Supplement to the paper

Inference for Linear Conditional Moment Inequalities

Isaiah Andrews Jonathan Roth Ariel Pakes

August 18, 2022

This supplement provides proofs and additional results for the paper "Inference for Linear Conditional Moment Inequalities." Appendix A proves the results stated in the main text. Appendix B proves validity of our tests in the finite-sample normal model when the dual problem has a non-unique solution. Appendix C discusses an estimator for the variance $\Omega(P_{D|Z},\beta_0)$, and provides sufficient conditions for it to be uniformly consistent. Appendix D provides sufficient conditions for Assumption 4 in the main text. Appendix E discusses how to quickly compute the bounds $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$ used by the conditional and hybrid tests. Finally, Appendix F discusses connections to LICQ conditions considered in the previous literature, while Appendix G provides further details on our simulations.

A Proofs for Results in Main Text

Proof of Lemma 1 Observe that $\hat{\gamma}=\gamma$ only if $Y_{n,0}$ lies in the polyhedron $\{y:(\gamma-\tilde{\gamma})'y\geq 0,\forall\tilde{\gamma}\in V(X_{n,0},\sigma_0)\}$. The result is then immediate from Lemma 5.1 in Lee et al. (2016).

Proof of Lemma 2 Let

$$V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)=\left\{\gamma\in\mathbb{R}^k:e_j'\gamma=0,\gamma^{-j}\in V\left(X_{n,0}^{-j},\sigma_0^{-j}\right)\right\}$$

be the k-dimensional version of $V\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$, and note that $V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)\subseteq V(X_{n,0},\sigma_0)$ by construction. Let $F(X_{n,0},\sigma_0)=\{\gamma\,|\,\gamma\geq 0,\gamma'X_{n,0}=0,\gamma'\sigma_0=1\}$ denote the dual feasible set using $(X_{n,0},\sigma_0)$, and define $F(X_{n,0}^{-j},\sigma_0^{-j})$ analogously. Observe that for any $\gamma\in V(X_{n,0},\sigma_0)\setminus V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$, either $e_j'\gamma>0$ or $\gamma^{-j}\in F\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$.

We first show that $\hat{\eta}_{n,0}^{j,d}\to\hat{\eta}_{n,0}^{-j}$. To this end, consider $\gamma\in V(X_{n,0},\sigma_0)\setminus V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$. If $e_j'\gamma>0$, then $\gamma'Y_{n,0}^{j,d}\to-\infty$ as $d\to\infty$. Hence, if $V\left(X_{n,0}^{-j},\sigma_0^{-j}\right)\neq\emptyset$ (i.e. if the dual problem for $(X_{n,0}^{-j},\sigma_0^{-j})$ is feasible) then for d sufficiently large we must have $\gamma\notin\operatorname{argmax}_{\gamma\in V(X_{n,0},\sigma_0)}\gamma'Y_{n,0}^{j,d}$.

If instead $e_j'\gamma=0$ then $\gamma^{-j}\in F\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$, so $\gamma' Y_{n,0}^{j,d}\le \max_{\gamma\in V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)}\gamma' Y_{n,0}^{j,d}=\hat\eta_{n,0}^{-j}$ for all d, and either $\hat\gamma^{j,d}\in V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$ for d sufficiently large or there exists $\tilde\gamma\in V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$ such that $\gamma' Y_{n,0}=\tilde\gamma' Y_{n,0}$, which we rule out by assumption. Hence, either $\hat\eta_{n,0}^{j,d}=\hat\eta_{n,0}^{-j}$ and $\hat\gamma^{j,d}\in V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$ for d sufficiently large or the dual is infeasible and $\hat\eta_{n,0}^{j,d}\to-\infty$. Infeasibility of the dual corresponds to unboundedness of the primal, so in this case $\hat\eta_{n,0}^{-j}=-\infty$ and we again have $\hat\eta_{n,0}^{j,d}\to\hat\eta_{n,0}^{-j}$.

By the definition of the conditional test, if $\hat\eta_{n,0}^{j,d}\to\hat\eta_{n,0}^{-j}=-\infty$ then $\phi_C^{j,d}\to\phi_C^{-j}=0$. Hence, for the remainder of the proof we consider the case with $\hat\eta_{n,0}^{-j}>-\infty$. In this case, the argument above implies that $e_j'\hat\gamma^{j,d}=0$ for d sufficiently large. It is straightforward to verify that if $\hat\gamma^{j,d}\in V^*(X_{n,0}^{-j},\sigma_0^{-j})$, then $S_{n,0,\hat\gamma^{-j}}^{-j}=M_{-j}S_{n,0,\hat\gamma^{j,d}}^{j,d}$, where M_{-j} is the matrix that selects all of the rows except row j. It follows that

$$\begin{aligned}\mathcal{V}_{n,0}^{lo,-j}&=\max_{\gamma^{-j}\in V(X_{n,0}^{-j},\sigma_0^{-j}):(\hat\gamma^{-j})'\Sigma_0^{-j}(\hat\gamma^{-j})>(\hat\gamma^{-j})'\Sigma_0^{-j}(\gamma^{-j})}\frac{(\hat\gamma^{-j})'\Sigma_0(\hat\gamma^{-j})\cdot(\gamma^{-j})'S_{n,0,\hat\gamma^{-j}}^{-j}}{(\hat\gamma^{-j})'\Sigma_0(\hat\gamma^{-j})-(\hat\gamma^{-j})'\Sigma_0(\gamma^{-j})}\\&=\max_{\gamma\in V^*(X_{n,0},\sigma_0):\hat\gamma_{jd}'\Sigma_0\hat\gamma_{jd}>\hat\gamma_{jd}'\Sigma_0\gamma}\frac{\hat\gamma_{jd}'\Sigma_0\hat\gamma_{jd}\cdot\gamma' S_{n,0,\hat\gamma_{jd}}^{j,d}}{\hat\gamma_{jd}'\Sigma_0\hat\gamma_{jd}-\hat\gamma_{jd}'\Sigma_0\gamma}\end{aligned}$$

for d sufficiently large, where for brevity of notation we write $\hat\gamma_{jd}$ instead of $\hat\gamma^{j,d}$. Considering $\gamma\in V(X_{n,0},\sigma_0)\setminus V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$, note that if $e_j'\gamma>0$ then $\gamma' S_{n,0,\hat\gamma_{jd}}^{j,d}\to-\infty$ as $d\to\infty$, which implies that either

$$\gamma\notin \operatorname{argmax}_{\tilde\gamma\in V(X_{n,0},\sigma_0):\hat\gamma_{jd}'\Sigma_0\hat\gamma_{jd}>\hat\gamma_{jd}'\Sigma_0\tilde\gamma}\frac{\hat\gamma_{jd}'\Sigma_0\hat\gamma_{jd}\cdot\tilde\gamma' S_{n,0,\hat\gamma_{jd}}^{j,d}}{\hat\gamma_{jd}'\Sigma_0\hat\gamma_{jd}-\hat\gamma_{jd}'\Sigma_0\tilde\gamma}$$

for d sufficiently large or $\mathcal{V}_{n,0}^{lo,j,d}\to\mathcal{V}_{n,0}^{lo,-j}=-\infty$, and similarly for $\mathcal{V}_{n,0}^{up,j,d}$.

If instead $e_j'\gamma=0$, then as noted above $\gamma^{-j}\in F\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$, so for any $y\in\mathbb{R}^k$

$$\gamma' y\le \max_{\tilde\gamma\in V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)}\tilde\gamma' y=\max_{\tilde\gamma\in V\left(X_{n,0}^{-j},\sigma_0^{-j}\right)}\tilde\gamma' y^{-j}.$$

Lemma 5.1 of Lee et al. (2016) implies, however, that

$$\mathcal{V}_{n,0}^{lo,j,d}=\min_y(\hat\gamma^{j,d})' y,\text{ s.t. }(\hat\gamma^{j,d})' y\ge\max_{\tilde\gamma\in V(X_{n,0},\sigma_0)}\tilde\gamma' y\text{ and }S(y,\hat\gamma^{j,d})=S_{n,0,\hat\gamma^{j,d}}^{j,d},$$

where $S(y,\hat{\gamma})=\left(I-\frac{\Sigma_0\hat{\gamma}\hat{\gamma}'}{\hat{\gamma}'\Sigma_0\hat{\gamma}}\right)y$. The previous two displays together imply that

$$\mathcal{V}_{n,0}^{lo,j,d}=\min_y(\hat{\gamma}^{j,d})'y, \text{ s.t. } (\hat{\gamma}^{j,d})'y\geq\max_{\tilde{\gamma}\in V(X_{n,0},\sigma_0)\setminus\{\gamma\}}\tilde{\gamma}'y \text{ and } S(y,\hat{\gamma}^{j,d})=S_{n,0,\hat{\gamma}^{j,d}}^{j,d}.$$

Applying Lemma 5.1 of Lee et al. (2016) in the opposite direction,

$$\max_{\tilde{\gamma}\in V(X_{n,0},\sigma_0):\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}>\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}}\frac{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}\cdot\tilde{\gamma}'S_{n,0,\hat{\gamma}_{jd}}^{j,d}}{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}-\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}}=\max_{\tilde{\gamma}\in V(X_{n,0},\sigma_0)\setminus\{\gamma\}:\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}>\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}}\frac{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}\cdot\tilde{\gamma}'S_{n,0,\hat{\gamma}_{jd}}^{j,d}}{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}-\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}}.$$

Iterating this argument, we obtain that

$$\max_{\tilde{\gamma}\in V(X_{n,0},\sigma_0):\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}>\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}}\frac{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}\cdot\tilde{\gamma}'S_{n,0,\hat{\gamma}_{jd}}^{j,d}}{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}-\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}}=\max_{\tilde{\gamma}\in V^*(X_{n,0},\sigma_0):\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}>\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}}\frac{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}\cdot\tilde{\gamma}'S_{n,0,\hat{\gamma}_{jd}}^{j,d}}{\hat{\gamma}_{jd}'\Sigma_0\hat{\gamma}_{jd}-\hat{\gamma}_{jd}'\Sigma_0\tilde{\gamma}},$$

where we showed above that the expression on the right-hand side is equal to $\mathcal{V}_{n,0}^{lo,-j}$ for d sufficiently large. A similar argument applies for $\mathcal{V}_{n,0}^{up,j,d}$. We have thus shown that $\left(\mathcal{V}_{n,0}^{lo,j,d},\mathcal{V}_{n,0}^{up,j,d}\right)\to\left(\mathcal{V}_{n,0}^{lo,-j},\mathcal{V}_{n,0}^{up,-j}\right)$ as $d\to\infty$.

This convergence, combined with the fact that $\hat{\gamma}^{j,d}\in V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$ for d sufficiently large and the fact that for $\gamma\in V^*\left(X_{n,0}^{-j},\sigma_0^{-j}\right)$, $\gamma'\Sigma_0\gamma=\gamma^{-j}\Sigma_0^{-j}\gamma^{-j}$, implies that $c_{\alpha,C}\left(Y_{n,0}^{j,d},X_{n,0},\Sigma_0\right)\to c_{\alpha,C}\left(Y_{n,0}^{-j},X_{n,0}^{-j},\Sigma_0^{-j}\right)$. Hence, so long as $\hat{\eta}_{n,0}^{-j}\neq c_{\alpha,C}\left(Y_{n,0}^{-j},X_{n,0}^{-j},\Sigma_0^{-j}\right)$, $\phi_C^{j,d}\to\phi_C^{-j}$, as desired. □

Proof of Lemma 3 Towards contradiction, suppose the conclusion of the lemma fails. Then there exists a sequence of distributions, null parameter values, and sample sizes $\left\{P_{D|Z,n_m},\beta_{0,n_m},n_m\right\}$ with $\beta_{0,n_m}\in B_I(P_{D|Z,n_m})$ for all m, and a constant $\varepsilon>0$ such that

$$\liminf_{m\to\infty}\sup_{f\in BL_1}\left|E_{P_{D|Z,n_m}}[f(U_{n_m,0}-\pi_{n_m,0})]-E\left[f\left(\xi_{P_{D|Z,n_m}}\right)\right]\right|>\varepsilon. \tag{17}$$

Since the set of possible variances Ω consistent with Assumption 1 is compact, there exists a subsequence $\left\{P_{D|Z,n_l},\beta_{0,n_l},n_l\right\}\subseteq\left\{P_{D|Z,n_m},\beta_{0,n_m},n_m\right\}$ along which $\Omega\left(P_{D|Z,n_l},\beta_{0,n_l}\right)\to\Omega^*$ for some Ω^*. Under this subsequence, however, the Lindeberg-Feller Central Limit Theorem (see e.g. Proposition 2.27 in Van der Vaart (2000)), along with the assumptions of the lemma, implies that

$$U_{n_l,0}-\pi_{n_l,0}\to_d N(0,\Omega^*),$$

and thus that

$$\lim_{l\to\infty}\sup_{f\in BL_1}\left|E_{P_{D|Z,n_l}}[f(U_{n_l,0}-\pi_{n_l,0})]-E\left[f\left(\xi_{P_{D|Z,n_l}}\right)\right]\right|=0.$$

This contradicts (17), completing the proof. □

The following result characterizes the vertices of the dual vertex set.

Lemma A.1 *Suppose $\gamma\in F(X,\sigma)$. Then $\gamma\in V(X,\sigma)$ if and only if $\gamma=A_B(X,\sigma)^{-1}e_1$, for e_1 the first standard basis vector in $\mathbb{R}^k$,*

$$A(X,\sigma)=\begin{pmatrix}\sigma'\\ X'\\ -I\end{pmatrix},$$

and $B\subset\{1,\ldots,p+k+1\}$ with $|B|=k$ and $1\in B$, where M_B denotes the rows of the matrix M contained in B.

Proof of Lemma A.1 From Theorem 8.4 and statement (23) in Section 8.5 in Schrijver (1986), $v\in\{x\in\mathbb{R}^k:Wx\leq b\}$ is a vertex of $\{x\in\mathbb{R}^k:Wx\leq b\}$ if and only if there exists $B\subset\{1,\ldots k\}$ such that W_B is invertible and $W_Bx=b_B$, where W_B denotes the rows of W corresponding with the indices in B, and b_B is defined analogously. Observe that $F(X,\sigma)$ takes the form $\{\gamma\in\mathbb{R}^k:W\gamma\leq b\}$, where

$$W=\begin{pmatrix}\sigma'\\ -\sigma'\\ X'\\ -X'\\ -I\end{pmatrix}\text{ and }b=\begin{pmatrix}1\\ -1\\ 0\\ 0\\ 0\end{pmatrix},$$

where W is $(2(p+1)+k)\times k$ and b is $(2(p+1)+k)\times 1$. Thus, $\gamma\in F(X,\sigma)$ is a vertex if and only if $\gamma=W_B^{-1}b_B$ for some index set $B\subset\{1,\ldots,2(p+1)+k\}$ with $|B|=k$ such that W_B is invertible.

Next, observe that $\gamma\in F(X,\sigma)$ satisfies $\gamma'\sigma=1$ and thus must be non-zero. Since $b_B=0$ unless B contains an index corresponding with a row of W containing either σ' or $-\sigma'$, it follows that if there is a vertex corresponding with B then B must always contain one such index. Moreover, it's clear that B can select at most one of each pair of inequalities of the opposite sign, since W_B is full-rank. Further, we claim that every vertex corresponds with an index B that only selects from the rows of the matrix $Q:=(\sigma, X)'$ and not from the matrix $-(\sigma, X)'$. To show this, let $B \subset \{1,...,2(p+1)+k\}$ with $|B|=k$ such that W_B is invertible, and suppose there is a vertex corresponding to B. Let $\tilde{B}$ be the analogous index that replaces all the indices of B corresponding to rows of $-Q$ with the analogous rows of Q. By the preceeding argument, B selects exactly one of the rows of Q corresponding to σ' or $-\sigma'$. Suppose first that B selects the row corresponding to $-\sigma$. Without loss of generality, order the remaining rows of W so that B and $\tilde{B}$ differ in the first w positions and agree otherwise. Then we can write

$$W_B = \begin{pmatrix} -I_w & 0 \\ 0 & I_{k-w} \end{pmatrix} W_{\tilde{B}}.$$

It follows that

$$W_B^{-1} = W_{\tilde{B}}^{-1} \begin{pmatrix} -I_w & 0 \\ 0 & I_w \end{pmatrix}^{-1} = W_{\tilde{B}}^{-1} \begin{pmatrix} -I_w & 0 \\ 0 & I_w \end{pmatrix}.$$

However, $b_{\tilde{B}} = e_1$ while $b_B = -e_1$, which combined with the previous display implies that $W_B^{-1} b_B = W_{\tilde{B}}^{-1} b_{\tilde{B}}$. Similarly, suppose that B selects the row corresponding with σ'. Order the remaining elements of W so that B differs from $\tilde{B}$ in positions $2,...,w+1$. Then we can write

$$W_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -I_w & 0 \\ 0 & 0 & I_{k-w-1} \end{pmatrix} W_{\tilde{B}}$$

and hence

$$W_B^{-1} = W_{\tilde{B}}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -I_w & 0 \\ 0 & 0 & I_{k-w-1} \end{pmatrix}$$

But $b_B = e_1 = b_{\tilde{B}}$, which together with the previous display implies that $W_B^{-1} b_B = W_{\tilde{B}}^{-1} W_{\tilde{B}}$, as we wished to show. We have thus established that $\gamma \in F(X,\sigma)$ is a vertex if and only if it takes the form $A_B^{-1} e_1$, where

$$A = \begin{pmatrix} \sigma' \\ X' \\ -I \end{pmatrix},$$

and $B \subset \{1,...,p+k+1\}$ with $|B| = p+1$ and $1 \in B$. $\square$

To prove our remaining results it is helpful to introduce some additional notation. Let $\Gamma(X,\sigma)$ be a matrix whose rows collect the elements of $V(X,\sigma)$,

$$V(X,\sigma)=\left\{\gamma\in\mathbb{R}^k:\gamma'=e_j'\Gamma(X,\sigma) \text{ for some } j\in\{1,\ldots,\dim(\Gamma(X,\sigma)\sigma)\}\right\}.$$

We first prove a lemma describing how $\Gamma(X,\sigma)$ varies with σ.

Lemma A.2 *Suppose Assumption 1 holds. For $\upsilon=\sqrt{Diag(TT')}$ and $\sigma=\sqrt{Diag(T\Omega T')}$ for some positive-definite Ω, $\Gamma(X,\sigma)=\Lambda(X,\sigma)\Gamma(X,\upsilon)$ where $\Lambda(X,\sigma)$ is a diagonal matrix with $\Lambda_{jj}(X,\sigma)=\frac{1}{e_j'\Gamma(X,\upsilon)\sigma}$.*

Proof of Lemma A.2 This follows by an argument as in Lemma A.1 of Rambachan & Roth (2022), but is included for completeness. Recall that the elements of $\Gamma(X,\sigma)$ take the form $A_B(X,\sigma)^{-1}e_1$ for B such that $A_B(X,\sigma)$ is invertible and $A_B(X,\sigma)^{-1}e_1\geq 0$. Fix a B corresponding to a vertex in $V(X,\sigma)$. Write

$$A_B(X,\sigma)=\begin{pmatrix}\sigma' \\ (X')_{B_1} \\ -I_{B_2}\end{pmatrix}$$

where B_1 and B_2 are the subsets of B corresponding to the rows of X' and $-I$ respectively. Since $A_B(X,\sigma)$ has rank k, it follows that $L:=\begin{bmatrix}(X')_{B_1} \\ -I_{B_2}\end{bmatrix}$ has rank $k-1$. Thus, the space of vectors v such that $Lv=0$ is a 1-dimensional linear subspace. Note, however, that by construction if $\vartheta=A_B(X,\tilde{\sigma})^{-1}e_1$ for some $\tilde{\sigma}$ such that $A_B(X,\tilde{\sigma})$ is full-rank, then $A_B(X,\tilde{\sigma})\vartheta=e_1$ and hence $L\vartheta=0$. It follows that if $A_B(X,\upsilon)$ is also full rank then $A_B(X,\sigma)\propto A_B(X,\upsilon)$. Note further that from the definition of the vertex set, we must have that $(A_B(X,\sigma)^{-1}e_1)'\sigma=1$. Thus, if $A_B(X,\sigma)$ and $A_B(X,\upsilon)$ both have full rank then

$$A_B(X,\sigma)^{-1}e_1=\frac{(A_B(X,\sigma)^{-1}e_1)'\sigma}{(A_B(X,\upsilon)^{-1}e_1)'\sigma}A_B(X,\upsilon)^{-1}e_1=\frac{1}{(A_B(X,\upsilon)^{-1}e_1)'\sigma}A_B(X,\upsilon)^{-1}e_1.$$

Note that Lemma A.1 implies that $A_B(X,\upsilon)^{-1}e_1\in V(X,\upsilon)$, since $A_B(X,\upsilon)\propto A_B(X,\sigma)\geq 0$ and $A_B(X,\upsilon)\upsilon=1$ by construction. By an analogous argument reversing the roles of σ and υ, we can show that if B corresponds to a vertex of $V(X,\upsilon)$, then a re-scaling of $A_B(X,\upsilon)^{-1}e_1$ is also a vertex of $V(X,\sigma)$ provided that $A_B(X,\upsilon)$ is full-rank.

