Supplement to the paper

Inference for Linear Conditional Moment Inequalities

Isaiah Andrews Jonathan Roth Ariel Pakes

September 17, 2019

This supplement contains proofs and additional results for the paper "Inference for Linear Conditional Moment Inequalities." Section A discusses results for an alternative formulation of the conditional approach based on the dual linear program, which allows the possibility of non-unique or degenerate solutions. Section B develops some additional results for the dual problem used in Section A. All proofs for the finite-sample normal model are collected in Section C. Section D states our asymptotic results, while proofs for these results are given in Section E. Section F provides simulation results for our tests in a simple example without nuisance parameters, while Section G provides additional details and results for the simulation designs discussed in Section 7 of the main text. Finally, Section H discusses a bisection algorithm for computing bounds used in the dual conditioning approach.

A Conditional Inference Based on the Dual

This section describes a conditioning approach based on a dual linear program which can be applied even in settings where the linear program (12) has a non-unique or degenerate solution, but which is equivalent to the primal conditioning approach described in the main text when the solution to (12) is unique and non-degenerate. To formally describe the dual approach, we first define the dual linear program.

Lemma 8 *When $\hat{\eta}$ as defined in Lemma 3 is finite, it is equal to*

$$\begin{array}{c}\max_\gamma \gamma' Y_n \\ \textit{subject to } \gamma \geq 0,\ W_n'\gamma = e_1.\end{array} \tag{21}$$

for W_n the matrix with row j equal to $W_{n,j} = \begin{pmatrix} \sqrt{\Sigma_{jj}} & X_{n,j} \end{pmatrix}$ and $e_1 = (1, 0, ..., 0)'$ the first standard basis vector.

The set of solutions to the dual linear program is

$$\widehat{\Gamma} = \left\{ \gamma : \hat{\eta} = \gamma' Y_n,\, \gamma \geq 0,\, W_n' \gamma = e_1 \right\}.$$

This set is defined by a collection of linear equalities and inequalities and so is a polytope. Our dual approach conditions on the set of vertices $\widehat{V}$ of $\widehat{\Gamma}$. Results in the next section show that this set of solution vertices has finite support, and that any pair of possible vertices γ_1, γ_2 arise together with probability either zero or one

$$Pr_{\mu_n} \left\{ \{\gamma_1, \gamma_2\} \subseteq \widehat{V} \right\} \in \{0, 1\}.$$

Thus, conditioning on a given value for the set of vertices, $\widehat{V} = V$, is equivalent to conditioning on $\gamma \in \widehat{V}$ for any $\gamma \in V$, up to sets of measure zero. We thus consider inference conditional on $\gamma \in \widehat{V}$. We further discuss the set of vertices $\widehat{V}$ and its properties in the next section.

As before, the distribution of $\hat{\eta}$ conditional on $\gamma \in \widehat{V}$ will in general depend on the full vector μ_n, rather than just on $\gamma' Y_n$. To eliminate dependence on μ_n other than through $\gamma' Y_n$ we again condition on a sufficient statistic for the rest of the vector μ_n, $S_{n,\gamma} = \left(I - \frac{\Sigma \gamma \gamma'}{\gamma' \Sigma \gamma} \right) Y_n$, which coincides with $S_{n,B}$ defined in the main text for $\gamma = \gamma_{n,B}$. We obtain the following conditional distribution for $\hat{\eta}$:

Lemma 9 *The conditional distribution of $\hat{\eta}$ given $\gamma \in \widehat{V}$ and $S_{n,\gamma} = s$ is truncated normal,*

$$\hat{\eta} | \left\{ S_{n,\gamma} = s \;\&\; \gamma \in \widehat{V} \right\} \sim \xi | \xi \in \left[\mathcal{V}^{lo}(s), \mathcal{V}^{up}(s) \right]$$

for $\xi \sim N(\gamma' \mu_n, \gamma' \Sigma \gamma)$,

$$\mathcal{V}^{lo}(s) = \min \left\{ c : \begin{array}{c} c = \max_{\tilde{\gamma}} \tilde{\gamma}' \left(s + \frac{\Sigma \gamma}{\gamma' \Sigma \gamma} c \right) \\ \text{subject to } \tilde{\gamma} \geq 0,\; W_n' \tilde{\gamma} = e_1 \end{array} \right\} \tag{22}$$

and

$$\mathcal{V}^{up}(s) = \max \left\{ c : \begin{array}{c} c = \max_{\tilde{\gamma}} \tilde{\gamma}' \left(s + \frac{\Sigma \gamma}{\gamma' \Sigma \gamma} c \right) \\ \text{subject to } \tilde{\gamma} \geq 0,\; W_n' \tilde{\gamma} = e_1 \end{array} \right\}, \tag{23}$$

provided s is such that the set on the right hand side of (22) is nonempty.

