An Honest Approach to Parallel Trends

Online Supplementary Materials

Ashesh Rambachan   Jonathan Roth

November 12, 2020

This document provides additional, supplementary materials for "An Honest Approach to Parallel Trends" by Ashesh Rambachan and Jonathan Roth. Sections C-D provide statements and proofs of uniform asymptotic results. Section E provides additional simulation results.

C Uniform asymptotic results

The main text of the paper considers a finite sample normal model, which is motivated as an asymptotic approximation to a variety of econometric settings of interest. In this section, we show that our main results for the conditional approach translate to uniform asymptotic results for a large class of data-generating processes. We refer the reader to Appendix C of Armstrong and Kolesar (2020) for uniformity results for fixed length confidence intervals.[44]

C.1 Assumptions

Throughout this section, we fix $\Delta = \{A\delta \leqslant d\}$ for some A with all non-zero rows, and assume that Δ is non-empty. We consider a class of data-generating processes, indexed by $P \in \mathcal{P}$, under which $\sqrt{n}(\hat{\beta}_n - \beta_P)$ is asymptotically normal, where the asymptotic mean β_P can be decomposed as the sum of $\delta_P \in \Delta$ and $M_{post}\tau_P$ with $\tau_P \in \mathbb{R}^{\bar{T}}$.[45] The parameter of interest is $\theta_P := l'\tau_P$, for some fixed $l \neq 0$.

Assumption 6. *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*

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

where $\xi_P \sim \mathcal{N}(0, \Sigma_P)$, and $\beta_P = \delta_P + M_{post}\tau_P$ for $\delta_P \in \Delta$ and $\tau_P \in \mathbb{R}^{\bar{T}}$.

[44] We note, however, that the setting of Armstrong and Kolesar (2020) differs from ours in that they consider a local-to-0 setting in which Δ shrinks with sample size.

[45] To avoid notational clutter, we drop the additional subscript "post" on τ and simply index τ by the underlying data generating process P.

Convergence in distribution is equivalent to convergence in bounded Lipschitz metric (see Theorem 1.12.4 in van der Vaart and Wellner (1996)), so Assumption 6 formalizes the notion of uniform convergence in distribution of $\sqrt{n}(\hat{\beta}_n - \beta_P)$ to a $\mathcal{N}(0, \Sigma_P)$ variable under P.

Our next assumption requires that the eigenvalues of the asymptotic variance of the event-study coefficients be bounded above and away from zero.

Assumption 7. *Let* $\mathbf{S}$ *denote the set of matrices with eigenvalues bounded below by* $\underline{\lambda} > 0$ *and above by* $\bar{\lambda} \geqslant \underline{\lambda}$*. For all* $P \in \mathcal{P}$, $\Sigma_P \in \mathbf{S}$.

Next, we assume that there is a uniformly consistent estimator of the variance of $\hat{\beta}$.

Assumption 8. *We have an estimator* $\hat{\Sigma}_n$ *that 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 $\epsilon > 0$.

In order to more clearly articulate our next assumption, it is useful to first present the following result, which characterizes the set of dual vertices under Assumption 7.

Lemma C.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 there exists a finite set of distinct, non-zero vectors* $\bar{\gamma}_1, ..., \bar{\gamma}_J$ *such that* $||\gamma_j|| = 1$ *and* $\gamma_j \geqslant 0$ *for all* j*, and for any* Σ *positive definite*

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

where $c_j(\Sigma) = (\bar{\gamma}_j'\tilde{\sigma}(\Sigma))^{-1}$.

For ease of notation, we define $\gamma_j(\Sigma) := c_j(\Sigma)\bar{\gamma}_j$. With this notation in hand, we can then state our next assumption.

Assumption 9. *Suppose* $\bar{\gamma}_j'A \neq 0$*. Then for all* $i \neq j$ *and all* $P \in \mathcal{P}$,

$$(\gamma_j(\Sigma_P) - \gamma_i(\Sigma_P))' A\Sigma_P A' (\gamma_i(\Sigma_P) - \gamma_j(\Sigma_P)) > c,$$

for some constant $c > 0$.

Assumption 9 guarantees that there are not two vertices of the feasible set that produce non-degenerate objective values in the dual problem (18) and are perfectly correlated asymptotically. Assumption 9 holds trivially if the minimal eigenvalue of $A\Sigma_P A'$ is bounded from below. Note that under Assumption 6, $A\Sigma_P A'$ is the asymptotic variance of $\sqrt{n}A\hat{\beta}_n$, and thus corresponds with the asymptotic variance of $\sqrt{n}\tilde{Y}_n(\bar{\theta})$, the moments used in the conditional and hybrid tests scaled by $\sqrt{n}$. Assumption 9 can be dispensed with if we use a modified version of the conditional and hybrid tests that adds full-rank normal noise to $\tilde{Y}_n$, which ensures that the asymptotic covariance of the scaled moments is positive definite.

C.2 Size control

We now establish uniform asymptotic size control for the conditional test. ARP establish uniform asymptotic size control under high-level conditions, whereas here we show size control in our setting under the lower-level conditions introduced above. These conditions are somewhat weaker than the higher-level conditions in ARP. For instance, we allow for the possibility that $\hat{\eta}$ has zero variance conditional on a set of optimal multipliers, which is ruled out by assumptions in ARP but can be shown to arise in our context, e.g. for $\Delta = \Delta^{SDPB}$.

As in ARP, we show size control for a modified version of the conditional and hybrid tests that never rejects if the critical value is below a certain finite value $-\underline{C}$. That is, we consider $\psi^C_{*,\alpha} = \psi^C_\alpha \cdot 1[\hat{\eta} \geqslant -\underline{C}]$, for ψ^C_α an indicator for whether the α-level conditional test rejects and $\hat{\eta}$ the solution to the linear program (17). We do this for technical reasons to avoid complications related to sequences where both $\hat{\eta}$ and the critical values diverge to $-\infty$. However, this modification is reasonable on substantive grounds, since when $\hat{\eta}$ is very small all of the moments are satisfied in the data, and the conditional test (potentially) rejects only due to extreme realizations of the critical values. Moreover, we show in Section C.4 below that the modified tests retain desirable asymptotic power properties.

Under the assumptions stated in the previous section, the modified conditional test uniformly controls size.

Proposition C.1. *Suppose Assumptions 6 to 9 hold. Then*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\psi^C_{*,\alpha}(\hat{\beta}_n, A, d, \theta_P, \frac{1}{n}\hat{\Sigma}_n)\right] \leqslant \alpha.$$

C.3 Consistency

We now provide conditions under which the conditional test is uniformly consistent. Specifically, we establish a uniform asymptotic version of the consistency result given in Proposition 4.1 in the context of the finite sample normal model.

To show uniform consistency for the conditional test, we require some additional assumptions on the asymptotic distribution of the estimated covariance matrix $\hat{\Sigma}$.

Assumption 10. *Let* $W_n = ((\hat{\beta}_n - \beta_P)', (vec(\hat{\Sigma}_n) - vec(\Sigma_P))')'$, *where* $vec(\Sigma)$ *is the vector of the elements of the matrix* Σ. *We assume*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{f\in BL_1} \left|\left|\mathbb{E}_P\left[f(\sqrt{n}W_n)\right] - \mathbb{E}\left[f(\xi_P^+)\right]\right|\right| = 0,$$

where $\xi_P^+ \sim \mathcal{N}(0, V_P)$, $V_P = \begin{pmatrix} \Sigma_P & V_{P,\beta\Sigma} \\ V_{P,\Sigma\beta} & V_{P,\Sigma} \end{pmatrix}$ *and* $\beta_P = \delta_P + M_{post}\tau_P$ *for* $\delta_P \in \Delta$ *and* $\tau_P \in \mathbb{R}^{\bar{T}}$.

Assumption 11. *For all* $P \in \mathcal{P}$, *the matrix* V_P *defined in Assumption 10 lies in a compact set* $\mathbf{V}$. *Additionally,* Σ_P *has eigenvalues bounded between* $\underline{\lambda} > 0$ *and* $\bar{\lambda}$, *and* $(\Sigma_P - V_{P,\beta\Sigma}V_{P,\Sigma}^{-1}V_{P,\Sigma\beta})$ *has eigenvalues bounded below by* $\tilde{\lambda} > 0$.

Assumption 10 strengthens Assumption 6 to require that the pair $(\hat{\beta}, \hat{\Sigma})$ converge uniformly to a joint normal distribution centered at their respective means. Although somewhat more restrictive, we note that event-study estimates are often estimated via OLS, and standard covariance estimators for OLS, including cluster-robust variance estimators, produce asymptotically normal estimates as the number of clusters grows large (Hansen, 2007; Stock and Watson, 2008; Hansen and Lee, 2019). Note that we do not impose that the asymptotic distributions of $\hat{\beta}$ and $\hat{\Sigma}$ are independent, as would occur in linear models if the linear model is properly specified. Likewise, Assumption 11 strengthens Assumption 7 to require that the asymptotic variance matrix of the pair $(\hat{\beta}, \hat{\Sigma})$ lies in a compact set, and that the error in $\hat{\beta}$ is not perfectly colinear with the error in $\hat{\Sigma}$. The latter condition can be ensured to hold by adding full-rank noise to $\hat{\beta}$. With these added conditions, we obtain asymptotic consistency for the (modified) conditional test.

Proposition C.2. *Suppose Assumptions 8 to 11 hold. Then for any* $x > 0$,

$$\lim_{n\to\infty} \inf_{P\in\mathcal{P}} \mathbb{E}_P\left[\psi^C_{*,\alpha}(\hat{\beta}_n, A, d, \theta_P^{ub} + x, \frac{1}{n}\hat{\Sigma}_n)\right] = 1.$$

C.4 Local Asymptotic Power

We now establish conditions under which the power of the conditional test converges uniformly to the power envelope.

Recall that in the finite sample normal model, we showed that the local power of the conditional test converged to the power envelope under Assumption 5, which intuitively guaranteed that the "right" number of moments bind at the edge of the identified set. We define $\mathcal{P}_\epsilon$ to be the set of distributions for which this condition holds and the non-binding moments are slack by at least ϵ.

Definition 2. For $\epsilon > 0$, let $\mathcal{P}_\epsilon$ denote the set of distributions $P \in \mathcal{P}$ such that Assumption 5 holds when setting $\delta_A = \delta_P$, and for which all elements of the vectors $\epsilon_{B(\delta*)}$ and $\epsilon_{B(\delta**)}$ as defined in Assumption 5 are bounded below by ϵ.

Recall from Appendix A.2 that our Assumption 5 is implied by linear independence constraint qualification (LICQ). Assuming that $P \in \mathcal{P}_\epsilon$ is thus similar to a uniform LICQ assumption, as in e.g., Gafarov (2019) and Cho and Russell (2018). We note, however, that we require this assumption only for our uniform local asymptotic power results, and not for uniform asymptotic size control.

Our next result states that the local power of the conditional test converges to the power envelope in the limiting model uniformly over $\mathcal{P}_\epsilon$. This can be viewed as an asymptotic version of Proposition 4.2.

Proposition C.3. *Suppose Assumptions 6 to 8 hold. Let $\theta_P^{ub} = \sup \mathcal{S}(\Delta, \beta_P)$. Then for any $\epsilon > 0$ and $x > 0$,*

$$\lim_{n\to\infty} \sup_{P \in \mathcal{P}_\epsilon} \left| \mathbb{E}_P \left[\psi^C_{*,\alpha}(\hat{\beta}_n, A, d, \theta_P^{ub} + \frac{1}{\sqrt{n}} x, \frac{1}{n}\hat{\Sigma}_n) \right] - \rho^*(P, x) \right| = 0,$$

where

$$\rho^*(P, x) = \lim_{n\to\infty} \sup_{\mathcal{C}_{\alpha,n} \in \mathcal{I}_\alpha(\Delta, \frac{1}{n}\Sigma_P)} \mathbb{P}_{(\delta_P, \tau_P, \frac{1}{n}\Sigma_P)} \left((\theta_P^{ub} + \frac{1}{\sqrt{n}} x) \notin \mathcal{C}_{\alpha,n} \right)$$

is the optimal limiting power of a size-α test in the finite sample normal model using $(\delta_A, \tau_A, \Sigma^*) = (\delta_P, \tau_P, \Sigma_P)$, provided that $-\underline{C}$, the threshold for the modified conditional test, is set sufficiently small.

If $\alpha \in (0, .5]$, then $\bar{C} = 0$ is sufficient for the conclusion of Proposition C.3 to hold.

Proposition C.3 shows that the power of the conditional test converges to the power of the optimal test in the limit of the finite sample normal model as $n \to \infty$. Using results from Müller (2011), we next show that that the power bound $\rho^*(P, x)$ from the limiting model is an upper bound on the asymptotic power of a large class of confidence sets that control size asymptotically. In particular, we consider the set of confidence sets that i) can be written as functions of $\sqrt{n}\hat{\beta}_n$ and $\hat{\Sigma}_n$, ii) control size asymptotically over all sequences of distributions that induce a normal limit, and iii) are invariant to transformations that preserve the identified set for all values of β. To formalize iii), let $A^\perp = \{v : Av = 0\}$ denote the null space of A and let G be the group of transformations of the form $g_v : \beta \mapsto \beta + v$ for $v \in A^\perp$. It is then immediate from the definition of the identified set, $\mathcal{S}(\Delta, \beta) = \{\theta : \exists \delta \in \Delta, \tau_{post} \text{ s.t. } \beta = \delta + M_{post}\tau_{post}, l'\tau_{post} = \theta\}$, that $\mathcal{S}(\Delta, \beta) = \mathcal{S}(\Delta, g_v\beta)$

for any β and $g_v \in G$. By iii) we mean that we will consider the class of confidence sets such that $\mathcal{C}(\sqrt{n}\hat{\beta}, \hat{\Sigma}) = \mathcal{C}(g_v(\sqrt{n}\hat{\beta}), \hat{\Sigma})$ for all $g_v \in G$ and all $\hat{\beta}$.

Proposition C.4. *Suppose that $\mathcal{C}_n(\cdot,\cdot)$ is such that*

$$\limsup_{n\to\infty} \mathbb{P}_{P_n}\left(\theta_{P_n} \notin \mathcal{C}_n(\sqrt{n}\hat{\beta}_n, \hat{\Sigma}_n)\right) \leqslant \alpha$$

for any sequence of distributions P_n such that $\sqrt{n}(\hat{\beta}_n - \beta_{P_n})) \xrightarrow{P_n}_d \mathcal{N}(0, \Sigma^*)$, $\hat{\Sigma}_n \xrightarrow{P_n}_p \Sigma^*$, where $\beta_{P_n} = \delta_{P_n} + M_{post}\tau_{P_n}$ and $\theta_{P_n} = l'\tau_{P_n}$ for some sequences $\tau_{P_n} \in \mathbb{R}^{\bar{T}}$ and $\delta_{P_n} \in \Delta$.

Suppose that for some distribution P^*, $\sqrt{n}(\hat{\beta}_n - \beta_{P*}) \xrightarrow{P^*}_d \mathcal{N}(0, \Sigma^*)$ and $\hat{\Sigma}_n \xrightarrow{P^*}_p \Sigma^*$, where $\beta_{P*} = \delta_{P*} + M_{post}\tau_{P*}$ for $\delta_{P*} \in \Delta$ satisfying Assumption 5. Let $\theta^{ub}_{P*} := \sup \mathcal{S}(\Delta, \beta_{P*})$ be the upper bound of the identified set given β_{P*}. Then, for any $x > 0$,

$$\limsup_{n\to\infty} \mathbb{P}_{P*}\left(\theta^{ub}_{P*} + \frac{1}{\sqrt{n}}x \notin \mathcal{C}_n(\sqrt{n}\hat{\beta}_n, \hat{\Sigma}_n)\right) \leqslant \rho^*(P^*, x),$$

where $\rho^*(P^*, x)$ is defined in Proposition C.3.

D Proofs of uniform asymptotic results

D.1 Proofs and Auxiliary Lemmas for Uniform Size Control

Proof of Lemma C.1

Proof. Recall from Section 8.5 of Schrijver (1986) that v is a vertex of the polyhedron $P = \{x \in \mathbb{R}^K : Wx \leqslant b\}$ iff $v \in P$ and $W_{(\mathcal{J},\cdot)}x = b_{\mathcal{J}}$ for $\mathcal{J}$ a set of indices such that $W_{(\mathcal{J},\cdot)}$ has K independent rows. It follows that $v \in V(\Sigma)$ iff $v \geqslant 0$ and there exists $\mathcal{J}$ such that

$$W_{\mathcal{J}} := \begin{pmatrix} \tilde{A}'_{(\cdot,-1)} \\ -I_{(\mathcal{J},\cdot)} \\ \tilde{\sigma}' \end{pmatrix}$$

has row rank equal to K, and $W_{\mathcal{J}}v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, where K is the number of rows of A.

Now, let $\boldsymbol{\mathcal{J}}$ be the set of indices $\mathcal{J}$ such that $\tilde{W}_{\mathcal{J}} := \begin{pmatrix} \tilde{A}'_{(\cdot,-1)} \\ -I_{(\mathcal{J},\cdot)} \end{pmatrix}$ has exactly $K-1$ linearly independent rows and there exists a vector $v_{\mathcal{J}} \neq 0$ such that $\tilde{W}_{\mathcal{J}}v = 0$ and $v_{\mathcal{J}} \geqslant 0$. Since by construction $\tilde{W}_{\mathcal{J}}$ has rank $K-1$ and K columns, its nullspace is 1-dimensional. It is then immediate that for each $\mathcal{J} \in \boldsymbol{\mathcal{J}}$, there is a unique vector $\bar{v}_{\mathcal{J}} \geqslant 0$ such that $||\bar{v}_{\mathcal{J}}|| = 1$ and $\tilde{W}_{\mathcal{J}}\bar{v}_{\mathcal{J}} = 0$. Moreover, $\boldsymbol{\mathcal{J}}$ is finite, since there are a finite number of possible subindices of I, and thus we can write $\{\bar{v}_{\mathcal{J}} : \mathcal{J} \in \boldsymbol{\mathcal{J}}\} = \{\bar{v}_1, ..., \bar{v}_J\}$ for distinct vectors $\bar{v}_1, ..., \bar{v}_J$.

It now remains to show that $V(\Sigma) = \{c_1(\Sigma)\bar{v}_1, ..., c_J(\Sigma)\bar{v}_J\}$, for c_j as defined above. First, suppose that $v = c_j(\Sigma)\bar{v}_j$ for some j. By construction, $\tilde{A}'_{(\cdot,-1)}v = 0$, $v \geqslant 0$, and $\tilde{\sigma}'v = (\tilde{\sigma}'v_j)^{-1}(\tilde{\sigma}'v_j) = 1$, and so $v \in F$. Additionally, there exists $\mathcal{J}$ such that $\tilde{W}_{\mathcal{J}} = \begin{pmatrix} \tilde{A}'_{(\cdot,-1)} \\ -I_{(\mathcal{J},\cdot)} \end{pmatrix}$ has rank $K-1$ and $\tilde{W}_{\mathcal{J}}v = 0$. From the fact that $\tilde{W}_{\mathcal{J}}v = 0$, whereas $\tilde{\sigma}'v = 1$, we see that $\tilde{\sigma}'$ must be linearly independent from the rows of $\tilde{W}_{\mathcal{J}}$, and thus $W_{\mathcal{J}} = \begin{pmatrix} \tilde{W}_{\mathcal{J}} \\ \tilde{\sigma}' \end{pmatrix}$ has rank K. It follows that $v \in V(\Sigma)$.