It thus remains to show that $A_B(X,\sigma)$ has full rank and satisfies $A_B(X,\sigma)^{-1}e_1\geq 0$ if and only if $A_B(X,\upsilon)$ does. To this end, suppose that $A_B(X,\upsilon)$ has full rank and

$A_B(X,\upsilon)^{-1}e_1 \geq 0$. Let $\vartheta = A_B(X,\upsilon)^{-1}e_1$ and note that by construction $\vartheta \geq 0$, $\upsilon'\vartheta = 1$, and $L\vartheta = 0$. Note, however, that the structure of σ implies that $\upsilon_j = 0$ if and only if $\sigma_j = 0$, so $\upsilon'\vartheta = 1$ and $\vartheta \geq 0$ implies that $\sigma'\vartheta > 0$. Hence, since $L\vartheta = 0$ while $\sigma'\vartheta > 0$, we see that σ' is linearly independent of L, and thus $A_B(X,\sigma)$ has full rank. Moreover, by the argument above, we have that $A_B(X,\sigma)^{-1}e_1$ is a positive rescaling of $A_B(X,\upsilon)e_1$, and thus $A_B(X,\sigma)^{-1}e_1 \geq 0$, as needed. Since we can repeat the same argument reversing the roles of σ and υ, we have established the desired result. □

Proof of Lemma 4 The first part of the Lemma follows immediately from Lemma A.2 above. To show the second part, let $\hat{\eta}_\dagger = \max_{\gamma \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})} \gamma' Y_{n,0}$ denote the analog to $\hat{\eta}_{n,0}$ using $V_\dagger$ instead of V, and define other variables subscripted with † analogously. Observe that by construction, $\hat{\eta}_\dagger = \hat{\eta}_{n,0}$ unless $\hat{\eta}_{n,0} \leq 0$. Next, consider the modified least favorable critical value, $c_{\alpha,LF,\dagger}$, which is the $1-\alpha$ quantile of $\max_{\gamma \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi$, for $\xi \sim N(0,\widehat{\Sigma}_{n,0})$. By construction, $\max_{\gamma \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi = \max_{\gamma \in V(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi$ unless $\max_{\gamma \in V(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi \leq 0$. Now, for any $\gamma_{1,\dagger} \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$, we have that $\gamma_{1,\dagger}'\xi \leq \max_{\gamma \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi$, and $\gamma_{1,\dagger}'\xi \sim N(0,\gamma_{1,\dagger}'\widehat{\Sigma}_{n,0}\gamma_{1,\dagger})$, which has median of zero. It follows that for $\alpha < 0.5$, the $1-\alpha$ quantile of $\max_{\gamma \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})} \gamma'\xi$ is weakly positive, and hence that $c_{\alpha,LF} = c_{\alpha,LF,\dagger}$. We have thus established the result for the LF test.

Next consider the conditional test. By construction the conditional test never rejects when $\hat{\eta}_{n,0} \leq 0$, so we will consider the case where $\hat{\eta}_{n,0} > 0$. As argued above, in this case $\hat{\eta}_{n,0} = \hat{\eta}_\dagger$, and moreover, $\hat{\gamma} = \hat{\gamma}_\dagger$ from the definition of $V_\dagger(X_{n,0},\hat{\sigma}_{n,0})$. Finally, recall that Lemma 5.1 in Lee et al. (2016) implies that $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$ are the minimum and maximum of the set

$$\left\{ \hat{\gamma}'y | y \text{ s.t. } \hat{\gamma}'y \geq \max_{\tilde{\gamma} \in V(X_{n,0},\hat{\sigma}_{n,0})} \tilde{\gamma}'y \text{ and } S(y,\hat{\gamma}) = S_{n,0,\hat{\gamma}} \right\}.$$

Since $\max_{\tilde{\gamma} \in V(X_{n,0},\hat{\sigma}_{n,0})} \tilde{\gamma}'y$ is equal to $\max_{\tilde{\gamma} \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})} \tilde{\gamma}'y$ whenever the former is positive, we see that $\mathcal{V}^{up}_{n,0} = \mathcal{V}^{up}_\dagger$, since $\mathcal{V}^{up}_{n,0} \geq \hat{\eta}_{n,0} > 0$. Further, since $V_\dagger(X_{n,0},\hat{\sigma}_{n,0}) \subseteq V(X_{n,0},\hat{\sigma}_{n,0})$, we have that $\hat{\gamma}'y \geq \max_{\tilde{\gamma} \in V_\dagger(X_{n,0},\hat{\sigma}_{n,0})} \tilde{\gamma}'y$ whenever $\hat{\gamma}'y \geq \max_{\tilde{\gamma} \in V(X_{n,0},\hat{\sigma}_{n,0})} \tilde{\gamma}'y$. It follows that $\mathcal{V}^{lo}_\dagger \leq \mathcal{V}^{lo}_{n,0}$. Note, however, that the critical value for the conditional test is increasing in the value of $\mathcal{V}^{lo}_{n,0}$, and thus $c_{\alpha,C} \geq c_{\alpha,C,\dagger}$. It follows that $\hat{\eta}_{n,0} > c_{\alpha,C}$ only if $\hat{\eta}_\dagger > c_{\alpha,C,\dagger}$, as we wished to show. The desired result for the hybrid test follows immediately from the arguments for the LF and conditional tests. □

Following D. Andrews et al. (2019), we establish size control using a subsequencing argument.

Lemma A.3 *Under Assumptions 1, 2, and 3, to show that a test ϕ which (i) depends on the data through $\left(Y_{n,0}, X_{n,0}, \widehat{\Sigma}_{n,0}\right)$ and (ii) does not reject when $\hat{\eta}_{n,0} = -\infty$ has uniformly correct asymptotic size,*

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} E_{P_{D|Z}}[\phi] \leq \alpha,$$

it suffices to show that $\limsup_{l\to\infty} E_{P_{D|Z,n_l}}[\phi] \leq \alpha$ *for all subsequences* $\{n_l\}\subseteq\{n\}$, $\left\{P_{D|Z,n_l}\right\} \in \mathcal{P}^{\infty}_{D|Z} = \times_{l=1}^{\infty}\mathcal{P}_{D|Z}$, $\{\beta_{0,n_l}\} \in \times_{l=1}^{\infty} B_I(P_{D|Z,n_l})$ *with*

1. $\min_\delta \max_j e_j' X_{n_l,0}\delta > -\infty$ *and* $\Omega\left(P_{D|Z,n_l}, \beta_{0,n_l}\right) \to \Omega^* $ *for some* $\Omega^* \in \mathbf{\Omega}_{\bar{\lambda}}$

2. *For each j and* $\psi_{j,n_l} = \sqrt{e_j'\Gamma(X_{n_l,0},\upsilon)TT'\Gamma(X_{n_l,0},\upsilon)e_j}$, *either* $\psi_{j,n_l} = 0$ *for all l or* $\psi_{j,n_l} \neq 0$ *for all l*

3. *If* $\psi_{j,n_l} > 0$ *for some j then for* $\psi_{n_l} = \max_j \psi_{j,n_l}$, $\psi_{n_l}^{-1}\Gamma(X_{n_l,0},\upsilon)T \to \Pi^*$ *for* $\Pi^* \neq 0$

4. *If* $\psi_{n_l} > 0$, *then* $\psi_{n_l}^{-1}\Gamma(X_{n_l,0},\upsilon)\mu_{n_l,0} \to \nu^* \in [-\infty, 0]^{\dim(Y_{n,0})}$

5. *For* $\sigma(\Omega) = \sqrt{Diag(T'\Omega T)}$ *and* $\Lambda(X,\sigma)$ *as defined in Lemma A.2,* $\Lambda(X_{n_l,0}, \sigma(\Omega(P_{D|Z,n_l}, \beta_{0,n_l}))) \to \Lambda^*$ *for Λ^* a diagonal, positive-definite matrix. Likewise,* $\Lambda(X_{n_l,0}, \hat{\sigma}_{n_l,0}) \to_p \Lambda^*$ *for* $\hat{\sigma}_{n_l,0} = \sigma(\hat{\Omega}_{n_l,0})$.

Proof of Lemma A.3 We establish that if size control fails, then there always exists a sequence satisfying the conditions of the lemma under which size control also fails.

If size control fails, then

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} E_{P_{D|Z}}[\phi] \geq \alpha + 2\varepsilon$$

for some $\varepsilon > 0$. This implies that there exists a subsequence $\{n_t^1\}\subseteq\{n\}$, $\left\{P_{D|Z,n_t^1}\right\}\in\mathcal{P}^{\infty}_{D|Z}$, $\{\beta_{0,n_t^1}\} \in \times_{t=1}^{\infty} B_I(P_{D|Z,n_t^1})$ such that $\liminf_{t\to\infty} E_{P_{D|Z,n_t^1}}[\phi] \geq \alpha+\varepsilon$. Since ϕ is assumed not to reject when $\hat{\eta}_{n,0} = -\infty$, it must be that $\min_\delta \max_j e_j' X_{n_t,0}\delta$ is finite for all t, since otherwise $\hat{\eta}_{n_t,0} = -\infty$ with probability 1 and the test never rejects. Since $\Omega\left(P_{D|Z,n_t^1}, \beta_{0,n_t^1}\right) \in \mathbf{\Omega}_{\bar{\lambda}}$ for all t by assumption, and $\mathbf{\Omega}_{\bar{\lambda}}$ is compact, there exists a further subsequence $\{n_t^2\}\subseteq\{n_t^1\}$ with $\Omega\left(P_{D|Z,n_t^2}, \beta_{0,n_t^2}\right) \to \Omega^* \in \mathbf{\Omega}_{\bar{\lambda}}$.

For each t, $\Gamma\left(X_{n_t^2,0},\upsilon\right)$ is a matrix with $\dim(Y_{n,0})$ columns, and a uniformly bounded number of rows. Hence there exists a subsequence $\{n_t^3\}\subseteq\{n_t^2\}$ along which the dimension of $\Gamma\left(X_{n_t^3,0},\upsilon\right)$ is constant. For each j and any subsequence $\{n_r\}\subseteq\{n\}$, either $\psi_{j,n_r} = 0$ infinitely often or not. We can thus extract a further subsequence $\{n_t^4\}\subseteq\{n_t^3\}$ along which part (2) of the lemma holds. If $\psi_{j,n_t^4}=0$ for all j then part (3) of the lemma is vacuous, while if $\psi_{j,n_t^4}>0$ for some j, $\psi_{j,n_t^4}^{-1}\left\|e_j'\Gamma\left(X_{n_t^4,0},\upsilon\right)T\right\|=1$ by construction, so $\psi_{n_t^4}^{-1}\left\|e_j'\Gamma\left(X_{n_t^4,0},\upsilon\right)T\right\|\leq 1$ for all j, and there exists a subsequence $\{n_t^5\}\subseteq\{n_t^4\}$ along which $\psi_{n_t^5}^{-1}\Gamma\left(X_{n_t^5,0},\upsilon\right)T\to\Pi^*$, where $\Pi^*\neq 0$ since $\psi_{n_t^5}^{-1}\left\|e_j'\Gamma\left(X_{n_t^5,0},\upsilon\right)T\right\|=1$ for at least one j, thus establishing part (3) of the lemma.

Part (4) of the lemma is again vacuous if $\psi_{n_l}=0$. Otherwise, note that since

$$\max_j e_j'\Gamma(X_{n,0},\upsilon)\mu_{n,0}=\min_\delta\max_j e_j'(\mu_{n,0}-X_{n,0}\delta)$$

whenever the solution is finite, $\Gamma\left(X_{n_t^5,0},\upsilon\right)\mu_{n_t^5,0}\leq 0$ for all t. For any subsequence $\{n_r\}\subseteq\{n_t^5\}$ and any j, $\psi_{n_r}^{-1}e_j'\Gamma(X_{n_r,0},\upsilon)\mu_{n_r,0}$ is either bounded or unbounded as $r\to\infty$, allowing us to extract a further subsequence $\{n_t^6\}\subseteq\{n_t^5\}$ along which $\psi_{n_t^6}^{-1}e_j'\Gamma\left(X_{n_t^6,0},\upsilon\right)\mu_{n_t^6,0}\to\nu_j^*\in[-\infty,0]$. Starting from $\{n_t^5\}$ and iterating this argument over the rows of $\psi_{n_t^5}^{-1}\Gamma\left(X_{n_t^5,0},\upsilon\right)\mu_{n_t^5,0}$ delivers a subsequence $\{n_s\}$ satisfying properties (1)-(4) of the lemma.

Next, let M be the matrix that selects the non-zero rows of T, and observe that M also selects the non-zero elements of υ and of $\sigma(\Omega)$ for any positive definite Ω. Let $\gamma_{n,j}' = e_j'(\Gamma(X_{n,0},\upsilon))$. By construction, $\gamma_{n,j}'\upsilon = (M\gamma_{n,j})'(M\upsilon) = 1$. Since $M\upsilon > 0$ and $M\gamma_{n,j} \geq 0$ by construction, it follows that $||M\gamma_{n,j}||$ is bounded. However, for $\sigma_{n,0}=\sigma(\Omega(P_{D|Z,n},\beta_{0,n}))$, we have $|\gamma_{n,j}'\sigma_{n,0}|=|(M\gamma_{n,j})'(M\sigma_{n,0})|\leq||M\gamma_{n,j}||\cdot||M\sigma_{n,0}||$, where part (ii) of Assumption 1 implies that $||M\sigma_{n,0}||$ is also bounded. It follows that there exists a subsequence $\left\{n_l^j\right\}\subseteq\{n_s\}$ such that $\gamma_{n_l^j,j}'\sigma_{n_l^j,0}$ converges. Moreover, the limit must be strictly positive, since by construction $\gamma_{n_l^j,j}'\upsilon=1$ and $\gamma_{n_l^j,j}\geq 0$, whereas the fact that the eigenvalues of $\Omega_{n_l^j,0}$ are bounded from below implies $\sigma_{n_l^j,0}\geq c\upsilon$ for some $c>0$. Iterating this argument for each j, we obtain a subsequence $\{n_l\}\subseteq\{n_s\}$ such that $\gamma_{n_l,j}'\sigma_{n_l,0}$ converges to a positive limit for all j. The jth diagonal element of $\Lambda\left(X_{n_l,0},\sigma(\Omega(P_{D|Z,n_l},\beta_{0,n_l}))\right)$ is $1/(\gamma_{n_l,j}'\sigma_{n_l,0})$, and hence $\Lambda\left(X_{n_l,0},\sigma(\Omega(P_{D|Z,n_l},\beta_{0,n_l}))\right)\to\Lambda^*$ for Λ^* a positive-definite and diagonal matrix, which establishes that the sequence also meets the first part of condition (5). To establish the second part of condition (5), observe that

$$|\gamma_{n,j}'\hat{\sigma}_{n_l,0}-\gamma_{n,j}'\sigma_{n_l,0}|=|(M\gamma_{n,j})'M(\hat{\sigma}_{n_l,0}-\sigma_{n_l,0})|\leq||M\gamma_{n,j}||\cdot||M(\hat{\sigma}_{n_l,0}-\sigma_{n_l,0})||\to_p 0.$$

However, the jth diagonal element of $\Lambda(X_{n_l,0},\sigma_{n_l,0})$ is equal to $1/(\gamma_{n,j}'\sigma_{n_l,0})$, which we showed above converges to a positive constant $e_j'\Lambda^*e_j$. The continuous mapping theorem thus implies that $e_j'\Lambda(X_{n_l,0},\hat{\sigma}_{n_l,0})e_j=1/(\gamma_{n,j}'\hat{\sigma}_{n_l,0})\to_p e_j'\Lambda^*e_j$.

We have thus established that there exists a sequence satisfying the conditions of the lemma under which size control fails, as we wished to show. □

Proof of Proposition 1 By construction, the least favorable test never rejects when $\hat{\eta}_{n,0} = -\infty$. Hence, by Lemma A.3, it suffices to show size control for sequences $\{n_l, P_{D|Z,n_l}, \beta_{0,n_l}\}$ satisfying the conditions of the lemma.

Note that by Lemma A.2 we can write

$$\hat{\eta}_{n_l,0} = \max_j \{e_j' \Gamma(X_{n_l,0}, \hat{\sigma}_{n_l,0}) Y_{n_l,0}\} = \max_j \{e_j' \Lambda(X_{n_l,0}, \hat{\sigma}_{n_l,0}) \Gamma(X_{n_l,0}, \upsilon) Y_{n_l,0}\}$$

$$= \max_j \{e_j' \Lambda(X_{n_l,0}, \hat{\sigma}_{n_l,0})(\Gamma(X_{n_l,0}, \upsilon)(Y_{n_l,0} - \mu_{n_l,0}) + \Gamma(X_{n_l,0}, \upsilon)\mu_{n_l,0})\}.$$

Assumption 1 implies that we can re-write $Y_{n_l,0} - \mu_{n_l,0}$ as $T(U_{n_l,0} - \pi_{n_l,0})$. Hence,

$$\hat{\eta}_{n_l,0} = \max_j \{e_j' \Lambda(X_{n_l,0}, \hat{\sigma}_{n_l,0})(\Gamma(X_{n_l,0}, \upsilon) T(U_{n_l,0} - \pi_{n_l,0}) + \Gamma(X_{n_l,0}, \upsilon)\mu_{n_l,0})\}.$$

First consider the case where $\psi_{n_l} = 0$. This implies that $\Gamma(X_{n_l,0}, \upsilon) T = 0$ for all l, which in turn implies that $\Gamma(X_{n_l,0}, \upsilon) Y_{n_l,0} \leq 0$ with probability one since $\beta_{0,n_l} \in B_I(P_{D|Z,n_l})$ by construction and thus $\Gamma(X_{n_l,0}, \upsilon)\mu_{n_l,0} \leq 0$. The least favorable test never rejects in this case, since $\alpha < \frac{1}{2}$ implies that $c_{\alpha,LF}\left(X_{n,0}, \widehat{\Sigma}_{n,0}\right) \geq 0$.

Next consider the case where $\psi_{n_l} > 0$. Assumption 3 implies that $Y_{n_l,0} - \mu_{n_l,0} \to_d N(0, T\Omega^* T')$. Parts (3) and (4) of Lemma A.3 thus imply that

$$\psi_{n_l}^{-1}(\Gamma(X_{n_l,0}, \upsilon) T(U_{n_l,0} - \pi_{n_l,0}) + \Gamma(X_{n_l,0}, \upsilon)\mu_{n_l,0}) \to N\left(\nu^*, \Pi^* \Omega^* \Pi^{*'}\right)$$

By part (5) of Lemma A.3, $\Lambda(X_{n_l,0}, \hat{\sigma}_{n_l,0}) \to_p \Lambda^*$, for Λ^* diagonal and positive definite, so by the continuous mapping theorem,

$$\psi_{n_l}^{-1} \Lambda(X_{n_l,0}, \hat{\sigma}_{n_l,0})(\Gamma(X_{n_l,0}, \upsilon) T(U_{n_l,0} - \pi_{n_l,0}) + \Gamma(X_{n_l,0}, \upsilon)\mu_{n_l,0})$$

$$\to_d G^* \sim N\left(\Lambda^* \nu^*, \Lambda^* \Pi^* \Omega^* \Pi^{*'} \Lambda^*\right).$$

Hence, by another application of the continuous mapping theorem, $\psi_{n_l}^{-1} \hat{\eta}_{n_l,0} \to_d \max_j e_j' G^*$, where since $\Lambda^* \nu^* \leq 0$, the limiting distribution is continuous at all strictly positive values.