Thus, we see that the dual conditioning approach yields conditional distributions of the same form as those based on primal approach, up to the difference in the truncation points. Unlike our result in Lemma 7 for the primal case, however, Lemma 9 does not impose any conditions regarding uniqueness or nondegeneracy of the solution to either the dual or primal problems.

Our next result shows that when the solution γ to the dual satisfies additional conditions, the truncation points $\mathcal{V}^{lo}$ and $\mathcal{V}^{up}$ in Lemma 9 are the same as those obtained in the primal problem.

Proposition 5 *Suppose there exists $\gamma \in \widehat{V}$ with exactly $p+1$ strictly positive entries. Let B denote the set of rows for these entries, and suppose that B corresponds to linearly independent rows of W_n. Then there exists a solution to the primal problem (12) with the moments B binding, $\gamma = \gamma_{n,B}$ as defined in Lemma 6, and the definition of $\mathcal{V}^{lo}$ and $\mathcal{V}^{up}$ in equations (14) and (15) coincides with that in equations (22) and (23).*

The conditions on γ in this proposition are implied by existence of a unique, non-degenerate solution to the primal problem.

Lemma 10 *If there is a unique, non-degenerate solution $(\hat{\eta}, \hat{\delta}')'$ to the primal problem (12), any solution $\hat{\gamma} \in \widehat{\Gamma}$ to the dual problem satisfies the conditions of Proposition 5.*

This result suggests a straightforward way to proceed in practice. The widely-used dual-simplex algorithm for solving the primal problem (12) automatically generates a vertex $\hat{\gamma} \in \widehat{V}$ of the dual solution set as well. To determine how to calculate the truncation points $\mathcal{V}^{lo}$ and $\mathcal{V}^{up}$, we can thus simply check whether the conditions of Proposition 5 hold at this solution. If they do we can calculate $\mathcal{V}^{lo}$ and $\mathcal{V}^{up}$ using the closed-form expressions given in Lemma 6, while otherwise we can use (22) and (23).[34]

Going forward we consider the conditional test

$$\phi_C = 1\left\{\hat{\eta} > c_{\alpha,C}\left(\hat{\gamma}, \mathcal{V}^{lo}(S_{n,\hat{\gamma}}), \mathcal{V}^{up}(S_{n,\hat{\gamma}}), \Sigma\right)\right\}.$$

If the solution to (12) is unique and non-degenerate this test coincides with (17).

[34]In particular, one can show that the set on the right hand side of (22) is convex, so we can quickly find lower and upper bounds using e.g. the bisection method (see Section H). See Section 6 in the main text for further discussion of implementation.

Conditional and Unconditional Size Control Now that we have formulated the conditional test in the general case, we can establish conditional and unconditional size control.

Proposition 6 *Under Assumption 1, the conditional test ϕ_C has size α both conditional on $\gamma \in \widehat{V}$,*

$$\sup_{\mu_n \in \mathcal{M}_0} E_{\mu_n}\left[\phi_C | \gamma \in \widehat{V}\right] = E_0\left[\phi_C | \gamma \in \widehat{V}\right] = \alpha$$

for all γ such that $Pr_{\mu_n}\left\{\gamma \in \widehat{V}\right\} > 0$, and unconditionally,

$$\sup_{\mu_n \in \mathcal{M}_0} E_{\mu_n}\left[\phi_C\right] = E_0\left[\phi_C\right] = \alpha.$$

Size Control for Hybrid Tests We can likewise show that the hybrid test based on the dual formulation controls size. As before, hybrid tests reject when $\hat{\eta} > c_{\kappa,LF}(X_n, \Sigma)$, and otherwise modify the upper bound to

$$\mathcal{V}^{up,H}(s) = \min\left\{\mathcal{V}^{up}(s), c_{\kappa,LF}(X_n, \Sigma)\right\},$$

yielding the test

$$\phi_H = \left\{\hat{\eta} > c_{\frac{\alpha-\kappa}{1-\kappa},C}\left(\hat{\gamma}, \mathcal{V}^{lo}(S_{n,\hat{\gamma}}), \mathcal{V}^{up,H}(S_{n,\hat{\gamma}}), \Sigma\right)\right\}.$$

Proposition 7 *Under Assumption 1, the hybrid test ϕ_H has size $\frac{\alpha-\kappa}{1-\kappa}$ conditional on $\hat{\eta} \leq c_{\kappa,LF}(X_n, \Sigma)$ and $\gamma \in \widehat{V}$ for all γ such that $Pr_{\mu_n}\left\{\hat{\eta} \leq c_{\kappa,LF}(X_n, \Sigma), \gamma \in \widehat{V}\right\} > 0$,*

$$\sup_{\mu_n \in \mathcal{M}_0} E_{\mu_n}\left[\phi_H | \hat{\eta} \leq c_{\kappa,LF}(X_n, \Sigma), \gamma \in \widehat{V}\right] = E_0\left[\phi_H | \hat{\eta} \leq c_{\kappa,LF}(X_n, \Sigma), \gamma \in \widehat{V}\right] = \frac{\alpha - \kappa}{1 - \kappa},$$

and has unconditional size α,

$$\sup_{\mu_n \in \mathcal{M}_0} E_{\mu_n}\left[\phi_H\right] = E_0\left[\phi_H\right] = \alpha.$$