Next, suppose that $v \in V(\Sigma)$. Then $v \geqslant 0$, and there exists $\mathcal{J}$ such that

$$W_{\mathcal{J}} := \begin{pmatrix} \tilde{A}'_{(\cdot,-1)} \\ -I_{(\mathcal{J},\cdot)} \\ \tilde{\sigma}' \end{pmatrix}$$

has row rank equal to K, and $W_{\mathcal{J}}v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Let $\tilde{W}_{\mathcal{J}} = \begin{pmatrix} \tilde{A}'_{(\cdot,-1)} \\ -I_{(\mathcal{J},\cdot)} \end{pmatrix}$. Note that since $\tilde{W}_{\mathcal{J}}v = 0$, whereas $\tilde{\sigma}'v = 1$, $\tilde{\sigma}'$ must be linearly independent of the other rows of $W_{\mathcal{J}}$, from which it follows that $\tilde{W}$ has row rank $K-1$. Thus, $\mathcal{J} \in \boldsymbol{\mathcal{J}}$, and so $v = c\bar{v}_j$ for some j and $c > 0$. Since $\tilde{\sigma}'v = 1$, we have $c\tilde{\sigma}'\bar{v}_j = 1$, which implies $c = (\tilde{\sigma}'\bar{v}_j)^{-1}$, which gives the desired result. □

Proof of Proposition C.1

Proof. First, note that by Lemma B.2, $\psi^C_\alpha(\hat{\beta}_n, A, d, \theta_P, \frac{1}{n}\hat{\Sigma}_n) = \psi^C_\alpha(\sqrt{n}\hat{\beta}_n, A, \sqrt{n}d, \sqrt{n}\theta_P, \hat{\Sigma}_n)$. Additionally, we show in the proof to Lemma B.2 that the values of $\hat{\eta}$ for these two problems are the same, from which it follows that the modified tests are tests are equivalent as well, $\psi^C_{*,\alpha}(\hat{\beta}_n, A, d, \theta_P, \frac{1}{n}\hat{\Sigma}_n) = \psi^C_{*,\alpha}(\sqrt{n}\hat{\beta}_n, A, \sqrt{n}d, \sqrt{n}\theta_P, \hat{\Sigma}_n)$. It thus suffices to show that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\psi^C_{*,\alpha}(\sqrt{n}\hat{\beta}_n, A, \sqrt{n}d, \sqrt{n}\theta_P, \hat{\Sigma}_n)\right] \leqslant \alpha.$$

Towards contradiction, suppose the proposition is false. Then, following Andrews, Cheng and Guggenberger (2020), there exists a sequence of distributions P_m and an increasing sequence of sample sizes n_m such that

$$\liminf_{m\to\infty} \mathbb{E}_{P_m}\left[\psi^C_{*,\alpha}(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m}d, \sqrt{n_m}\theta_P, \hat{\Sigma}_{n_m})\right] \geqslant \alpha + \omega, \tag{61}$$

for some $\omega > 0$.

Define $Y_m := \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - \tilde{A}_{(\cdot,-1)}\theta_{P_m}\right)$ and $X := \tilde{A}_{(\cdot,-1)}$. Then,

$$\psi^C_{*,\alpha}(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m}d, \sqrt{n_m}\theta_P, \hat{\Sigma}_{n_m}) = \psi^C_{*,\alpha}(Y_m, X, A\hat{\Sigma}_{n_m}A').$$

Further, define $\tilde{Y}_m := Y_m - \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}(\sqrt{n_m}\tau_{P_m})$ For notational convenience, let $\Sigma_m := \Sigma_{P_m}$ and $\hat{\Sigma}_m := \hat{\Sigma}_{n_m}$. By Lemma 16 in ARP, $\psi^C_{*,\alpha}(Y_m, X, A\hat{\Sigma}_m A') = \psi^C_{*,\alpha}(\tilde{Y}_m, X, A\hat{\Sigma}_m A')$. Additionally, observe that

$$\begin{aligned}
\tilde{Y}_m &= \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - \tilde{A}_{(\cdot,1)}\theta_{P_m} - \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}\tau_{P_m}\right) \\
&= \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - \tilde{A}_{(\cdot,1)}l'\tau_{P_m} - \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}\tau_{P_m}\right) \\
&= \sqrt{n_m}\left(A\left(\hat{\beta}_{n_m} - \begin{pmatrix} 0 \\ \tau_{P_m} \end{pmatrix}\right) - d\right),
\end{aligned}$$

where the first equality uses the definition of $\theta_{P_m} = l'\tau_{P_m}$ and the second equality follows from Lemma D.5. This implies that

$$\tilde{Y}_m = A\sqrt{n_m}\left(\hat{\beta}_{n_m} - \delta_{P_m} - \begin{pmatrix} 0 \\ \tau_{P_m} \end{pmatrix}\right) + \sqrt{n_m}\left(A\delta_{P_m} - d\right). \tag{62}$$

Next, observe that by Assumption 6, $\delta_P \in \Delta = \{\delta : A\delta \leqslant d\}$ for all P, and so $\sqrt{n_m}\left(A\delta_{P_m} - d\right) \leqslant 0$. We can therefore extract a subsequence m_1 such that

$$\sqrt{n_{m_1}}\left(A\delta_{P_{m_1}} - d\right)_1 \to \mu_1^* \in \mathbb{R} \cup \{-\infty\}.$$

Passing to further subsequences, we can extract a subsequence m_K (for K the number of rows of A) along which

$$\sqrt{n_{m_K}}\left(A\delta_{P_{m_K}} - d\right) \to \mu^* \in \{\mathbb{R} \cup \{-\infty\}\}^K.$$

Additionally, by Assumption 7, Σ_{P_m} is contained within a compact set, and so we can extract a further subsequence m_{K+1} along which $\Sigma_{m_{K+1}} \to \Sigma^*$ for some $\Sigma^* \in \mathbf{S}$. For notational ease, we will assume that these convergences hold for the original sequence (m, n_m) for the remainder of the proof

Now, equation (62) along with Assumptions 6 and 8 and the continuous mapping theorem imply that

$$(\tilde{Y}_m, \hat{\Sigma}_m) \xrightarrow{d} (\xi + \mu^*, \Sigma^*),$$

for $\xi \sim \mathcal{N}(0, A\Sigma^* A')$. Observe from (62) that for all m, $\tilde{Y}_m \in col(A) + \{-a \cdot d : a > 0\}$, where $col(A)$ is the column space of A and $+$ represents the Minkowski sum. Likewise, if $\xi \sim \mathcal{N}(0, A\Sigma^* A')$, then $\xi = A\xi_{\Sigma^*}$ for $\xi_{\Sigma^*} \sim \mathcal{N}(0, \Sigma^*)$, and so ξ is supported on $col(A)$. Thus, $\xi + \mu^*$ is supported on $col(A) + \mu^*$. We then see that both $\tilde{Y}_m$ and $\xi + \mu^*$ are supported on $\Omega := col(A) + (\{-a \cdot d : a \in \mathbb{R}\} \cup \{\mu^*\})$.

Suppose first that $\max_{\gamma \in V(\Sigma^*)} \gamma' \mu^* = -\infty$. Note that $\hat{\eta}_m = \max_{\gamma \in V(\hat{\Sigma}_m)} \gamma' \tilde{Y}_m$. From Lemma C.1, $V(\Sigma) = \{c_1(\Sigma)\bar{\gamma}_1, ..., c_J(\Sigma)\bar{\gamma}_J\}$, where the functions $c_j(\Sigma)$ are continuous and by Lemma D.1, $c_j(\Sigma)^* \geqslant -\underline{c} > 0$ for all j. Since $\max_{\gamma \in V(\Sigma^*)} \gamma' \mu^* = -\infty$, we have $c_j(\Sigma^*)\bar{\gamma}_j' \mu^* = -\infty$ for all j. But the continuous mapping theorem then implies that for all j, $c_j(\hat{\Sigma}_m)\bar{\gamma}_j' \tilde{Y}_m \to_d c_j(\Sigma^*)\bar{\gamma}_j'(\xi + \mu^*) = -\infty$, and hence $\hat{\eta}_m \to_p -\infty$. Thus, $\mathbb{P}(\hat{\eta}_m < -C) \to 1$, and so our tests never reject asymptotically, which contradicts size control failing. For the remainder of the proof, we assume that $\max_{\gamma \in V(\Sigma^*)} \gamma' \mu^*$ is finite. (Note that since $\bar{\gamma}_j \geqslant 0$ and $\mu^* \leqslant 0$, we cannot have $\max_{\gamma \in V(\Sigma^*)} \gamma' \mu^* = \infty$.)

Next, note that it follows readily from the construction of the (unmodified) conditional test in Section 4.2 that the unmodified conditional test rejects iff

$$p(Y, \Sigma) := \mathbb{P}_\zeta \left(\zeta < \hat{\eta}(Y, \Sigma) \,|\, \zeta \in [v^{lo}(Y, \Sigma), v^{up}(Y, \Sigma)], \zeta \sim \mathcal{N}\left(0, \sigma_\eta^2(Y, \Sigma)\right)\right) > 1 - \alpha,$$

where the functions $\hat{\eta}, \sigma_\eta^2$, v^{lo} and v^{up} are defined as follows. We define $\hat{\eta}(Y, \Sigma)$ to be the conditional test statistic using Y and Σ,

$$\hat{\eta}(Y, \Sigma) := \max_{\gamma \in V(\Sigma)} \gamma' Y,$$

We define $\sigma_\eta^2(Y, \Sigma)$ to be the estimated variance of $\gamma_*' Y$ for $\gamma_* \in \arg\max_{\gamma \in V(\Sigma)} \gamma' Y$. That is,

$$\sigma_\eta^2(Y, \Sigma) = \gamma_*' A \Sigma A' \gamma_*,$$

Note that $\sigma_\eta^2(Y, \Sigma)$ is only well-defined if $\gamma_*' A \Sigma A' \gamma_*$ is the same for all $\gamma_* \in \arg\max_{\gamma \in V(\Sigma)} \gamma' Y$. We will show below, however, that this occurs with probability 1 in the limiting model.

If $\sigma_\eta^2(Y, \Sigma) > 0$, then we define $v^{lo}(Y, \Sigma)$ and $v^{up}(Y, \Sigma)$ to be the minimum and maximum of the set

$$C = \{c : \max_{\gamma \in V(\Sigma)} \gamma' \left(S_{\gamma_*} + \frac{\hat{\Sigma}\gamma_*}{\gamma_* \hat{\Sigma} \gamma_*} c \right)\},$$

where as before γ_* is an element of $\arg\max_{\gamma \in V(\Sigma)} \gamma' Y$ and we define

$$S_{\gamma*} = \left(I - \frac{\hat{\Sigma}\gamma_* \gamma_*'}{\gamma_*' \hat{\Sigma} \gamma_*} \right) Y.$$

On the other hand, if $\sigma_\eta^2(Y, \Sigma) = 0$, then we define $v^{lo} = -\infty$ and $v^{up} = \infty$. This is a notational convenience that allows us to capture the fact that when $\sigma_\eta^2 = 0$, the unmodified conditional test rejects iff $\hat{\eta}(Y, \Sigma) > 0$, since $\mathbb{P}(\zeta < \hat{\eta} \,|\, \zeta \sim \mathcal{N}(0, 0)) = 1[\hat{\eta} > 0]$.

Since the modified conditional test rejects only if the unmodified conditional test rejects, (61) thus implies that

$$\liminf_{m \to \infty} \mathbb{P}_{P_m} \left(p(\tilde{Y}_m, \hat{\Sigma}) > 1 - \alpha \right) \geqslant \alpha + \omega. \tag{63}$$

Lemma D.3 shows that the function $p(\cdot, \cdot)$ is continuous at $(\xi + \mu^*, \Sigma^*)$ for almost every $\xi \sim \mathcal{N}(0, A\Sigma^* A')$. The continuous mapping theorem then implies that

$$p(\tilde{Y}_m, \hat{\Sigma}) \xrightarrow{d} p(\xi + \mu^*, \Sigma^*).$$

Moreover, Lemma D.4 implies that the distribution of $p(\xi + \mu^*, \Sigma^*)$ does not have a mass point at $1 - \alpha$, and hence

$$\mathbb{P}_{P_m} \left(p(\tilde{Y}_m, \hat{\Sigma}) > 1 - \alpha \right) \to \mathbb{P}\left(p(\xi + \mu^*, \Sigma^*) > 1 - \alpha \right).$$

However, since the conditional test controls size in the finite-sample normal model,

$$\mathbb{P}_\xi \left(p(\xi + \mu^*, \Sigma^*) > 1 - \alpha \right) \leqslant \alpha,$$

and thus

$$\liminf_{m \to \infty} \mathbb{P}_{P_m} \left(p(\tilde{Y}_m, \hat{\Sigma}) > 1 - \alpha \right) \leqslant \alpha,$$

which contradicts (63).

□

Lemma D.1. *Suppose Assumption 7 holds. Then for any x and $\Sigma \in \mathbf{S}$, $\underline{\lambda} x'x \leqslant x'\Sigma x \leqslant \bar{\lambda} x'x$. Additionally, there exist constants $\underline{c} > 0$ and $\bar{c}$ such that for all $\Sigma \in \mathbf{S}$ and all $j = 1, ..., J$, $\underline{c} \leqslant c_j(\Sigma) \leqslant \bar{c}$, for $c_j(\Sigma)$ as defined in Lemma C.1.*

Proof. By the singular value decomposition, we can write $\Sigma = U\Lambda U'$, where U is a unitary matrix ($UU' = I$) and Λ is the diagonal matrix with the eigenvalues of Σ on the diagonal. By Assumption 7, these eigenvalues are bounded between $\underline{\lambda} > 0$ and $\bar{\lambda} \geqslant \underline{\lambda}$. Thus, for any x, we have $x'\Sigma x = (U'x)'\Lambda(U'x)' = \sum_i \lambda_i (U'x)_i^2$. It follows that $x'\Sigma x \leqslant \sum_i \bar{\lambda}(U'x)_i^2 = \bar{\lambda}x'UU'x = \bar{\lambda}x'x$. It can be shown analogously that $x'\Sigma x \geqslant \underline{\lambda}x'x$. Now, recall that $c_j(\Sigma) = (\bar{\gamma}_j'\tilde{\sigma}(\Sigma))^{-1}$, where $\tilde{\sigma}_i^2 = A_{(i,\cdot)}'\Sigma A_{(i,\cdot)}$. Let $\bar{m}_A = \max_i A_{(i,\cdot)}'A_{(i,\cdot)}$ and $\underline{m}_A = \min_i A_{(i,\cdot)}'A_{(i,\cdot)}$, and note that both $\bar{m}$ and $\underline{m}$ are strictly positive since A is assumed to have no all-zero rows. It then follows from the previous discussion that $\tilde{\sigma}_i \in [\sqrt{\underline{\lambda}\underline{m}_A}, \sqrt{\bar{\lambda}\bar{m}_A}] := [\tilde{\sigma}_{lb}, \tilde{\sigma}_{ub}]$. Moreover, since $\bar{\gamma}_j \geqslant 0$ and $\bar{\gamma}_j \neq 0$ for all j, we have that $\bar{\gamma}_j'\tilde{\sigma} \geqslant \max\{\bar{\gamma}_j\}\tilde{\sigma}_{lb} \geqslant \min_j\{\max\{\bar{\gamma}_j\}\}\tilde{\sigma}_{lb} > 0$, where the last inequality uses the fact that the set $\bar{\gamma}_1, ...\bar{\gamma}_J$ is finite. Likewise, for K the dimension of $\bar{\gamma}_j$, we have $\bar{\gamma}_j'\tilde{\sigma} \leqslant K\max\{\bar{\gamma}_j\}\tilde{\sigma}_{ub} \leqslant \max_j\{\max\{\bar{\gamma}_j\}\tilde{\sigma}_{ub}\} < \infty$. We have thus shown that $\bar{\gamma}_j'\tilde{\sigma}(\Sigma)$ is bounded between two positive finite values, and thus the same is true of its inverse, which suffices for the result. □

Lemma D.2. *Let μ^*, Σ^*, and Ω be as defined in the proof to Proposition C.1, and assume $\max_{\gamma \in V(\Sigma^*)} \gamma'\mu^*$ is finite and Assumption 9 holds. Let $N(\Sigma^*)$ be an open set containing Σ^*. Then $\hat{\eta}(Y,\Sigma)$, $\sigma_\eta^2(Y,\Sigma)$, $v^{lo}(Y,\Sigma)$, $v^{up}(Y,\Sigma)$ – when viewed as functions over $\Omega \times N(\Sigma^*)$ – are continuous in (Y,Σ) at $(\xi + \mu^*, \Sigma^*)$ for almost every $\xi \sim \mathcal{N}(0, A\Sigma^*A')$. Additionally, for almost every ξ, one of the following holds:*

1) There is a neighborhood of $(\xi + \mu^*, \Sigma^*)$ on which $\sigma_\eta^2(Y,\Sigma) > 0$ and $v^{lo}(Y,\Sigma) < v^{up}(Y,\Sigma)$.

2) There is a neighborhood of $(\xi+\mu^*, \Sigma^*)$ on which $\hat{\eta}(Y,\Sigma) \leqslant 0$, $\sigma_\eta^2(Y,\Sigma) = 0$ and $v^{lo}(Y,\Sigma) = -\infty$, $v^{up}(Y,\Sigma) = \infty$.

Proof. We first show that $\hat{\eta}(Y,\Sigma)$ is continuous. Lemma C.1 implies that

$$\hat{\eta}(Y,\Sigma) := \max_{\gamma \in V(\Sigma)} \gamma'Y = \max\{c_1(\Sigma)\bar{\gamma}_1'Y, ..., c_J(\Sigma)\bar{\gamma}_J'Y\},$$

where the functions $c_j(\Sigma)$ are continuous. We claim that each of the functions in the max above are continuous in (Y,Σ) at $(\xi + \mu^*, \Sigma^*)$. If Y were finite-valued, then this would hold trivially. However, since some elements of Y may be equal to $-\infty$, we additionally need to show that there is a neighborhood of Σ^* such that for all Σ in this neighborhood and all j, the elements of $c_j(\Sigma)\bar{\gamma}_j$ do not change from 0 to non-zero or vice versa. However, by Lemma D.1, $c_j(\Sigma^*) \geqslant \underline{c} > 0$ for all j, and so for Σ sufficiently close to Σ^*, $c_j(\Sigma) > 0$, and thus each element of $c_j(\Sigma)\bar{\gamma}_j$ has the same sign (0 or positive) as the corresponding element of $\bar{\gamma}_j$, as we desired to show.