To show size control for the least favorable test, we must further show convergence of the critical value. To this end, note that Assumptions 1 and 2, together with convergence of $\Lambda(X_{n_l,0},\hat{\sigma}_{n_l,0})$, imply that

$$\psi_{n_l}^{-2}\Gamma(X_{n_l,0},\hat{\sigma}_{n_l,0})\widehat{\Sigma}_{n,0}\Gamma(X_{n_l,0},\hat{\sigma}_{n_l,0})'\to_p \Lambda^*\Pi^*\Omega^*\Pi^{*'}\Lambda^*,$$

where the limit is nonzero. Note, moreover, that

$$c_{\alpha,LF}\left(X_{n_l,0},\widehat{\Sigma}_{n,0}\right)=\psi_{n_l}\cdot c_{\alpha,LF}\left(X_{n_l,0},\psi_{n_l}^{-2}\cdot\widehat{\Sigma}_{n,0}\right).$$

Hence, $c_{\alpha,LF}\left(X_{n_l,0},\psi_{n_l}^{-2}\cdot\widehat{\Sigma}_{n,0}\right)$ converges in probability to $c^*_{\alpha,LF}$, the $1-\alpha$ quantile of $\max_j e_j'\tilde{G}$ for $\tilde{G}\sim N\left(0,\Lambda^*\Pi^*\Omega^*\Pi^{*'}\Lambda^*\right)$, where $c^*_{\alpha,LF}>0$ for $\alpha<\frac{1}{2}$. Note further that

$$\phi_{LF}=1\left\{\hat{\eta}_{n_l,0}>c_{\alpha,LF}\left(X_{n_l,0},\widehat{\Sigma}_{n,0}\right)\right\}=1\left\{\psi_{n_l}^{-1}\hat{\eta}_{n_l,0}>c_{\alpha,LF}\left(X_{n_l,0},\psi_{n_l}^{-2}\cdot\widehat{\Sigma}_{n,0}\right)\right\},$$

so by another application of the continuous mapping theorem,

$$\phi_{LF}\to_d 1\left\{\left(\max_j e_j'G^*\right)>c^*_{\alpha,LF}\right\},$$

which implies that $\limsup_{s\to\infty}E_{P_{D|Z,n_l}}[\phi_{LF}]\leq\alpha$, as we wanted to show. $\square$

Proof of Proposition 2 We first prove the result for the conditional test. As in Lemma A.3, we use a subsequencing argument. Specifically, begin with sequences of sample sizes, data generating processes, and null parameter values $\{n_s\}\subseteq\{n\}$, $\{P_{D|Z,n_s}\}\in\mathcal{P}_{D|Z}^{\infty}$, and $\{\beta_{0,n_s}\}\in\times_{s=1}^{\infty}B_I(P_{D|Z,n_s})$. Observe that whether $V_{\dagger}(X_{n_s,0},\sigma_{n_s,0})$ is empty depends only on $X_{n_s,0}$. If $X_{n_s,0}$ is such that $V_{\dagger}(X_{n,0},\hat{\sigma}_{n_s,0})$ is empty, then $\hat{\eta}_{n,0}\leq 0$ with probability 1, and thus the conditional and hybrid tests never reject. For the remainder of the proof, we therefore consider sequences where $X_{n_s,0}$ is such that $V_{\dagger}(X_{n_s,0},\hat{\sigma}_{n_s,0})$ is non-empty, which implies that $\min_\delta\max_j e_j'X_{n,0}\delta>-\infty$, and thus $\hat{\eta}_{n_s,0}$ is finite with probability 1. It then suffices to establish size control for the test $\phi_{C,\dagger}$, since $\phi_C\leq\phi_{C,\dagger}$ with probablity 1 by Lemma 4.

Let M be the selection matrix such that $M'T$ picks out the nonzero rows of T, and note that by construction $\Gamma_{\dagger}(X_{n,0},\upsilon)MM'\upsilon=\iota$, where $\Gamma_{\dagger}$ denotes the subset of rows of Γ corresponding with vertices in $V_{\dagger}(X_{n,0},\upsilon)$ and ι is the vector of ones. Since $M'\upsilon$ is strictly positive, $\Gamma_{\dagger}(X_{n,0},\upsilon)M$ is a non-negative matrix with a uniformly bounded number of rows and uniformly bounded row-sums. There thus exists a subsequence of sample sizes $\{n_r\}\subseteq\{n_s\}$ such that $\Gamma_{\dagger}(X_{n_r,0},\upsilon)M$ has fixed dimensions and $\Gamma_{\dagger}(X_{n_r,0},\upsilon)M\to\Gamma_{\dagger}^*M$ for $\Gamma_{\dagger}^*$ a non-negative matrix with $\Gamma_{\dagger}^*\upsilon=\iota$. Since $\Omega\left(P_{D|Z,n_r},\beta_{0,n_r}\right)\in\mathbf{\Omega}_{\bar{\lambda}}$ for all r by assumption, and $\mathbf{\Omega}_{\bar{\lambda}}$

is compact, there exists a further subsequence $\{n_t\} \subseteq \{n_r\}$ with $\Omega\left(P_{D|Z,n_t}, \beta_{0,n_t}\right) \to \Omega^* \in \mathbf{\Omega}_{\bar{\lambda}}$.

Note, next, that

$$\begin{aligned}\Gamma_\dagger(X_{n_t,0}, \upsilon) Y_{n_t,0} &= \Gamma_\dagger(X_{n_t,0}, \upsilon)(Y_{n_t,0} - \mu_{n_t,0}) + \Gamma_\dagger(X_{n_t,0}, \upsilon)\mu_{n_t,0} \\ &= \Gamma_\dagger(X_{n_t,0}, \upsilon) M M' T (U_{n_t,0} - \pi_{n_t,0}) + \Gamma_\dagger(X_{n_t,0}, \upsilon)\mu_{n_t,0},\end{aligned} \tag{18}$$

where $\Gamma_\dagger(X_{n_t,0}, \upsilon)\mu_{n_t,0} \leq 0$ for all t since $\beta_{0,n_t} \in B_I(P_{D|Z,n_t})$. Assumptions 1 and 3 imply that

$$U_{n_t,0} - \pi_{n_t,0} \to_d N(0, \Omega^*),$$

so for $\Sigma^* = T \Omega^* T'$,

$$\Gamma_\dagger(X_{n_t,0}, \upsilon) M M' T (U_{n_t,0} - \pi_{n_t,0}) \to_d N\left(0, \Gamma_\dagger^* M M' \Sigma^* M M' \Gamma_\dagger^{*'}\right) = N\left(0, \Gamma_\dagger^* \Sigma^* \Gamma_\dagger^{*'}\right) \tag{19}$$

by the continuous mapping theorem, where Assumption 4 implies that the diagonal elements of $\Gamma_\dagger^* T \Omega^* T' \Gamma_\dagger^{*'} = \Gamma_\dagger^* \Sigma^* \Gamma_\dagger^{*'}$ are bounded away from zero. As argued in the proof of Lemma A.3, we can extract a further subsequence $\{n_l\}$ where

$$\Gamma_\dagger(X_{n_l,0}, \upsilon)\mu_{n_l,0} \to \nu^* \in [-\infty, 0]^{\dim\left(\Gamma_\dagger^* \upsilon\right)}.$$

By an argument analogous to that for part (5) of Lemma A.3, we can also choose $\{n_l\}$ such that, for $\sigma_{n_l,0} = \sigma(\Omega(P_{D|Z,n_l}, \beta_{0,n_l}))$ and $\hat{\sigma}_{n_l,0} = \sigma(\hat{\Omega}_{n_l,0})$, $\Lambda_\dagger(X_{n_l,0}, \sigma_{n_l,0}) \to \Lambda_\dagger^*$ and $\Lambda_\dagger(X_{n_l,0}, \hat{\sigma}_{n_l,0}) \to_p \Lambda_\dagger^*$ for $\Lambda_\dagger^*$ diagonal and positive definite.

Note next that if $\hat{\eta}_\dagger \to_p -\infty$ (because $\nu_j^* = -\infty$ for all j) then the rejection probability of the test $\phi_{C,\dagger}$ converges to zero. If instead $\hat{\eta}_\dagger \not\to_p -\infty$, then it must be that $\nu_j^* > -\infty$ for some j. Let M_+ be a selection matrix such that $M_+ \nu^*$ picks out the finite elements of ν^*. Note that for any γ corresponding to a row of $\Gamma_\dagger(X_{n_l,0}, \hat{\sigma}_{n,0})$ not selected by M_+, $Pr_{P_{D|Z,n_l}}\{\hat{\gamma}_\dagger = \gamma\} \to 0$, and thus asymptotically neither $\hat{\gamma}_\dagger$ nor $\hat{\eta}_\dagger$ is affected by $\gamma' Y_{n_l,0}$. By an argument analogous to that in the proof to Lemma 2, one can also show that asymptotically $\gamma' Y_{n_l,0}$ does not affect the values of $\mathcal{V}_{n,0,\dagger}^{lo}$ or $\mathcal{V}_{n,0,\dagger}^{lo}$. The asymptotic behavior of the $\phi_{C,\dagger}$ test is thus determined by $(M_+ \Gamma_\dagger(X_{n_l,0}, \hat{\sigma}_{n,0}) Y_{n_l,0}, M_+ \Gamma_\dagger(X_{n_l,0}, \hat{\sigma}_{n,0}) \widehat{\Sigma}_{n,0} \Gamma_\dagger(X_{n_l,0}, \hat{\sigma}_{n,0})' M_+')$.

Next, observe from equations (18) and (19), combined with the fact that $\Gamma_\dagger(X_{n,0}, \hat{\sigma}_{n,0}) = \Lambda_\dagger(X_{n,0}, \hat{\sigma}_{n,0}) \Gamma_\dagger(X_{n,0}, \upsilon)$, that

$$M_+ \Gamma_\dagger(X_n, \hat{\sigma}_{n,0})(Y_{n,0} - \mu_{n,0}) \to_d N(0, M_+ \Lambda_\dagger^* \Gamma_\dagger^* \Sigma^* \Gamma_\dagger^{*'} \Lambda_\dagger^* M_+').$$

Further, since $M_+\Gamma_\dagger(X_{n_l,0},\upsilon)\mu_{n_l,0}$ converges to a finite vector by construction, we have that

$$M_+(\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n,0})-\Gamma_\dagger(X_{n_l,0},\sigma_{n_l,0}))\mu_{n_l,0}=M_+(\Lambda_\dagger(X_{n_l,0},\hat{\sigma}_{n_l,0})-\Lambda_\dagger(X_{n_l,0},\sigma_{n_l,0}))\Gamma_\dagger(X_{n_l,0},\upsilon)\mu_{n_l,0}\to_p 0,$$

where we use the fact that $\Lambda_\dagger(X_{n_l,0},\sigma_{n_l,0})\to\Lambda^*_\dagger$ and $\Lambda_\dagger(X_{n_l,0},\hat{\sigma}_{n_l,0})\to_p\Lambda^*_\dagger$. Hence,

$$M_+\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n_l,0})Y_{n_l,0}-M_+\Gamma_\dagger(X_{n_l,0},\sigma_{n_l,0})\mu_{n_l,0}\to_d G^*\sim N(0,M_+\Lambda^*_\dagger\Gamma^*_\dagger\Sigma^*\Gamma^{*\prime}_\dagger\Lambda^*_\dagger M_+'),$$

where Assumption 4 implies (i) that the diagonal elements of the limiting variance are nonzero and (ii) that no two rows of G^* are perfectly positively correlated. Further, by the continuous mapping theorem

$$M_+\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n_l,0})\widehat{\Sigma}_{n,0}\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n_l,0})'M_+'\to_p M_+\Lambda^*_\dagger\Gamma^*_\dagger\Sigma^*\Gamma^{*\prime}_\dagger\Lambda^*_\dagger M_+'.$$

These are precisely the conditions assumed in Andrews et al. (2021), which we shorthand as AKM, to establish uniform asymptotic size control, so we can use their results to establish size control in our setting.

Specifically, to connect our setting to that in AKM, let X_n and Y_n in the notation of AKM both be equal to $M_+\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n,0})Y_{n_l,0}$, and let $\mu_{X,n}$ and $\mu_{Y,n}$ both be equal to $M_+\Gamma_\dagger(X_{n_l,0},\sigma_{n_l,0})\mu_{n_l,0}$. Let $\hat{j}$ be the row of $M_+\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n,0})$ corresponding to $\hat{\gamma}_\dagger$, and let $\hat{\gamma}_{\dagger,*}$ be the $\hat{j}$th row of $M_+\Gamma_\dagger(X_{n_l,0},\sigma_{n_l,0})$. We have established that Assumptions 2-4 of AKM hold under the sequence $\{n_l,P_{D|Z,n_l},\beta_{0,n_l}\}$, so Proposition 10 in AKM establishes that for $\hat{\mu}_{\alpha,n_l}$ the α-quantile unbiased estimator for $\hat{\gamma}'_{\dagger,*}\mu_{n_l,0}$ (see AKM for details),

$$\limsup_{l\to\infty}\left|Pr_{P_{D|Z,n_l}}\{\hat{\mu}_{\alpha,n}\geq\hat{\gamma}'_{\dagger,*}\mu_{n_l,0}\}-\alpha\right|=0.$$

The quantile unbiased estimator is closely related to our conditional test, however: the $\phi_{C,\dagger}$ test rejects if and only if $\hat{\mu}_{\alpha,n_l}>0$ and $\hat{\eta}_\dagger>0$, provided that the test statistic and critical value for the $\phi_{C,\dagger}$ test are determined only by the vertices in $M_+\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n_l,0})$, which we have established occurs w.p.a. 1. Since $\hat{\gamma}'_{\dagger,*}\mu_{n_l,0}\leq 0$ under the null hypothesis, this suffices to establish that $\limsup_{l\to\infty}Pr_{P_{D|Z,n_l}}\{\phi_{C,\dagger}=1\}\leq\alpha$, as we wanted to show. As in the proof of Lemma A.3, this implies size control for the conditional test.

Next consider the hybrid test. For $\hat{\mu}^H_{\alpha,n_l}$ the α-quantile hybrid estimator of AKM with conditioning event $\left\{\hat{\eta}\leq c_{\kappa,LF,\dagger}(X_{n_l,0},\widehat{\Sigma}_{n_l,0}),\hat{\gamma}_\dagger=\gamma\right\}$, Proposition 12 of AKM implies that

$$\limsup_{l\to\infty}\left|Pr_{P_{D|Z,n_l}}\left\{\hat{\mu}^H_{\alpha,n_l}\geq\hat{\gamma}'_{\dagger,*}\mu_{n_l,0}|\hat{\eta}_\dagger\leq c_{\kappa,LF,\dagger}(X_{n_l,0},\widehat{\Sigma}_{n_l,0}),\hat{\gamma}_\dagger=\gamma\right\}-\alpha\right|Pr_{P_{D|Z,n_l}}\left\{\hat{\eta}_\dagger\leq c_{\kappa,LF,\dagger}(X_{n_l,0}\widehat{\Sigma}_{n_l,0}),\hat{\gamma}_\dagger=\gamma\right\}$$

is equal to 0. Since the vertex set is finite, it follows that

$$\limsup_{l\to\infty}\left|Pr_{P_{D|Z,n_l}}\left\{\hat{\mu}^H_{\alpha,n_l}\geq\hat{\gamma}'_{\dagger,*}\mu_{n_l,0}|\hat{\eta}_\dagger\leq c_{\kappa,LF,\dagger}(X_{n_l,0},\widehat{\Sigma}_{n_l,0})\right\}-\alpha\right|Pr_{P_{D|Z,n_l}}\left\{\hat{\eta}_\dagger\leq c_{\kappa,LF,\dagger}(X_{n_l,0}\widehat{\Sigma}_{n_l,0})\right\}=0.$$

Note, however, that the $\phi_{H,\dagger}$ test rejects only if $\hat{\eta}_\dagger>c_{\kappa,LF,\dagger}$ or $\hat{\mu}^H_{\frac{\alpha-\kappa}{1-\kappa},n_l}>0$ (again, assuming the test is determined only by the vertices of $M_+\Gamma_\dagger(X_{n_l,0},\hat{\sigma}_{n_l,0})$), and $0\geq\hat{\gamma}'_{\dagger,*}\mu_{n_l,0}$, so

$$Pr_{P_{D|Z,n_l}}\{\phi_{H,\dagger}=1\}\leq Pr_{P_{D|Z,n_l}}\left\{\hat{\eta}_\dagger>c_{\alpha,LF,\dagger}(X_{n_l,0},\widehat{\Sigma}_{n_l,0})\right\}+$$

$$Pr_{P_{D|Z,n_l}}\left\{\hat{\mu}^H_{\frac{\alpha-\kappa}{1-\kappa},n}\geq\hat{\gamma}'_{\dagger,*}\mu_{n_l,0}|\hat{\eta}_\dagger\leq c_{\alpha,LF,\dagger}(X_{n_l,0},\widehat{\Sigma}_{n_l,0})\right\}Pr_{P_{D|Z,n_l}}\left\{\hat{\eta}_\dagger\leq c_{\alpha,LF,\dagger}(X_{n_l,0},\widehat{\Sigma}_{n_l,0})\right\}.$$

Proposition 1 establishes that $\liminf_{l\to\infty}Pr_{P_{D|Z,n_l}}\{\hat{\eta}_\dagger\leq c_{\kappa,LF,\dagger}\}\geq1-\kappa$, so

$$\limsup_{l\to\infty}Pr_{P_{D|Z,n_l}}\{\phi_{H,\dagger}=1\}\leq\kappa+\frac{\alpha-\kappa}{1-\kappa}(1-\kappa)=\alpha,$$

implying size control for the hybrid test. □

B Non-Unique Dual Solutions

We now consider the behavior of the conditional test in the finite sample normal model without assuming that the dual solution is unique. Recall that we define $\hat{\gamma}$ as the argmax in the dual problem, so $\hat{\gamma}$ is set-valued when the dual solution is non-unique. We show that a version of the conditional test which chooses an arbitrary dual solution when there is multiplicity is well-defined with probability 1 in the finite-sample normal model and also controls size.

We first show that we can partition the set of vertices into disjoint subsets $V_1,...,V_m$ such that the set of optimal vertices is one of the V_j with probability 1.

Lemma B.1 *For every $(\mu_{n,0},X_{n,0},\Sigma_0)$, there exists a finite collection of disjoint sets $\mathbf{V}=\{V_1,...,V_m\}$ such that $V(X_{n,0},\sigma_0)=V_1\cup...\cup V_m$ and $Pr\{\hat{\gamma}\in\mathbf{V}\}=1$ under the finite-sample normal model (9).*

Proof of Lemma B.1 Let $\gamma,\tilde{\gamma},\check{\gamma} \in V(X_{n,0},\sigma_0)$. Observe that $\gamma,\tilde{\gamma} \in \hat{\gamma}$ only if $\gamma' Y_{n,0} = \tilde{\gamma}' Y_{n,0}$. However, for $Y_{n,0} \sim N(\mu_{n,0},\Sigma_0)$,

$$Pr\{\gamma' Y_{n,0} = \tilde{\gamma}' Y_{n,0}\} \in \{0,1\}.$$

Moreover, $Pr\{\gamma' Y_{n,0} = \tilde{\gamma}' Y_{n,0}\} = 1$ and $Pr\{\gamma' Y_{n,0} = \check{\gamma}' Y_{n,0}\} = 1$ if and only if $Pr\{\gamma' Y_{n,0} = \tilde{\gamma}' Y_{n,0} = \check{\gamma}' Y_{n,0}\} = 1$. It follows that we can partition $V(X_{n,0},\sigma_0)$ into distinct equivalence classes $V_1,...,V_m$ where $\gamma,\tilde{\gamma} \in V(X_{n,0},\sigma)$ are contained in the same V_j if and only if $Pr\{\gamma' Y_{n,0} = \tilde{\gamma}' Y_{n,0}\} = 1$. Towards contradiction, suppose that $Pr\{\hat{\gamma} \in \mathbf{V}\} < 1$. Then it must be that either (i) there exists $\gamma,\tilde{\gamma} \in V_j$ such that $Pr\{\gamma \in \hat{\gamma}, \tilde{\gamma} \notin \hat{\gamma}\} > 0$, or (ii) there exists $\gamma \in V_j$, $\tilde{\gamma} \in V_{j'}$ for $j \neq j'$ such that $Pr\{\gamma \in \hat{\gamma}, \tilde{\gamma} \in \hat{\gamma}\} > 0$. Note, however, that $\gamma \in \hat{\gamma}, \tilde{\gamma} \notin \hat{\gamma}$ only if $\gamma' Y_{n,0} \neq \tilde{\gamma}' Y_{n,0}$, and by construction if $\gamma,\tilde{\gamma} \in V_j$ then $Pr\{\gamma' Y_{n,0} \neq \tilde{\gamma}' Y_{n,0}\} = 0$ so (i) cannot be satisfied. Likewise, $\gamma \in \hat{\gamma}, \tilde{\gamma} \in \hat{\gamma}$ only if $\gamma' Y_{n,0} = \tilde{\gamma}' Y_{n,0}$, and by construction if $\gamma \in V_j, \tilde{\gamma} \in V_{j'}$ then $Pr\{\gamma' Y_{n,0} = \tilde{\gamma}' Y_{n,0}\} = 0$ so (ii) cannot be satisfied. We have thus reached a contradiction. $\square$

Our next result establishes that if one computes the conditional test using the formulas for $\mathcal{V}_{n,0}^{lo}, \mathcal{V}_{n,0}^{up}$ in (14), then one obtains the same values regardless of which element of V_j one chooses. Together with the previous lemma, this result implies that a modified version of the conditional test which chooses arbitrarily among the optimal vertices is well-defined with probability 1 in the finite sample normal model.