B Properties of the Dual Solution Verticies $\widehat{V}$

In this section we further discuss the set of solution vertices $\widehat{V}$ used in the dual conditioning approach. As noted above, the set of solutions γ to the dual problem is the polytope

$$\widehat{\Gamma} = \{\gamma : \hat{\eta} = \gamma' Y_n,\, \gamma \geq 0,\, W_n' \gamma = e_1\}.$$

Letting $\widehat{V} = \widehat{V}(Y_n, W_n)$ denote the set of vertices of $\widehat{\Gamma}$, and

$$\widehat{C} = C(Y_n, W_n) = \{\gamma : \gamma' Y_n = 0, \gamma \geq 0, W_n' \gamma = 0\},$$

the characteristic cone of $\widehat{\Gamma}$, we can write $\widehat{\Gamma} = CH(\widehat{V}) + \widehat{C}$ for $CH(A)$ the convex hull of a set A, where we use $B + D$ to denote the Minkowski sum of sets B and D (see e.g. Chapter 8.2 of Schrijver (1986)). Let us further define the set of values γ satisfying the constraints in (21) (often called the feasible set) as

$$F = \{\gamma : \gamma \geq 0, W_n' \gamma = e_1\}.$$

The set F is again a polytope. Let V_F denote the vertices of F, often called the basic feasible solutions to the linear program (21). Any vertex of $\widehat{\Gamma}$ must also be a vertex of F (see e.g. Chapter 8.3 of Schrijver (1986)), so $\widehat{V} \subseteq V_F$. We can view $\widehat{V}$ as a random variable with support contained in the (finite) power set of V_F.

Lemma 10 and Proposition 5 above show that when the primal problem has a unique and non-degenerate solution, conditioning on the set of vertices $\widehat{V}$ is equivalent to conditioning on the set of binding moments in the primal problem. In more general cases, however, conditioning on $\widehat{V}$ rather than the set of binding moments resolves a number of difficulties. Specifically, when there are multiple solutions to the primal problem, approaches that condition on the set of binding moments face the question of which set(s) of binding moments to use. By contrast, our results show that the presence of multiple solutions to the dual raises no difficulties when we condition on $\widehat{V}$. As another alternative, rather than conditioning on $\widehat{V}$, one might instead condition on the full solution set $\widehat{\Gamma}$ or, equivalently, on $\widehat{C}$ in addition to $\widehat{V}$. Such conditioning is unnecessary to obtain tractable tests, however, and would further reduce the variation in the data usable for inference. We thus do not pursue this possibility.

The problem of conditioning on $\widehat{V}$ is greatly simplified by the fact that the support of $\widehat{V}$ is finite and disjoint.

Lemma 11 *There is a finite collection of sets* $\mathcal{V} = \{V_1, V_2, ..., V_m\}$ *, with* $V_j \subseteq V_F$ *for all* j*, such that* $Pr_{\mu_n}\left\{\widehat{V} \in \mathcal{V}\right\} = 1$*,* $Pr_{\mu_n}\left\{\widehat{V} = V_j\right\} > 0$ *for all* j*, and* $V_j \cap V_k = \emptyset$ *for all* $j \neq k$.

This result simplifies the problem of conditioning on $\widehat{V}$, since for any $\gamma \in V_j \in \mathcal{V}$ the event $\gamma \in \widehat{V}$ is equivalent to the event $\widehat{V} = V_j$. Thus, in order for us to construct conditional tests it will be enough for us to find a single vertex $\hat{\gamma}$ of $\widehat{V}$, rather than fully characterizing $\widehat{V}$. The widely used dual-simplex method for solving linear programs finds such a vertex.

