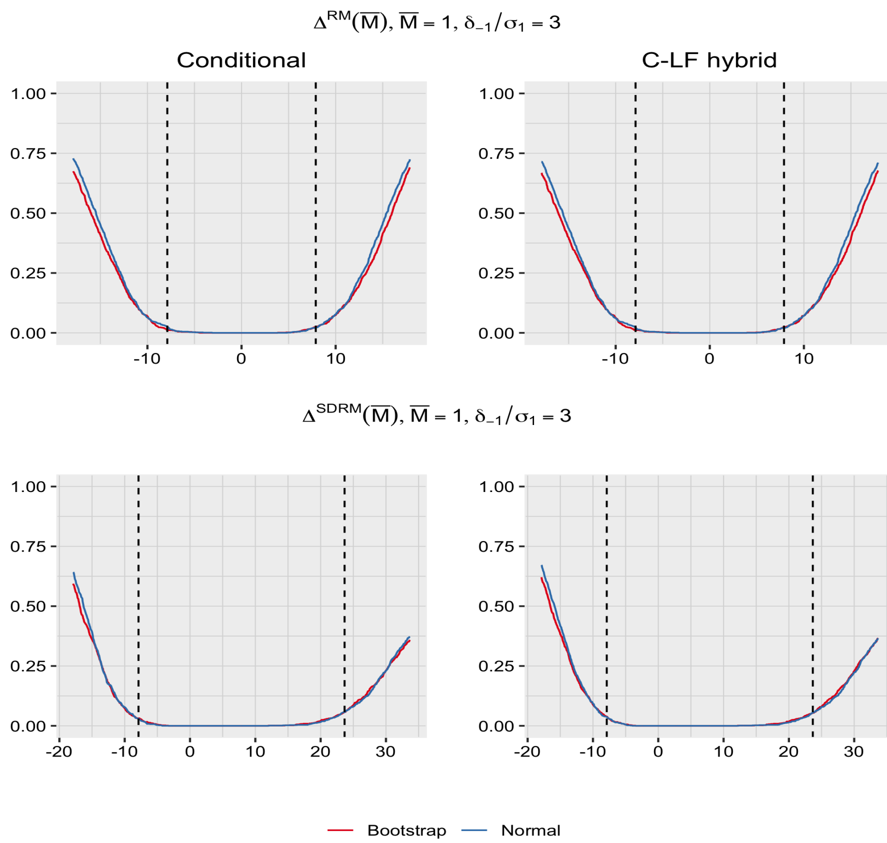


Figure I7: Comparison of rejection probabilities using bootstrap and normal simulations for $\Delta^{SDRM}(\bar{M})$ and $\Delta^{RM}(\bar{M})$. Results are shown for $\theta = \tau_1$, $\bar{M} = 1$ and $\delta_{-1}/\sigma_1 = 3$. The average rejection rate for the non-normal simulations are in red and the average rejection rate for the normal simulations are in blue; the dashed black lines indicate the identified set bounds. Results are averaged over 1000 simulations.



C Comparing Confidence Sets to Empirical Analogs to the Identified Set

In this section we report the estimated identified set, $\mathcal{S}(\hat{\beta}, \Delta)$, for our applications, in addition to the confidence sets described in the paper. We show below that the estimated identified set is respectively Hausdorff consistent and inner consistent for the true identified set for the Benzarti and Carloni (2019) and Lovenheim and Willen (2019) applications. Thus, comparing the estimated identified set to the confidence sets is informative about the extent to which the width of the confidence sets is driven by sampling uncertainty versus the width of the identified set.

Empirical Results. Figure I8 shows the estimated identified sets for the Benzarti and Carloni application. The figure suggests that both sampling uncertainty and the length of the identified set can be important. In the left panel, for example, the confidence set is 78% larger than the estimated identified set for $\bar{M} = 0.5$. This number decreases as $\bar{M}$ increases (and the estimated identified set becomes longer), reaching about 40% at $\bar{M} = 2$.

Figure I9 shows the estimated identified sets for the Lovenheim and Willen application. We note that the estimated identified set is empty for values of M close to zero.[44] Intuitively, the estimated identified set can be empty when $\Delta = \Delta^{SD}(M)$ for $M \approx 0$, since if $\delta \in \Delta^{SD}(M)$ for $M \approx 0$, the true pre-trend β_{pre} must be very close to linear. However, even if β_{pre} is exactly linear, the sample analog $\hat{\beta}_{pre}$ will typically be non-linear owing to sampling variation, leading the estimated identified set to be empty. This occurs for values of M less than 1.49 for the specification for men and 2.01 for the specification for women, as shown in Figure I9. However, we note that when $\Delta = \Delta^{SD}(M)$, the length of the true identified set (assuming it is non-empty) is only a function of M and not of β: specifically, for $\theta = \tau_t$, the identified set has length $t(t+1)M$. We can thus compare the length of our confidence set relative to the length of the true identified set for each value of M (even though the true identified set is unknown). When $M = 0.005$, for example, the confidence sets are 3.96 and 2.59 times longer than the length of the identified set for women and men, respectively, suggesting that the sampling variation is large relative to the length of the identified set. The relative importance of the sampling variation decreases as M increases, and thus the identified set widens. When $M = 0.04$, for example, these ratios are 1.59 and 1.43, and they decrease to 1.28 and 1.26 when $M = 0.1$.