Lemma B.2 *Let $V_1,...,V_m$ be as defined in Lemma B.1. Suppose $Y_{n,0}$ follows the finite sample normal model (9). If $\gamma_{(1)},\gamma_{(2)} \in V_j$ for some j, then with probability 1 the values for $\mathcal{V}_{n,0}^{lo}$ and $\mathcal{V}_{n,0}^{up}$ given in (14) are the same if one sets $\gamma = \gamma_{(1)}$ or $\gamma = \gamma_{(2)}$.*

Proof of Lemma B.2 By construction, if $\gamma_{(1)},\gamma_{(2)} \in V_j$ then $Pr\{\gamma_{(1)}' Y_{n,0} = \gamma_{(2)}' Y_{n,0}\} = 1$ for $Y_{n,0} \sim N(\mu_{n,0},\Sigma_0)$. It follows that $(\gamma_{(1)} - \gamma_{(2)})'\Sigma_0 = 0$ and $\gamma_{(1)}' \Sigma \gamma_{(1)} = \gamma_{(2)}' \Sigma \gamma_{(2)}$. It is then immediate that for any $\tilde{\gamma} \in V(X_{n,0},\sigma_0)$, $\gamma_{(1)}' \Sigma_0 \tilde{\gamma} = \gamma_{(2)}' \Sigma_0 \tilde{\gamma}$. Note, however, that the formulas for $\mathcal{V}_{n,0}^{lo}$ and $\mathcal{V}_{n,0}^{up}$ in (14) depend on γ only through the expressions $\gamma' \Sigma_0 \gamma, \gamma' \Sigma_0 \tilde{\gamma}, \Sigma_0 \gamma$, and $\gamma' Y_{n,0}$. Since we have shown that with probability 1 all of these expressions obtain the same value if we set $\gamma = \gamma_{(1)}$ as if we set $\gamma = \gamma_{(2)}$, the result follows. $\square$

Finally, we establish that the conditional test which chooses arbitrarily among the optimal dual vertices controls size in the finite-sample normal model.

Proposition B.1 *Consider a version of the conditional test where the critical values are determined by the formulas for* $\mathcal{V}^{lo}_{n,0},\mathcal{V}^{up}_{n,0}$ *in (14) setting* $\gamma = h(\hat{\gamma})$ *for any arbitrary (possibly randomized) function* $h(\cdot)$ *that selects among the elements of* $\hat{\gamma}$. *Let* ϕ^h_C *denote the indicator for whether the test rejects. Then under the finite sample normal model (9),* $E[\phi^h_C] \le \alpha$ *whenever* $\mu_{n,0} \in \mathcal{M}_{n,0}$.

Proof of Proposition B.1 Observe that the proof to Lemma 1 does not rely on uniqueness of the dual, and thus the statement of Lemma 1 holds replacing the conditioning event $\hat{\gamma} = \gamma$ with $\gamma \in \hat{\gamma}$. Moreover, by Lemma B.1, there is some j such that $Pr\{1\{\gamma \in \hat{\gamma}\} = 1\{\hat{\gamma} = V_j\}\} = 1$. It follows that the statement of Lemma 1 also holds if we replace the conditioning event $\hat{\gamma} = \gamma$ with $\hat{\gamma} = V_j$. Additionally, by Lemma B.2, the values of $\mathcal{V}^{lo}_{n,0},\mathcal{V}^{up}_{n,0}$ are the same for all $\gamma \in V_j$. Thus, the conclusion of Lemma 1 holds if we condition on $\hat{\gamma} = V_j$ and replace all instances of γ with $h(\hat{\gamma})$. By the same argument as in Section 3.3 for the unique-solution case, it then follows that $E[\phi^h_C | \hat{\gamma} = V_j] \le \alpha$ for $\mu_{n,0} \in \mathcal{M}_{n,0}$. But Lemma B.1 implies that $E[\phi^h_C] = \sum_j E[\phi^h_C | \hat{\gamma} = V_j] P\{\hat{\gamma} = V_j\}$, from which unconditional size control is immediate. □

By analogous arguments, one can also establish that the hybrid test is well-defined with probability 1 and controls size in the finite sample normal model when there is multiplicity in the dual.

C Asymptotic Variance Estimation

Assumption 2 requires the existence of a uniformly consistent estimator $\widehat{\Omega}_{n,0}$ for the conditional variance $\Omega(P_{D|Z},\beta_0)$. Here, we establish the uniform consistency of the matching estimator discussed in Section 5.3 under mild conditions. For brevity, we shorthand $U_i(\beta_0)$ as $U_{i,0}$.

Following Abadie et al. (2014), we consider the nearest-neighbor variance estimator given in (16). The intuition for the estimator $\widehat{\Omega}_{n,0}$ is straightforward: provided the conditional mean and variance of $U_{i,0}$ given $Z_i = z$ are smooth in z, if $Z_{\ell_Z(i)}$ is close to Z_i, then the mean and variance of $U_{i,0}|Z_i$ will be nearly the same as the mean and variance of $U_{\ell_Z(i),0}|Z_{\ell_Z(i)}$. Hence, the variance of $U_{i,0} - U_{\ell_Z(i),0}$ will be approximately twice the variance of $U_{i,0}|Z_i$, and the approximation error will vanish as $Z_{\ell_Z(i)}$ approaches Z_i. If the support of Z_i is compact, however, then with a large enough sample we are guaranteed to have observations quite "close" to almost all of our observations, and $\widehat{\Omega}_{n,0}$ will converge to the average conditional variance $\Omega(P_{D|Z},\beta_0)$. The next assumption formalizes the conditions needed for this argument.

Assumption C.1 *For $\lambda_{\max}(A)$ the maximal eigenvalue of a matrix A, the following conditions hold*

1. $\{Z_i\}_{i=1}^{\infty} \subseteq \mathcal{Z}$ *for $\mathcal{Z}$ a compact set*
2. $\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} \frac{1}{n}\sum E_{P_{D|Z}}\left[\|U_{i,0}\|^4 | Z_i\right]$ *is finite*
3. $\mu_{P_{D|Z}}(z,\beta_0) = E_{P_{D|Z}}[U_{i,0}|Z_i = z]$ *is Lipschitz in z with Lipschitz constant uniformly bounded over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$, and is uniformly bounded over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$*
4. $V_{P_{D|Z}}(z,\beta_0) = E_{P_{D|Z}}\left[U_{i,0}U_{i,0}'|Z_i = z\right]$ *is Lipschitz in z with Lipschitz constant uniformly bounded over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$*
5. $\sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} \sup_{z\in\mathcal{Z}} \lambda_{\max}\left(Var_{P_{D|Z}}(U_{i,0}|Z_i = z)\right)$ *is finite*
6. *For $\widehat{\Sigma}_Z = \widehat{Var}(Z_i)$ the sample variance of Z_i, $\widehat{\Sigma}_Z \to \Sigma_Z$ for a positive-definite limit Σ_Z*

Assumption C.1(1) is used only to establish that the average distance between Z_i and $Z_{\ell_Z(i)}$ converges to zero, $\frac{1}{n}\sum\left\|Z_i - Z_{\ell_Z(i)}\right\| \to 0$. Hence, one may instead assume this condition directly. Assumption C.1(2) and (5) restrict the variance and fourth moment of $U_{i,0}$, and are satisfied under a wide range of data generating processes. Assumption C.1(3) and (4) impose Lipschitz continuity on the mean and second moment of $U_{i,0}$, consistent with the heuristic argument given above. Finally, Assumption C.1(6) requires only that $\widehat{\Sigma}_Z$ converge to a positive-definite limit.

Proposition C.1 *Under Assumptions 1 and C.1, for $\widehat{\Omega}_{n,0}$ as defined in (16) and all $\varepsilon > 0$*

$$\lim_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{\beta_0\in B_I(P_{D|Z})} Pr_{P_{D|Z}}\left\{\left\|\widehat{\Omega}_{n,0} - \Omega\left(P_{D|Z},\beta_0\right)\right\| > \varepsilon\right\} = 0,$$

so Assumption 2 holds.

C.1 Proof of Variance Consistency

We first prove two auxiliary lemmas, which we then use to prove Proposition C.1.

Lemma C.1 *Under Assumption C.1,*

$$\frac{1}{n}\sum_{i=1}^{n}\left(U_{\ell_Z(i),0}U_{\ell_Z(i),0}' - V_{P_{D|Z}}(Z_i,\beta_0)\right) \to_p 0$$

uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$.

Proof of Lemma C.1 Note that we can write

$$\frac{1}{n}\sum_{i=1}^{n}\left(U_{\ell_Z(i),0}U'_{\ell_Z(i),0}-V_{P_{D|Z}}(Z_i,\beta_0)\right)=$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(U_{\ell_Z(i),0}U'_{\ell_Z(i),0}-V_{P_{D|Z}}(Z_{\ell_Z(i)},\beta_0)\right)+\frac{1}{n}\sum_{i=1}^{n}\left(V_{P_{D|Z}}(Z_{\ell_Z(i)},\beta_0)-V_{P_{D|Z}}(Z_i,\beta_0)\right),$$

so to prove the result it suffices to show that both terms tend to zero. To show that the second term tends to zero, note that by the triangle inequality and Assumption C.1(4),

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left(V_{P_{D|Z}}(Z_{\ell_Z(i)},\beta_0)-V_{P_{D|Z}}(Z_i,\beta_0)\right)\right\|\leq\frac{1}{n}\sum_{i=1}^{n}\left\|V_{P_{D|Z}}(Z_{\ell_Z(i)},\beta_0)-V_{P_{D|Z}}(Z_i,\beta_0)\right\|$$

$$\leq\frac{K}{n}\sum_{i=1}^{n}\left\|Z_i-Z_{\ell_Z(i)}\right\|$$

for K the upper bound on the Lipschitz constant. Note, next, that since $\mathcal{Z}$ is compact by Assumption C.1(1), the proof of Lemma 1 of Abadie & Imbens (2008) implies that

$$\frac{1}{n}\sum_{i=1}^{n}\left\|Z_i-Z_{\ell_Z(i)}\right\|\to 0.$$

Thus, we immediately see that $\frac{1}{n}\sum_{i=1}^{n}\left(V_{P_{D|Z}}(Z_{\ell_Z(i)},\beta_0)-V_{P_{D|Z}}(Z_i,\beta_0)\right)\to 0$ uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$ and $\beta_0\in B_I(P_{D|Z})$.

We next show that

$$\frac{1}{n}\sum_{i=1}^{n}\left(U_{\ell_Z(i),0}U'_{\ell_Z(i),0}-V_{P_{D|Z}}(Z_{\ell_Z(i)},\beta_0)\right)\to_p 0.$$

To do so, note first that the number of observations that can be matched to a given Z_i, $|\{j:\ell_Z(j)=i\}|$, is bounded above by the so-called "kissing number" which is a finite function $\mathcal{K}(\dim(Z_i))$ of the dimension of Z (see Abadie et al. (2014)). Since $U_{i,0}$ is independent across i, this implies that for $(A)_{jk}$ the (j,k) element of a matrix A,

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(U_{\ell_Z(i),0}U'_{\ell_Z(i),0}-V_{P_{D|Z}}(Z_{\ell_Z(i)},\beta_0)\right)_{jk}|\{Z_i\}_{i=1}^{\infty}\right)$$

$$\leq \mathcal{K}(\dim(Z_i))^2 Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(U_{i,0}U_{i,0}'\right)_{jk}|\{Z_i\}_{i=1}^{\infty}\right)$$

$$=\frac{\mathcal{K}(\dim(Z_i))^2}{n^2}\sum_{i=1}^{n}Var\Big(\left(U_{i,0}U_{i,0}'\right)_{jk}|Z_i\Big).$$

By Assumption C.1(2) and Chebyshev's inequality, however, this implies that

$$\frac{1}{n}\sum_{i=1}^{n}\Big(U_{\ell_Z(i),0}U_{\ell_Z(i),0}'-V_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)\Big)\to_p 0,$$

uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$ and $\beta_0\in B_I(P_{D|Z})$, which completes the proof. □

Lemma C.2 *Under Assumption C.1,*

$$\frac{1}{n}\sum_{i=1}^{n}\Big(U_{i,0}U_{\ell_Z(i),0}'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'\Big)\to_p 0,$$

uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$ *and* $\beta_0\in B_I(P_{D|Z})$.

Proof of Lemma C.2 Note that we can write

$$\frac{1}{n}\sum_{i=1}^{n}\Big(U_{i,0}U_{\ell_Z(i),0}'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'\Big)$$

$$=\frac{1}{n}\sum_{i=1}^{n}\Big(U_{i,0}U_{\ell_Z(i),0}'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'\Big)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\Big(\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'\Big).$$

We first show the initial term converges in probability to zero, and then do the same for the second term.

By independence,

$$E\Big[U_{i,0}U_{\ell_Z(i),0}'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'|Z_i,Z_{\ell_Z(i)}\Big]=0,$$

while the variance of the jkth element is

$$Var_{P_{D|Z}}\Big(\Big(U_{i,0}U_{\ell_Z(i),0}'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'\Big)_{jk}|Z_i,Z_{\ell_Z(i)}\Big)$$

$$=E_{P_{D|Z}}\left[\left(U_{i,0,j}U_{\ell_Z(i),0,k}-\mu_{P_{D|Z},j}(Z_i,\beta_0)\mu_{P_{D|Z},k}\left(Z_{\ell_Z(i)},\beta_0\right)\right)^2|Z_i,Z_{\ell_Z(i)}\right]$$

$$=\begin{array}{c}\mu^2_{P_{D|Z},j}(Z_i,\beta_0)Var_{P_{D|Z}}\left(U_{\ell_Z(i),0,k}|Z_{\ell_Z(i)}\right)+Var_{P_{D|Z}}(U_{i,0,j}|Z_i)\mu^2_{P_{D|Z},k}\left(Z_{\ell_Z(i)},\beta_0\right)\\ +Var_{P_{D|Z}}(U_{i,0,j}|Z_i)Var_{P_{D|Z}}\left(U_{\ell_Z(i),0,k}|Z_{\ell_Z(i)}\right).\end{array}$$

Assumption C.1(5) thus implies that for some constant C,

$$\begin{array}{c}Var_{P_{D|Z}}\left(\left(U_{i,0}U'_{\ell_Z(i),0}-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'\right)_{jk}|Z_i,Z_{\ell_Z(i)}\right)\\ \le\left(\mu^2_{P_{D|Z},j}(Z_i,\beta_0)+\mu^2_{P_{D|Z},k}\left(Z_{\ell_Z(i)},\beta_0\right)+C\right)C\end{array},$$

which, together with Assumption C.1(3) and the finiteness of the "kissing number" $\mathcal{K}(\dim(Z_i))$ (see the proof of Lemma C.1 above) implies that

$$\limsup_{n\to\infty}\sup_{P_{D|Z}\in\mathcal{P}_{D|Z}}\sup_{\beta_0\in B_I(P_{D|Z})}Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(U_{i,0}U'_{\ell_Z(i),0}-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'\right)|\{Z_i\}_{i=1}^{\infty}\right)=0,$$

and thus by Chebyshev's inequality that

$$\frac{1}{n}\sum_{i=1}^{n}\left(U_{i,0}U'_{\ell_Z(i),0}-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'\right)\to_p 0,$$

uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$, as we wanted to show.

To complete the proof, we need only show that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'\right).$$

converges to zero uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$. Note, however, that by the triangle inequality and Assumption C.1(3),

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'\right)\right\|$$

$$\le\frac{1}{n}\sum_{i=1}^{n}\left\|\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)'-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'\right\|$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}\left\|\mu_{P_{D|Z}}(Z_i,\beta_0)\right\|\cdot\left\|\mu_{P_{D|Z}}\left(Z_{\ell_Z(i)},\beta_0\right)-\mu_{P_{D|Z}}(Z_i,\beta_0)\right\|$$

$$\leq \frac{K}{n}\sum_{i=1}^{n}\left\|\mu_{P_{D|Z}}(Z_i,\beta_0)\right\|\cdot\left\|Z_{\ell_Z(i)}-Z_i\right\| \leq \frac{KC}{n}\sum_{i=1}^{n}\left\|Z_{\ell_Z(i)}-Z_i\right\| \tag{20}$$

for K a Lipschitz constant and C a constant. As above, since $\mathcal{Z}$ is compact by Assumption C.1(1), the proof of Lemma 1 of Abadie & Imbens (2008) implies that

$$\frac{1}{n}\sum_{i=1}^{n}\left\|Z_i-Z_{\ell_Z(i)}\right\|\to 0,$$

and thus that (20) converges to zero uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$. □

Proof of Proposition C.1 Following proof of Lemma A.3 in Abadie et al. (2014), note that

$$\widehat{\Omega}_{n,0}=\frac{1}{2n}\sum_{i=1}^{n}\left(U_{i,0}-U_{\ell_Z(i),0}\right)\left(U_{i,0}-U_{\ell_Z(i),0}\right)'$$

$$=\frac{1}{2n}\sum_{i=1}^{n}U_{i,0}U_{i,0}'+\frac{1}{2n}\sum_{i=1}^{n}U_{\ell_Z(i),0}U_{\ell_Z(i),0}'-\frac{1}{2n}\sum_{i=1}^{n}\left(U_{i,0}U_{\ell_Z(i),0}'+U_{\ell_Z(i),0}U_{i,0}'\right).$$

Assumption C.1(2) together with Chebyshev's inequality implies that

$$\frac{1}{2n}\sum_{i=1}^{n}\Big(U_{i,0}U_{i,0}'-V_{P_{D|Z}}(Z_i,\beta_0)\Big)\to_p 0$$

uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$, $\beta_0\in B_I(P_{D|Z})$. Since

$$Var(U_{i,0}|Z_i)=V_{P_{D|Z}}(Z_i,\beta_0)-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)',$$

however, we see that

$$\frac{1}{n}\sum_{i=1}^{n}Var_{P_{D|Z}}(U_{i,0}|Z_i)=\frac{1}{n}\sum_{i=1}^{n}V_{P_{D|Z}}(Z_i,\beta_0)-\frac{1}{n}\sum_{i=1}^{n}\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'.$$

Thus, to prove that

$$\widehat{\Omega}_{n,0}-\frac{1}{n}\sum_{i=1}^{n}Var_{P_{D|Z}}(U_{i,0}|Z_i)\to_p 0,$$

it suffices to prove that

$$\frac{1}{n}\sum_{i=1}^{n}\left(U_{\ell_Z(i),0}U'_{\ell_Z(i),0}-V_{P_{D|Z}}(Z_i,\beta_0)\right)\to_p 0$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\left(U_{i,0}U'_{\ell_Z(i),0}-\mu_{P_{D|Z}}(Z_i,\beta_0)\mu_{P_{D|Z}}(Z_i,\beta_0)'\right)\to_p 0,$$

where the first statement follows from Lemma C.1 and the second from Lemma C.2. Since

$$\frac{1}{n}\sum_{i=1}^{n}Var_{P_{D|Z}}(U_{i,0}|Z_i)-\Omega(P_{D|Z},\beta_0)\to 0$$

uniformly over $P_{D|Z}\in\mathcal{P}_{D|Z}$ and $\beta_0\in B_I(P_{D|Z})$ by Assumption 1, however, the result follows by the triangle inequality. $\square$

D Sufficient Conditions for Assumption 4

We now provide lower-level sufficient conditions for Assumption 4 for the case where the degeneracy in Σ_0 arises from moment equalities represented as inequalities, or other moment pairs which cannot bind simultaneously. This setting is similar to that in Assumption E.3.2 in Kaido et al. (2018).

Assumption D.1 *We can write $Y_i(\beta_0)=TU_i(\beta_0)+\zeta_i(\beta_0)$, where $\zeta_i(\beta_0)$ is non-stochastic conditional on Z_i, and $U_i(\beta_0)$ satisfies the conditions of Assumption 1. Further, we can decompose $U_{n,0}=\frac{1}{\sqrt{n}}\sum U_i(\beta_0)$ as $U_{n,0}=(U'_{n,0,1},U'_{n,0,2})'$, where the matrix T takes the form*

$$T=\begin{bmatrix} I_{dim(U_{n,0,1})} & 0 \\ -I_{dim(U_{n,0,1})} & 0 \\ 0 & I_{dim(U_{n,0,2})} \end{bmatrix},$$

while $\zeta_i(\beta_0)=[\zeta_{i1}(\beta_0)'\ \zeta_{i2}(\beta_0)'\ \zeta_{i3}(\beta_0)']'$ with $\zeta_{i1}(\beta_0)+\zeta_{i2}(\beta_0)\le 0$ (elementwise).[35] *We can likewise decompose $X_{n,0}=TQ_{n,0}$ for a comformable matrix $Q_{n,0}$.*

We note that Assumption D.1 is trivially satisfied with $T=I$ when $E[Var(Y_i(\beta_0)|Z_i)]$ is guaranteed to be full rank.

[35] Observe that $e'_jE[U_i(\beta_0)+\zeta_{i1}-Q\delta|Z_i]+e'_jE[-U_i(\beta_0)+\zeta_{2i}+Q\delta|Z_i]=\zeta_{1i}+\zeta_{2i}$, regardless of $E[U_i(\beta_0)|Z_i]$, and thus the null hypothesis can only possibly be satisfied if $\zeta_{i1}+\zeta_{i2}\le 0$.