C Proofs for Finite-Sample Normal Model

Proof of Lemma 1 Follows immediately from the Lindeberg-Feller central limit theorem (see e.g. Proposition 2.27 in Van der Vaart (2000))). □

Proof of Lemma 2 Immediate from the central limit theorem for iid data (see e.g. Proposition 2.17 in Van der Vaart (2000)). □

Proof of Lemma 3 By the definition of the maximum, $S(Y_n - X_n\delta, \Sigma)$ is equal to the smallest value η satisfying

$$(Y_{n,j} - X_{n,j}\delta)/\sqrt{\Sigma_{jj}} \leq \eta \ \ \forall j.$$

The result of the lemma follows immediately. □

Proof of Proposition 1 To prove this result, we note first that $\min_\delta S(Y_n - X_n\delta, \Sigma)$ is invariant to shifts of Y_n by $X_n\tilde{\delta}$, in the sense that

$$\min_\delta S(Y_n - X_n\delta, \Sigma) = \min_\delta S\left(Y_n + X_n\tilde{\delta} - X_n\delta, \Sigma\right) \text{ for all } \tilde{\delta}.$$

From this, we see immediately that $c_\alpha(\mu_n, X_n, \Sigma)$ is also invariant, in the sense that

$$c_\alpha(\mu_n, X_n, \Sigma) = c_\alpha\left(\mu_n + X_n\tilde{\delta}, X_n, \Sigma\right) \text{ for all } \tilde{\delta}. \tag{24}$$

Next, we note that $\min_\delta S(Y_n - X_n\delta, \Sigma)$ is elementwise nondecreasing in Y_n, and thus that $c_\alpha(\mu_n, X_n, \Sigma)$ is elementwise nondecreasing in μ_n.

To complete the proof, we first argue that

$$\{c_\alpha(\mu_n, X_n, \Sigma) : \mu_n \in \mathcal{M}_0\} = \{c_\alpha(\mu_n, X_n, \Sigma) : \mu_n \leq 0\}, \tag{25}$$

so the set of of critical values for μ_n consistent with the null is equal to the set of critical values consistent with $\mu_n \leq 0$. To see that this is the case, consider any $\mu_n \in \mathcal{M}_0$, and note that by the definition of $\mathcal{M}_0$ there exists $\delta(\mu_n)$ such that $\mu_n - X_n\delta(\mu_n) \leq 0$. By (24) above, however, this means that

$$c_\alpha(\mu_n, X_n, \Sigma) = c_\alpha(\mu_n - X_n\delta(\mu_n), X_n, \Sigma).$$

Since $\mu_n - X_n\delta(\mu_n) \leq 0$, and we can repeat this argument for all $\mu_n \in \mathcal{M}_0$, we see that

$$\{c_\alpha(\mu_n, X_n, \Sigma) : \mu_n \in \mathcal{M}_0\} \subseteq \{c_\alpha(\mu_n, X_n, \Sigma) : \mu_n \leq 0\}.$$

On the other hand, $\{\mu_n \leq 0\} \subseteq \mathcal{M}_0$, so (25) follows immediately. Finally, note that since we showed above that $c_\alpha(\mu_n, X_n, \Sigma)$ is elementwise nondecreasing in μ_n,

$$\sup_{\mu_n \leq 0} c_\alpha(\mu_n, X_n, \Sigma) = c_\alpha(0, X_n, \Sigma)$$

which completes the proof. □

Proof of Lemma 4 This result follows from Lemma 10 below. In particular, note that by Lemma 10 any solution γ to the dual linear program has exactly $p+1$ nonzero elements. By complementary slackness the corresponding constraints in the primal problem (12) must bind, and Lemma 10 implies that the corresponding rows of W_n have full rank. Further, no additional constraints can bind since this would imply degeneracy of the solution. □

Proof of Lemma 5 To prove this result, note that since (12) is a linear program, the Kuhn-Tucker conditions are necessary and sufficient for a solution. By arguments in the text, if there exists a solution with the moments B binding, then we can write the optimal values as $(\hat{\eta}, \hat{\delta}')' = W_{n,B}^{-1} Y_{n,B}$, which sets $Y_{n,B} - W_{n,B}(\hat{\eta}, \hat{\delta}')' = 0$. If the remaining inequalities fail to hold when evaluated at $(\hat{\eta}, \hat{\delta}')'$ then $(\hat{\eta}, \hat{\delta}')'$ is infeasible and so not a solution. If, on the other hand, the remaining inequalities hold when evaluated at $(\hat{\eta}, \hat{\delta}')'$, then if we take the corresponding Kuhn-Tucker multipliers to be zero while setting the multipliers on the binding moments equal to $M_B \gamma_{n,B} = W_{n,B}^{-1\prime} e_1$, one can verify that the Kuhn-Tucker conditions hold. □

Proof of Lemma 6 Follows immediately from Lemma 5 together with Lemma 5.1 of Lee et al. (2016). □

Proof of Lemma 7 Follows immediately from Lemma 9 together with Lemma 10 and Proposition 5. □

Proof of Proposition 2 Follows immediately from Lemma 10, together with Propositions 5 and 6. □