[44] We expand the range of the x-axis relative to the main text to show larger values of M, where the estimated identified set is non-empty.

Consistency proofs. We next formally establish the Hausdorff and inner consistency of the estimated identified set in our applications. We first review the notions of Hausdorff and inner consistency for set estimators. For a point $x \in \mathbb{R}$ and set $B \subseteq \mathbb{R}$, let $d(x, B) := \inf_{b \in B} |x - b|$. For two non-empty sets B and C, the Hausdorff metric is then defined as

$$d_H(B, C) := \max\{ \sup_{b \in B} d(b, C), \sup_{c \in C} d(c, B) \},$$

and we define $d_H(B, C) = \infty$ if either A or B is empty. An estimator $\hat{B}$ is said to be *Hausdorff consistent* for a non-empty set B if $d_H(\hat{B}, B) \to_p 0$. Likewise, we say that $\hat{B}$ is *inner consistent* for B if $\sup_{b \in \hat{B}} d(b, B) \to_p 0$, where $\sup_{b \in \hat{B}} d(b, B) = 0$ if $\hat{B} = \varnothing$. Intuitively, Hausdorff consistency requires that asymptotically every point in $\hat{B}$ is close to a point in B and vice versa; whereas inner consistency requires only that every point in $\hat{B}$ is close to a point in B.

Proposition C.1. *Suppose $\theta = \tau_t$ for some $t = 1, \ldots, \bar{T}$ or $\theta = \bar{T}^{-1}(\tau_1 + ... + \tau_{\bar{T}})$. Assume that $\hat{\beta} \to_p \beta$, where β satisfies (3) with $\delta \in \Delta$.*

1. *If $\Delta = \Delta^{RM}(\bar{M})$, then $\mathcal{S}(\hat{\beta}, \Delta)$ is Hausdorff consistent for $\mathcal{S}(\beta, \Delta)$.*
2. *If $\Delta = \Delta^{SD}(M)$, then $\mathcal{S}(\hat{\beta}, \Delta)$ is inner consistent for $\mathcal{S}(\beta, \Delta)$.*

Proof. For ease of exposition, we provide a proof for the case where $\theta = \tau_1$; the cases where $\theta = \tau_t$ for $t > 1$ or $\theta = \bar{T}^{-1}(\tau_1 + ... + \tau_{\bar{T}})$ can be handled analogously.

When $\Delta = \Delta^{RM}(\bar{M})$ observe that

$$\mathcal{S}(\hat{\beta}, \Delta) = \hat{\beta}_1 \pm \bar{M} \max_{s<0} |\hat{\beta}_{s+1} - \hat{\beta}_s|.$$

Since $\hat{\beta} \to_p \beta$, it follows from the continuous mapping theorem that

$$\hat{\beta}_1 + \bar{M} \max_{s<0} |\hat{\beta}_{s+1} - \hat{\beta}_s| \to_p \beta_1 + \bar{M} \max_{s<0} |\beta_{s+1} - \beta_s|,$$

so that the upper bound of $\mathcal{S}(\hat{\beta}, \Delta)$ converges in probability to the upper bound of $\mathcal{S}(\beta, \Delta)$. Convergence of the lower bound can be shown analogously, from which the result follows.

When $\Delta = \Delta^{SD}(M)$, observe that

$$\mathcal{S}(\hat{\beta}, \Delta) = \begin{cases} (\hat{\beta}_1 + \hat{\beta}_{-1}) \pm M & \text{if } \hat{\beta}_{pre} \in \Delta_{pre} \\ \varnothing & \text{if } \hat{\beta}_{pre} \notin \Delta_{pre} \end{cases},$$

where recall that $\Delta_{pre} = \{\delta_{pre} : \exists \delta_{post} \text{ s.t. } (\delta'_{pre}, \delta'_{post})' \in \Delta\}$. (Note that we did not need to consider the case where $\hat{\beta}_{pre} \notin \Delta_{pre}$ when $\Delta = \Delta^{RM}$, since in that case $\Delta_{pre} = \mathbb{R}^{T}$.) It follows that $\mathcal{S}(\hat{\beta}, \Delta) \subseteq (\hat{\beta}_1 + \hat{\beta}_{-1}) \pm M$. However, by the continuous mapping theorem, the upper and lower bounds of $(\hat{\beta}_1 + \hat{\beta}_{-1}) \pm M$ converge in probability to $\beta_1 + \beta_{-1} \pm M$, which are the upper and lower bounds of $\mathcal{S}(\beta, \Delta)$, from which the inner consistency of $\mathcal{S}(\hat{\beta}, \Delta)$ is immediate. □