Our second primitive condition ensures that for n sufficiently large, $X_{n,0}$ lies in a set on which the distance between distinct vertices of $V(X,\upsilon)$ is bounded away from zero (where $\upsilon = \sqrt{diag(TT')}$). Let $\mathcal{B}$ denote the set of $B \subset \{1,...,k+p+1\}$ with $|B| = k$ and $1 \in B$.

Assumption D.2 *For n sufficiently large and all β_0, $X_{n,0}$ is contained in a set $\mathcal{X}$ such that for some constant $\omega > 0$ and any distinct $B, B' \in \mathcal{B}$, either*

1. *$A_B(X,\upsilon)^{-1}e_1 = A_{B'}(X,\upsilon)^{-1}e_1$ for all $X \in \mathcal{X}$ such that $A_B(X,\upsilon)$ and $A_{B'}(X,\upsilon)$ are full-rank, OR*
2. *$||A_B(X,\upsilon)^{-1}e_1 - A_{B'}(X,\upsilon)^{-1}e_1|| \geq \omega$ for all $X \in \mathcal{X}$ such that $A_B(X,\upsilon)$ and $A_{B'}(X,\upsilon)$ are full-rank*

where the matrix $A_B(X,\upsilon)$ is as defined as in Lemma A.1.

Recall from Lemma A.1 that each vertex in $V(X,\upsilon)$ corresponds to $A_B(X,\upsilon)^{-1}e_1$ for some B, so Assumption D.1 guarantees that the distance between distinct vertices of $V(X,\upsilon)$ is bounded from below over $X \in \mathcal{X}$. We note that Assumption D.2 is satisfied trivially if $X_{n,0}/||X_{n,0}||$ is constant, since in that case $V(X_{n,0},\upsilon)$ is constant.

Proposition D.1 *Assumptions D.1 and D.2 imply Assumption 4.*

To prove Proposition D.1, we first establish some auxilliary lemmas. In the following results, we partition a vertex $\gamma \in V(X,\upsilon)$ as $(\gamma_1', \gamma_2', \gamma_3')'$ comformably with the blocks of T in Assumption D.1. We also define $V_{\mathcal{B}^*}(X,\upsilon) \subset V(X,\upsilon)$ to be the subset of $V(X,\upsilon)$ such that $\max\{e_j'\gamma_1, e_j'\gamma_2\} = 0$ for each $j = 1,...,\dim(\gamma_1)$. Intuitively, $V_{\mathcal{B}^*}(X,\upsilon)$ is the set of vertices that have at most one positive entry corresponding with each pair of matching moments of opposite signs.

Lemma D.1 *If Assumption D.1 holds, then for any $\gamma, \tilde{\gamma} \in V_{\mathcal{B}^*}(X,\sigma)$ and $c \geq 0$,*

$$||(\gamma - c \cdot \tilde{\gamma})'T|| \geq k^{-\frac{1}{2}} ||\gamma - c \cdot \tilde{\gamma}||.$$

Proof of Lemma D.1 To establish the result, it suffices to show that

$$||(\gamma - c \cdot \tilde{\gamma})'T||_\infty \geq ||\gamma - c \cdot \tilde{\gamma}||_\infty, \tag{21}$$

where $||x||_\infty = \max\{|x_1|,...,|x_k|\}$ is the ℓ_∞ norm. The desired result then follows from the fact that for any $x \in \mathbb{R}^k$, $||x|| \geq ||x||_\infty \geq k^{-\frac{1}{2}}||x||$.

Clearly, the inequality (21) holds trivially when $\gamma - c\cdot\tilde{\gamma} = 0$, so for the remainder of the proof we consider the case where $||\gamma - c\cdot\tilde{\gamma}||_\infty = m > 0$. Write

$$(\gamma - c\cdot\tilde{\gamma})'T = \begin{pmatrix} \gamma_1 - \gamma_2 \\ \gamma_3 \end{pmatrix}' - c\cdot \begin{pmatrix} \tilde{\gamma}_1 - \tilde{\gamma}_2 \\ \tilde{\gamma}_3 \end{pmatrix}'.$$

It is clear from the previous display that if $|\gamma_{3,j} - c\cdot\tilde{\gamma}_{3,j}| = m$ for some j, then $||(\gamma - c\cdot\tilde{\gamma})'T||_\infty \geq m$. Consider next the case where $|\gamma_{1,j} - c\cdot\tilde{\gamma}_{1,j}| = m$ for some j. Suppose first that $\gamma_{1,j} > c\cdot\tilde{\gamma}_{1,j} \geq 0$. By the definition of $V_{\mathcal{B}^*}(X,\sigma)$, this implies that $\gamma_{2,j} = 0$. Hence the jth element of $(\gamma - c\cdot\tilde{\gamma})'T$ is equal to

$$\underbrace{\gamma_{1,j} - \tilde{\gamma}_{1,j}}_{=m} + \underbrace{c\cdot\tilde{\gamma}_{2,j}}_{\geq 0} \geq m,$$

which implies that $||(\gamma - c\cdot\tilde{\gamma})'T||_\infty \geq m$. Likewise, if $c\cdot\tilde{\gamma}_{1,j} > \gamma_{1,j} \geq 0$, then we know that $\tilde{\gamma}_{2,j} = 0$, and thus the jth element of $(\gamma - c\cdot\tilde{\gamma})'T$ is equal to

$$\underbrace{\gamma_{1,j} - c\cdot\tilde{\gamma}_{1,j}}_{=-m} - \underbrace{\gamma_{1,j}}_{\geq 0} \leq -m,$$

which implies that $||(\gamma - c\cdot\tilde{\gamma})'T||_\infty \geq m$. We have thus established that $||(\gamma - c\cdot\tilde{\gamma})'T||_\infty \geq m$ when $|\gamma_{1,j} - c\tilde{\gamma}_{1,j}| = m$ for some j. The case where $|\gamma_{2,j} - c\tilde{\gamma}_{2,j}| = m$ for some j can be handled analogously. □

Lemma D.2 *If Assumption D.1 holds, then there exists a constant $c_\lambda > 0$ such that $c_\lambda^{-1} \leq \lambda_j(X,\sigma(\Omega)) \leq c_\lambda$ for all $\Omega \in \mathbf{\Omega}_{\bar{\lambda}}$ and for all j and X, where the function $\lambda_j(X,\sigma)$ is as given in Lemma 4.*

Proof of Lemma D.2 Recall from the proof of Lemma A.2 that $\lambda_j(X,\sigma) = 1/((A_B(X,\upsilon)^{-1}e_1)'\sigma(\Omega))$ for some index set B. Since by construction $(A_B(X,\upsilon)^{-1}e_1)'\upsilon = 1$, we have that

$$\lambda_j(X,\sigma) = \frac{(A_B(X,\upsilon)^{-1}e_1)'\upsilon}{(A_B(X,\upsilon)^{-1}e_1)'\sigma(\Omega)}.$$

Since $A_B(X,\upsilon)^{-1}e_1$, υ, and $\sigma(\Omega)$ are all non-negative vectors by construction, it thus suffices to establish that $c_\lambda^{-1}\upsilon \leq \sigma(\Omega) \leq c_\lambda \upsilon$ (where the inequalities hold elementwise). Observe, however, that $\upsilon_j = ||T_j||$, whereas $\sigma(\Omega)_j = \sqrt{T_j \Omega T_j'}$. However, since the eigenvalues of Ω are bounded above and below by $\bar{\lambda}$ and $\bar{\lambda}^{-1}$ respectively, we have that for every j, $||T_j||^2\bar{\lambda}^{-1} \leq T_j\Omega T_j' \leq \bar{\lambda}||T_j||^2$, and hence $c_\lambda^{-1}\upsilon_j \leq \sigma(\Omega)_j \leq c_\lambda \upsilon_j$ for $c_\lambda = \bar{\lambda}^{\frac{1}{2}}$. □

Proof of Proposition D.1 First, we show that $V^{\dagger}(X,\sigma)\subseteq V_{\mathcal{B}_*}(X,\sigma)$ for all σ. Suppose that $\gamma\in V^{\dagger}(X,\sigma)$. By part 1 of Lemma 4, $\gamma=\lambda(\sigma)\bar{\gamma}$ for a scalar function $\lambda(\sigma)$ and vector $\bar{\gamma}$ (both depending on X). Under the structure imposed by Assumption D.1, the fact that $\gamma\in V^{\dagger}(X,\sigma)$ implies that for some $\tilde{\sigma}$, $\tilde{\gamma}=\lambda(\tilde{\sigma})\bar{\gamma}$ is a Lagrange multiplier for the primal linear program

$$\hat{\eta}=\min_{\eta,\delta}\eta \text{ subject to } \left(Tu+\begin{pmatrix} \zeta_1' & \zeta_2' & \zeta_3' \end{pmatrix}'-TQ\delta\leq\eta\cdot\tilde{\sigma}\right)$$

for some u such that $\hat{\eta}>0$. Observe, however, that the constraints in the linear program corresponding with $\tilde{\gamma}_{1,j}$ and $\tilde{\gamma}_{2,j}$ can bind simultaneously only if

$$e_j'(u-Q\delta^*)+e_j'\zeta_1=\hat{\eta}e_j'\tilde{\sigma}=-e_j'(u-Q\delta^*)+e_j'\zeta_2,$$

for δ^* an optimizer to the linear program for $\hat{\eta}$. This implies that $\hat{\eta}=\frac{1}{2e_j'\tilde{\sigma}}e_j'(\zeta_1+\zeta_2)\leq 0$. Since $\hat{\eta}>0$, it must be that at most one of the moments corresponding with $\tilde{\gamma}_{1,j}$ and $\tilde{\gamma}_{2,j}$ is binding. Hence, complementary slackness implies that $\min\{e_j'\tilde{\gamma}_1,e_j'\tilde{\gamma}_2\}=0$, and thus that $\min\{e_j'\gamma_1,e_j'\gamma_2\}=0$ since $\gamma\propto\tilde{\gamma}$. It follows that $\gamma\in V_{\mathcal{B}_*}(X,\sigma)$, as we wished to show.

Next, note that since every $\Omega\in\mathbf{\Omega}_{\bar{\lambda}}$ has eigenvalues bounded below by assumption, Assumption 4 can fail only if there exists a sequence of $\Omega_m\in\mathbf{\Omega}_{\bar{\lambda}}$, $X_m\in\mathcal{X}$, distinct vertices $\gamma_m,\tilde{\gamma}_m\in V_{\dagger}(X_m,\sigma(\Omega_m))$, and values $c_m\geq 0$ such that $||(\gamma_m-c_m\cdot\tilde{\gamma}_m)'T||\to 0$ as $m\to\infty$. From Lemma D.1 combined with the argument in the previous paragraph, it follows that Assumption 4 can fail only if there exist a sequence of distinct vertices $\gamma_m,\tilde{\gamma}_m\in V_{\mathcal{B}^*}(X_m,\sigma(\Omega_m))$ and values $c_m\geq 0$ such that $||\gamma_m-c_m\cdot\tilde{\gamma}_m||\to 0$ as $m\to\infty$. Towards contradiction, suppose that such a sequence exists. Since by construction $\gamma_m'\sigma_m=\tilde{\gamma}_m'\sigma_m=1$, where $\sigma_m=\sigma(\Omega_m)$, we have that $|\sigma_m'(\gamma_m-c_m\cdot\tilde{\gamma}_m)|=|1-c_m|$. By the Cauchy-Schwarz inequality, it follows that $||\gamma_m-c_m\cdot\tilde{\gamma}_m||\geq|1-c_m|/||\sigma_m||$. However, since Ω_m has eigenvalues bounded above, $||\sigma_m||$ is bounded above, and thus it must be that $c_m\to 1$. Note further that $\sigma_{m,j}^2=T_j\Omega_mT_j'$, where by Assumption D.1, $||T_j||=1$, and thus $\sigma_{m,j}^2\geq\bar{\lambda}^{-1}$. Since the elements of $\sigma_m>0$ are bounded away from zero while $\gamma_m,\tilde{\gamma}_m\geq 0$ and $\gamma_m'\sigma_m=\tilde{\gamma}'\sigma_m=1$, we know that $||\gamma_m||$ and $||\tilde{\gamma}_m||$ are both bounded above. It follows that we can find a convergent subsequence indexed by r such that $\gamma_r\to\gamma$. This, together with the fact that $||\gamma_r-c_r\cdot\tilde{\gamma}_r||\to 0$ and $c_r\to 1$ implies that $\tilde{\gamma}_r\to\gamma$ as well. Thus, we see that Assumption 4 can be violated only if we can find a sequence of distinct vertices γ_r and $\tilde{\gamma}_r$ in $V_{\mathcal{B}^*}(X_r,\sigma_r)$ such that $\gamma_r-\tilde{\gamma}_r\to 0$.

The fact that $\gamma_r-\tilde{\gamma}_r\to 0$ further implies that there exist a sequence of distinct vertices ϑ_s and $\tilde{\upsilon}_s$ in $V_{\mathcal{B}^*}(X_s,\upsilon)$ such that $\vartheta_s-\tilde{\vartheta}_s\to 0$. To see this, recall that we can write

$\gamma_r = \lambda_{B_r}(X_r, \sigma_r)\gamma_{B_r}(X_r, \upsilon)$, where $\gamma_{B_r}(X, \upsilon) = A_{B_r}(X, \upsilon)^{-1}e_1$ and $\lambda_B(\cdot,\cdot)$ is a scalar which we showed to be bounded both above and away from zero in Lemma D.2. Since the set of possible values for B_r is finite, we can extract a subsequence r_1 on which B_{r_1} is constant. We can likewise extract a further subsequence r_2 on which $\tilde{B}_{r_2}$ is constant, where $\tilde{B}_r$ is defined analogously to B_r, i.e. $\tilde{\gamma}_r = \lambda_{\tilde{B}_r}(X_r, \sigma_r)\gamma_{\tilde{B}_r}(X_r, \upsilon)$. Since the values of the $\lambda(\cdot)$ functions are bounded both above and away from zero, we can extract a further subsequence s along which $\lambda_{B_s}(X_s, \sigma_s) \to \lambda^* > 0$ and $\lambda_{\tilde{B}_s}(X_s, \sigma_s) \to \tilde{\lambda}^* > 0$. Since $\gamma_s \to \gamma$ and $\lambda_{B_s}(X_s, \sigma_s) \to \lambda^*$, it follows that $\vartheta_s = \gamma_{B_s}(X_s, \upsilon) \to \frac{1}{\lambda^*}\gamma$. Likewise, we have that $\tilde{\vartheta}_s = \gamma_{\tilde{B}_s}(X_s, \upsilon) \to \frac{1}{\tilde{\lambda}^*}\gamma$. However, by construction $\vartheta_s'\upsilon = \tilde{\vartheta}_s'\upsilon = 1$, which implies

$$1 = \lim_{s\to\infty} \vartheta_s'\upsilon = \frac{1}{\lambda^*}\gamma'\upsilon = \lim_{s\to\infty} \tilde{\vartheta}_s\upsilon = \frac{1}{\tilde{\lambda}^*}\gamma'\upsilon,$$

and hence $\lambda^* = \tilde{\lambda}^*$. It follows that $\vartheta_s - \tilde{\vartheta}_s \to 0$.

However, by construction $\vartheta_s = A_B(X_s, \upsilon)^{-1}e_1$ and $\tilde{\vartheta}_s = A_{\tilde{B}}(X_s, \upsilon)^{-1}e_1$ with $\vartheta_s \neq \tilde{\vartheta}_s$. It follows that $||A_B(X_s, \upsilon)^{-1}e_1 - A_{\tilde{B}}(X_s, \upsilon)^{-1}e_1|| \to 0$, which contradicts Assumption D.2. □

E Computation of $\mathcal{V}_{n,0}^{lo}$ and $\mathcal{V}_{n,0}^{up}$

We now provide additional details on the computation of the truncation points $\mathcal{V}_{n,0}^{lo}$ and $\mathcal{V}_{n,0}^{up}$ for the conditional and hybrid tests. Equation (14) gives formulas for $\mathcal{V}_{n,0}^{lo}$ and $\mathcal{V}_{n,0}^{up}$ that require taking a maximum/minimum over all of the dual vertices, which may be computationally challenging in practice. To facilitate computation, we provide two results which together allow for rapid calculation of these endpoints even when the number of dual vertices is large.

Our first result provides conditions under which $\mathcal{V}_{n,0}^{lo}$ and $\mathcal{V}_{n,0}^{up}$ can be calculated as the maximum/minimum over sets with at most k elements.

Lemma E.1 *Suppose the primal problem (10) has a solution (η^*, δ^*). Let $B \subset \{1,...,k\}$ denote the set of binding moments at (η^*, δ^*).*[36] *Let $W_{n,0} = (\widehat{\sigma}_{n,0},\ X_{n,0})$ and let M_B be the matrix so that $M_B W_{n,0}$ selects the rows of $W_{n,0}$ corresponding with the index set B. If $|B| = p+1$, $W_{n,0,B}$ is invertible (i.e., the primal solution is non-degenerate), and $e_1' W_{n,0,B}^{-1} \geq 0$, then the vector γ with $M_B\gamma = (e_1' W_{n,0,B}^{-1})'$ and remaining elements equal to 0 is a solution to the dual problem. Moreover, for $L = (I - W_{n,0}W_{n,0,B}^{-1}M_B)$ and $\Delta = \widehat{\Sigma}_{n,0}\gamma/(\gamma'\widehat{\Sigma}_{n,0}\gamma)$, we have that*

$$\mathcal{V}_{n,0}^{lo} = \max_{j:(L\Delta)_j < 0} -\frac{(LS_{n,0,\gamma})_j}{(L\Delta)_j} \text{ and } \mathcal{V}_{n,0}^{up} = \min_{j:(L\Delta)_j > 0} -\frac{(LS_{n,0,\gamma})_j}{(L\Delta)_j} \tag{22}$$

[36] That is, $Y_{n,0,B} - X_{n,0,B}\delta^* = \eta^* \cdot \hat{\sigma}_{n,0,B}$ and $Y_{n,0,-B} - X_{n,0,-B}\delta^* < \eta^* \cdot \hat{\sigma}_{n,0,-B}$, where we use the notation $-B$ to denote rows not contained in B.

for $\mathcal{V}^{lo}_{n,0},\mathcal{V}^{up}_{n,0}$ *as defined in (14).*

Proof of Lemma E.1 It is straightforward to verify that γ satisfies the Karush-Kuhn-Tucker (KKT) conditions at (η^*,δ^*). The KKT conditions are necessary and sufficient for the solution to a linear program, and thus γ is a solution to the dual problem. (In fact, if the primal is non-degenerate, then the dual is unique (e.g. Wachsmuth 2013, Theorem 1(v)), so γ must be the unique dual solution, $\hat{\gamma}=\gamma$.) Observe that when (η^*,δ^*) is a solution to the primal problem with rows indexed by B binding, then $(\eta^*,\delta^{*\prime})'=W^{-1}_{n,0,B}M_BY_{n,0}$. Since the KKT conditions are necessary and sufficient, it follows that $\gamma'y=\max_{\tilde{\gamma}\in V(X_{n,0},\hat{\sigma}_{n,0})}\tilde{\gamma}'y$ if and only if $Ly=y-W_{n,0}W^{-1}_{n,0,B}M_By\leq 0$. But we argued in the proof to Lemma 4 that when $\hat{\gamma}=\gamma$, $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$ are respectively the minimum and maximum of the set

$$\left\{\gamma'y|y \text{ s.t. } \gamma'y\geq \max_{\tilde{\gamma}\in V(X_{n,0},\hat{\sigma})}\tilde{\gamma}'y \text{ and } S(y,\gamma)=S_{n,0,\gamma}\right\},$$

which by the preceeding argument is equivalent to the set

$$\{\gamma'y|y \text{ s.t. } Ly\leq 0 \text{ and } S(y,\gamma)=S_{n,0,\gamma}\}.$$

The result then follows from Lemma 5.1 in Lee et al. (2016). □

Since the dual-simplex method naturally returns the solution η^* and optimizer δ^*, it is straightforward to verify that $W_{n,0,B}$ is invertible and $e_1'W^{-1}_{n,0,B}\geq 0$. If these conditions are met, then $\mathcal{V}^{lo}_{n,0},\mathcal{V}^{up}_{n,0}$ can be calculated using (22), which is computationally straightforward since it involves a maximum/minimum over sets of at most k elements. For cases where the conditions for Lemma E.1 are not met, the following result provides a useful alternative method for computing $\mathcal{V}^{lo}_{n,0},\mathcal{V}^{up}_{n,0}$.