Proof of Proposition 3 We prove this result for the dual conditioning approach introduced in Section A. That these results also hold in the primal conditioning approach discussed in Section 5.2 when the solution to the linear program (12) is unique and non-degenerate is immediate from Lemma 10 and Proposition 5. Our assumptions imply that the set of feasible vertices V_F in the dual problem based on $(Y_{n,m}, W_n, \Sigma)$ is non-empty, and that the set of optimal vertices $\widehat{V}$ is likewise non-empty. Since the primal is feasible by construction, we further know that the dual is bounded. We begin by showing that $\widehat{V}$ converges to the set $\widehat{V}_{\mathcal{B}}$ of solution vertices in the dual problem based on $(Y_{n,m,\mathcal{B}}, W_{n,\mathcal{B}}, \Sigma_{\mathcal{B}})$. In particular, let

$$V_{F,\mathcal{B}}^{\mathcal{B}} = \left\{\gamma \in \mathbb{R}^k : \gamma_{\mathcal{B}} \in V_{F,\mathcal{B}},\, \gamma_j = 0\, \forall j \notin \mathcal{B}\right\} = \{\gamma \in V_F : \gamma_j = 0\, \forall j \notin \mathcal{B}\}$$

denote the set of vertices in V_F corresponding to vertices $V_{F,\mathcal{B}}$ of the feasible region in the problem restricted to the moments $\mathcal{B}$, and $\widehat{V}_{\mathcal{B}}^{\mathcal{B}} \subseteq V_{F,\mathcal{B}}^{\mathcal{B}}$ the analog for $\widehat{V}_{\mathcal{B}}$,

$$\widehat{V}_{\mathcal{B}}^{\mathcal{B}} = \left\{\gamma \in \mathbb{R}^k : \gamma_{\mathcal{B}} \in \widehat{V}_{\mathcal{B}},\, \gamma_j = 0\, \forall j \notin \mathcal{B}\right\}.$$

We will show that $Pr_{\mu_{n,m}}\left\{\widehat{V} = \widehat{V}_{\mathcal{B}}^{\mathcal{B}}\right\} \to 1$.

To establish this result, recall that the dual problem (restricted to $\gamma \in V_F$) is

$$\max_{\gamma \in V_F} \gamma' Y_{n,m}.$$

For any $\gamma \in V_F$ with $\gamma_j \neq 0$ for some $j \notin \mathcal{B}$, $\gamma' Y_{n,m} \to_p -\infty$ as $m \to \infty$. Our assumption that there exists $\gamma_{\mathcal{B}} \geq 0$ with $W'_{n,\mathcal{B}} \gamma_{\mathcal{B}} = e_1$ implies that there exists at least one $\tilde{\gamma} \in V_F$ such that $\tilde{\gamma}_j = 0$ for all $j \notin \mathcal{B}$. Thus, for any $\gamma \in V_F$ with $\gamma_j \neq 0$ for some $j \notin \mathcal{B}$, since $\tilde{\gamma}' Y_{n,m} = O_p(1)$ as $m \to \infty$,

$$Pr \left\{ \tilde{\gamma}' Y_{n,m} > \gamma' Y_{n,m} \right\} \to 1.$$

Thus, all $\gamma \in V_F$ with $\gamma_j > 0$ for some $j \notin \mathcal{B}$ yield a value of the objective smaller than that for $\tilde{\gamma}$ with probability tending to one. This implies that $Pr \left\{ \widehat{V} \subseteq V^{\mathcal{B}}_{F,\mathcal{B}} \right\} \to 1$. However, for any $\gamma \in \widehat{V}$ such that $\gamma \in V^{\mathcal{B}}_{F,\mathcal{B}}$, $\gamma \in \widehat{V}^{\mathcal{B}}_{\mathcal{B}}$ as well. Thus, we see that $Pr \left\{ \widehat{V} = \widehat{V}^{\mathcal{B}}_{\mathcal{B}} \right\} \to 1$, as we wanted to show.

For $\hat{\eta}$ the optimal value of η based on $(Y_{n,m}, W_n, \Sigma)$, and $\hat{\eta}^{\mathcal{B}}$ the optimal value based on $(Y_{n,m,\mathcal{B}}, W_{n,\mathcal{B}}, \Sigma_{\mathcal{B}})$, we see that $\widehat{V} = \widehat{V}^{\mathcal{B}}_{\mathcal{B}}$, implies $\hat{\eta} = \hat{\eta}^{\mathcal{B}}$. Thus, the argument above shows that $\hat{\eta} \to_p \hat{\eta}^{\mathcal{B}}$ as $m \to \infty$.