Figure I8: Comparison of confidence sets and estimated identified set for Benzarti and Carloni (2019) application

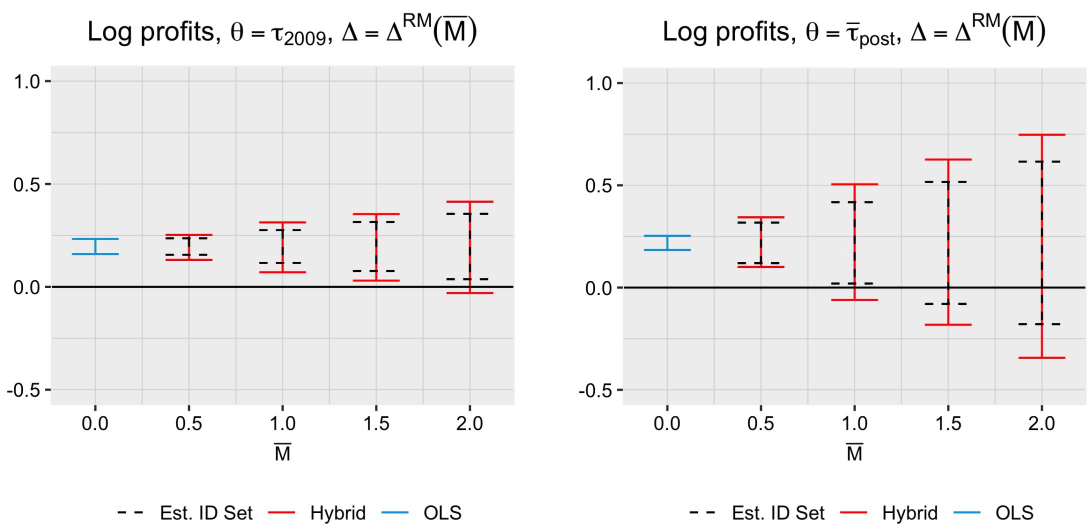
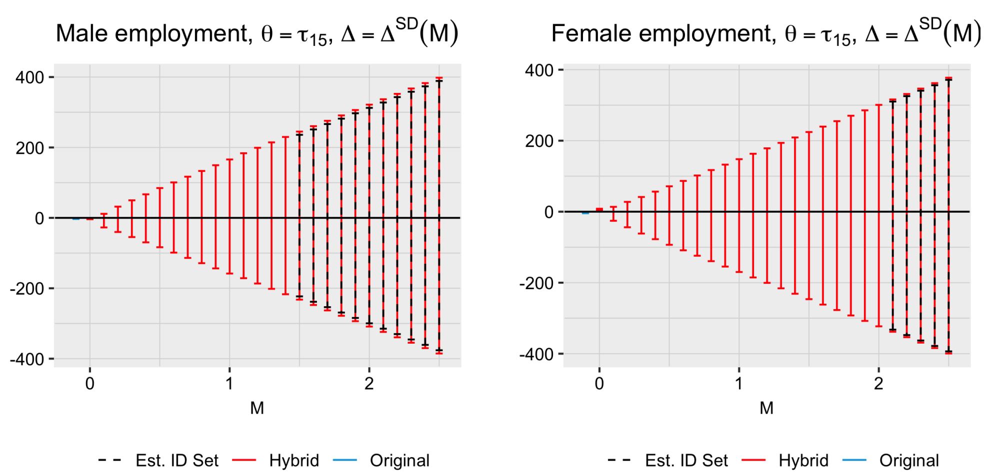


Figure I9: Comparison of confidence sets and estimated identified sets for Lovenheim and Willen (2019) application



Appendix References

Armstrong, Timothy and Michal Kolesár, "Optimal Inference in a Class of Regression Models," *Econometrica*, 2018, *86*, 655–683.

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

Benzarti, Youssef and Dorian Carloni, "Who Really Benefits from Consumption Tax Cuts? Evidence from a Large VAT Reform in France," *American Economic Journal: Economic Policy*, February 2019, *11* (1), 38–63.

Ingster, Yuri and I. A. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models* Lecture Notes in Statistics, New York: Springer-Verlag, 2003.

Lovenheim, Michael F. and Alexander Willen, "The Long-Run Effects of Teacher Collective Bargaining," *American Economic Journal: Economic Policy*, 2019, *11* (3), 292–324.

Nam, Nguyen Mau, "Convex Analysis: An introduction to convexity and nonsmooth analysis," https://maunamn.wordpress.com/ 2019.