Lemma E.2 *Suppose* γ *is a solution to the dual problem and* $\gamma'\widehat{\Sigma}_{n,0}\gamma>0$. *Then the values of* $\mathcal{V}^{lo}_{n,0}$ *and* $\mathcal{V}^{up}_{n,0}$ *associated with* γ *correspond, respectively, to the minimum and maximum of the convex set*

$$C=\left\{c|c=\max_{\tilde{\gamma}\in V(X_{n,0},\hat{\sigma}_{n,0})}\tilde{\gamma}'\left(S_{n,0,\gamma}+\frac{c}{\gamma'\widehat{\Sigma}_{n,0}\gamma}\widehat{\Sigma}_{n,0}\gamma\right)\right\}.$$

Proof of Lemma E.2 Recall that the values of $\mathcal{V}^{lo}_{n,0}$ and $\mathcal{V}^{up}_{n,0}$ associated with γ are the minimum and maximum of the set

$$\tilde{C} = \left\{ \gamma' y | y \text{ s.t. } \gamma' y \geq \max_{\tilde{\gamma} \in V(X_{n,0}, \hat{\sigma}_{n,0})} \tilde{\gamma}' y \text{ and } S(y,\gamma) = S_{n,0,\gamma} \right\}.$$

From the definition of $S(y,\gamma) = \left(I - \left(\gamma' \widehat{\Sigma}_{n,0} \gamma \right)^{-1} \widehat{\Sigma}_{n,0} \gamma \gamma' \right) y$, we have that $y = S(y,\gamma) + (\gamma' y) / \left(\gamma' \widehat{\Sigma}_{n,0} \gamma \right) \cdot \widehat{\Sigma}_{n,0} \gamma$, from which it follows that

$$\tilde{C} = \left\{ \gamma' y | y \text{ s.t. } \gamma' y \geq \max_{\tilde{\gamma} \in V(X_{n,0}, \hat{\sigma}_{n,0})} \tilde{\gamma}' \left(S_{n,0,\gamma} + \frac{\gamma' y}{\gamma' \widehat{\Sigma}_{n,0} \gamma} \widehat{\Sigma}_{n,0} \gamma \right) \text{ and } S(y,\gamma) = S_{n,0,\gamma} \right\}.$$

To establish that $\tilde{C} = C$, it thus suffices to show that $\{\gamma' y | S(y,\gamma) = S_{n,0,\gamma}\} = \mathbb{R}$, which follows from the assumption that $\gamma' \widehat{\Sigma}_{n,0} \gamma > 0$ along with the fact that if $S(y,\gamma) = s$ then $S\left(y + a \cdot \widehat{\Sigma}_{n,0} \gamma, \gamma \right) = s$ for any $a \in \mathbb{R}$ (which follows immediately from the definition of $S(y,\gamma)$). Finally, convexity follows immediately from the form of $\tilde{C}$ and the fact that $\max_{\tilde{\gamma} \in V(X_{n,0}, \hat{\sigma}_{n,0})} \tilde{\gamma}' y$ is convex in y. $\square$

Lemma E.2 implies that $\mathcal{V}^{lo}_{n,0}, \mathcal{V}^{up}_{n,0}$ can be calculated via a bisection method. The intuition for the algorithm is as follows. By construction, $\hat{\eta}_{n,0} \in C$. If there is some large value M such that $M \notin C$, then we know that $\mathcal{V}^{up}_{n,0}$ lies between $\hat{\eta}_{n,0}$ and M. We start by testing whether the midpoint between $\hat{\eta}_{n,0}$ and M falls in the set C by solving the linear program in the definition of C. If this point lies within C, then we can test the midpoint between the previously tested value and M, whereas if it does not, then we can test the midpoint between $\hat{\eta}_{n,0}$ and the previous midpoint. We can proceed in this way to narrow down the range in which $\mathcal{V}^{up}_{n,0}$ must fall. This tends to be computationally efficient, since the range in which $\mathcal{V}^{up}_{n,0}$ can lie is reduced by a factor of 2 in each step. Algorithm E.1 below formally describes the algorithm used for bisection (and is implemented in our Matlab code). We recommend initializing the value of M to some large value such that, for computational purposes, if $\mathcal{V}^{up}_{n,0} > M$ then it would suffice to set $\mathcal{V}^{up}_{n,0} = \infty$.[37] Note that the formulas in Lemma E.2 require knowledge of a dual solution γ. Fortunately, the dual-simplex method returns a dual solution by default, and thus γ can be obtained at no additional computational cost.

We note that whenever the conditions of Lemma E.1 are met, the dual solution is

[37] In our implementation, we set $M = \max\left(100, \hat{\eta}_{n,0} + 20\sqrt{\gamma' \widehat{\Sigma} \gamma} \right)$, which guarantees that M is at least 20 standard deviations above $\hat{\eta}_{n,0}$.

unique, since non-degeneracy in the primal implies uniqueness in the dual (e.g. Wachsmuth 2013, Theorem 1(v)). If the conditions of Lemma E.1 are not met, then the dual may or may not be unique. A researcher interested in testing whether the dual is unique can use the algorithm suggested by Appa (2002) to verify the uniqueness of a linear program. We note, however, that as described in Appendix B, uniqueness of the dual is not needed for the validity of the our tests in the finite-sample normal model. Tests based on the formulas given in Lemma E.2 using an arbitrarily-chosen dual solution therefore remain valid in the finite-sample normal model. Our conditions for asymptotic size control do imply, however, that the dual will be unique with probability tending to one.

Algorithm E.1 Bisection Method for Calculating $V^{up}_{n,0}$

procedure COMPUTEVUP
    if CheckIfInC(M) **then**
        $V^{up}_{n,0} \leftarrow \infty$
    else
        $lb \leftarrow \hat{\eta}_{n,0}$
        $ub \leftarrow M$
        while $ub - lb > TolV$ **do**
            mid $\leftarrow \frac{1}{2}(lb + ub)$
            if CheckIfInC(mid) **then**
                $lb \leftarrow$ mid
            else
                $ub \leftarrow$ mid
        $V^{up}_{n,0} \leftarrow \frac{1}{2}(lb + ub)$

where we define the functions:

function LPVALUE(c)
    return

$$\max_{\tilde{\gamma}} \tilde{\gamma}' \left(S_{n,0,\gamma} + \frac{\widehat{\Sigma}_{n,0}\gamma}{\gamma'\widehat{\Sigma}_{n,0}\gamma} c \right)$$
$$\text{subject to } \tilde{\gamma} \geq 0, W'_{n,0}\tilde{\gamma} = e_1$$

function CHECKIFINC(c)
    if $|\, c - LPValue(c)| < TolLP$ **then**
        return True
    else
        return False

F Connections to LICQ

We now briefly discuss the connections and differences between Assumption 4 and linear independence constraint qualification (LICQ) conditions that have been imposed in the literature. We refer the reader to Kaido et al. (2021) for detailed discussion of constraint qualifications in the moment inequality literature, and Section 3 of Rambachan & Roth (2022) for additional results for our conditional test under LICQ.

We focus on the special case where the target parameter is scalar ($\beta \in \mathbb{R}$) and enters the moments linearly, which simplifies exposition and facilitates comparisons to other papers that consider the LICQ or closely related assumptions in the linear case (e.g. Cho & Russell 2021, Gafarov 2019, Kaido & Santos 2014). That is, we consider moments of the form $Y_i - X_{i,\beta}\beta - X_{i,\delta}\delta$, where $Y_i \in \mathbb{R}^k$, $X_{i,\beta} \in \mathbb{R}^k$, $X_{i,\delta} \in \mathbb{R}^{k \times p}$, and $(Y_i, X_{i,\delta}, X_{i,\beta})$ doesn't depend on β or δ.

To give a formal definition of LICQ, we introduce the following notation. Let $X_i = (X_{i,\beta}, X_{i,\delta})$ and $\tau = (\beta, \delta')'$, so that we can write the moments as $Y_i - X_i\tau$. Define $\mathbb{T} = \{\tau | E_P[Y_i - X_i\tau] \leq 0\}$ to be the set of values for τ such that the unconditional moments are satisfied, and define the set of support points in direction p by $S(p) = \{\tau | p'\tau = \sup_{\tilde{\tau} \in \mathbb{T}} p'\tilde{\tau}\}$. We will be most interested in the support points in the directions e_1 and $-e_1$, so that the optimization in the definition of $S(p)$ corresponds with the upper and lower bounds for β. We say that LICQ holds in the direction p if for all $\tau^* \in S(p)$, the matrix X_B has full row rank, where $X = E_P[X_i]$ and B is the set of rows such that $E_P[Y_{i,B} - X_{i,B}\tau^*] = 0$.[38]

We now show that LICQ implies uniqueness in a "population version" of the dual problem for our test statistic. Specifically, for any $\sigma \in \mathbb{R}^k$ with $\sigma > 0$, let

$$\eta(Y, X, \beta, \sigma) = \min_{\eta, \delta} \eta \text{ s.t. } Y - X_\beta\beta - X_\delta\delta \leq \sigma \cdot \eta.$$

We then have the following result for the dual problem to $\eta(Y, X, \beta, \sigma)$.

Lemma F.1 *Let $\beta^{ub} = \sup_{\tau \in \mathbb{T}} e_1'\tau$ and $\mu = E_P[Y_i]$. If LICQ holds in the direction e_1, then for any $\sigma > 0$, $\eta(\mu, X, \beta^{ub}, \sigma)$ has a unique dual solution, i.e. there is a unique solution to*

$$\max_{\gamma \in V(X_\delta, \sigma)} \gamma'(\mu - X_\beta\beta^{ub}).$$

[38] LICQ is typically defined in terms of the Jacobian of the expectation of the moments with respect to τ, but in our linear setting the Jacobian of $E_P[Y_i - X_i\tau]$ is simply $-X$.

Proof of Lemma F.1 We first show that $\eta(\mu,X,\beta^{ub},\sigma)=0$. Since $\beta^{ub}=\sup_{\tau\in\mathbb{T}}e_1'\tau$ by definition, we must have that $\eta(\mu,X,\beta^{ub},\sigma)\leq 0$. Towards contradiction, suppose that $\eta(\mu,X,\beta^{ub},\sigma)<0$. Then there exists δ^* such that $\mu-X_\beta\beta^{ub}-X_\delta\delta^*<0$. But then for some $\epsilon>0$, $\mu-X_\beta(\beta^{ub}+\epsilon)-X_\delta\delta^*<0$, which is a contradiction, since it implies that $\sup_{\tau\in\mathbb{T}}e_1'\tau>\beta$.

We thus see that if δ^* is a solution for $\eta(\mu,X,\beta^{ub},\sigma)$, then $(\beta^{ub},\delta^{*\prime})'\in S(e_1)$. Hence, LICQ implies that for B the set of binding moments at δ^*, we have that $X_B=(X_{\beta,B},X_{\delta,B})$ has rank $|B|$. It follows that $X_{\delta,B}$ has rank $|B|-1$. However, observe that there can be no $\tilde{\delta}$ such that $X_{\delta,B}\tilde{\delta}>0$, since if there were, then for $\epsilon>0$ sufficiently small we would have that $\mu_B-X_{\beta,B}\beta^{ub}-X_{\delta,B}(\delta^*+\epsilon\tilde{\delta})<0$ while the remaining moments are still slack, and thus $\eta(\mu,X,\beta^{ub},\sigma)<0$. Since $\sigma_B>0$, it follows that $W_B=(\sigma_B,X_{\delta,B})$ has rank $|B|$. Note that W_B is the gradient of the binding constraints at the optimum to $\eta(\mu,X,\beta^{ub},\sigma)$. Since the gradient of the binding constraints has full-rank, Theorem 1(v) in Wachsmuth (2013) implies that $\eta(\mu,X,\beta^{ub},\sigma)$ has a unique Lagrangian, i.e. a unique dual solution. □

It is worth noting that uniqueness of $\max_{\gamma\in V(X_\delta,\sigma)}\gamma'(\mu-X_\beta\beta^{ub})$ can imply restrictions on the possible values of μ — for example, if $X_\delta=0$ and $X_\beta=\sigma=\iota$, then it implies that μ has a unique maximal element. By comparison, Assumption 4 implies that with probability approaching 1, the *sample* dual problem (i.e., the dual to $\eta(Y_{n,0},X_{n,0},\beta_0,\hat{\sigma}_{n,0})$) has a unique solution. When $X_\delta=0$ and $X_\beta=\sigma=\iota$, this is satisfied if Σ is full-rank, regardless of the value of μ. More generally, as shown in Section D, for a wide variety of settings Assumption 4 can be guaranteed to holds under restrictions on $X_{n,0}$ and Σ only, without imposing restrictions on μ.

G Simulation Details

G.1 Moment Inequality Specification

We adopt the notation of Example 3 in the main text, so $J_{f,i,t}$ is the set of products marketed by firm f in market i in period t, and $\Delta\pi(J_{f,i,t},J'_{f,i,t})$ is the difference in expected profits from marketing $J_{f,i,t}$ rather then $J'_{f,i,t}$. Following Wollmann (2018), and as discussed in the main text, the fixed cost to firm f of marketing product j at time t is $\beta(\delta_{c,f}+\delta_g g_j)$ if the product was marketed last year ($j\in J_{f,i,t-1}$), and $\delta_{c,f}+\delta_g g_j$ otherwise. Here $\delta_{c,f}$ is a per-product cost which is constant across products but may differ across firms, while g_j is the gross weight rating of product j.

If we begin with the case where fixed costs are constant across firms ($\delta_{c,f}=\delta_c$ for all f) and again let $1\{\cdot\}$ denote the indicator function, we obtain four conditional moment inequalities by adding and subtracting one product at a time from the set marketed. For instance, similar to the Example 3, if firm f markets product j at both $t-1$ and t, then for

$$m^1(\theta)_{j,f,i,t} \equiv -[\Delta\pi(J_{f,i,t},J_{f,i,t}\setminus j)-(\delta_c+\delta_g g_j)\beta]\times 1\{j\in J_{f,i,t},j\in J_{f,i,t-1}\},$$

we must have $E[m^1(\theta)_{j,f,i,t}|V_{f,i,t}]\leq 0$ for all variables $V_{f,i,t}$ in the firm's information set when time-t production decisions were made, since otherwise the firm would have chosen not to market product j in period t. We can analogously obtain moments $m^2(\theta)_{j,f,i,t},\ldots,m^4(\theta)_{j,f,i,t}$ corresponding with the cases where a firm markets product j only at period t, only at period $t-1$, or in neither period.

We obtain two further conditional moment inequalities by considering the case where a firm markets a product of a given weight g_j but not a higher or lower weight $g_{j'}$. For example, we obtain the moment

$$m^5_{j,f,i,t}(\theta)\equiv$$
$$-\left(\frac{\sum_{j'\in J^-(j,f,i,t)}[\Delta\pi(J_{f,i,t},(J_{f,i,t}\setminus j)\cup j')-\delta_g(g_j-g_{j'})]}{\#J^-(j,f,i,t)}\right)\times 1\{j\in J_{f,i,t},j\notin J_{f,i,t-1}\},$$

where $J^-(j,f,i,t)$ is the set of products not marketed by firm f at time t or $t-1$ with weight below g_j. We likewise construct a moment for heavier products that were not marketed.

As in Wollmann, there are nine firms ($F=9$). To generate data we model the expected and observed profits for firm f from marketing product j in market i in period t, denoted by $\pi^*_{j,f,i,t}$ and $\pi_{j,f,i,t}$ respectively, as

$$\pi^*_{j,f,i,t}=\eta_{j,i,t}+\epsilon_{j,f,i,t},\text{ and }\pi_{j,f,i,t}=\pi^*_{j,f,i,t}+\nu_{j,i,t}+\nu_{j,f,i,t},$$

where the ν terms are mean zero disturbances that arise from expectational and measurement error and the η and ϵ terms represent product-, market-, and firm-specific profit shifters known to the firm when marketing decisions are made. The distributions of these errors are calibrated to match moments in Wollmann's data, as described in the next section.[39]

As described below, each simulated dataset is a cross-section containing data on one period for 500 markets following the sequential process described above. The moments

[39] The terms $\eta_{j,i,t}$ and $\nu_{j,i,t}$ reflect product/market/time "shocks" that are known and unknown to the firms, respectively, when they make their decisions. Shocks of this sort are an important aspect of Wollmann's setting. Note that Wollmann also estimates (point-identified) demand and variable cost parameters in a first step, while for simplicity we treat the variable profits $\pi_{j,f,i,t}$ as known to the econometrician.

used in our simulations are then averages (over markets i) of

$$\frac{1}{J}\sum_{j}\left(m^{l}_{j,f,i}(\theta)\otimes\tilde{Z}_{j,f,i}\right)', \tag{23}$$

where we also average over all firms f assumed to share the same fixed cost $\delta_{f,c}$. Since we consider a single period for each market i in cross-section, we suppress the time subscript. We present results both for the case where $\tilde{Z}_{j,f,i}$ includes only a constant, and for the case where all moments are interacted with a constant and the first four moments are additionally interacted with the common profit-shifters η,

$$\tilde{Z}_{j,f,i}=(1,\eta^{+}_{j,i},\eta^{-}_{j,i}),$$

for $q^{+}=\max\{q,0\}$ and $q^{-}=-\min\{q,0\}$. In the model with a single constant term, $\delta_{c,f}=\delta_c$ for all f, this generates 6 and 14 moment inequalities. We also present results when the nine firms are divided into three groups each with a separate constant term, and when each firm has a separate constant term. For each specification we consider the first four moments separately for the firm(s) associated with distinct parameters $\delta_{c,f}$, but average the last two moments across all firms as they do not depend on the constant terms. This generates 14 and 38 moments for the three group classification, and 38 and 110 moments when each firm has a separate constant term. To estimate the conditional variance $\Sigma=\Omega$, in each specification we define the value of the instrument Z_i in market i as the Jacobian of (23) with respect to the linear parameters $(\delta_g,\{\delta_{c,f}\})$.

G.2 Data-generating Process Details

G.2.1 Competition and Firm Decisions

We now describe the data-generating process for a single market, suppressing the i subscript for notational brevity. We consider competition between F firms, who in each period decide which set of products to offer. Firm f estimates that marketing product j in period t will earn variable profits π^{*}_{jft}, and chooses to market the product if and only if the expected profits exceed the fixed costs. Thus, if a firm marketed product j in period $t-1$, then the firm chooses to market j in period t if and only if

$$\pi^{*}_{jft}-\beta\theta_c-\beta\theta_g g_j>0.$$

If the firm did not market the product j in period $t-1$, then it chooses to add product j if and only if

$$\pi^*_{jft} - \theta_c - \theta_g g_j > 0.$$

G.2.2 Distributional Assumptions

We set $\pi^*_{jft} = \eta_{jt} + \epsilon_{jft}$, the sum of a product-level shock that is common to all firms and a firm-product idiosyncratic shock. We assume that $\eta_{jt} \sim \mathcal{N}(0, \sigma^2_\eta)$. If j was not marketed in the previous period, then $\epsilon_{jft} \sim \mathcal{N}(\beta\mu_f + \beta\theta_g g_j, \sigma^2_\epsilon)$; if the product was marketed previously, then $\epsilon_{jft} \sim \mathcal{N}(\mu_f + \theta_g g_j, \sigma^2_\epsilon)$. Note that the mean profitability of marketing a product depends on a firm-specific mean, μ_f, which allows us to match the firm-level market shares observed in Wollmann's data. We also construct the mean of the ϵ_{jft} term to depend on the product's weight and whether it was marketed in the previous period in a way that guarantees that all simulated products will be offered with the same probability in our simulations.

While firms make their decisions using π^*_{jft}, we assume that the econometrician observes only $\pi_{jft} = \pi^*_{jft} + \nu_{jt} + \nu_{jft}$. The ν terms represent measurement or expectational errors. We assume that ν_{jt} and ν_{jft} are independently drawn from a normal distribution with mean 0 and variance σ^2_ν.

G.3 Calibration

We calibrate our parameters to estimates and moments reported in the November 2014 version of Wollmann. We set $F = 9$ to match the number of firms in Wollmann's data, and $G = 22$ to match the number of unique values of GWR. We use $\theta_c = 129.73$, $\theta_g = -21.38$, and $\beta = 0.386$ to match the results from the estimates in Table VII in Wollmann.[40] We set the values of g to be 22 evenly spaced points between 12,700 and 54,277 to match the lowest and highest GWR figures reported in Table II, which gives the average GWR for different buyer types.

To calibrate the remaining parameters, we simulate data according to the process described above, and set the parameters to match moments of the simulated data to those in Wollmann's data. In order to simulate the data for the calibration, we first fix standard normal draws that are used to construct the η, ϵ, and ν shocks. These standard normals draws are then scaled by the desired variance parameters in each simulation. Letting J_{ft} denote the set of products offered by firm f in period t, the simulations begin in state 0 with $J_{f0} = \emptyset$ for all firms. We then simulate J_{ft} and π^* going forward using the dynamics

[40] Note that Wollmann denotes by $-\frac{1}{\lambda}$ what we have been calling β.

described above. We discard the first 1,000 periods as burnout so as to obtain draws from the stationary distribution, and calibrate the model using 27,000 subsequent periods. After discarding 1,000 draws, we obtain essentially identical results if we begin from the state where all products are in the market in rather than all products out of the market.