Next, define $\hat{V}(Y,\Sigma) := \arg\max_{\gamma \in V(\Sigma)} \gamma'Y$. We claim that with probability 1, either $\hat{V}(\xi + \mu^*, \Sigma^*)$ is unique, or $\gamma_*'A = 0$ for all $\gamma_* \in \hat{V}(Y,\Sigma)$. Observe that since ξ is finite with probability 1 and $\max_{\gamma \in V(\Sigma^*)} \gamma'\mu^*$ is finite by assumption, it follows that $max_{\gamma \in V(\Sigma^*)}\gamma'(\xi + \mu^*)$ is finite with probability 1. Let $\gamma_1, \gamma_2 \in V(\Sigma^*)$. Note that $\gamma_1, \gamma_2 \in \hat{V}(\xi, \Sigma^*)$ only if

$(\gamma_1 - \gamma_2)'\xi = (\gamma_2 - \gamma_1)'\mu^*$. Observe further that $(\gamma_1 - \gamma_2)'\xi$ is normally distributed with variance $(\gamma_1 - \gamma_2)' A \Sigma^* A' (\gamma_1 - \gamma_2)'$. Thus, $(\gamma_1 - \gamma_2)'\xi$ is equal to any particular constant with positive probability only if $(\gamma_1 - \gamma_2)' A \Sigma^* A' (\gamma_1 - \gamma_2)' = 0$. Since Σ^* is positive definite, $(\gamma_1 - \gamma_2)' A \Sigma^* A' (\gamma_1 - \gamma_2)' = 0$ iff $(\gamma_1 - \gamma_2)' A = 0$. However, by Assumption 9, $(\gamma_1 - \gamma_2)' A = 0$ only if $\gamma_1' A = \gamma_2' A = 0$. It follows that at most one of γ_1 and γ_2 are in $\hat{V}$ with probability 1, or $\gamma_1' A = \gamma_2' A = 0$. Since the set $V(\Sigma^*)$ is finite, it follows that either $\hat{V}(\xi + \mu^*, \Sigma^*)$ is unique, or all of its elements have $\gamma_*' A = 0$, as needed.

Suppose first that every $\gamma_* \in \hat{V}(\xi + \mu^*, \Sigma^*)$ satisfies $\gamma_*' A = 0$. Without loss of generality, assume that $\hat{V}(\xi + \mu^*) = \{c_1(\Sigma^*)\bar{\gamma}_1, ..., c_{J_1}(\Sigma^*)\bar{\gamma}_{J_1}\}$, where $1 \leqslant J_1 \leqslant J$. We first claim that there is a neighborhood of $(\xi + \mu^*, \Sigma^*)$ on which $\max_{\gamma \in V(\Sigma)} \gamma' Y = c_j(\Sigma)\bar{\gamma}_j' Y$ for some $j \leqslant J_1$. This is trivial if $J_1 = J$. If not, let $j \leqslant J_1$ and $i > J_1$. Since $c_j(\Sigma^*)\bar{\gamma}_j'(\xi + \mu^*) \in \hat{V}(\xi + \mu^*, \Sigma^*)$ and $c_i(\Sigma^*)\bar{\gamma}_i'(\xi + \mu^*) \notin \hat{V}(\xi + \mu^*, \Sigma^*)$, we must have $c_j(\Sigma^*)\bar{\gamma}_j'(\xi + \mu^*) > c_i(\Sigma^*)\bar{\gamma}_i'(\xi + \mu^*)$. We showed above that the functions on both sides of the inequality are continuous in (Y, Σ) at $(\xi + \mu^*, \Sigma^*)$, and thus there exists a neighborhood of $(\xi + \mu^*, \Sigma^*)$ on which the inequality is preserved, and hence $\max_{\gamma \in V(\Sigma)} \gamma' Y > c_i(\Sigma)\bar{\gamma}_j'(\xi + \Sigma)$. Additionally, since there are finitely many $i > J_1$, we can choose a neighborhood such that this holds simultaneously for all $i > J_1$, which implies that in this neighborhood $\hat{V}(Y, \Sigma) \subseteq \{c_1(\Sigma)\bar{\gamma}_1, ..., c_{J_1}(\Sigma)\bar{\gamma}_{J_1}\}$, as needed. It follows that $\sigma_\eta^2(Y, \Sigma) = 0$ for all (Y, Σ) in this neighborhood, since $\bar{\gamma}_j' A = 0$ for all $j \leqslant J_1$, which implies $\bar{\gamma}_j' A \Sigma A' \bar{\gamma}_j = 0$. Additionally, note that by definition, $v^{lo}(Y, \Sigma) = -\infty$ and $v^{up}(Y, \Sigma) = \infty$ whenever $\sigma_\eta^2(Y, \Sigma) = 0$. Thus, $\sigma_\eta^2(Y, \Sigma)$, $v^{lo}(Y, \Sigma)$, and $v^{up}(Y, \Sigma)$ are continuous at $(\xi + \mu^*, \Sigma^*)$.

To show that $\eta(Y, \Sigma) \leqslant 0$ in a neighborhood of $(\xi + \mu^*, \Sigma^*)$, observe that it is immediate from the definition of Ω that any $Y \in \Omega$ can be written as $Av - a_1 \cdot d + a_2 \mu^*$, for $v \in \mathbb{R}^K$ and $a_1, a_2 \geqslant 0$. For any $j \in \{1, ..., J_1\}$, $\bar{\gamma}_j' A = 0$, and thus $\bar{\gamma}_j' Y = -a_1 \bar{\gamma}_j' d + a_2 \bar{\gamma}_j' \mu^*$. However, since $\bar{\gamma}_j \geqslant 0$ and $\mu^* \leqslant 0$, we have that $a_2 \bar{\gamma}_j' \mu^* \leqslant 0$. Likewise, since Δ is assumed to be non-empty, there exists some δ such that $A\delta - d \leqslant 0$. Since $\bar{\gamma}_j' A = 0$ and $\bar{\gamma}_j \geqslant 0$, it follows that $\bar{\gamma}_j'(-d) \leqslant 0$. Hence, $\bar{\gamma}_j' Y \leqslant 0$ for any $Y \in \Omega$, and thus, in a neighborhood of Σ^* sufficiently small such that $c_j(\Sigma) \geqslant 0$, $c_j(\Sigma)\bar{\gamma}_j' Y \leqslant 0$. Since we've shown that in a neighborhood of $(\xi + \mu^*, \Sigma^*)$, $\hat{\eta}(Y, \Sigma) = c_j(\Sigma)\bar{\gamma}_j' Y$ for some j, it follows that $\eta(Y, \Sigma) \leqslant 0$ for (Y, Σ) sufficiently close to $(\xi + \mu^*, \Sigma^*)$.

Next, suppose that $\hat{V}(\xi + \mu^*)$ has a single element $\gamma_* = c_j(\Sigma^*)\bar{\gamma}_j'(\xi + \mu^*)$ for some $j \in \{1, ..., J\}$ such that $\bar{\gamma}_j' A \neq 0$. Without loss of generality, suppose $j = 1$. We first show that $\hat{V}(Y, \Sigma) = c_1(\Sigma)\bar{\gamma}_1$ in a neighborhood of $(\xi + \mu^*)$. Indeed, since $\hat{V}(\xi + \mu^*) = c_1(\Sigma^*)\bar{\gamma}_1'(\xi + \mu^*)$, for all $i > 1$, $c_1(\Sigma^*)\bar{\gamma}_1'(\xi + \mu^*) > c_i(\Sigma^*)\bar{\gamma}_i'(\xi + \mu^*)$. However, since we've shown the functions on both sides of this inequality to be continuous in (Y, Σ) at $(\xi + \mu^*, \Sigma^*)$, there is a neighborhood of $(\xi + \mu^*, \Sigma^*)$ such that for all $i > 1$, $c_1(\Sigma)\bar{\gamma}_1' Y > c_i(\Sigma)\bar{\gamma}_i' Y$, and hence $\hat{V}(Y,\Sigma) = c_1(\Sigma)\bar{\gamma}_1$ in this neighborhood. It follows that in a neighborhood of $(\xi+\mu^*)$, $\sigma_\eta^2(Y,\Sigma) = c_1(\Sigma)\bar{\gamma}_1' A\Sigma A' c_1(\Sigma)\bar{\gamma}_1$, which is clearly continuous in Σ. Additionally, by Lemma D.1, $c(\Sigma^*) \geqslant \underline{c} > 0$, and so $\sigma_\eta^2 \geqslant \underline{c}^2\bar{\gamma}_1' A\Sigma^* A'\bar{\gamma}_1$, which is positive since $\gamma_1' A \neq 0$ and Σ^* is positive definite. From the continuity of σ_η^2, it follows that there is a neighborhood of $(\xi+\mu^*,\Sigma^*)$ such that $\sigma_\eta^2(Y,\Sigma) > 0$.

Next, consider $v^{lo}(Y,\Sigma)$. Let $\gamma_*(\Sigma) = c_1(\Sigma)\bar{\gamma}_1$. For ease of notation, we will make the dependence of γ_* on Σ implicit where it is clear below. The results above imply that in a neighborhood of $(\xi+\mu^*,\Sigma^*)$, $v^{lo}(Y,\Sigma)$ is the minimum of the set

$$C(Y,\Sigma) = \{c \,:\, \max_{\gamma\in V(\Sigma)} \gamma' \left(S_{\gamma_*}(Y) + \frac{\Sigma\gamma_*}{\gamma_*\Sigma\gamma_*} c \right) = c\},$$

for

$$S_{\gamma_*}(Y,\Sigma) = \left(I - \frac{\Sigma\gamma_*\gamma_*'}{\gamma_*'\Sigma\gamma_*} \right) Y.$$

Rearranging terms, we see that

$$C = \{c \,:\, 0 = \max_{\gamma\in V(\Sigma)} a_{\gamma,\gamma_*}(Y) + b_{\gamma,\gamma_*} c\},$$

where $a_{\gamma,\gamma_*}(Y) := \gamma' S_{\gamma_*}(Y)$ and $b_{\gamma,\gamma_*} := \dfrac{\gamma'\Sigma\gamma_*}{\gamma_*'\Sigma\gamma_*} - 1$. Note that $a_{\gamma_*,\gamma_*}(Y) = 0 = b_{\gamma_*,\gamma_*}$, so $0 \leqslant \max_{\gamma\in V(\Sigma)} a_{\gamma,\gamma_*}(Y) + b_{\gamma,\gamma_*} c$ for all c. Moreover, for $c = \gamma_*' Y$, the max is attained at γ_* by construction. Hence, the set C is non-empty.

Intuitively, if we plot $a_{\gamma,\gamma_*}(Y) + b_{\gamma,\gamma_*}$ as a function of c, then each $\gamma \in V(\Sigma)$ defines a line, and the set C represents the values of c for which 0 is the upper envelope of this set. It follows that the lower bound of C is the maximal x-intercept of a line of the form $a_{\gamma,\gamma_*}(Y) + b_{\gamma,\gamma_*} c$ with $b_{\gamma,\gamma_*} < 0$. Hence,

$$v^{lo}(Y,\Sigma) = \max_{\{\gamma\in V(\Sigma)\setminus\{\gamma_*\} \,:\, b_{\gamma,\gamma_*}<0\}} \frac{-a_{\gamma,\gamma_*}(Y)}{b_{\gamma,\gamma_*}}.$$

Recall that by Lemma C.1, $V(\Sigma) := \{\gamma_1(\Sigma), ..., \gamma_J(\Sigma)\}$, where $\gamma_j(\Sigma) := c_j(\Sigma)\bar{\gamma}_j$ and $c_j(\Sigma)$ is continuous. Additionally, we showed earlier in the proof that for all j, $c_j(\Sigma)\bar{\gamma}_j' Y$ is continuous in a neighborhood of $(\xi+\mu^*,\Sigma^*)$. It is then immediate from the definitions of $a_{\gamma,\gamma_*}(Y)$ and b_{γ,γ_*} that for all j, $a_{\gamma_j(\Sigma),\gamma_*(\Sigma)}(Y)$ and $b_{\gamma_j(\Sigma),\gamma_*(\Sigma)}$ are continuous in (Y,Σ). Without loss of generality, suppose that for $2 \leqslant k \leqslant k_1$, $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*)} < 0$; for $k_1 < k \leqslant k_2$, $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*)} = 0$; and for $k > k_2$, $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*)} > 0$. From the continuity of $b_{\gamma_j(\Sigma),\gamma_*(\Sigma)}$, it is clear that in a neighborhood of $(\xi+\mu^*,\Sigma^*)$, $b_{\gamma_k(\Sigma),\gamma_*(\Sigma)} > 0$ for all $2 \leqslant k \leqslant k_1$ and $b_{\gamma_k(\Sigma),\gamma_*(\Sigma)} < 0$ for all $k > k_2$. Hence, in this neighborhood,

$$v^{lo}(Y,\Sigma) = \max\left\{ \max_{\gamma_k(\Sigma)\,:\,2\leqslant k\leqslant k_1} \frac{-a_{\gamma_k(\Sigma),\gamma_*(\Sigma)}}{b_{\gamma_k(\Sigma),\gamma_*(\Sigma)}}, \max_{\gamma\in V^0(\Sigma)} \frac{-a_{\gamma,\gamma_*(\Sigma)}}{b_{\gamma,\gamma_*(\Sigma)}} \right\}, \tag{64}$$

where

$$V^0(\Sigma) := \{\gamma_k(\Sigma) \,:\, k_1 < k \leqslant k_2, b_{\gamma_k(\Sigma),\gamma_*(\Sigma)} < 0\}$$

and we define the max of an empty set to be $-\infty$. It is clear from the continuity of the functions a and b that the inner max on the left side of (64) is continuous. To show that v^{lo} is continuous at $(\xi+\mu^*,\Sigma^*)$, it suffices to show that for any sequence $(Y,\Sigma)\to(\xi+\mu^*,\Sigma^*)$, the max on the right hand side of (64) converges to $-\infty$. To do this, observe that by construction, $a_{\gamma,\gamma_*}(Y) + b_{\gamma,\gamma_*}\cdot\gamma_*'Y = \gamma'Y - \gamma_*'Y$. Since for any $k>1$, $\gamma_*(\Sigma^*)'(\xi+\mu^*) > \gamma_k(\Sigma^*)'(\xi+\mu^*)$, it follows that $a_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*)}(\xi+\mu^*) + b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*)}\cdot(\xi+\mu^*) < 0$. Additionally, $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*)}(\xi+\mu^*) = 0$ for $k\in(k_1,k_2]$, and so for such values of k, $a_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*)}(\xi+\mu^*) < 0$. However, this implies that for any sequence $(Y,\Sigma)\to(\xi+\mu^*)$ and $k\in(k_1,k_2]$, we have $-a_{\gamma_k(\Sigma),\gamma_*(\Sigma)}(Y)$ approaching a positive limit, and $b_{\gamma_k(\Sigma),\gamma_*(\Sigma)}$ approaching 0. For values of (Y,Σ) where $b_{\gamma_k(\Sigma),\gamma_*(\Sigma)} > 0$, it follows that $-a_{\gamma_k(\Sigma),\gamma_*(\Sigma)}(Y)/b_{\gamma_k(\Sigma),\gamma_*(\Sigma)}$ becomes arbitrarily negative, whereas for values of (Y,Σ) where $b_{\gamma_k(\Sigma),\gamma_*(\Sigma)} \geqslant 0$, γ_k is not included in V^0. It is then immediate that the max on the right hand side of (64) converges to $-\infty$, which suffices to establish the continuity of v^{lo} at $(\xi+\mu^*,\Sigma^*)$. The continuity of v^{up} can be shown analogously.

To complete the proof, we now demonstrate that in a neighborhood of $(\xi+\mu^*)$, $v^{lo}(Y,\Sigma) < v^{up}(Y,\Sigma)$ for almost every ξ. Note that since we have shown v^{lo} and v^{up} to be continuous, it suffices to show that $v^{lo}(\xi+\mu^*,\Sigma^*) < v^{up}(\xi+\mu^*,\Sigma^*)$. We showed above that for almost every ξ, either $\hat{V}(\xi+\mu^*)$ contains only elements such that $\gamma'A = 0$, or $\hat{V}(\xi+\mu^*)$ has a unique element such that $\gamma'A \neq 0$. In the former case, we showed that $v^{lo} = -\infty$ and $v^{up} = \infty$. Suppose we are in the latter case. We showed that $v^{lo}(\xi+\mu^*,\Sigma^*)$ is the x-intercept of a line of the form $a + b\cdot c$, where $b<0$ and $a+b\cdot\hat{\eta} < 0$. Hence, $v^{lo}(\xi+\mu^*,\Sigma) < \hat{\eta}(\xi+\mu^*,\Sigma)$. However, by construction $v^{lo}\leqslant\hat{\eta}\leqslant v^{up}$, and thus $v^{lo}<\hat{\eta}$ implies $v^{lo}<v^{up}$, which completes the proof.

□

Lemma D.3. *Let μ^*, Σ^*, and Ω be as defined in the proof to Proposition C.1, and assume $\max_{\gamma\in V(\Sigma^*)}\gamma'\mu^*$ is finite. Let $N(\Sigma^*)$ be an open set containing Σ^*. Define $p:\Omega\times N(\Sigma^*)\to$*

$[0,1]$ *by*

$$p(Y,\Sigma) := \mathbb{P}_\zeta\left(\zeta < \hat{\eta}(Y,\Sigma) \,|\, \zeta \in [v^{lo}(Y,\Sigma), v^{up}(Y,\Sigma)], \zeta \sim \mathcal{N}\left(0,\, \sigma^2_\eta(Y,\Sigma)\right)\right).$$

Then $p(Y,\Sigma)$ *is continuous in both arguments at* $(\xi+\mu^*,\Sigma^*)$ *for almost every* $\xi \sim \mathcal{N}(0,\, A\Sigma^* A')$ *and* $\Sigma^* \in \mathbf{S}$ *non-stochastic.*

Proof. From Lemma D.2, for almost every ξ, the functions $\hat{\eta}$, $v^{lo}, v^{up}, \sigma^2_\eta$ are continuous at $(\xi+\mu^*,\Sigma^*)$. Additionally, for almost every ξ, either

1) There is a neighborhood of $(\xi+\mu^*,\Sigma^*)$ on which $\sigma^2_\eta(Y,\Sigma) > 0$ and $v^{lo}(Y,\Sigma) < v^{up}(Y,\Sigma)$, or

2) There is a neighborhood of $(\xi+\mu^*,\Sigma^*)$ on which $\hat{\eta}(Y,\Sigma) \leqslant 0$, $\sigma^2_\eta(Y,\Sigma) = 0$ and $v^{lo}(Y,\Sigma) = -\infty$, $v^{up}(Y,\Sigma) = \infty$.

First, suppose 1) holds. Note that for $v^{lo} < v^{up}$ and $\sigma_\eta > 0$,

$$\mathbb{P}_\zeta\left(\zeta < \hat{\eta} \,|\, \zeta \in [v^{lo}, v^{up}], \zeta \sim \mathcal{N}\left(0,\, \sigma^2_\eta\right)\right) = \frac{\Phi(\hat{\eta}/\sigma_\eta) - \Phi(v^{lo}/\sigma_\eta)}{\Phi(v^{up}/\sigma_\eta) - \Phi(v^{lo}/\sigma_\eta)},$$

which is clearly continuous in $\hat{\eta}, v^{lo}, v^{up}$, and σ_η. The continuity of $p(Y,\Sigma)$ then follows from the continuity of $\hat{\eta}, v^{lo}, v^{up}$, and σ_η.