We next argue that the critical values $c_{\alpha,C}\left(\hat{\gamma}, \mathcal{V}^{lo}(S_{n,m,\hat{\gamma}}), \mathcal{V}^{up}(S_{n,m,\hat{\gamma}}), \Sigma\right)$ based on $(Y_{n,m}, W_n, \Sigma)$ converge to the critical values $c_{\alpha,C}\left(\hat{\gamma}_{\mathcal{B}}, \mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_{\mathcal{B}}}), \mathcal{V}^{up}(S_{n,\mathcal{B},\hat{\gamma}_{\mathcal{B}}}), \Sigma_{\mathcal{B}}\right)$ which limit attention to the moments $\mathcal{B}$. To do so, we will show that $\mathcal{V}^{lo}\left(S_{n,m,\hat{\gamma}}\right) \to_p \mathcal{V}^{lo}\left(S_{n,\mathcal{B},\hat{\gamma}_{\mathcal{B}}}\right)$, and likewise for $\mathcal{V}^{up}\left(S_{n,m,\hat{\gamma}}\right)$.

Recall, in particular, that

$$\mathcal{V}^{lo}(s) = \min \left\{ c : \begin{array}{c} c = \max_{\tilde{\gamma}} \tilde{\gamma}' \left(s + \frac{\Sigma\gamma}{\gamma'\Sigma\gamma} c \right) \\ \text{subject to } \tilde{\gamma} \geq 0,\, W'_n \tilde{\gamma} = e_1 \end{array} \right\}.$$

By the results above we know that $\hat{\gamma} \in \widehat{V}^{\mathcal{B}}_{\mathcal{B}}$ with probability approaching one. Note that for

$$S_{n,m,\hat{\gamma}} = \left(I - \frac{\Sigma\hat{\gamma}\hat{\gamma}'}{\hat{\gamma}'\Sigma\hat{\gamma}} \right) Y_{n,m},$$

the conditioning statistic based on $Y_{n,m}$, we have $\hat{\gamma}' Y_{n,m} = O_p(1)$, so $S_{n,m,\hat{\gamma},j} = O_p(1)$ for all $j \in \mathcal{B}$. By contrast $S_{n,m,j} \to -\infty$ for all $j \notin \mathcal{B}$.

Note, next, that by linearity of the problem we can restrict the optimization in the construction of $\mathcal{V}^{lo}$ to $\tilde{\gamma} \in V_F$, and so write

$$\mathcal{V}^{lo}(s) = \min \left\{ c : c = \max_{\tilde{\gamma} \in V_F} \tilde{\gamma}' \left(s + \frac{\Sigma\gamma}{\gamma'\Sigma\gamma} c \right) \right\}.$$

Using the divergence of $S_{n,m,\hat{\gamma}}$, for any $\tilde{\gamma} \in V_F$ such that $\tilde{\gamma}_j > 0$ for some $j \notin \mathcal{B}$ and any compact set C,

$$Pr\left\{\tilde{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)<\hat{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\ \forall c\in C\right\}\to 1.$$

From the finiteness of V_F, we thus see that for any compact set C

$$Pr\left\{\max_{\tilde{\gamma}\in V_F/V_{F,\mathcal{B}}^{\mathcal{B}}}\tilde{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)<\hat{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\ \forall c\in C\right\}\to 1. \tag{26}$$

Since $\hat{\gamma}\in V_F$, this implies

$$Pr\left\{\max_{\tilde{\gamma}\in V_F}\tilde{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)=\max_{\tilde{\gamma}\in V_{F,\mathcal{B}}^{\mathcal{B}}}\tilde{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\ \forall c\in C\right\}\to 1.$$

Note that by the definition of $\hat{\gamma}$,

$$\hat{\gamma}'Y_{n,m}=\max_{\tilde{\gamma}\in V_F}\tilde{\gamma}'Y_{n,m}.$$

Since $\hat{\gamma}\in V_F$, for any v we have

$$\hat{\gamma}'(Y_{n,m}+v)\le\max_{\tilde{\gamma}\in V_F}\tilde{\gamma}'(Y_{n,m}+v).$$

Note further that from the definition of $S_{m,n,\hat{\gamma}}$, $c=\hat{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)$ for any c and that $Y_{n,m}=S_{n,m}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}\hat{\gamma}'Y_{n,m}$. Setting $v=\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}(c-\hat{\gamma}'Y_{n,m})$, we then have

$$c=\hat{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\le\max_{\gamma\in V_F}\gamma'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\ \forall c.$$

Note, further, that for all c,

$$\max_{\gamma\in V_F}\gamma'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\ge\max_{\gamma\in V_{F,\mathcal{B}}^{\mathcal{B}}}\gamma'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right),$$

since the left hand side optimizes over a larger set. The fact that $Pr\left\{\widehat{V}\subseteq V_{F,\mathcal{B}}^{\mathcal{B}}\right\}\to 1$ implies that with probability approaching one

$$c=\hat{\gamma}'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\le\max_{\gamma\in V_{F,\mathcal{B}}^{\mathcal{B}}}\gamma'\left(S_{n,m,\hat{\gamma}}+\frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\ \forall c,$$

and thus that

$$\left\{c : c = \max_{\gamma\in V_F} \gamma' \left(S_{n,m} + \frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\right\} \subseteq \left\{c : c = \max_{\gamma\in V_{F,\mathcal{B}}^{\mathcal{B}}} \gamma' \left(S_{n,m} + \frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\right\}.$$

Hence, if $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$ is finite then with probability approaching one $\mathcal{V}^{lo}(S_{n,\hat{\gamma}})$ is finite as well.