The remaining parameter values to calibrate are $\{\mu_f\}$,σ_η,σ_ϵ,σ_ν. The intuition for the calibration is as follows. The firm-specific means μ_f affect the number of products each firm offers, and so we calibrate these to match the market shares and total number of products offered in Wollmann's data. The σ_ϵ and σ_η terms affect how often firms add and remove products, and so we calibrate these to match the variability of the number of products offered over time in Wollmann's data. Lastly, we calibrate σ_ν, which governs the variance of the expectational/measurement error. We do not have direct measures of the variability of firm profits in Wollmann's data, but if markups are constant, then the variance in firm profits is one-to-one with the variance of quantity sold, and so we calibrate σ_ν to match the variability of quantities sold assuming mark-ups are fixed at 35%.

Specifically, the calibration uses the following steps:

1) We first calibrate (σ_η,σ_ϵ) and the μ_f terms to match the market shares and variability of products offered in Wollmann. This calibration process involves an inner and outer loop, described below.

a) The inner loop for μ_f. Given a guess for (σ_η,σ_ϵ), we calibrate μ_f to match the market share and average number of products in Wollmann's data. Market shares are taken from Table III in Wollmann. Wollmann does not provide the mean number of products offered by year, only the min and max, so we approximate it by taking the midpoint between the two extremes, which gives 48 total products per year on average.

b) In the outer loop, we calibrate (σ_η,σ_ϵ) to match a measure of the variability of the number of products offered in Wollmann's data. In particular, Table I in Wollmann lists 9-year averages for the total number of products offered for three 9-year periods (he has 27 years of data). We run 1,000 simulations of 27 periods, and for each 27-year period we calculate the average number of products offered within each 9-year subinterval, just as Wollmann does. We then calibrate σ_η so that the average variance in the number of products offered across three consecutive 9 year periods matches that in Wollmann's data.

The simulated variance comes very close to the target variance whenever $\sigma_\eta = \sigma_\epsilon$, regardless of scaling. We therefore choose $\sigma_\eta = \sigma_\epsilon = 30$, which gives that the variance of π^* is roughly half of the variance of π.

2) Lastly, we calibrate σ_ν to match a moment implied by the variability in quantity

sold across time in Wollmann. If prices and markups are relatively constant, then the variance in quantities will be well-approximated by a constant times the variance in profits: $Var(\pi_{jft}) \approx \bar{p}^2 \bar{m}^2 Var(Q_{jft})$, where $\bar{p}$ and $\bar{m}$ are the average prices and markups.[41] For our calibration, we set $\bar{p}$ to be the average price in Wollmann's data ($66,722), and set $\bar{m}$ equal to 0.35. As with the number of products offered, Wollmann does not report annual quantities, but rather the average for three 9-year periods. We thus use a procedure analogous to that described in step 1b) to match the variance of the 9-year averages of quantity sold.

G.3.1 Calibrated Parameters

Tables G.1 and G.2 show the calibrated values for the μ_f and variance parameters, respectively.

Table G.1: Calibrated μ_f Parameters

Firm	μ_f
Chrysler	74.31
Ford	98.36
Daimler	114.69
GM	80.11
Hino	67.71
International	110.63
Isuzu	80.15
Paccar	114.63
Volvo	94.17

G.3.2 Sampling from the DGP

Wollmann's data involves observations of sequential periods from the same market. If we were to construct moments at the product-period level in this setting, then the sequential nature of the model would induce serial correlation in the realizations of the moments.

[41] This is because if prices and costs are constant across firms,

$$\begin{aligned}\pi_{jft} &= Q_{jft}(p-c)\\ &= Q_{jft}\frac{p-c}{p}p\\ &= Q_{jft}\times m\times p.\end{aligned}$$

Thus, $Var(\pi_{jft}) = m^2 p^2 Var(Q_{jft})$ when p and c are constant, and this holds approximately with averages if the variance in m and p is small relative to that in Q.

Table G.2: Calibrated Variance Parameters

Parameter	Value
σ_η	30.00
σ_ϵ	30.00
σ_ν	57.96

Although Σ can be estimated in this setting, accounting for serial correlation substantially complicates covariance estimation. Since covariance estimation is not the focus of this paper, and Wollmann (2018) performs inference assuming no serial correlation, we instead focus on a modified DGP corresponding to a cross-section of independent markets, a common setting in the industrial organization literature. To do this, we sample from the stationary distribution of the calibrated DGP described above as follows. We draw a 51,000 period sequential chain, and discard the first 1,000 observations as a burn-in period. For each simulated dataset, we then randomly subsample 500 periods from this chain. This cross-sectional set-up also allows us to consider specifications with more moments than in Wollmann.

G.4 Implementation Details

G.4.1 Parameter Grids

For procedures that require test inversion for the parameter of interest, we invert tests over a discretized parameter space.[42] For δ_g and the cost of the mean-weight truck, we use 1,001 gridpoints (plus estimates of the identified set bounds); for β, we use 100 gridpoints for our main simulations, and 1,000 gridpoints for timing comparisons.

G.4.2 Implementation of LF and LFP tests

To calculate the LFP critical values, we draw a fixed matrix Ξ of standard normal draws of size $k \times 10{,}000$, and we use these for all of our calculations. Since the LF procedure is more computationally intensive, we calculate it using a matrix of size $k \times 1000$.

In simulating the draws for the LF approach, in certain very rare cases we encountered computational issues in which the linear program for one of the draws did not converge. In these cases, we treat the draw as if it were infinity, which pushes the estimated critical value slightly higher. However, in all specifications this happens in no more than 0.01% of

[42]For the LF and LFP approaches, we do not need to discretize the parameter space when the parameter of interest enters the moments linearly, since the endpoints of the confidence set can be calculated analytically using linear programming, as discussed in Section 5.

cases (of approximately 50 million simulations), and is thus unlikely to have any substantial impact on our results.

G.4.3 Implementation of the sCC and sRCC tests

We implement the sCC and sRCC tests using code provided by the authors. The refinement needed for the sRCC test is difficult to compute with many moments and many parameters. Thus, when our specification has both 100+ moments and 10+ parameters, we instead report the results of a test that rejects whenever the sRCC test rejects. In particular, the refinement to the sRCC test can matter only when there is one active moment ($\hat{r}=1$) and the test statistic falls between the $1-\alpha$ and $1-\alpha/2$ quantile of the χ^2 distribution with 1 degree of freedom. For specifications with 100+ moments and 10+ parameters, we thus report the power of the test that rejects when either the sCC test rejects or the refinement could matter. The power and size of this test can thus be viewed as upper bounds on the power and size of the sRCC test, and its runtime is a lower bound on the runtime of the sRCC test.

G.4.4 Implementation of the AS and KMS tests

We next describe the implementation of the AS and KMS tests, which uses the Matlab package developed by Kaido et al. (2017). The Matlab package is developed for the case where the moments are additively separable in the data and the parameters, i.e. when the moments take the form $E[m(D_i)]-g(\theta)\leq 0$, where θ is a vector of parameters and the target parameter takes the form $l'\theta$. Note that in our first two simulation designs, where the target parameter is δ_g or the cost of the mean-weight truck (and β is known), the moments take the form $E[Y_i|X_i]-X_i\delta\leq 0$ and the target parameter is $l'\delta$. The moments thus take the form needed to use the Matlab package *conditional* on X_i. The Matlab package, however, uses a bootstrap procedure that samples from the unconditional distribution of the data, which is unsuitable for our setting. To use the package in our setting with conditional moments, we adopt the following procedure. Given $Y_{n,0},X_{n,0},\widehat{\Sigma}_{n,0}$, we draw $Y_i^*\sim N(n^{-\frac{1}{2}}Y_{n,0},\widehat{\Sigma}_{n,0})$ independently for $i=1,...,n$.[43] We then provide the Matlab package with the data $(Y_i^*)_{i=1}^n$ and set $m(Y_i^*)=Y_i^*$ and $g(\theta)=X_{n,0}\theta$. This ensures that the bootstrap distribution of the sample mean of Y_i^* (scaled by $\sqrt{n}$) within the Matlab package approximates the conditional distribution of $Y_{n,0}|X_{n,0}$.

We use the default tolerances in the Matlab package except we halve the default tolerance for the objective (i.e., we set EAM_obj_tol and EAM_thetadistort to 0.005/2). Tightening the objective tolerance appears to reduce numerical precision errors that can, for

[43] We re-center and re-scale the draws so that the sample mean of Y_i^* is exactly $n^{-\frac{1}{2}}Y_{n,0}$ and the sample covariance is $\widehat{\Sigma}_{n,0}$.

instance, lead the estimated bounds for the AS test to be tigher than for the KMS test. On the other hand, the tighter tolerances increase runtime and lead to some convergence issues. In the specification with the most moments and parameters, the KMS test fails to converge correctly in 6% of the cases with the tigher tolerances. We discard all such draws and report size and excess length conditional on the algorithm converging correctly. We obtain qualitatively similar results using the default tolerances, which have fewer convergence issues but are less numerically precise.

G.5 Additional Simulation Results

This appendix reports additional simulation results to complement the results reported in Section 6 of the main text. Figures G.1-G.2 show comparisons analogous to Figure 1 except for the alternative parameters δ_g and β. Figures G.3-G.5 show comparisons of the hybrid to the LFP, sCC, and sRCC tests, while Figures G.6-G.7 show comparisons to the AS and KMS tests.

Figure G.1: Rejection probabilities for 5% tests of θ_g

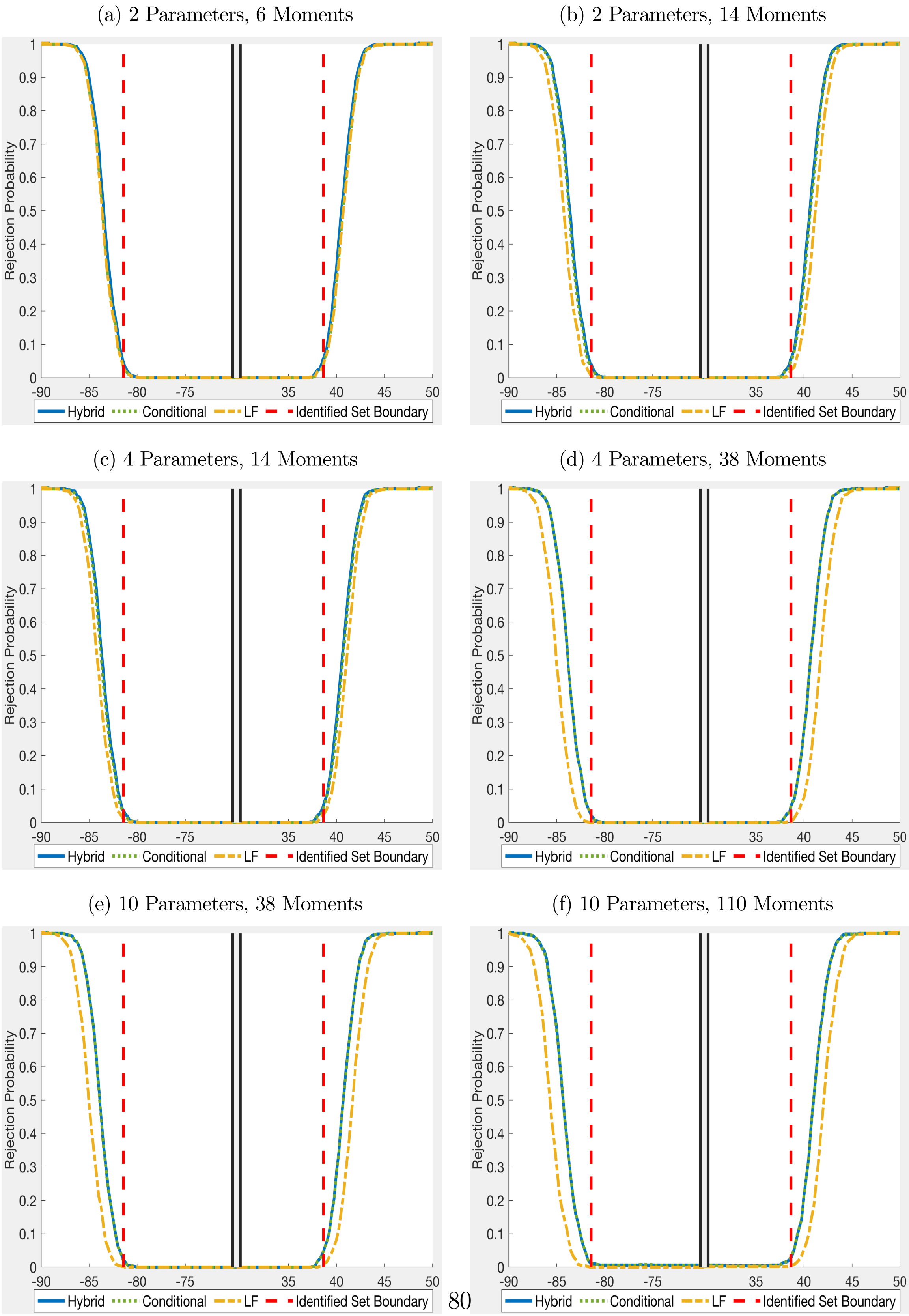


Figure G.2: Rejection probabilities for 5% tests of β

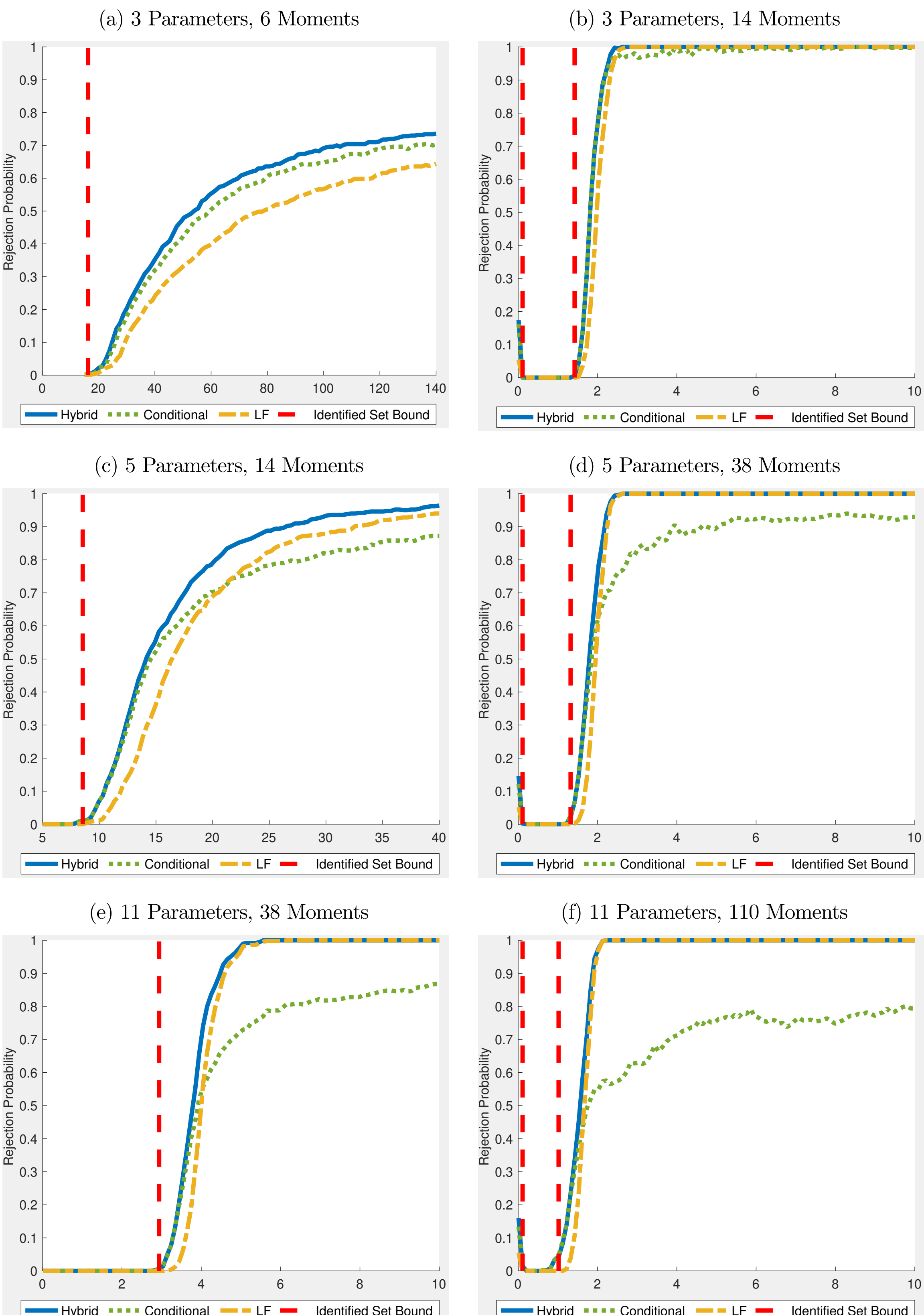
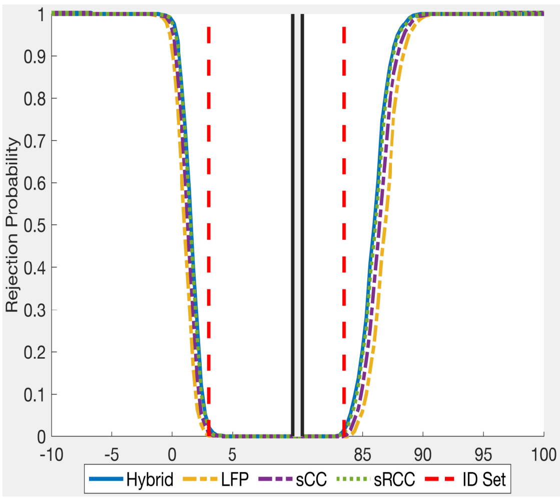
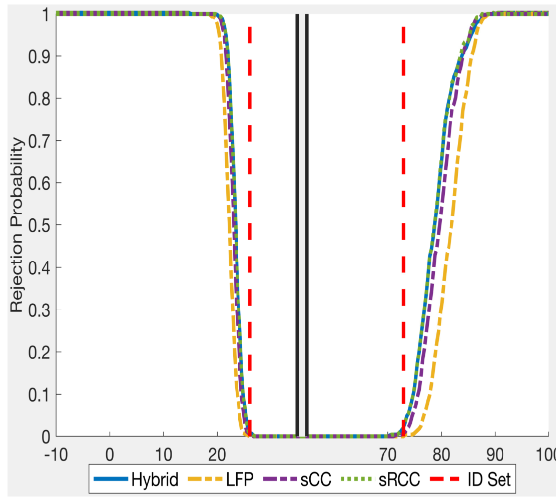


Figure G.3: Rejection Probabilities for 5% tests of Cost of Mean-Weight Truck: Comparisons to Cox & Shi (2022) and LFP tests