Next, suppose 2) holds. Note that

$$\mathbb{P}_\zeta\left(\zeta < \hat{\eta} \,|\, \zeta \in [-\infty, \infty], \zeta \sim \mathcal{N}(0,\, 0)\right) = 1[\hat{\eta} > 0].$$

It then follows that when 2) holds, $p(Y,\Sigma) = 0$ in a neighborhood of $(\xi+\mu^*,\Sigma^*)$, and thus is continuous. □

Lemma D.4. *Let* $p(Y,\Sigma)$ *be as defined in Lemma D.3, and suppose* $\max_{\gamma\in V(\Sigma^*)} \gamma'\mu^*$ *is finite. Let* $\xi \sim \mathcal{N}(0,\, A\Sigma^* A')$*. Then for any* $\alpha \in (0,1)$, $\mathbb{P}\left(p(\xi+\mu^*,\Sigma^*) = 1-\alpha\right) = 0$.

Proof. Note that for $v^{lo} < v^{up}$ and $\sigma_\eta > 0$,

$$\mathbb{P}_\zeta\left(\zeta < \hat{\eta} \,|\, \zeta \in [v^{lo}, v^{up}], \zeta \sim \mathcal{N}\left(0,\, \sigma^2_\eta\right)\right) = \frac{\Phi(\hat{\eta}/\sigma_\eta) - \Phi(v^{lo}/\sigma_\eta)}{\Phi(v^{up}/\sigma_\eta) - \Phi(v^{lo}/\sigma_\eta)}.$$

Thus, when $v^{lo} < v^{up}$ and $\sigma_\eta > 0$, $p(\xi+\mu^*,\Sigma^*) = 1-\alpha$ iff $\hat{\eta} = \sigma_\eta \cdot c_{1-\alpha}(v^{lo}, v^{up}, \sigma_\eta)$, where $c_{1-\alpha}(v^{lo}, v^{up}, \sigma_\eta)$ is the unique value that solves

$$\frac{\Phi(c_{1-\alpha}) - \Phi(v^{lo}/\sigma_\eta)}{\Phi(v^{up}/\sigma_\eta) - \Phi(v^{lo}/\sigma_\eta)} = 1-\alpha.$$

However, (19) implies that $\hat{\eta}(\xi+\mu^*,\Sigma^*)$ has a truncated normal distribution conditional on $v^{lo}(\xi+\mu^*,\Sigma^*)$, $v^{up}(\xi+\mu^*,\Sigma^*)$ and $\sigma_\eta^2(\xi+\mu^*,\Sigma^*)$, with truncation points $v^{lo}(\xi+\mu^*,\Sigma^*)$ and $v^{up}(\xi+\mu^*,\Sigma^*)$ and (untruncated) variance $\sigma_\eta^2(\xi+\mu^*,\Sigma^*)$, and hence is continuously distributed when $v_{lo}(\xi+\mu^*,\Sigma^*) < v_{up}(\xi+\mu^*,\Sigma^*)$ and $\sigma_\eta^2(\xi+\mu^*,\Sigma^*) > 0$. Thus, conditional on $v_{lo}(\xi+\mu^*,\Sigma^*) < v_{up}(\xi+\mu^*,\Sigma^*)$ and $\sigma_\eta^2(\xi+\mu^*,\Sigma^*) > 0$, $\hat{\eta}(\xi+\mu^*,\Sigma^*) = c_{1-\alpha}(v^{lo}, v^{up}, \sigma_\eta)$ with probability zero.

Additionally, observe that

$$\mathbb{P}\left(\zeta < \hat{\eta}) \,|\, \zeta \in [-\infty,\infty], \zeta \sim \mathcal{N}\left(0,\,0\right)\right) = 1[\hat{\eta} > 0].$$

Hence, whenever $\hat{\eta}(\xi+\mu^*,\Sigma^*) \leqslant 0$, $v^{lo}(\xi+\mu^*,\Sigma^*) = -\infty$, $v^{up}(\xi+\mu^*,\Sigma^*) = \infty$ and $\sigma_\eta(\xi+\mu^*,\Sigma^*) = 0$, we have $p(\xi+\mu^*,\Sigma^*) = 0 \neq 1-\alpha$ for almost every ξ.

However, from Lemma D.2, with probability 1 either i) $v_{lo}(\xi+\mu^*,\Sigma^*) < v_{up}(\xi+\mu^*,\Sigma^*)$ and $\sigma_\eta^2(\xi+\mu^*,\Sigma^*) > 0$, or ii) $\hat{\eta}(\xi+\mu^*,\Sigma^*) \leqslant 0$, $v^{lo}(\xi+\mu^*,\Sigma^*) = -\infty$ $v^{up}(\xi+\mu^*,\Sigma^*) = \infty$ and $\sigma_\eta(\xi+\mu^*,\Sigma^*) = 0$. The desired result then follows immediately.

□

Lemma D.5. *For any vector $v \in \mathbb{R}^{\bar{T}}$,*

$$\tilde{A}_{(\cdot,1)}(l'v) + \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}v = A\begin{pmatrix} 0 \\ I_{\bar{T}} \end{pmatrix} v.$$

Proof. By definition,

$$\begin{aligned}
\tilde{A}_{(\cdot,1)} &= A\begin{pmatrix} 0 \\ I \end{pmatrix}\Gamma^{-1} I_{(\cdot,1)} \\
\tilde{A}_{(\cdot,-1)} &= A\begin{pmatrix} 0 \\ I \end{pmatrix}\Gamma^{-1} I_{(\cdot,-1)} \\
\Gamma_{(-1,\cdot)} &= I_{(-1,\cdot)}\Gamma.
\end{aligned}$$

Additionally, the first row of Γ is assumed to be l', so $l' = I_{(1,\cdot)}\Gamma$. It follows that

$$\begin{aligned}
\tilde{A}_{(\cdot,1)} l'v &= A\begin{pmatrix} 0 \\ I \end{pmatrix}\Gamma^{-1} I_{(\cdot,1)} I_{(1,\cdot)}\Gamma v \\
\tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)} v &= A\begin{pmatrix} 0 \\ I \end{pmatrix}\Gamma^{-1} I_{(\cdot,-1)} I_{(-1,\cdot)}\Gamma v.
\end{aligned}$$

Noting that $I_{(\cdot,-1)}I_{(-1,\cdot)} + I_{(\cdot,1)}I_{(1,\cdot)} = I$, the two equations in the previous display imply that

$$\tilde{A}_{(\cdot,1)}(l'v) + \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}v = A\begin{pmatrix} 0 \\ I \end{pmatrix}\Gamma^{-1}I\Gamma v = A\begin{pmatrix} 0 \\ I \end{pmatrix}v,$$

as needed. $\square$

D.2 Proofs and auxiliary lemmas for uniform consistency results

Proof of Proposition C.2

Proof. As in the proof to Proposition C.1, $\psi^C_{*,\alpha}(\hat{\beta}_n, A, d, \theta^{ub}_P + x, \frac{1}{n}\hat{\Sigma}_n) = \psi^C_{*,\alpha}(\sqrt{n}\hat{\beta}_n, A, d, \sqrt{n}\theta^{ub}_P + \sqrt{n}x, \hat{\Sigma}_n)$, so it suffices to show that

$$\lim_{n\to\infty}\inf_{P\in\mathcal{P}} \mathbb{E}_P\left[\psi^C_{*,\alpha}(\sqrt{n}\hat{\beta}_n, A, \sqrt{n}d, \sqrt{n}\theta^{ub}_P + \sqrt{n}x, \hat{\Sigma}_n)\right] = 1.$$

Towards contradiction, suppose this is false. Then there exists an increasing sequence of distributions P_m and sample sizes n_m such that

$$\limsup_{m\to\infty} \mathbb{E}_{P_m}\left[\psi^C_{*,\alpha}(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m}d, \sqrt{n_m}\theta^{ub}_{P_m} + \sqrt{n_m}x, \hat{\Sigma}_{n_m})\right] \leqslant 1-\omega, \tag{65}$$

for some $\omega > 0$. Since $\mathbf{V}$ is compact, we can extract a subsequence m_1 along which $V_{P_{m_1}} \to V^* = \begin{pmatrix} \Sigma^* & V^*_{\beta\Sigma} \\ V^*_{\Sigma\beta} & V^*_{\Sigma} \end{pmatrix} \in \mathbf{V}$. For ease of notation, without loss of generality we assume that this holds for the original sequence m. Now, let

$$\begin{aligned}\tilde{Y}_m &:= \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_P + x)\right) \\ &= \sqrt{n_m}A\left(\hat{\beta}_{n_m} - \beta_{P_m}\right) + \sqrt{n_m}\left(A\beta_{P_m} - d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_P + x)\right),\end{aligned} \tag{66}$$

and observe that

$$\psi^C_{*,\alpha}(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m}d, \sqrt{n_m}\theta^{ub}_{P_m} + \sqrt{n_m}x, \hat{\Sigma}_{n_m}) = \psi^C_{*,\alpha}(\tilde{Y}_m, X, A\hat{\Sigma}_{n_m}A'),$$

where

$$\tilde{Y}_m = \sqrt{n_m}A\left(\hat{\beta}_{n_m} - \beta_{P_m}\right) + \sqrt{n_m}\underbrace{\left(A\beta_{P_m} - d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_P + x)\right)}_{=:\lambda_m}. \tag{67}$$

Now, from Lemma B.13, there exists a constant $c > 0$ such that $\eta(\beta_{P_m}, A, d, \theta^{ub}_{P_{n_m}} + x, \Sigma^*) \geqslant c \cdot x$ for $\eta(\cdot)$ defined in (38). Reformulating (38) in terms of its dual, and noting that the dual vertices are the same as in the dual problem for $\hat{\eta}$, we see that there is a dual vertex $\gamma_j(\Sigma^*) \in V(\Sigma^*)$ such that $\gamma_j(\Sigma^*)' \left(A\beta_{P_m} - d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_{P_m} + x)\right) \geqslant c \cdot x$. From Lemma C.1, $\gamma_j(\Sigma^*) = c_j(\Sigma^*)\bar{\gamma}_j$, and there is a vertex of $V(\hat{\Sigma}_{n_m})$ of the form $\gamma_j(\hat{\Sigma}_{n_m}) = c_j(\hat{\Sigma}_{n_m})\bar{\gamma}_j$, where the function $c_j(\cdot)$ is continuous. Since $\hat{\Sigma}_{n_m} \to_p \Sigma^*$, it follows that $\gamma_j(\hat{\Sigma}_{n_m}) \to_p \gamma_j(\Sigma^*)$, and hence $\gamma_j(\hat{\Sigma}_{n_m})' \left(A\beta_{P_m} - d - \tilde{A}_{(\cdot,1)}(\theta^{ub}_{P_m} + x)\right) \to_p c \cdot x > 0$. It is then clear from (67) that $\gamma_j(\hat{\Sigma}_{n_m})'\tilde{Y}_m \to_p \infty$, since the inner product of $\gamma_j(\hat{\Sigma}_{n_m})$ with the first term of (67) converges in distribution to a normal distribution with mean 0 and finite variance by Assumption 8 and Slutksy's lemma, and the second term converges in probability to ∞. Since $\gamma_j(\hat{\Sigma}_{n_m})'\tilde{Y}_m$ is feasible in the dual problem for $\hat{\eta}_{n_m}$, it follows that $\hat{\eta}_{n_m} \to_p \infty$. It follows that $\mathbb{P}_{P_m}\left(\hat{\eta}_{n_m} < -\underline{C}\right) \to 0$, so the modified test agrees with the unmodified test with probability approaching 1. For simplicity, we therefore consider the unmodified test for remainder of the proof.

Now, suppose $C > \max\{0, z_{1-\alpha}\}$. We showed in the proof to Lemma B.16 that if $\hat{\eta}(\tilde{Y}, \tilde{\Sigma}) > C$, then $\psi^C_\alpha(\tilde{Y}, \tilde{\Sigma}) = 1$ unless $\sigma_{\gamma_*} := \sqrt{\gamma_*'\tilde{\Sigma}\gamma_*} > 0$ and $\frac{1}{\sigma_{\gamma_*}}(\hat{\eta} - v^{lo}) < \zeta(C)$, where γ_* is an optimal solution to the dual problem and $\zeta(\cdot)$ is a function such that $\zeta(C) \to 0$ as $C \to \infty$. Additionally, by Lemma D.6, there exists some vertex γ such that $\frac{1}{\sigma_{\gamma_*}}(\hat{\eta} - v^{lo}) = \kappa(\gamma_*, \gamma)\left(\gamma_*'\tilde{Y} - \gamma'\tilde{Y}\right)$, where $\kappa(\gamma, \gamma_*) = \frac{\sqrt{\gamma_*'\tilde{\Sigma}\gamma_*}}{\gamma_*'\tilde{\Sigma}\gamma_* - \gamma'\tilde{\Sigma}\gamma_*}\left(\gamma_*'\tilde{Y} - \gamma'\tilde{Y}\right) > 0$.

To complete the proof, we will show that we can extract a subsequence of m, indexed by q, along with a constant $C > \max\{0, z_{1-\alpha}\}$ such that

$$\limsup_{q\to\infty} \mathbb{P}_{P_q}\left(\{\hat{\eta}_{n_q} < C\} \vee \left\{\{\hat{\sigma}_{\eta,n_q} > 0\} \wedge \left\{\frac{1}{\hat{\sigma}_{\eta,n_q}}(\hat{\eta}_{n_q} - v^{lo}_{n_q}) < \zeta(C)\right\}\right\}\right) \leqslant \omega/2.$$

This implies a contradiction of (65), since the event in the probability in the previous display is a necessary condition for the conditional test to not reject. Further, since we've shown that $\hat{\eta}_{n_m} \to_p \infty$, it suffices to construct a subsequence such that

$$\limsup_{q\to\infty} \mathbb{P}_{P_q}\left(\{\hat{\sigma}_{\eta,n_q} > 0\} \wedge \left\{\frac{1}{\hat{\sigma}_{\eta,n_q}}(\hat{\eta}_{n_q} - v^{lo}_{n_q}) < \zeta(C)\right\}\right) \leqslant \omega/2. \tag{68}$$

Now, recall from Lemma C.1 that we can write $V(\tilde{\Sigma}) = \{c_1(\tilde{\Sigma})\bar{\gamma}_1, ..., c_J(\tilde{\Sigma})\bar{\gamma}_J\}$ for positive continuous functions c_j and distinct non-zero vectors $\bar{\gamma}_j \geqslant 0$. For notational convenience, let $c_{i,m} = c_i(\hat{\Sigma}_{n_m}), c^*_i = c_i(\Sigma^*)$, $\gamma_{i,m} = c_{i,m}\bar{\gamma}_i$, and $\gamma^*_i = c^*_i\bar{\gamma}_i$. Likewise, for a pair (i, j) let $\kappa_{ij,m} = \kappa(\gamma_{i,m}, \gamma_{j,m})$ and $\kappa^*_{ij} = \kappa(\gamma^*_i, \gamma^*_j)$. Assumption 8 implies that $\hat{\Sigma}_{n_m} \to_p \Sigma^*$. By the continuous mapping theorem, we therefore have $c_{i,m} \to_p c_i^*$, $\gamma_{i,m} \to_p \gamma_i^*$, and $\kappa_{ij,m} \to_p \kappa_{ij}^*$.

Note that if $\gamma_{i,m}$ is optimal and $\bar{\gamma}_i' A = 0$, then $\hat{\sigma}_{\eta,n_m} = (c_{i,m}\bar{\gamma}_i)' A \hat{\Sigma}_{n_m} A' (c_{i,m}\bar{\gamma}_i) = 0$. Thus, we can only have $\hat{\sigma}_{\eta,n_m} > 0$ if the optimal vertex corresponds with an index i such that $\bar{\gamma}_i' A \neq 0$. To establish (68), it therefore suffices to extract a subsequence q such that for any pair (i,j) with $i \neq j$ and $\bar{\gamma}_i' A \neq 0$, either

$$\lim_{q\to\infty} \mathbb{P}_{P_q}\left(\hat{\eta}_{n_q} = \gamma_{i,q}'\tilde{Y}_m\right) = 0, \text{ OR} \tag{69}$$

$$\limsup_{q\to\infty} \mathbb{P}_{P_q}\left(\left\{\hat{\eta}_{n_q} = \gamma_{i,q}'\tilde{Y}_q\right\} \wedge \left\{|\kappa_{ij,q}(\gamma_{i,q} - \gamma_{j,q})'\tilde{Y}_q| < \zeta(C)\right\}\right) \leqslant \omega/(2m), \tag{70}$$

where m is the number of such pairs (i,j).

Consider any such pair (i,j). First, we claim that $\bar{\gamma}_i'\lambda_m \leqslant -\bar{\gamma}_i'\tilde{A}_{(\cdot,1)}x$. To show this, note that since $\theta_{P_m}^{ub} \in \mathcal{S}(\Delta, \beta_{P_m})$, $\exists \tilde{\tau} \in \mathbb{R}^{\bar{T}-1}$ such that $\lambda_m + \tilde{A}_{(\cdot,1)}x = A\beta_{n_m} - d - \tilde{A}_{(\cdot,1)}\theta_{P_m}^{ub} - \tilde{A}_{(\cdot,-1)}\tilde{\tau} \leqslant 0$. By construction (see the proof to Lemma C.1) $\bar{\gamma}_i'\tilde{A}_{(\cdot,-1)} = 0$ and $\bar{\gamma}_i \geqslant 0$, and hence $\bar{\gamma}_i'(\lambda_m + \tilde{A}_{(\cdot,1)}x) \leqslant 0$, which implies $\bar{\gamma}_i'\lambda_m \leqslant -\bar{\gamma}_i'\tilde{A}_{(\cdot,1)}x$, as desired.

Since $\bar{\gamma}_i'\lambda_m$ is bounded above, it follows that either i) $\bar{\gamma}_i'\lambda_m \to -\infty$, or ii) there exists a subsequence m_1 such that $\bar{\gamma}_i'\lambda_m \to \mu_1 \in \mathbb{R}$. If i) holds, then it is clear from (67) that $\gamma_{i,m}'\tilde{Y}_m \to_p -\infty$, since the inner product of $\gamma_{i,m}$ with the first term in (67) converges in distribution to a normal distribution with mean 0 and finite variance by Assumption 10 and Slutsky's lemma, and the second term converges in probability to $-\infty$. Since $\hat{\eta}_{n_m} \to_p \infty$, it follows that $\mathbb{P}\left(\hat{\eta}_{n_m} = \gamma_{i,m}'\tilde{Y}_m\right) \to 0$, so $\gamma_{i,m}$ is optimal with vanishing probability. Now, suppose ii) holds and consider the sequence m_1. By an analogous argument for $\gamma_{j,m}$, we can show that either ii.a) $\gamma_{j,m_1}'\tilde{Y}_m \to_p -\infty$ or ii.b) there exists a further subsequence m_2 such that $\bar{\gamma}_{j,m_2}'\lambda_{m_2} \to \mu_2 \in \mathbb{R}$. If ii.a) holds, then it is immediate that for any $\zeta > 0$, $\mathbb{P}\left(\left\{\hat{\eta}_{n_{m_2}} = \gamma_{i,m_2}'\tilde{Y}_{m_2}\right\} \wedge \left\{\kappa_{ij,m_2}(\gamma_{i,m_2} - \gamma_{j,m_2})'\tilde{Y}_{m_2} \in [-\zeta, \zeta]\right\}\right) \to 0$, since $\hat{\eta}_{n_{m_2}} \to \infty$, $\gamma_{j,m_1}'\tilde{Y}_m \to_p -\infty$, and $\kappa_{ij,m_2} \to \kappa_{ij}^* > 0$. Now, suppose ii.b) holds. Since $\sqrt{n_{m_2}}(\gamma_i^* - \gamma_j^*)'\lambda_{m_2}$ is non-stochastic, we can choose a subsequence m_3 such that $\sqrt{n_{m_3}}(\gamma_i^* - \gamma_j^*)'\lambda_{m_3} \to \mu_3 \in \mathbb{R} \cup \{\pm\infty\}$. Then

$$\begin{aligned}(\gamma_{i,m_3} - \gamma_{j,m_3})'\tilde{Y}_{m_3} = &\underbrace{(\gamma_{i,m_3} - \gamma_{j,m_3})'\sqrt{n_{m_3}}A(\hat{\beta}_{n_{m_3}} - \beta_{P_{m_3}})}_{=:Z_1} \\ &+ \underbrace{\sqrt{n_{m_3}}(\gamma_{i,m_3} - \gamma_i^*)'\lambda_{m_3}}_{=:Z_2} - \underbrace{\sqrt{n_{m_3}}(\gamma_{j,m_3} - \gamma_j^*)'\lambda_{m_3}}_{=:Z_3} \\ &+ \underbrace{\sqrt{n_{m_3}}(\gamma_i^* - \gamma_j^*)'\lambda_{m_3}}_{Z_4}\end{aligned}$$

By Assumption 10 along with Slutsky's lemma, $Z_1 \to_d (\gamma_i^* - \gamma_j^*)A\xi_\beta$, for $\xi_\beta \sim \mathcal{N}(0, \Sigma^*)$.