Note that the distribution of $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$ does not depend on m. Further, the distribution of $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$ conditional on $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$ being finite is trivially tight. Hence, conditional on the event that $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$ is finite, our argument above for compact sets C implies that

$$Pr\left\{\mathcal{V}^{lo}(S_{n,m,\hat{\gamma}}) = \mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}}) | \mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}}) \text{ finite}\right\} \to 1.$$

On the other hand, when $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$ is infinite, we know that

$$c = \hat{\gamma}' \left(S_{n,m,\hat{\gamma}} + \frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right) = \max_{\gamma\in V_{F,\mathcal{B}}^{\mathcal{B}}} \gamma' \left(S_{n,m,\hat{\gamma}} + \frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)$$

for all c sufficiently small. Hence, (26) implies that when $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}}) = -\infty$, $\mathcal{V}^{lo}(S_{n,m,\hat{\gamma}}) \to_p -\infty$ as well.

We can apply the same argument for $\mathcal{V}^{up}\left(S_{n,m,\hat{\gamma}}\right)$. Note, however, that the conditional critical value is a continuous function of $\mathcal{V}^{lo}\left(S_{n,m\hat{\gamma}}\right)$ and $\mathcal{V}^{up}\left(S_{n,m,\hat{\gamma}}\right)$, including at $\mathcal{V}^{lo}\left(S_{n,m,\hat{\gamma}}\right) = -\infty$ and $\mathcal{V}^{up}\left(S_{n,m,\hat{\gamma}}\right) = \infty$. Thus, by the continuous mapping theorem, we see that

$$(\hat{\eta}, c_{\alpha,C}\left(\hat{\gamma}, \mathcal{V}^{lo}(S_{n,m,\hat{\gamma}}), \mathcal{V}^{up}(S_{n,m,\hat{\gamma}}), \Sigma\right))$$

converge in distribution to their analogs calculated based on the moments $\mathcal{B}$ alone,

$$(\hat{\eta}^{\mathcal{B}}, c_{\alpha,C}\left(\hat{\gamma}_\mathcal{B}, \mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}}), \mathcal{V}^{up}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}}), \Sigma_\mathcal{B}\right)).$$

Assumption 1 implies that the variance of $\gamma' Y_n$ is strictly positive for all $\gamma_\mathcal{B} \in V_{F,\mathcal{B}}$. Hence, $\gamma_\mathcal{B}' Y_{n,\mathcal{B}}$ is continuously distributed and independent of $\mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$, $\mathcal{V}^{up}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}})$, from which it follows that $\mathcal{V}^{lo}(S_{n,\mathcal{B},\gamma_\mathcal{B}}) < \mathcal{V}^{up}(S_{n,\mathcal{B},\gamma_\mathcal{B}})$ with probability one. Hence, since $V_{F,\mathcal{B}}$ is finite,

$$\hat{\eta}^{\mathcal{B}} - c_{\alpha,C}\left(\hat{\gamma}_\mathcal{B}, \mathcal{V}^{lo}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}}), \mathcal{V}^{up}(S_{n,\mathcal{B},\hat{\gamma}_\mathcal{B}}), \Sigma_\mathcal{B}\right),$$

is continuously distributed, and the result follows from the continuous mapping theorem. □

Proof of Proposition 4 Follows immediately from Lemma 10 and Propositions 5 and 7.

Proof of Lemma 8 This result follows from standard duality results for linear programming. Note, in particular, that the primal problem (10) is equivalent to

$$-\hat{\eta} = \max_{\theta} -e_1'\theta$$
$$\text{subject to } Y_{n,j} - W_{n,j}\theta \leq 0 \ \ \forall j.$$

for $\theta = (\eta, \delta)$. The duality theorem for linear programming (see e.g. (24) in Chapter 7.4 of Schrijver (1986)) implies that if the optimum in this problem is finite, it is equal to the solution in the dual problem

$$-\hat{\eta} = \min_{\gamma} -\gamma' Y_n$$
$$\text{subject to } \gamma \geq 0, \ -W_n'\gamma = -e_1.$$

However, we see that the optimal value $\hat{\eta}$ in this problem is in turn equal to that in (21). □

Proof of Lemma 9 The result follows from the argument in Section 5.1 of Fithian et al. (2017), but we provide a separate proof for completeness.