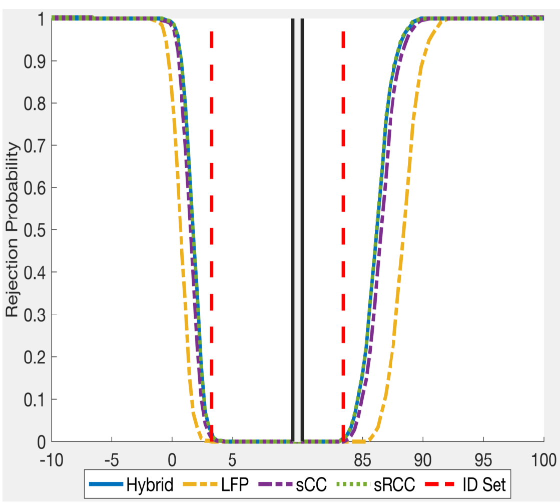
(a) 2 Parameters, 6 Moments



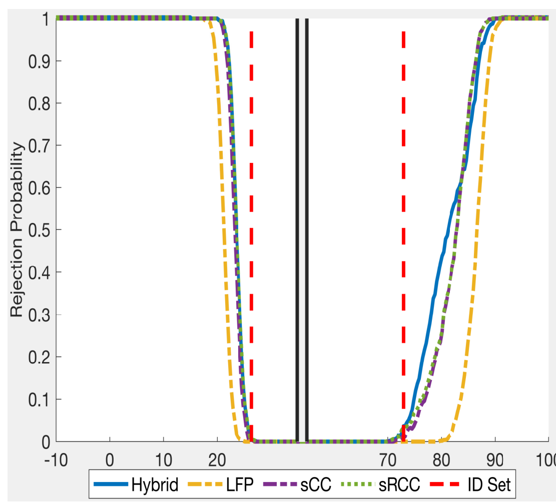
(b) 2 Parameters, 14 Moments



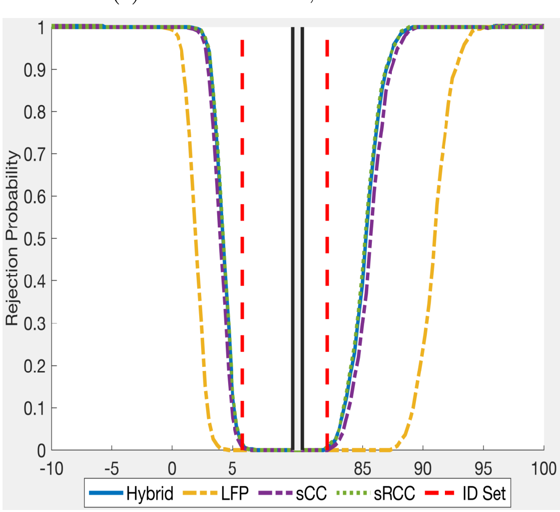
(c) 4 Parameters, 14 Moments



(d) 4 Parameters, 38 Moments



(e) 10 Parameters, 38 Moments



(f) 10 Parameters, 110 Moments

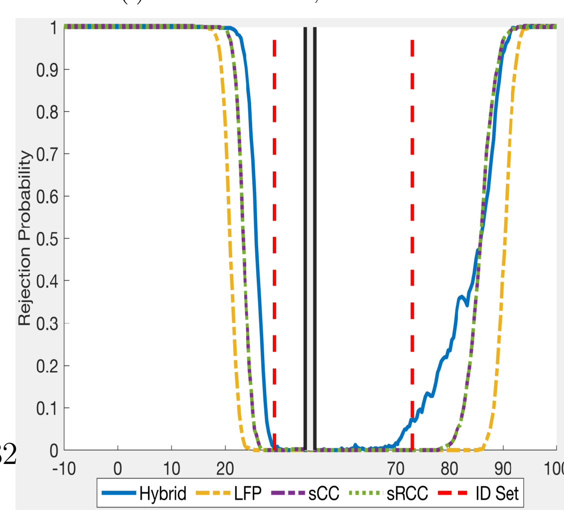


Figure G.4: Rejection Probabilities for 5% tests of θ_g: Comparisons to Cox & Shi (2022) and LFP tests

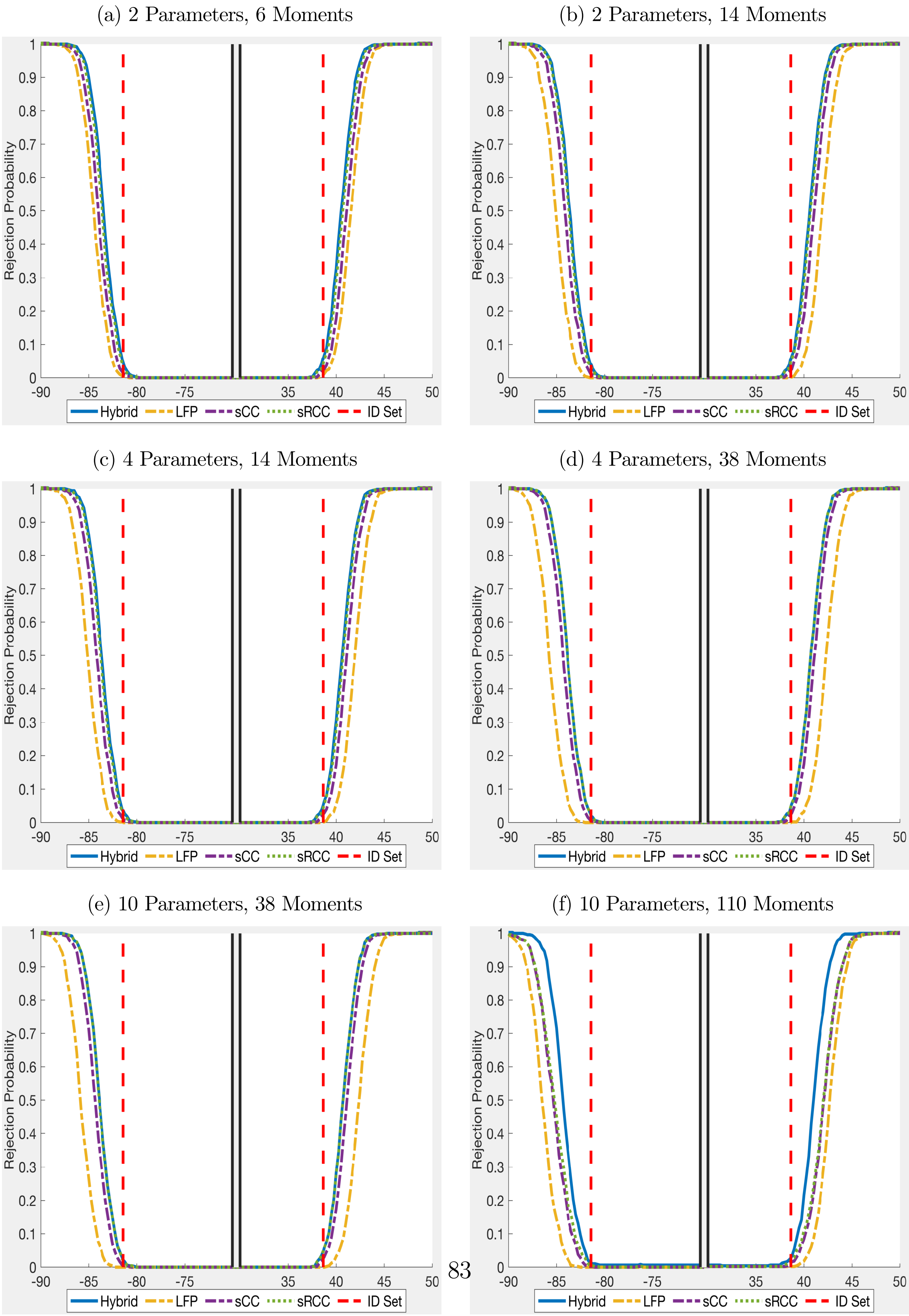
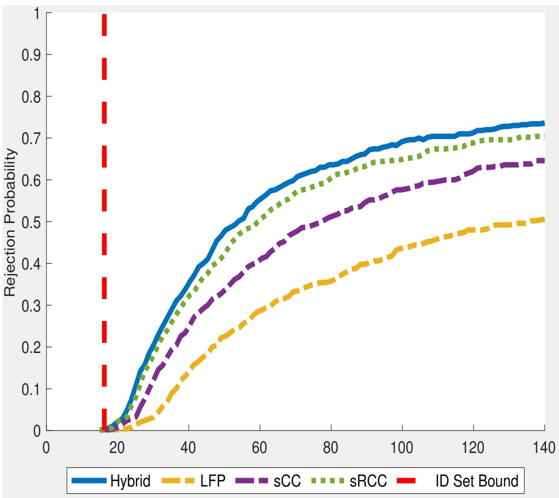
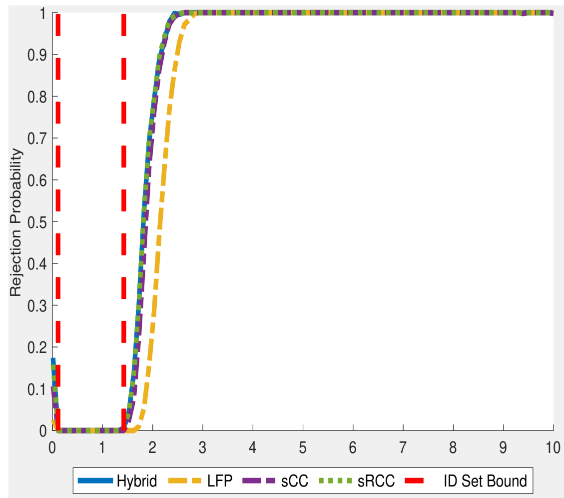


Figure G.5: Rejection Probabilities for 5% tests of β: Comparisons to Cox & Shi (2022) and LFP tests

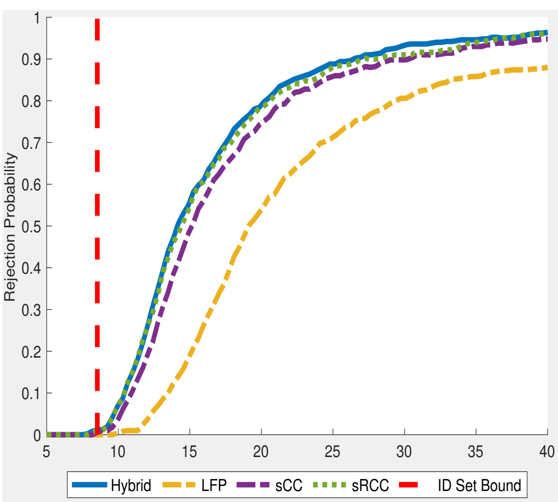
(a) 2 Parameters, 6 Moments



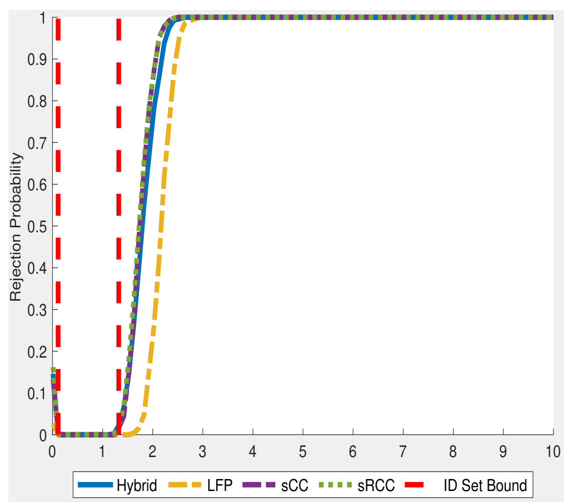
(b) 2 Parameters, 14 Moments



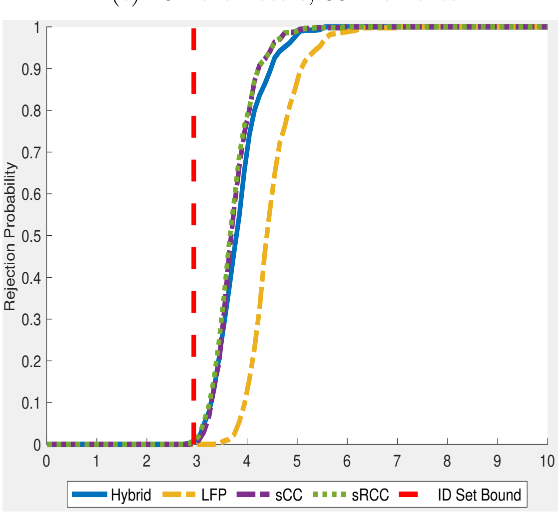
(c) 4 Parameters, 14 Moments



(d) 4 Parameters, 38 Moments



(e) 10 Parameters, 38 Moments



(f) 10 Parameters, 110 Moments

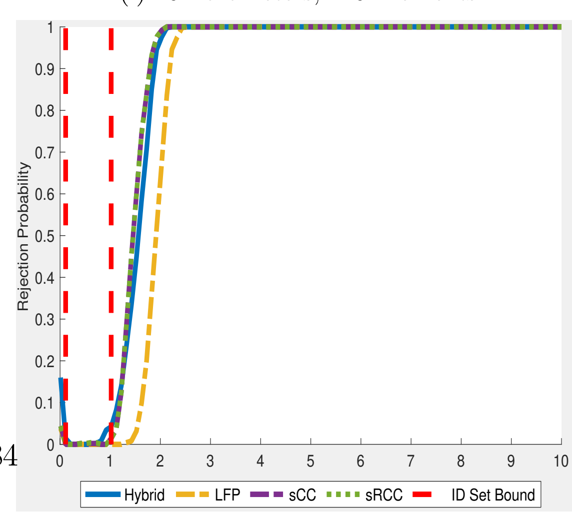


Figure G.6: Rejection Probabilities for 5% tests of Cost of Mean-Weight Truck: Comparisons to AS and KMS tests

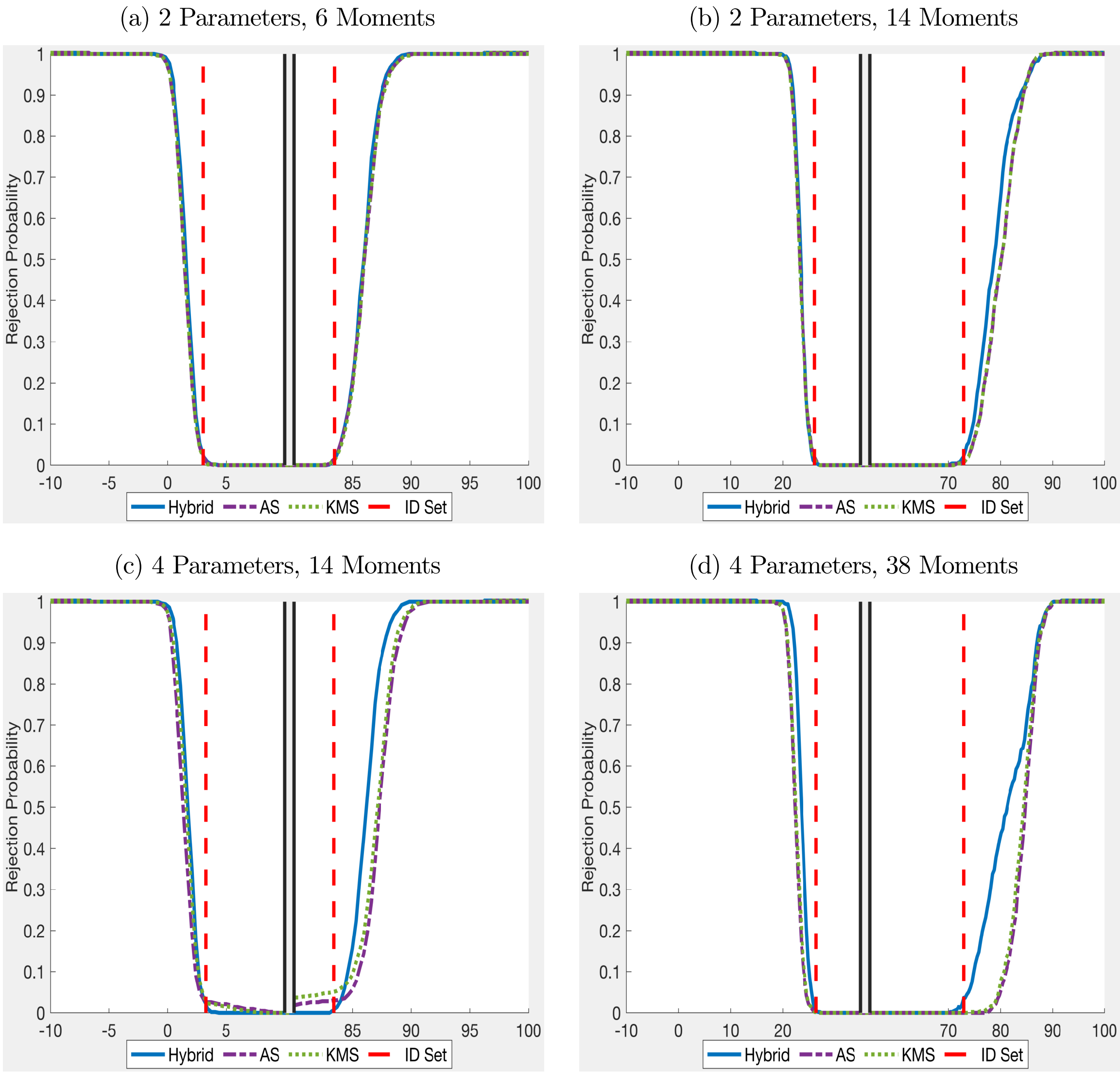
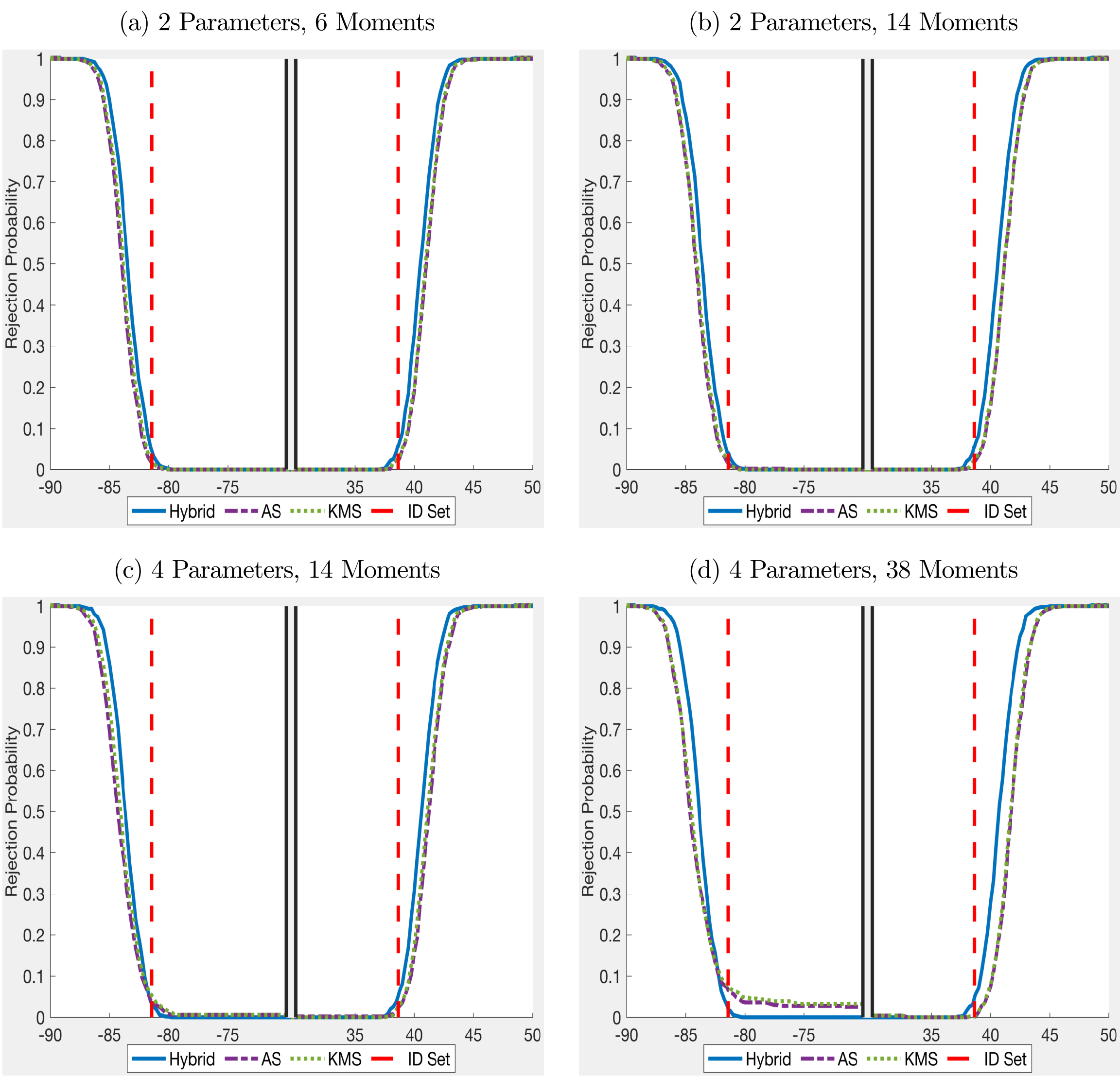


Figure G.7: Rejection Probabilities for 5% tests of θ_g: Comparisons to AS and KMS tests



Supplement References

Abadie, A. & Imbens, G. W. (2008), 'Estimation of the conditional variance in paired experiments', *Annales d'Économie et de Statistique* (91/92), 175–187.

Abadie, A., Imbens, G. W. & Zheng, F. (2014), 'Inference for misspecified models with fixed regressors', *Journal of the American Statistical Association* **109**(508), 1601–1614.

Andrews, D. W., Guggenberger, P. & Cheng, X. (2019), 'Generic results for establishing the asymptotic size of confidence sets and tests', *Journal of Econometrics* **Forthcoming**.

Chernozhukov, V., Chetverikov, D. & Kato, K. (2015), 'Comparison and anti-concentration bounds for maxima of gaussian random vectors', *Probability Theory and Related Fields* **162**(1-2), 47–70.

Fithian, W., Sun, D. L. & Taylor, J. E. (2017), Optimal inference after model selection. Working Paper.

Kaido, H., Molinari, F. & Stoye, J. (2019), Constraint qualifications in partial identification. Working Paper.

Tijssen, G. & Sierksma, G. (1998), 'Balinski-tucker simplex tableaus: dimensions, degeneracy degrees, and interior points of optimal faces', *Mathematical Programming* **81**, 349–372.

Van der Vaart, A. (2000), *Asymptotic Statistics*, Cambridge University Press.

Van der Vaart, A. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.