Next, note that we write $Z_2 = \sqrt{n}(c_i(\hat{\Sigma}_{n_{m_3}}) - c_i(\Sigma^*))\bar{v}_i'\lambda_{m_3}$. Since c_i is continuous, Assumption 10 along with the delta method imply that $\sqrt{n}(c_i(\hat{\Sigma}_{n_{m_3}}) - c_i(\Sigma^*)) \to_d G_i'\xi_\Sigma$, where $G_i = D_{vec(\Sigma)}c_i(\Sigma^*)$ is the gradient of c_i at Σ^*, and $\xi_\Sigma \sim \mathcal{N}(0, V_\Sigma)$. Since $\bar{v}_i'\lambda_m \to \mu_1$, by Slutsky's lemma, we have $Z_2 \to_d \mu_1 G_i'\xi_\Sigma$. By an analogous argument, we have that $Z_3 \to_d \mu_2 G_j'\xi_\Sigma$. Finally, recall that $Z_4 \to \mu_3$ by construction, and $\kappa_{ij,m_3} \to \kappa_{ij}^* > 0$. Combining these results, along with the fact that these convergences hold jointly by Assumption 10, we have that

$$\kappa_{ij,m_3}(\gamma_{i,m_3} - \gamma_{j,m_3})'Y_{m_3} \to_d \kappa_{ij}^*(\gamma_i^* - \gamma_j^*)'A\xi_\beta + \kappa_{ij}^*(\mu_1 G_i - \mu_2 G_j)'\xi_\Sigma + \kappa_{ij}^*\mu_3,$$

where $(\xi_\beta', \xi_\Sigma')' \sim \mathcal{N}(0, V^*)$. It is immediate that the limiting distribution in the previous display, which we will denote by ξ_{ij}, is normally distributed. We claim further that its variance is strictly positive. Indeed, note that $\xi_\beta \,|\, \xi_\Sigma$ is normally distributed with variance $\Sigma^* - V_{\beta\Sigma}^* V_\Sigma^{*-1} V_{\Sigma\beta}^*$, which is positive definite by Assumption 11. Further, Assumption 9 implies that $(\gamma_i^* - \gamma_j^*)'A \neq 0$, and thus $\kappa_{ij}^*(\gamma_i^* - \gamma_j^*)'A\xi_\beta$ has positive variance conditional on ξ_Σ. That the unconditional variance of ξ_{ij} is positive then follows from the law of total variance. Let σ_{ij}^2 denote the unconditional variance of ξ_{ij}. We then see that for any $\zeta > 0$, $\mathbb{P}(\xi_{ij} \in [-\zeta, \zeta]) \leqslant \Phi(\zeta/\sigma_{ij}) - \Phi(-\zeta/\sigma_{ij})$, since the normal distribution is single-peaked and symmetric about its mean, so the maximal probability that a normal variable falls in an interval of length 2ζ occurs when the interval is centered around the mean. Since $\zeta(C) \to 0$ as $C \to \infty$, we can choose C sufficiently large such $\Phi(\zeta/\sigma_{ij}) - \Phi(-\zeta/\sigma_{ij}) < \omega/(2m)$. Hence,

$$\limsup_{m_3 \to \infty} \mathbb{P}\left(|\kappa_{ij,n_{m_3}}(\gamma_{i,m_3} - \gamma_{j,m_3})'Y_{m_3}| < \zeta(C)\right) \leqslant \omega/(2m).$$

We have thus established that we can find a subsequence along which (69) or (70) holds for a single pair (i, j). However, since there are finitely many such pairs (i, j), we can use analogous arguments to further refine our subsequence and constant C such that this holds for all pairs (i, j).

□

Lemma D.6. *Let $\hat{\eta}(Y, \Sigma)$ be as defined in the proof to Proposition C.1, and γ_* an optimal solution to the dual problem for $\hat{\eta}(Y, \Sigma)$. Then, if $v^{lo}(Y, \Sigma)$ is finite,*

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

for some vertex $\gamma \in V(\Sigma)$ such that $\frac{\gamma_*'\Sigma\gamma_*}{\gamma_*'\Sigma\gamma_* - \gamma'\Sigma\gamma_*} > 0$.

Proof. We show in the proof to Lemma D.11 that

$$v^{lo} = \min_{\{\gamma \in V(\Sigma)\,:\, b_{\gamma,\gamma_*} < 0\}} \frac{-a_{\gamma,\gamma_*}(\tilde{Y})}{b_{\gamma,\gamma_*}},$$

where

$$b_{\gamma,\gamma_*} = \frac{\gamma'\Sigma\gamma_*}{\gamma_*'\Sigma\gamma_*} - 1$$
$$a_{\gamma,\gamma_*}(Y) = \gamma'(I - \frac{\Sigma\gamma_*'}{\gamma_*'\Sigma\gamma_*}\gamma_*')Y.$$

Noting that $\hat{\eta} = \gamma_*'Y$, the result then follows from applying the expressions above and cancelling like terms. □

D.3 Proofs and auxiliary lemmas for uniform local asymptotic power results

Proof of Proposition C.3

Proof. Let $\bar{\gamma}_1, ..., \bar{\gamma}_J$ be as defined in Lemma C.1. By Lemma D.16, there exists a value $C^* \in \mathbb{R}$ such that for any $\Sigma \in \mathbf{S}$ and any j such that $\bar{\gamma}_j' A \neq 0$,

$$\Phi\left(\frac{\hat{\eta}}{\sqrt{\gamma_j(\Sigma)' A \Sigma A' \gamma_j(\Sigma)}}\right) > 1 - \alpha$$

only if $\hat{\eta} > -C^*$. We suppose throughout the proof that $-\underline{C} \leqslant -C^*$.

Towards contraction, suppose that the proposition is false. Then there exists a sequence of distributions $P_m \in \mathcal{P}_\epsilon$ and an increasing sequence of sample sizes n_m such that

$$\liminf_{n\to\infty} \left| \mathbb{E}_{P_m}\left[\psi^C_{*,\alpha}(\hat{\beta}_{n_m}, A, d, \theta^{ub}_{P_m} + \frac{1}{\sqrt{n_m}}x, \frac{1}{n_m}\hat{\Sigma}_{n_m})\right] - \rho^*(P_m)\right| \geqslant \omega \tag{71}$$

for some $\omega > 0$. We showed in the proof to Proposition C.1 that $\psi^C_{*,\alpha}$ is invariant to scale, so this is equivalent to

$$\liminf_{n\to\infty} \left| \mathbb{E}_{P_m}\left[\psi^C_{*,\alpha}(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m}d, \sqrt{n}\theta^{ub}_{P_m} + x, \hat{\Sigma}_{n_m})\right] - \rho^*(P_m)\right| \geqslant \omega. \tag{72}$$

Define

$$Y_m = \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - \tilde{A}_{(\cdot,-1)}(\theta^{ub}_{P_m} + x)\right)$$

and $X := \tilde{A}_{(\cdot,-1)}$. Then

$$\psi^C_{*,\alpha}(\sqrt{n_m}\hat{\beta}_{n_m}, A, \sqrt{n_m}d, \sqrt{n}\theta^{ub}_{P_m} + x, \hat{\Sigma}_{n_m}) = \psi^C_{*,\alpha}(Y_m, X, A\hat{\Sigma}_{n_m}A').$$

For notational convenience, define $\tau_m := \tau_{P_m}$; define δ_m, δ^{**}_m and Σ_m analogously. Let $\tilde{Y}_m := Y_m - \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}\sqrt{n_m}(\tau_{P_m} - \delta_{P_m,post} + \delta^{**}_{P_m,post})$. By Lemma 16 in ARP, $\psi^C_{*,\alpha}(Y_m, X, A\hat{\Sigma}_{n_m}A') = \psi^C_{*,\alpha}(\tilde{Y}_m, X, A\hat{\Sigma}_{n_m}A')$. Additionally, recall from the proof of Lemma B.7 that $\theta^{ub}_P = l'(\tau_P + \delta_{P,post} - \delta^{**}_{P,post})$. From this, we see that

$$\begin{aligned}
\tilde{Y}_m &= \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - \tilde{A}_{(\cdot,1)}\theta^{ub}_{P_m} - \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}(\tau_{P_m} - \delta_{P_m,post} + \delta^{**}_{P_m,post})\right) - \tilde{A}_{(\cdot,1)}x \\
&= \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - \tilde{A}_{(\cdot,1)}l'(\tau_{P_m} + \delta_{P_m,post} - \delta^{**}_{P_m,post}) - \tilde{A}_{(\cdot,-1)}\Gamma_{(-1,\cdot)}(\tau_{P_m} - \delta_{P_m,post} + \delta^{**}_{P_m,post})\right) \\
&\quad - \tilde{A}_{(\cdot,1)}x \\
&= \sqrt{n_m}\left(A\hat{\beta}_{n_m} - d - A\begin{pmatrix} 0 \\ I \end{pmatrix}(\tau_{P_m} + \delta_{P_m,post} - \delta^{**}_{P_m,post})\right) - \tilde{A}_{(\cdot,1)}x,
\end{aligned}$$

where the last line follows from Lemma D.5. Additionally, note that by construction, $\delta_{P_m,pre} = \delta^{**}_{P_m,pre}$. Thus, $\delta_{P_m} - \delta^{**}_{P_m} = \begin{pmatrix} 0 \\ I \end{pmatrix}(\delta_{P_m,post} - \delta^{**}_{P_m,post})$. It follows that

$$\tilde{Y}_m = \sqrt{n_m}A\left(\hat{\beta}_{n_m} - \delta_{P_m} - \begin{pmatrix} 0 \\ \tau_{P_m} \end{pmatrix}\right) + \sqrt{n_m}\left(A\delta^{**}_{P_m} - d\right) - \tilde{A}_{(\cdot,1)}x. \tag{73}$$

Now, since $P_m \in \mathcal{P}_\epsilon$, by definition there exists an index B_m such that

$$\begin{aligned}
A_{(B_m,\cdot)}\delta^{**}_{P_m} - d_{B_m} &= 0 \\
A_{(-B_m,\cdot)}\delta^{**}_{P_m} - d_{-B_m} &< \epsilon,
\end{aligned}$$

and $A_{B_m,post}$ has rank $|B_m|$. Since there are finitely many possible subindices of the rows of A, we can choose a subsequence m_1 such that $B_{m_1} \equiv B$ for some index B such that $A_{B,post}$ has rank $|B|$. Additionally, since $\mathbf{S}$ is compact, we can choose a further subsequence m_2 along which $\Sigma_{P_{m_2}} \to \Sigma^*$ for some $\Sigma^* \in \mathbf{S}$. To avoid notational clutter, we will assume that these convergences hold for the original sequence (m, n_m). Additionally, without loss of generality, we will assume that B corresponds with the first $|B|$ rows of A. It follows that

$$\sqrt{n_m}\left(A\delta^{**}_{P_m} - d\right) - \tilde{A}_{(\cdot,1)}x = \begin{pmatrix} -\tilde{A}_{(B,1)}x \\ \sqrt{n_m}\left(A_{(-B,\cdot)}\delta^{**}_{P_m} - d_{-B}\right) - \tilde{A}_{(-B,1)}x \end{pmatrix}$$
$$\leqslant \begin{pmatrix} -\tilde{A}_{(B,1)}x \\ -\sqrt{n_m}\epsilon - \tilde{A}_{(-B,1)}x \end{pmatrix},$$

from which it is apparent that

$$\sqrt{n_m}\left(A\delta^{**}_{P_m} - d\right) - \tilde{A}_{(\cdot,1)}x \to \begin{pmatrix} -\tilde{A}_{(B,1)}x \\ -\infty \end{pmatrix} =: \bar{\mu}$$

as $m \to \infty$. Now, equation (73) along with Assumptions 6 and 8 and the continous mapping theorem imply that

$$(\tilde{Y}_m, \hat{\Sigma}_m) \to_d (\xi + \bar{\mu}, \Sigma^*),$$

for $\xi \sim \mathcal{N}(0, A\Sigma^* A')$.

Now, as in the proof to Proposition C.1, note that the (unmodified) conditional test rejects iff $p(Y,\Sigma) > 1-\alpha$ for

$$p(Y,\Sigma) := \mathbb{P}\left(\zeta < \hat{\eta}(Y,\Sigma) \,|\, \zeta \in [v^{lo}(Y,\Sigma), v^{up}(Y,\Sigma)], \zeta \sim \mathcal{N}\left(0,\, \sigma^2_\eta(Y,\Sigma)\right)\right) > 1-\alpha.$$

It follows that the modified conditional test rejects iff $\tilde{p}(Y,\Sigma) := p(Y,\Sigma)\cdot 1\left[\hat{\eta}(Y,\Sigma) \geqslant -\underline{C}\right] > 1-\alpha$. Thus, (72) implies that

$$\liminf_{n\to\infty} \left|\mathbb{P}_{P_m}\left(\tilde{p}(\tilde{Y}_m, \hat{\Sigma}_m) > 1-\alpha\right) - \rho^*(P_m)\right| \geqslant \omega.$$

Additionally, Proposition 4.2 implies that for all m, $\rho^*(P_m) = \Phi(c^*x - z_{1-\alpha})$, where $c^* = -\bar{\gamma}'_B\tilde{A}_{(B,1)}/\sigma_B$, for $\sigma_B = \sqrt{\bar{\gamma}'_B A_{(B,\cdot)}\Sigma A'_{(B,\cdot)}\bar{\gamma}_B}$ and $\bar{\gamma}_B$ the unique vector such that $\bar{\gamma}'_B\tilde{A}_{(B,-1)} = 0, \bar{\gamma}_B \geqslant 0, ||\bar{\gamma}_B|| = 1$. Thus,

$$\liminf_{n\to\infty} \left|\mathbb{P}_{P_m}\left(\tilde{p}(\tilde{Y}_m, \hat{\Sigma}_m) > 1-\alpha)\right) - \Phi(c^*x - z_{1-\alpha})\right| \geqslant \omega. \tag{74}$$

However, Lemma D.14 gives that $\tilde{p}(Y,\Sigma)$ is continuous at $(\xi+\bar{\mu},\Sigma^*)$ for almost every $\xi \sim \mathcal{N}(0, A\Sigma^* A')$, and so from the continuous mapping theorem,

$$\tilde{p}(\tilde{Y}, \hat{\Sigma}_m) \to_d \tilde{p}(\xi+\bar{\mu}, \Sigma^*).$$

Additionally, Lemma D.15 gives that the distribution of $\tilde{p}(\xi+\bar{\mu},\Sigma^*)$ is continuous at $1-\alpha$, and thus

$$\mathbb{P}_{P_m}\left(\tilde{p}(\tilde{Y}_m, \hat{\Sigma}_m) > 1-\alpha\right) \to \mathbb{P}\left(\tilde{p}(\xi + \bar{\mu}, \Sigma^*) > 1-\alpha\right).$$

Lemma D.12 implies that with probability 1,

$$p(\xi + \bar{\mu}, \Sigma^*) = \Phi\left(\frac{\gamma_j(\Sigma^*)'(\xi+\bar{\mu})}{\sqrt{\gamma_j(\Sigma^*)'A\Sigma^*A'\gamma_j(\Sigma^*)}}\right),$$

where $\gamma_j(\Sigma^*) = c_j(\Sigma)\bar{\gamma}_j$ for $\bar{\gamma}_j$ the unique element of $\{\bar{\gamma}_1, ..., \bar{\gamma}_J\}$ such that $\bar{\gamma}_{j,-B} = 0$. Additionally, Lemma D.9 gives that with probability 1, $\hat{\eta}(\xi + \bar{\mu}, \Sigma^*) = \gamma_j(\Sigma^*)'(\xi + \bar{\mu})$. Since $-\underset{\sim}{C} \leqslant -C^*$, $\Phi\left(\frac{\hat{\eta}}{\sqrt{\gamma_j(\Sigma^*)'A\Sigma^*A'\gamma_j(\Sigma^*)}}\right) > 1-\alpha$ only if $\hat{\eta} > -\bar{C}$, from which we see that $\mathbb{P}\left(\tilde{p}(\xi+\bar{\mu}, \Sigma^*) > 1-\alpha\right) = \mathbb{P}\left(p(\xi+\bar{\mu}, \Sigma^*) > 1-\alpha\right)$. It follows from the expression for $p(\xi + \bar{\mu}, \Sigma^*)$ in the previous display that with probability 1, $p(\xi + \bar{\mu}, \Sigma^*) > 1-\alpha$ iff

$$\frac{\gamma_j(\Sigma^*)'\xi}{\sqrt{\gamma_j(\Sigma^*)'A\Sigma^*A'\gamma_j(\Sigma^*)}} > z_{1-\alpha} - \frac{\gamma_j(\Sigma^*)'\bar{\mu}}{\sqrt{\gamma_j(\Sigma^*)'A\Sigma^*A'\gamma_j(\Sigma^*)}}.$$

The term on the left-hand side has the standard normal distribution, and thus

$$\mathbb{P}\left(p(\xi+\bar{\mu}, \Sigma^*) > 1-\alpha\right) = \Phi\left(\frac{\gamma_j(\Sigma^*)'\bar{\mu}}{\sqrt{\gamma_j(\Sigma^*)'A\Sigma^*A'\gamma_j(\Sigma^*)}} - z_{1-\alpha}\right).$$

Next, note that by definition $\gamma_j(\Sigma^*) = c_j(\Sigma^*)\bar{\gamma}_j$, where by construction $\bar{\gamma}_{j,-B} = 0$. Further, from Lemma D.7, $\bar{\gamma}_{j,B}$ is equal to the vector $\bar{\gamma}_B$ defined above (i.e. the unique vector satisfying the unique vector such that $\bar{\gamma}_B'\tilde{A}_{(B,-1)} = 0, \bar{\gamma}_B \geqslant 0, ||\bar{\gamma}_B|| = 1$). It is then immediate from the previous display and the fact that $\bar{\mu}_B = -\tilde{A}_{(B,1)}x$ that

$$\mathbb{P}\left(p(\xi+\bar{\mu}, \Sigma^*) > 1-\alpha\right) = \Phi\left(c^*x - z_{1-\alpha}\right).$$

But this implies that

$$\liminf_{n\to\infty}\left|\mathbb{P}_{P_m}\left(\tilde{p}(\tilde{Y}_m, \hat{\Sigma}_m) > 1-\alpha\right) - \Phi(c^*x - z_{1-\alpha})\right| = 0,$$

which contradicts (74). ☐

Lemma D.7. *Suppose Assumption 5 holds. Let $B = B(\delta^{**})$ be the index of the binding moments. Let $\bar{\gamma}_1, ..., \bar{\gamma}_J$ be as defined in Lemma C.1. Then $\bar{\gamma}_{j,-B} = 0$ for exactly one $j \in \{1, ..., J\}$. Additionally, $\bar{\gamma}_j'A \neq 0$, and $\bar{\gamma}_{j,B}$ is the unique vector in the set $\{\gamma_B : \gamma_B'\tilde{A}_{(B,-1)} = 0, \gamma_B \geqslant 0, ||\gamma_B|| = 1\}$.*

Proof. We first show that there can be at most one $\bar{\gamma}_j$ such that $\bar{\gamma}_{j,-B} = 0$. Recall from the proof to Lemma C.1 that for all j, $\bar{\gamma}_j' \tilde{A}_{(\cdot,-1)} = 0, \bar{\gamma}_j \geqslant 0$ and $||\bar{\gamma}_j|| = 1$. Thus, if $\bar{\gamma}_{j,-B} = 0$, we have $\bar{\gamma}_{j,B}' \tilde{A}_{(B,-1)} = 0$. However, from Lemma B.7, the set $\{\bar{\gamma}_B : \bar{\gamma}' \tilde{A}_{(B,-1)} = 0\} = \{c\gamma_B^* \,|\, c \in \mathbb{R}\}$ for some non-zero vector $\gamma_B^* \geqslant 0$. Thus, there is a single vector in the set $\{\gamma_B : \gamma_B' \tilde{A}_{(B,-1)} = 0, \gamma_B \geqslant 0, ||\gamma_B|| = 1\}$. In particular, its lone element is $c^*\gamma_B^*$, for $c^* = 1/||\gamma_B^*||$. Hence, if there is such a $\bar{\gamma}_j$, it has $c^*\gamma_B^*$ in the positions corresponding with B and zeros otherwise.

It thus remains to show that the vector with $c^*\gamma_B^*$ in the positions corresponding with B and zeros otherwise is in the set $\{\bar{\gamma}_1, ..., \bar{\gamma}_J\}$. Denote this vector γ^*. Note that by construction, $\gamma^{*\prime} \tilde{A}_{(\cdot,-1)} = 0$. Thus, for any Σ positive definite, $(\gamma^{*\prime}\tilde{\sigma})^{-1}\gamma^* \in F(\Sigma) = \{\gamma : \gamma' \tilde{A}_{(\cdot,-1)} = 0, \gamma'\tilde{\sigma} = 1\}$. Moreover, $(\gamma^{*\prime}\tilde{\sigma})^{-1}\gamma^*$ must be the unique vector in $F(\Sigma)$ with $\gamma_{-B} = 0$, since as discussed above, $\{\bar{\gamma}_B : \bar{\gamma}' \tilde{A}_{(B,-1)} = 0\} = \{c\gamma_B^* \,|\, c \in \mathbb{R}\}$ and so there is a unique vector with $\gamma_B' \tilde{A}_{(\cdot,-1)} = 0, \gamma \geqslant 0$, and $\gamma'\tilde{\sigma} = 1$. Let ν be the vector with -1 in the positions corresponding with $-B$ and zeros otherwise. Then $\nu'(\gamma^{*\prime}\tilde{\sigma})^{-1}\gamma^* = 0$, whereas $\nu'\gamma < 0$ for any other $\gamma \in F(\Sigma)$, since every $\gamma \in F(\Sigma)$ satisfies $\gamma \geqslant 0$ and $\gamma_{-B} \neq 0$. Thus, $(\gamma^{*\prime}\tilde{\sigma})^{-1}\gamma^*$ is a minimal face of $F(\Sigma)$, and hence a vertex (see Schrijver (1986), Section 8.5). By Lemma C.1, $F(\Sigma) = \{c_1(\Sigma)\bar{\gamma}_1, ..., c_J(\Sigma)\bar{\gamma}_J\}$ where $c_J > 0$. It follows that $(\gamma^{*\prime}\tilde{\sigma})^{-1}\gamma^* = c_j(\Sigma)\bar{\gamma}_j$ for some j, so γ^* is a constant multiple of $\bar{\gamma}_j$. However, since by construction γ^* and $\bar{\gamma}_j$ are both positive and have a norm of 1, they must be equal, which gives the first result.

Next, note that we showed in the proof to Lemma B.7 that $\gamma_B^{*\prime} \tilde{A}_{(B,\cdot)} = e_1'$. Since $\tilde{A}_{(B,\cdot)} = A_{(B,\cdot)} \begin{pmatrix} 0 \\ I \end{pmatrix} \Gamma^{-1}$ and Γ^{-1} is full rank, it follows that $\gamma_B^{*\prime} A_{(B,\cdot)} \neq 0$. Since $\bar{\gamma}_{j,B} = c^*\gamma_B^*$ and $\gamma_{j,-B} = 0$, we have that $\gamma_j' A = c^*\gamma_B^{*\prime} A_{(B,\cdot)} \neq 0$, which gives the second result.

□

Lemma D.8. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3. Let $\hat{V}(Y, \Sigma) = \arg\max_{\gamma \in V(\Sigma)} \gamma' Y$. By Lemma D.7, there is a unique index j such that $\bar{\gamma}_{j,-B} = 0$. Then for almost every $\xi \sim \mathcal{N}(0, A\Sigma^* A')$, there is a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$ such that $\hat{V}(Y, \Sigma) = c_j(\Sigma)\bar{\gamma}_j$ for almost every $\xi \sim \mathcal{N}(0, A\Sigma^* A')$.*

Proof. Without loss of generality, suppose that $\bar{\gamma}_{1,-B} = 0$. Lemma C.1 implies that

$$\hat{\eta}(Y, \Sigma) := \max_{\gamma \in V(\Sigma)} \gamma' Y = \max\{c_1(\Sigma)\bar{\gamma}_1' Y, ..., c_J(\Sigma)\bar{\gamma}_1' Y\},$$

where the functions $c_j(\Sigma)$ are continuous. Each of the elements of the max are continuous functions of (Y, Σ) in a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$ by an argument analogous to that in the proof to Lemma D.2 (replacing μ^* with $\bar{\mu}$). Note, however, that $\bar{\gamma}_1'(\bar{\mu} + \xi) = \bar{\gamma}_{1,B}'(\xi_B + \tilde{A}_{(B,1)}x)$, which is finite with probability 1. On the other hand, for $j > 1$, $\bar{\gamma}_j'(\xi + \bar{\mu}) = -\infty$, since

$\bar{\gamma}_j \geqslant 0$ and has at least one strictly positive element in the index $-B$, and $\mu_{-B} = -\infty$. Since $c_j(\Sigma^*) > 0$ for all j by Lemma D.1, it follows that $c_1(\Sigma^*)\bar{\gamma}_1'(\xi + \bar{\mu}) > c_j(\Sigma^*)\bar{\gamma}_j'(\xi + \bar{\mu})$ for all $j > 2$. Since the functions on both sides of the inequality are continuous at $(\xi + \bar{\mu}, \Sigma^*)$, this implies that $c_1(\Sigma)\bar{\gamma}_1'Y > c_j(\Sigma)\bar{\gamma}_j'Y$ in a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$, which gives the desired result.

□

Lemma D.9. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3. Let $\hat{\eta}(Y, \Sigma) = \max_{\gamma \in V(\Sigma)} \gamma' Y$. Then for almost every $\xi \sim \mathcal{N}(0,\ A\Sigma^* A')$, $\eta(Y, \Sigma)$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$. Further, there is a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$ such that $\hat{\eta}(Y, \Sigma) = c_j(\Sigma)\bar{\gamma}_j'Y$, where j is the unique index such that $\bar{\gamma}_{j,-B} = 0$ (which exists by Lemma D.7).*

Proof. Follows immediately from the proof to Lemma D.8. □

Lemma D.10. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3. Then for almost every $\xi \sim \mathcal{N}(0,\ A\Sigma^* A')$, $\sigma_\eta^2(Y, \Sigma)$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$. Further, there is a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$ such that $\sigma_\eta(Y, \Sigma) = c_j(\Sigma)^2 \bar{\gamma}_j' A \Sigma A' \bar{\gamma}_j > 0$.*

Proof. By Lemma D.7, there is a unique index j such that $\bar{\gamma}_{j,-B} = 0$, and this $\bar{\gamma}_j$ satisfies $\bar{\gamma}_j' A \neq 0$. Lemma D.8 implies that $\hat{V}(Y, \Sigma) = c_j(\Sigma)\bar{\gamma}_j$ in a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$. Thus, in that neighborhood, $\hat{\sigma}_\eta^2(Y, \Sigma) = c_j(\Sigma)^2 \bar{\gamma}_j' A \Sigma A' \bar{\gamma}_j$, which is clearly continuous in Σ. Additionally, $c_j(\Sigma^*) > 0$ by Lemma D.1, and Σ^* is positive definite, so $\hat{\sigma}_\eta^2(\xi + \bar{\mu}, \Sigma^*) = c_j(\Sigma^*)^2 \bar{\gamma}_j' A \Sigma^* A' \bar{\gamma}_j > 0$. Since $\hat{\sigma}_\eta^2$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$, it is also positive in a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$. □

Lemma D.11. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3. Then for almost every $\xi \sim \mathcal{N}(0,\ A\Sigma^* A')$, $v^{lo}(\xi + \bar{\mu}, \Sigma^*) = -\infty$, $v^{up}(\xi + \bar{\mu}, \Sigma^*) = \infty$, and the functions v^{lo} and v^{up} are continuous at $(\xi + \bar{\mu}, \Sigma^*)$.*

Proof. By Lemma D.7, there is a unique index j such that $\bar{\gamma}_{j,-B} = 0$, and this $\bar{\gamma}_j$ satisfies $\bar{\gamma}_j' A \neq 0$. Without loss of generality, assume this holds for $j = 1$. Lemmas D.8 and D.10 then imply that $\hat{V}(Y, \Sigma) = c_1(\Sigma)\bar{\gamma}_1$ and $\hat{\sigma}_\eta^2(Y, \Sigma) > 0$ in a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$.

The proof of the continuity of v^{lo} and v^{up} is then similar to that in Lemma D.2. Let $\gamma_*(\Sigma) = c_1(\Sigma)\bar{\gamma}_1$. For ease of notation, we will make the dependence of γ_* on Σ implicit where it is clear below. Since in a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$, $\hat{\sigma}_\eta^2(Y, \Sigma) > 0$ and $\hat{V}(Y, \Sigma) = \{\gamma_*(\Sigma)\}$, in that neighborhood $v^{lo}(Y, \Sigma)$ is the minimum of the set

$$C = \{c\ :\ \max_{\gamma \in V(\Sigma)} \gamma' \left(S_{\gamma_*}(Y, \Sigma) + \frac{\Sigma \gamma_*}{\gamma_* \Sigma \gamma_*} c \right) \},$$

for

$$S_{\gamma_*}(Y, \Sigma) = \left(I - \frac{\Sigma\gamma_*\gamma_*'}{\gamma_*'\Sigma\gamma_*} \right) Y.$$

Rearranging terms, we see that

$$C = \{c : 0 = \max_{\gamma \in V(\Sigma)} a_{\gamma,\gamma_*,Y,\Sigma} + b_{\gamma,\gamma_*,\Sigma} c\},$$

where $a_{\gamma,\gamma_*,Y,\Sigma} := \gamma' S_{\gamma_*}(Y)$ and $b_{\gamma,\gamma_*,\Sigma} := \frac{\gamma'\Sigma\gamma_*}{\gamma_*'\Sigma\gamma_*} - 1$. Note that $a_{\gamma_*,\gamma_*,Y} = 0 = b_{\gamma_*,\gamma_*}$, so $0 \leqslant \max_{\gamma \in V(\Sigma)} a_{\gamma,\gamma_*,Y} + b_{\gamma,\gamma_*} c$ for all c. Moreover, for $c = \gamma_*' Y$, the max is attained at γ_* by construction. Hence, the set C is non-empty.

Intuitively, if we plot $a_{\gamma,\gamma_*,Y,\Sigma} + b_{\gamma,\gamma_*,\Sigma}$ as a function of c, then each $\gamma \in V(\Sigma)$ defines a line, and the set C represents the values of c for which 0 is the upper envelope of this set. It follows that the lower bound of C is the maximal x-intercept of the lines of the form $a_{\gamma,\gamma_*,Y,\Sigma} + b_{\gamma,\gamma_*,\Sigma} c$ with $b_{\gamma,\gamma_*,\Sigma} < 0$. Hence,

$$v^{lo}(Y, \Sigma) = \max_{\{\gamma \in V(\Sigma) \backslash \{\gamma_*\} : b_{\gamma,\gamma_*,\Sigma} < 0\}} \frac{-\hat{a}_{\gamma,\gamma_*,Y,\Sigma}}{\hat{b}_{\gamma,\gamma_*,\Sigma}}.$$

Now, let $\gamma_{**} = \gamma_*(\Sigma^*)$. Observe that for any $\gamma \in V(\Sigma^*) \backslash \gamma_{**}$,

$$\gamma' \left(I - \frac{\Sigma^*\gamma_{**}\gamma_{**}'}{\gamma_{**}'\Sigma^*\gamma_{**}} \right) (\xi + \bar{\mu}) = \gamma'(\xi + \bar{\mu}) - \frac{\gamma'\Sigma^*\gamma_{**}}{\gamma_{**}'\Sigma^*\gamma_{**}} \gamma_{**}'(\xi + \bar{\mu}).$$

Since $\gamma_{-B} \leqslant 0$ and has at least one strictly positive element, $\gamma'(\xi + \bar{\mu}) = -\infty$ with probability 1. On the other hand, $\gamma_{**,B} = 0$, and so $\gamma_{**}'(\xi + \bar{\mu})$ is finite with probability one. It follows that $a_{\gamma,\gamma_{**},\xi+\bar{\mu},\Sigma^*} = -\infty$ with probability 1. Hence, $v^{lo}(\xi + \bar{\mu}, \Sigma^*) = -\infty$.

Next, recall that by Lemma C.1, $V(\Sigma) := \{\gamma_1(\Sigma), ..., \gamma_J(\Sigma)\}$, where $\gamma_j(\Sigma) := c_j(\Sigma)\bar{\gamma}_j$ and $c_j(\Sigma)$ is continuous. Additionally, we showed in the proof to Lemma D.8 that for all j, $c_j(\Sigma)\bar{\gamma}_j' Y$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$. It is then immediate from the definitions of the functions $a_{\gamma,\gamma_*,Y,\Sigma}$ and $b_{\gamma,\gamma_*,\Sigma}$ that for all j, $a_{\gamma_j(\Sigma),\gamma_*(\Sigma),Y,\Sigma}$ and $b_{\gamma_j(\Sigma),\gamma_*(\Sigma),\Sigma}$ are continuous in (Y, Σ) as well. Without loss of generality, suppose that for $2 \leqslant k \leqslant k_1$, $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*),\Sigma^*} < 0$; for $k_1 < k \leqslant k_2$, $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*),\Sigma^*} = 0$; and for $k > k_2$, $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*),\Sigma^*} > 0$. From the continuity of $b_{\gamma_j(\Sigma),\gamma_*(\Sigma),\Sigma}$, it is clear that in a neighborhood of $(\xi + \mu^*, \Sigma^*)$, $b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma} > 0$ for all $2 \leqslant k \leqslant k_1$ and $b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma} < 0$ for all $k > k_2$. Hence, in this neighborhood,

$$v^{lo}(Y, \Sigma) = \max \left\{ \max_{\gamma_k(\Sigma) : 2 \leqslant k \leqslant k_1} \frac{-a_{\gamma_k(\Sigma),\gamma_*(\Sigma),Y,\Sigma}}{b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma}}, \max_{\gamma \in V^0(\Sigma)} \frac{-a_{\gamma,\gamma_*(\Sigma),Y,\Sigma}}{b_{\gamma,\gamma_*(\Sigma),\Sigma}} \right\}, \quad (75)$$

where

$$V^0(\Sigma) := \{\gamma_k(\Sigma) : k_1 < k \leqslant k_2, b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma} < 0\}$$

and we define the max of an empty set to be $-\infty$. It is clear from the continuity of the functions a and b that the inner max on the left side of (75) is continuous and converges to $-\infty$. To show that v^{lo} is continuous at $(\xi + \bar{\mu}, \Sigma^*)$, it thus suffices to show that for any sequence $(Y, \Sigma) \to (\xi + \bar{\mu}, \Sigma^*)$, the max on the right hand side of (75) converges to $-\infty$. To do this, note that by construction $b_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*),\Sigma^*} = 0$ for $k \in (k_1, k_2]$, and so along any sequence $(Y, \Sigma) \to (\xi + \bar{\mu}, \Sigma^*)$, $b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma} \to 0$ since b is continuous in (Y, Σ). Additionally, since a is continuous, along such a sequence, $a_{\gamma_k(\Sigma),\gamma_*(\Sigma),Y,\Sigma} \to a_{\gamma_k(\Sigma^*),\gamma_*(\Sigma^*),\xi+\bar{\mu},\Sigma} = -\infty$. For values of (Y, Σ) where $b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma} > 0$, it follows that $-a_{\gamma_k(\Sigma),\gamma_*(\Sigma),Y,\Sigma}/b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma}$ becomes arbitrarily negative, whereas for values of (Y, Σ) where $b_{\gamma_k(\Sigma),\gamma_*(\Sigma),\Sigma} \geqslant 0$, γ_k is not included in V^0. It is then immediate that the max on the right hand side of (75) converges to $-\infty$, which suffices to establish the continuity of v^{lo} at $(\xi + \bar{\mu}, \Sigma^*)$. The continuity of v^{up} can be shown analogously.

□

Lemma D.12. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3. Define $p(Y, \Sigma)$ as in Lemma D.3. Then for almost every $\xi \sim \mathcal{N}(0, A\Sigma^* A')$, $p(Y, \Sigma)$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$, and $p(\xi + \bar{\mu}, \Sigma^*) = \Phi\left(\frac{\gamma_j(\Sigma^*)'(\xi + \bar{\mu})}{\sqrt{\gamma_j(\Sigma^*)' A\Sigma^* A' \gamma_j(\Sigma^*)}}\right)$, where j is the unique index such that $\bar{\gamma}_{j,B} = 0$ (which exists by Lemma D.7).*

Proof. Lemmas D.9 to D.11 imply that for almost every ξ, $\hat{\eta}(Y, \Sigma)$, $\sigma_\eta^2(Y, \Sigma)$, $v^{lo}(Y, \Sigma)$ and $v^{up}(Y, \Sigma)$ are continuous at $(\xi+\bar{\mu}, \Sigma^*)$, and when evaluated at $(\xi+\bar{\mu}, \Sigma^*)$, $\hat{\eta} = c_j(\Sigma^*)\bar{\gamma}_j'(\xi+\bar{\mu})$, $\hat{\sigma}_\eta^2 = c_j(\Sigma^*)^2 \bar{\gamma}_j' A\Sigma A' \bar{\gamma}_j > 0$, $v^{lo} = -\infty$, and $v^{up} = \infty$. Thus, $\hat{\sigma}_\eta > 0$ and $v^{lo} < v^{up}$ in a neighborhood of $(\xi + \bar{\mu}, \Sigma^*)$. When $\hat{\sigma}_\eta^2 > 0$ and $v^{lo} < v^{up}$,

$$p(Y, \Sigma) = \frac{\Phi(\hat{\eta}/\hat{\sigma}_\eta) - \Phi(v^{lo}/\hat{\sigma}_\eta)}{\Phi(v^{up}/\hat{\sigma}_\eta) - \Phi(v^{lo}/\hat{\sigma}_\eta)},$$

which is clearly continuous in $\hat{\eta}, v^{lo}, v^{up}$, and $\hat{\sigma}_\eta$, including when $v^{lo} = -\infty$ and $v^{up} = \infty$. The continuity of $p(Y, \Sigma)$ thus follows from the continuity of $\hat{\eta}, v^{lo}, v^{up}$, and $\hat{\sigma}_\eta$.

Additionally, when evaluated at $(Y, \Sigma) = (\xi + \bar{\mu}, \Sigma^*)$, we have

$$p(Y, \Sigma) = \frac{\Phi\left(\frac{\gamma_j(\Sigma^*)'(\xi + \bar{\mu})}{\sqrt{\gamma_j(\Sigma^*)' A\Sigma^* A' \gamma_j(\Sigma^*)}}\right) - \Phi(-\infty)}{\Phi(\infty) - \Phi(-\infty)} = \Phi\left(\frac{\gamma_j(\Sigma^*)'(\xi + \bar{\mu})}{\sqrt{\gamma_j(\Sigma^*)' A\Sigma^* A' \gamma_j(\Sigma^*)}}\right).$$

□

Lemma D.13. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3. For any $\bar{C} \in \mathbb{R}$, the function $1\left[\hat{\eta}(Y, \Sigma) \geqslant -\underline{C}\right]$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$ for almost every $\xi \sim \mathcal{N}(0, A\Sigma^* A')$.*

Proof. By Lemma D.9, for almost every ξ, the function $\hat{\eta}(Y, \Sigma)$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$. It thus suffices to show that for almost every ξ, $\hat{\eta}(\xi + \bar{\mu}, \Sigma^*) \neq -\bar{C}$. Lemma D.9 gives that $\hat{\eta}(\xi + \bar{\mu}, \Sigma^*) = c_j(\Sigma^*)\bar{\gamma}_j'(\xi + \bar{\mu})$ where $\bar{\gamma}_j$ is the unique element of $\{\bar{\gamma}_1, ..., \bar{\gamma}_J\}$ such that $\bar{\gamma}_{j,-B} = 0$. Thus, $\hat{\eta}(\xi + \bar{\mu}, \Sigma^*) = -\underline{C}$ only if $c_j(\Sigma^*)\bar{\gamma}_j'\xi = -\underline{C} - c_j(\Sigma^*)\bar{\gamma}_j'\bar{\mu}$, where the right-hand side of the previous equation is finite since $\bar{\mu}_B$ is finite and $\bar{\gamma}_{j,-B} = 0$. Observe further that $c_j(\Sigma^*)\bar{\gamma}_j'\xi$ is normally distributed with variance $c_j(\Sigma^*)^2\bar{\gamma}_j'A\Sigma^*A'\bar{\gamma}_j > 0$. Since $c_j(\Sigma^*)\bar{\gamma}_j'\xi$ is continuously distributed, it follows that $c_j(\Sigma^*)\bar{\gamma}_j'\xi = -\underline{C} - c_j(\Sigma^*)\bar{\gamma}_j'\bar{\mu}$ with probability zero, which suffices for the result. □

Lemma D.14. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3. Let the function $p(Y, \Sigma)$ be as defined in Lemma D.12. For any $\bar{C} \in \mathbb{R}$, the function $\tilde{p}(Y, \Sigma) := p(Y, \Sigma) \cdot 1\left[\hat{\eta}(Y, \Sigma) \geqslant -\underline{C}\right]$ is continuous at $(\xi + \bar{\mu}, \Sigma^*)$ for almost every $\xi \sim \mathcal{N}(0, A\Sigma^* A')$.*

Proof. Follows immediately from Lemmas D.12 and D.13 and the fact that the product of continuous functions is continuous. □

Lemma D.15. *Let $\bar{\mu}$ and Σ^* be as defined in the proof to Proposition C.3 and $\tilde{p}(Y, \Sigma)$ as defined in Lemma D.14. For $\xi \sim \mathcal{N}(0, A\Sigma^* A')$, $\tilde{p}(\xi + \bar{\mu}, \Sigma^*) = 1 - \alpha$ with probability 0.*

Proof. Note that $\tilde{p}(Y, \Sigma) := p(Y, \Sigma)1[\hat{\eta}(Y, \Sigma) \geqslant -\underline{C}]$ can equal $1 - \alpha$ only if $1[\hat{\eta}(Y, \Sigma) \geqslant -\underline{C}] = 1$ and $p(Y, \Sigma) = 1 - \alpha$. It thus suffices to show that $p(\xi + \bar{\mu}, \Sigma^*) = 1 - \alpha$ with probability zero. From Lemma D.12, for almost every ξ, $p(\xi + \bar{\mu}, \Sigma^*) = \Phi\left(\frac{\gamma_j(\Sigma^*)'(\xi + \bar{\mu})}{\sqrt{\gamma_j(\Sigma^*)'A\Sigma^*A'\gamma_j(\Sigma^*)}}\right)$, where $\gamma_j(\Sigma^*) := c_j(\Sigma^*)\bar{\gamma}_j$ and $\bar{\gamma}_j$ is the unique element of $\{\bar{\gamma}_1, ..., \bar{\gamma}_J\}$ such that $\bar{\gamma}_{j,-B} = 0$. Thus, $p(\xi + \bar{\mu}, \Sigma^*) = 1 - \alpha$ iff $\gamma_j(\Sigma^*)'\xi = z_{1-\alpha}\sqrt{\gamma_j(\Sigma^*)'A\Sigma^*A'\gamma_j(\Sigma^*)} - \gamma_j(\Sigma^*)'\bar{\mu}$. However, we showed in the proof to Lemma D.13 that $\gamma_j(\Sigma^*)'\xi$ is continuously distributed, and thus this occurs with probability 0. □

Lemma D.16. *Let $\bar{\gamma}_1, ..., \bar{\gamma}_J$ be as defined in Lemma C.1, and $\gamma_j(\Sigma) := c_j(\Sigma)\bar{\gamma}_j$. There exists a value $C^* \in \mathbb{R}$ such that for any $\Sigma \in \mathbf{S}$ and any j such that $\bar{\gamma}_j'A \neq 0$,*

$$\Phi\left(\frac{\hat{\eta}}{\sqrt{\gamma_j(\Sigma)'A\Sigma A'\gamma_j(\Sigma)}}\right) > 1 - \alpha$$

only if $\hat{\eta} > C^*$.

Proof. Observe that

$$\Phi\left(\frac{\hat{\eta}}{\sqrt{\gamma_j(\Sigma)'A\Sigma A'\gamma_j(\Sigma)}}\right) > 1-\alpha$$

iff

$$\hat{\eta} > z_{1-\alpha}\sqrt{\gamma_j(\Sigma)'A\Sigma A'\gamma_j(\Sigma)}.$$

If $z_{1-\alpha} \geqslant 0$, then the lower bound in the previous display is weakly greater than zero. On the other hand if $z_{1-\alpha} < 0$, then the lower bound is weakly greater than $z_{1-\alpha}$ times the maximum possible value of $\sqrt{\gamma_j(\Sigma)'A\Sigma A'\gamma_j(\Sigma)}$. Note, however, that $\sqrt{\gamma_j(\Sigma)'A\Sigma A'\gamma_j(\Sigma)} = \sqrt{c_j(\Sigma)^2\bar{\gamma}_j'A\Sigma A'\bar{\gamma}_j}$ by Lemma C.1. By Lemma D.1, $c_j(\Sigma) \leqslant \bar{c}$. Additionally, since the set $\{\bar{\gamma}_1, ..., \bar{\gamma}_J\}$ is finite, $\max_j ||\bar{\gamma}_j'A||^2$ is finite. It then follows from Lemma D.1 that $\bar{\gamma}_j'A\Sigma A'\bar{\gamma}_j \leqslant \bar{\lambda}\max_j ||\bar{\gamma}_j'A||^2 < \infty$, and so we obtain a finite upper bound on $\sqrt{\gamma_j(\Sigma)'A\Sigma A'\gamma_j(\Sigma)}$, which suffices for the result.

□

Proof of Proposition C.4

Proof. We first claim that the function $m(\beta) = A\beta$ is a maximal invariant of the group G. Since by definition $Av = 0$ for any $v \in A^{\perp}$, it is immediate that $m(\beta) = m(g_v\beta)$ for any $g_v \in G$. To show that m is a maximal invariant, consider β_1 and β_2 such that $m(\beta_1) = m(\beta_2)$. Then $A(\beta_1 - \beta_2) = 0$ and hence $(\beta_1 - \beta_2) \in A^{\perp}$. From this we see that $\beta_1 = \beta_2 + (\beta_1 - \beta_2) = g_{(\beta_1-\beta_2)}(\beta_2)$, and thus $m(\beta)$ is a maximal invariant. Note further that $A\beta_1 = A\beta_2$ iff $A\beta_1 + h = A\beta_2 + h$ for any constant vector h, and so the same argument applies to show that $m_n(\beta) = A\beta + h_n$ is maximal for any h_n. It follows from Theorem 1 in Lehmann (1986, p. 285) that $\mathcal{C}_n$ can be written as a function of $(m_n(\beta), \hat{\Sigma})$ only, so that $\mathcal{C}_n(\sqrt{n}\hat{\beta}_n, \hat{\Sigma}_n) = \tilde{\mathcal{C}}_n(m_n(\sqrt{n}\hat{\beta}_n), \hat{\Sigma}_n)$. From Lemma B.7, there exists a vector $\tilde{\tau}$ such that

$$A_{(B,\cdot)}\beta_{P*} - d_B - \tilde{A}_{(B,1)}\theta^{ub}_{P*} - \tilde{A}_{(B,-1)}\tilde{\tau} = 0 \tag{76}$$

$$A_{(-B,\cdot)}\beta_{P*} - d_{-B} - \tilde{A}_{(-B,1)}\theta^{ub}_{P*} - \tilde{A}_{(-B,1)}\tilde{\tau} = -\epsilon < 0. \tag{77}$$

We set the constant $h_n = -\sqrt{n}[d - \tilde{A}_{(\cdot,1)}\theta^{ub}_{P*} - \tilde{A}_{(\cdot,-1)}\tilde{\tau}]$, so that $\tilde{\mathcal{C}}$ is a function of $Y_n := \sqrt{n}[A\hat{\beta}_n - d - \tilde{A}_{(\cdot,1)}\theta^{ub} - \tilde{A}_{(\cdot,-1)}\tilde{\tau}]$ and $\hat{\Sigma}_n$.

Observe that

$$Y_n = \sqrt{n}A(\hat{\beta} - \beta_{P*}) - \sqrt{n}[A\beta_{P*} - d - \tilde{A}_{(\cdot,-1)}\tilde{\tau}].$$

It follows immediately from (76) and (77) that $\sqrt{n}[A\beta_{P*} - d - \tilde{A}_{(\cdot,-1)}\tilde{\tau}] \to \bar{\mu}$, where $\bar{\mu}_B = 0$ and $\bar{\mu}_{-B} = -\infty$. Since by assumption $\sqrt{n}(\hat{\beta}_n - \beta_{P*}) \to_d \mathcal{N}(0, \Sigma^*)$ under P^*, the continuous mapping theorem along with Slutsky's lemma imply that $Y_n \xrightarrow{P^*}_d \xi + \bar{\mu}$ for $\xi \sim \mathcal{N}(0, A\Sigma^* A')$.

Similarly, suppose $\beta_{P_n} = \beta_{P^*} + \frac{1}{\sqrt{n}}(\tilde{\beta} - \beta_{P^*})$ for some fixed $\tilde{\beta}$. Suppose further that $\sqrt{n}(\hat{\beta}_n - \beta_{P_n}) \xrightarrow{P_n}_d \mathcal{N}(0, \Sigma^*)$. Observe that

$$Y_n = \sqrt{n}A(\hat{\beta} - \beta_{P_n}) + A(\tilde{\beta} - \beta_{P^*}) - \sqrt{n}[A\beta_{P^*} - d - \tilde{A}_{(\cdot,-1)}\tilde{\tau}].$$

Thus, $Y_n \xrightarrow{P_n}_d \xi + A(\tilde{\beta} - \beta_{P^*}) + \bar{\mu}$.

Now, as in Lemma B.12, let $\mathcal{B}_0(\bar{\theta}) := \{\beta : \exists \tau \text{ s.t. } l'\tau = \bar{\theta}, A\beta - d - A \begin{pmatrix} 0 \\ \tau \end{pmatrix} \leqslant 0\}$ be the set of values β consistent with $\theta = \bar{\theta}$, and $\mathcal{B}_0^B(\bar{\theta}) = \{\beta : \exists \tau \text{ s.t. } l'\tau = \bar{\theta}, A_{(B,\cdot)}\beta - d_B - A_{(B,\cdot)} \begin{pmatrix} 0 \\ \tau \end{pmatrix} \leqslant 0\}$ be the analogous set using only the moments B. Suppose that $\tilde{\beta} \in \mathcal{B}_0^B(\theta^{ub} + x)$. We claim that for n sufficiently large, $\beta_n := \beta_{P^*} + \frac{1}{\sqrt{n}}(\tilde{\beta} - \beta_{P^*}) \in \mathcal{B}_0(\theta^{ub} + \frac{1}{\sqrt{n}}x)$. It follows from the definition of $\mathcal{B}_0^B(\theta^{ub} + x)$ and the construction of the matrix $\tilde{A}$ that there exists $\breve{\tau}$ such that $A_{(B,\cdot)}\tilde{\beta} - d_B - \tilde{A}_{(B,1)}(\theta^{ub}_{P^*} + x) - \tilde{A}_{(B,-1)}\breve{\tau} \leqslant 0$. This, combined with (76), implies that

$$A_{(B,\cdot)}\beta_n - d_B - \tilde{A}_{(B,1)}(\theta^{ub}_{P^*} + \frac{1}{\sqrt{n}}x) - \tilde{A}_{(B,-1)}((1 - \frac{1}{\sqrt{n}})\tilde{\tau} + \frac{1}{\sqrt{n}}\breve{\tau}) \leqslant 0.$$

However, from (77), it follows that

$$A_{(-B,\cdot)}\beta_n - d_{-B} - \tilde{A}_{(-B,1)}(\theta^{ub}_{P^*} + \frac{1}{\sqrt{n}}x) - \tilde{A}_{(-B,1)}((1 - \frac{1}{\sqrt{n}})\tilde{\tau} + \frac{1}{\sqrt{n}}\breve{\tau}) =$$
$$(1 - \frac{1}{\sqrt{n}})(-\epsilon) + \frac{1}{\sqrt{n}}\left(A_{(-B,\cdot)}\tilde{\beta} - d_B - \tilde{A}_{(-B,1)}(\theta^{ub}_{P^*} + x) - \tilde{A}_{(-B,1)}\breve{\tau}\right),$$

which is negative for n sufficiently large since $-\epsilon < 0$. The previous two displays imply that for n sufficiently large, $\beta_n \in \mathcal{B}_0(\theta^{ub}_{P^*} + \frac{1}{\sqrt{n}}x)$, as we desired to show. Hence, for n sufficiently large, there exists $\delta_n \in \Delta$ and τ_n such that $\beta_n = \delta_n + \begin{pmatrix} 0 \\ \tau_n \end{pmatrix}$ and $l'\tau_n = \theta^{ub} + \frac{1}{\sqrt{n}}x$.

Now, let $\varphi_n(Y_n, \hat{\Sigma}_n) = 1[\theta^{ub}_{P^*} + \frac{1}{\sqrt{n}}x \in \tilde{\mathcal{C}}_n(Y_n, \hat{\Sigma}_n)]$. It follows from the previous paragraph along with the assumptions of the proposition that for any sequence P_n such that $\sqrt{n}(\hat{\beta}_n - \beta_{P_n}) \xrightarrow{P_n}_d \mathcal{N}(0, \Sigma^*)$, $\hat{\Sigma}_n \xrightarrow{P_n}_p \Sigma^*$, and $\beta_{P_n} = \beta_{P^*} + \frac{1}{\sqrt{n}}(\tilde{\beta} - \beta_{P^*})$ for $\tilde{\beta} \in \mathcal{B}_0^B(\theta^{ub}_{P^*} + x)$, we have that

$$\limsup_{n\to\infty} \mathbb{E}_{P_n}\left[\varphi_n(Y_n, \hat{\Sigma}_n)\right] \leqslant \alpha.$$

It then follows from Theorem 1 in Müller (2011) that

$$\limsup_{n\to\infty} \mathbb{E}_{P*}\left[\varphi_n(Y_n, \hat{\Sigma}_n)\right] \leqslant \bar{\rho},$$

for $\bar{\rho}$ the power of the most powerful test between

$$H_0 : \tilde{\beta} \in \mathcal{B}_0^B(\theta^{ub} + x) \text{ vs. } H_1 : \tilde{\beta} = \beta_{P*}$$

given a single observation $Y \sim \mathcal{N}\left(\bar{\mu} + A(\tilde{\beta} - \beta_{P*}),\, A\Sigma^* A'\right)$.[46] Since $\mu_{-B} = -\infty$, $Y_{-B} = -\infty$ with probability 1 under both the null and alternative, so it suffices to consider tests of H_0 vs H_1 given an observation $Y_B \sim \mathcal{N}\left(\bar{\mu}_B + A_{(B,\cdot)}(\tilde{\beta} - \beta_{P*}),\, A_{(B,\cdot)}\Sigma^* A'_{(B,\cdot)}\right)$. Recalling that $\bar{\mu}_B = 0$ by construction, we see that $\bar{\rho}$ is the power of the most powerful test between $H_0 : \mu \in M_0 := \{A_{(B,\cdot)}(\tilde{\beta} - \beta_{P*}) \,:\, \tilde{\beta} \in \mathcal{B}_0^B(\theta^{ub}_{P*} + x)\}$ and $H_1 : \mu = 0$ given $Y \sim \mathcal{N}\left(\mu,\, A_{(B,\cdot)}\Sigma^* A'_{(B,\cdot)}\right)$.

Now, it follows from the proof to Lemma B.12 that

$$\mathcal{B}_0^B(\theta^{ub}_{P*} + x) = \{\beta \,:\, \bar{\gamma}'_B\left(A_{(B,\cdot)}\beta - d_B - \tilde{A}_{(B,1)}(\theta^{ub}_{P*} + x)\right) \leqslant 0\},$$

for $\bar{\gamma}_B$ the unique vector such that $\bar{\gamma}'_B\tilde{A}_{(B,-1)} = 0$, $\bar{\gamma}_B \geqslant 0$, $||\bar{\gamma}_B|| = 1$. This, combined with (76) and the fact that $\bar{\gamma}'\tilde{A}_{(B,-1)} = 0$, implies that $\mathcal{B}_0^B(\theta^{ub}_{P*} + x) = \{\beta \,:\, \bar{\gamma}'_B\left(A_{(B,\cdot)}(\beta - \beta_{P*})\right) \leqslant \bar{\gamma}'_B\tilde{A}_{(B,1)}x\}$. It is then immediate that $M_0 \subseteq \{v \,:\, \bar{\gamma}'_B v \leqslant \bar{\gamma}'_B\tilde{A}_{(B,1)}x\}$. Additionally, since δ_{P*} satisfies Assumption 5, $A_{(B,\cdot)}$ has rank B, and thus its image is $\mathbb{R}^{|B|}$. This implies inclusion in the opposite direction, and hence $M_0 = \{v \,:\, \bar{\gamma}'_B v \leqslant \bar{\gamma}'_B\tilde{A}_{(B,1)}x\}$. It then follows from Lemma B.11 that $\bar{\rho} = \Phi\left(-\bar{\gamma}'_B\tilde{A}_{(B,1)}x/\sigma^*_B - z_{1-\alpha}\right)$, for $\sigma^*_B = \sqrt{\bar{\gamma}'_B A_{(B,\cdot)}\Sigma^* A'_{(B,\cdot)}\bar{\gamma}_B}$. This accords with the formula for $\rho^*(P^*, x)$ given in Proposition 4.2, which completes the proof. □

E Additional Simulation Results

This section contains additional simulation results that complement the simulations presented in the main text. Section E.1 describes the computation of the optimal bound for expected excess length. Section E.2 contains additional results from the normal data-generating process considered in the main text. Section E.3 presents results from a non-normal data-generating process in which the covariance matrix is estimated from the data.

[46] See also Section 3.2 of Müller (2011) on applying Theorem 1 to invariant tests.

E.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 (9). In Section 6, 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 Kolesar (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 E.1. *Suppose that Δ is convex. Let $\mathcal{I}_\alpha$ denote the set of confidence sets that satisfy the coverage requirement (9). Then, for any $\delta_A \in \Delta$ and $\tau_A \in \mathbb{R}^{\bar{T}}$,*

$$\inf_{\mathcal{C}\in\mathcal{I}_\alpha} \mathbb{E}_{(\delta_A,\tau_A,\Sigma_n)}\left[\lambda(\mathcal{C})\right] = (1-\alpha)\mathbb{E}\left[\bar{\omega}(z_{1-\alpha}-Z) - \underline{\omega}(z_{1-\alpha}-Z) \,|\, Z < z_{1-\alpha}\right],$$

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

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

for $\beta_A := \delta_A + M_{post}\tau_A$, and $||x||_\Sigma = x'\Sigma^{-1}x$.

The proof of this result follows from observing that the confidence set that optimally directs power against (δ_A, τ_A) inverts Neyman-Pearson tests of $H_0 : \delta \in \Delta, \theta = \bar{\theta}$ against $H_A : (\delta,\tau) = (\delta_A,\tau_A)$ 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 $\bar{\theta} \in \mathcal{S}(\Delta, \beta_A)$, 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 $(1-\alpha)\cdot LID(\Delta, \delta_{A,pre})$ shorter than the lowest achievable expected length, where as in Section 3, LID denotes the length of the identified set.

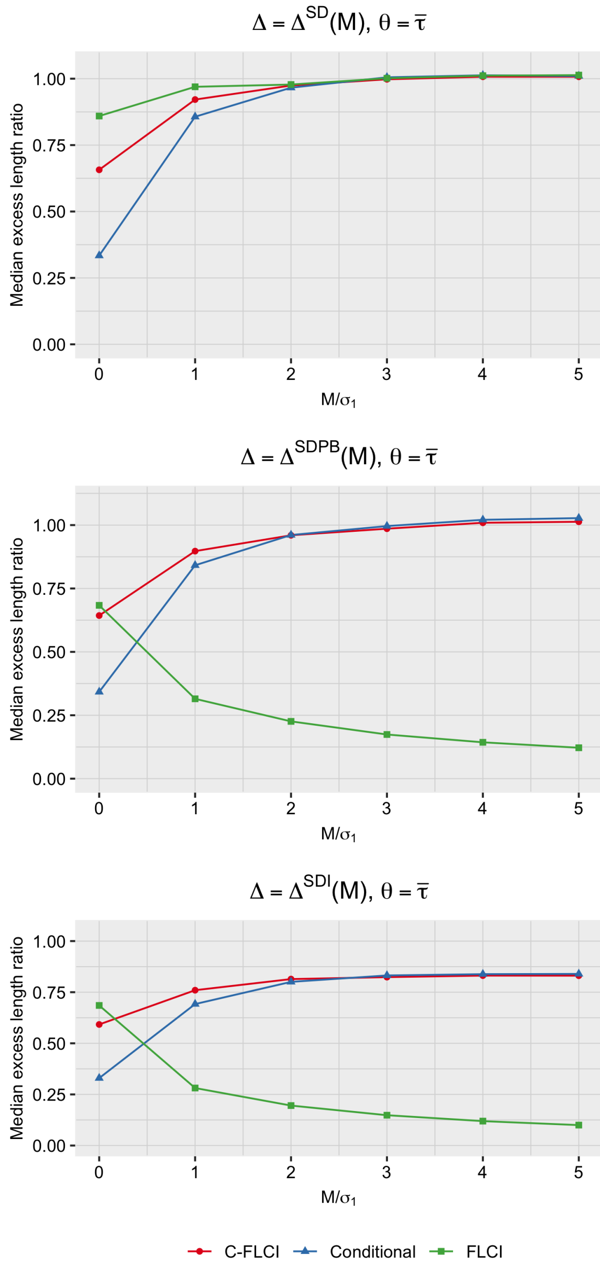
Corollary E.1. *Under the conditions of Lemma E.1,*

$$\inf_{\mathcal{C}\in\mathcal{I}_\alpha} \mathbb{E}_{(\delta_A,\tau_A,\Sigma_n)}\left[EL(\mathcal{C};\delta_A,\tau_A)\right] = \inf_{\mathcal{C}\in\mathcal{I}_\alpha} \mathbb{E}_{(\delta_A,\tau_A,\Sigma_n)}\left[\lambda(\mathcal{C})\right] - (1-\alpha)LID(\Delta,\delta_{A,pre}).$$

E.2 Additional Results for Normal Simulations

In the main text, we report efficiency in terms of excess length for the median paper considered in our simulations. Figures I1 show results using the average of the post-period causal effects as the target parameter, rather than the first period after treatment; that is, $\theta = \bar{\tau}_{post}$.

Figure I1: Median efficiency ratios for proposed procedures when $\theta = \bar{\tau}_{post}$.



Note: This figure shows the median efficiency ratios for our proposed confidence sets for $\theta = \bar{\tau}_{post}$. 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 green, the results for the conditional-FLCI hybrid confidence interval in red and the results for the conditional confidence interval in blue. Results are averaged over 1000 simulations for each of the 12 papers surveyed, and the median across papers is reported here.

E.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 considerations 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.

E.3.1 Simulation design

The simulations are calibrated using the empirical distribution of the data in Bailey and Goodman-Bacon (2015).[47] 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 parallel trends holds, $\hat{\beta}_b^{centered} = \hat{\beta}_b - \hat{\beta}$. We then construct our proposed confidence sets for bootstrap draw b using the pair $(\hat{\beta}_b^{centered}, \hat{\Sigma}_b)$.

We focus on three choices of Δ to highlight the performance of the proposed confidence sets under a range of conditions: $\Delta^{SD}(M), \Delta^{SDPB}(M)$ and $\Delta^{SDI}(M)$. The parameter of interest in these simulations is the causal effect in the first post-period ($\theta = \tau_1$). We report the performance of the FLCI, conditional confidence set, and conditional-FLCI hybrid confidence set. All results are averaged over 1000 bootstrap samples.

E.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 $\mathcal{S}(\Delta, 0)$. We report results for each choice of Δ and $M = 0, 1, 2, 3, 4, 5$. The table shows that all our procedures approximately control size, with null rejection rates never substantially exceeding the nominal rate of 0.05.

[47] 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.

Δ	M	Conditional	FLCI	C-F Hybrid
$\Delta^{SD}(M)$				
	0	0.06	0.08	0.08
	1	0.05	0.04	0.05
	2	0.05	0.05	0.05
	3	0.05	0.07	0.04
	4	0.04	0.06	0.04
	5	0.04	0.06	0.04
$\Delta^{SDPB}(M)$				
	0	0.06	0.08	0.08
	1	0.05	0.04	0.05
	2	0.05	0.04	0.04
	3	0.05	0.08	0.06
	4	0.04	0.05	0.04
	5	0.04	0.05	0.04
$\Delta^{SDI}(M)$				
	0	0.07	0.08	0.08
	1	0.06	0.04	0.06
	2	0.06	0.04	0.07
	3	0.08	0.08	0.08
	4	0.07	0.05	0.07
	5	0.08	0.08	0.08

Table 2: Maximum null rejection probability over the identified set using the empirical distribution from Bailey and Goodman-Bacon (2015).

E.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. Figure I3 shows the rejection probabilities at different values of the parameter θ using both simulation methods. Specifically, we plot results for each choice of Δ using $M = 0$ and $M = 5$. (The results are quite similar for all values of M considered, and we thus omit the intermediate values.) As can be seen, 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.

Figure I2: Comparison of rejection probabilities using bootstrap and normal simulations. Results are shown for $\theta = \tau_1$, and each choice of $\Delta = \Delta^{SD}(M), \Delta^{SDPB}(M), \Delta^{SDI}(M)$, and $M = 0$. 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.

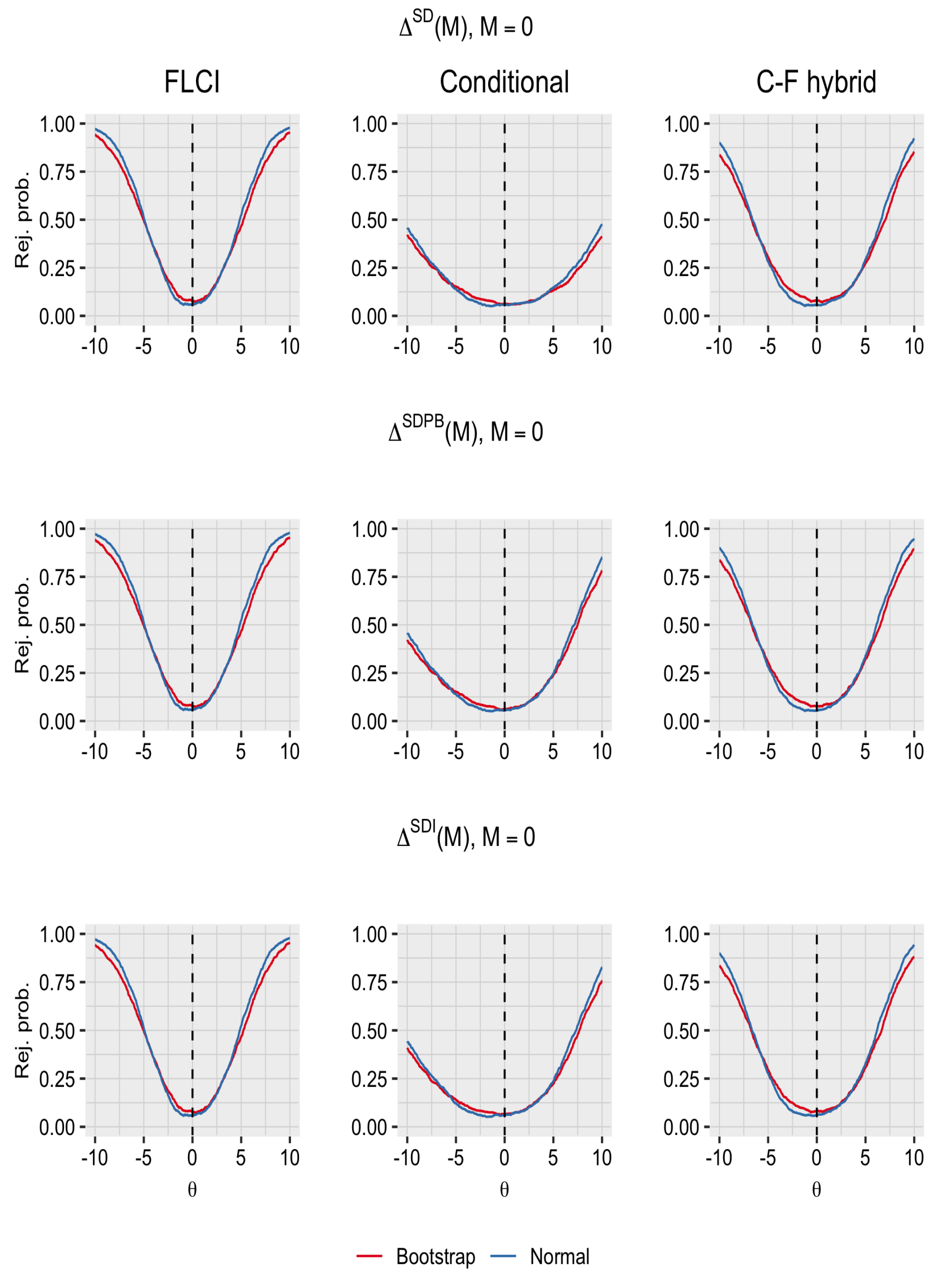
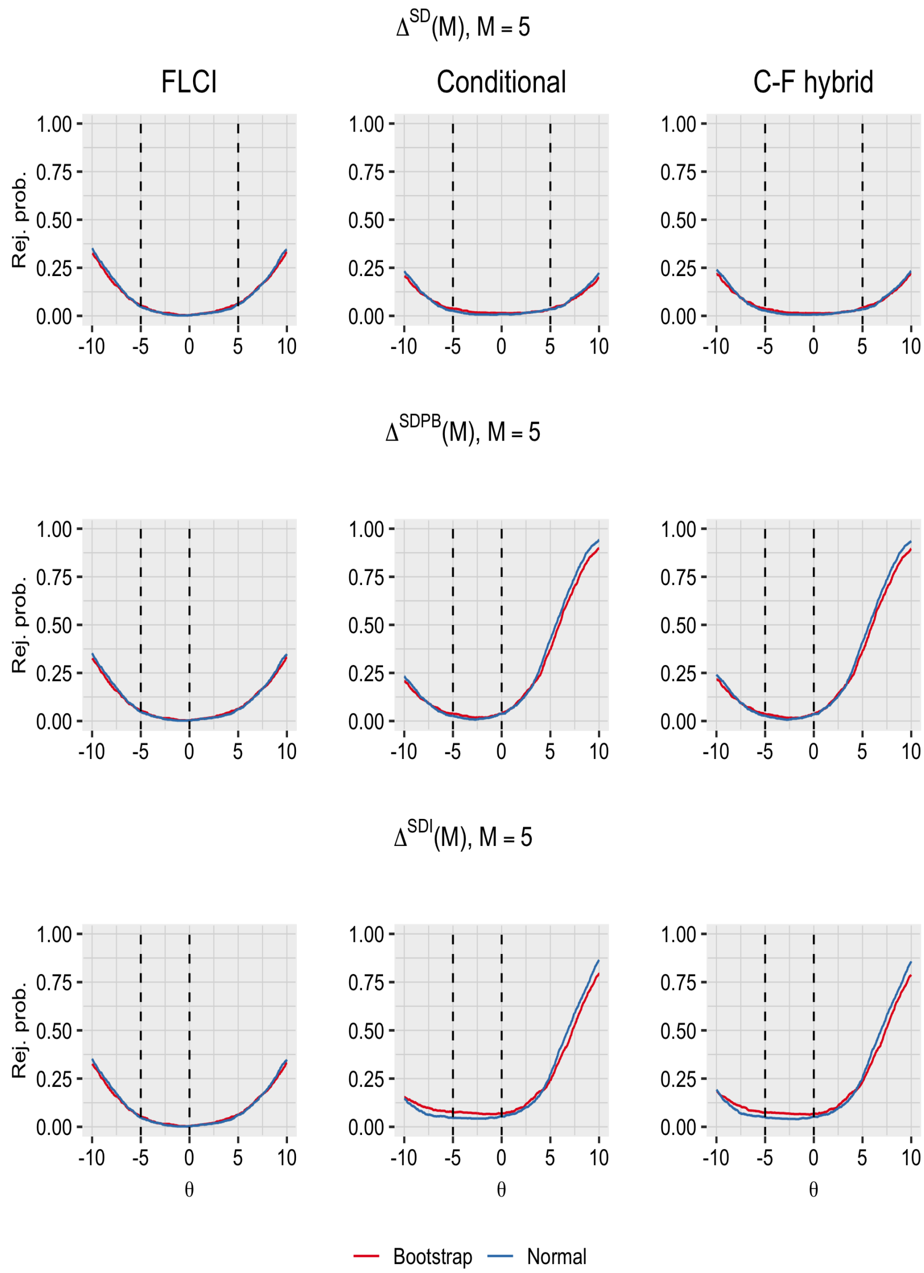


Figure I3: Comparison of rejection probabilities using bootstrap and normal simulations. Results are shown for $\theta = \tau_1$, and each choice of $\Delta = \Delta^{SD}(M), \Delta^{SDPB}(M), \Delta^{SDI}(M)$, and $M = 5$. The average rejection rate for the non-normal simulations are in red and the 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.



Online Supplement References

Andrews, Donald W. K., Xu Cheng, and Patrik Guggenberger, "Generic results for establishing the asymptotic size of confidence sets and tests," *Journal of Econometrics*, October 2020, *218* (2), 496–531.

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

_ **and** _ , "Sensitivity Analysis using Approximate Moment Condition Models," *Quantitative Economics*, 2020. Forthcoming.

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.

Cho, JoonHwan and Thomas M. Russell, "Simple Inference on Functionals of Set-Identified Parameters Defined by Linear Moments," October 2018.

Gafarov, Bulat, "Inference in high-dimensional set-identified affine models," *arXiv:1904.00111 [econ.EM]*, 2019.

Hansen, Bruce E. and Seojeong Lee, "Asymptotic theory for clustered samples," *Journal of Econometrics*, June 2019, *210* (2), 268–290.

Hansen, Christian B., "Asymptotic properties of a robust variance matrix estimator for panel data when T is large," *Journal of Econometrics*, December 2007, *141* (2), 597–620.

Lehmann, Erich Leo, *Testing Statistical Hypotheses*, Wiley, January 1986.

Müller, Ulrich K, "Efficient tests under a weak convergence assumption," *Econometrica*, 2011, *79* (2), 395–435.

Schrijver, Alexander, *Theory of Linear and Integer Programming*, Wiley-Interscience, 1986.

Stock, James and Mark Watson, "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression," *Econometrica*, 2008, *76* (1), 155–174.

van der Vaart, Aad W and Jon A Wellner, "Weak Convergence and Empirical Processes: With Applications to Statistics," 1996.