The set of values Y_n such that

$$Y_n'\gamma = \begin{array}{c} \max_{\tilde{\gamma}} \tilde{\gamma}' Y_n \\ \text{subject to } \tilde{\gamma} \geq 0, \ W_n'\tilde{\gamma} = e_1 \end{array} \tag{27}$$

is convex. This follows from the fact that if (27) holds for both Y_n and Y_n^*, then we know that both $Y_n'\gamma \geq Y_n'\tilde{\gamma}$ and $Y_n^{*\prime}\gamma \geq Y_n^{*\prime}\tilde{\gamma}$ for all $\tilde{\gamma} \geq 0$ with $W_n'\tilde{\gamma} = e_1$, which implies that $(\alpha Y_n + (1-\alpha) Y_n^*)' \gamma \geq (\alpha Y_n + (1-\alpha) Y_n^*)' \tilde{\gamma}$ as well.

Thus, once we condition on S_n, the set of values $\gamma' Y_n$ such that (27) holds is an interval. To derive the form of the endpoints, note that

$$\mathcal{V}^{lo}(s) = \min_{Y_n : S_n = s} \left\{ Y_n'\gamma : \begin{array}{c} Y_n'\gamma = \max_{\tilde{\gamma}} \tilde{\gamma}' Y_n \\ \text{subject to } \tilde{\gamma} \geq 0, \ W_n'\tilde{\gamma} = e_1 \end{array} \right\}.$$

Using the definition of S_n, this is equivalent to:

$$\mathcal{V}^{lo}(s) = \min_{Y_n : S_n = s} \left\{ Y_n'\gamma : \begin{array}{c} Y_n'\gamma = \max_{\tilde{\gamma}} \tilde{\gamma}' \left(s + \frac{\Sigma\gamma}{\gamma'\Sigma\gamma} Y_n'\gamma \right) \\ \text{subject to } \tilde{\gamma} \geq 0,\ W_n'\tilde{\gamma} = e_1 \end{array} \right\}.$$

Finally, this is equivalent to

$$\mathcal{V}^{lo}(s) = \min \left\{ c : \begin{array}{c} c = \max_{\tilde{\gamma}} \tilde{\gamma}' \left(s + \frac{\Sigma\gamma}{\gamma'\Sigma\gamma} c \right) \\ \text{subject to } \tilde{\gamma} \geq 0,\ W_n'\tilde{\gamma} = e_1 \end{array} \right\},$$

if the support of $Y_n'\gamma$ equals $\mathbb{R}$. The linear structure of the problem implies that this holds if and only if $\gamma \neq 0$, which we know to be the case since $W_n'\gamma = e_1 \neq 0$. The expression for $\mathcal{V}^{up}$ follows by the same argument.

Independence of $\gamma'Y_n$ and S_n then implies that the conditional distribution of $Y_n'\gamma$ given S_n and (27) is truncated normal. $\square$

Proof of Proposition 5 The Kuhn-Tucker conditions for optimality of γ in the dual problem (which are necessary and sufficient since the problem is a linear program) are that there exist $\left(\hat{\theta}, \hat{\lambda}\right)$ such that

$$Y_n + \hat{\lambda} - W_n\hat{\theta} = 0$$

$$\hat{\lambda} \geq 0,\ \hat{\lambda}_j\gamma_j = 0\,\forall j.$$

From the complementary slackness conditions $\hat{\lambda}_j\gamma_j = 0\,\forall j$, we see that $\hat{\lambda}_j = 0$ for all $j \in B$. Thus, for M_B again the matrix which selects rows B, and M_{B^c} which selects the remaining rows, $Y_{n,B} = M_BY_n$ and $Y_{n,B^c} = M_{B^c}Y_n$, $Y_{n,B} - W_{n,B}\hat{\theta} = 0$. Since the strictly positive elements of γ correspond to linearly independent rows of W_n by assumption, we know that $W_{n,B}$ has full rank. Thus, $\hat{\theta} = W_{n,B}^{-1}Y_{n,B}$. For such $\hat{\theta}$, however, there exists $\hat{\lambda}$ satisfying the conditions above if and only if

$$Y_{n,B^c} - W_{n,B^c}\hat{\theta} \leq 0.$$

Note that any such $\hat{\theta}$ is a solution to the primal problem, with $\hat{\theta} = (\hat{\eta}, \hat{\delta}')'$. In particular, in the dual problem we know that $\gamma'Y_n = \gamma_B'M_B'Y_{n,B}$ and $W_{n,B}'M_B\gamma_B = e_1$, so $M_B\gamma_B = \left(W_{n,B}'\right)^{-1} e_1$ (where γ_B is as defined in Lemma 6) and the optimal value in the dual problem is $e_1' W_{n,B}^{-1} Y_{n,B}$. If we consider the value implied by $\hat{\theta}$, we again obtain

$$\hat{\eta} = e_1'\hat{\theta} = e_1 W_{n,B}^{-1} Y_{n,B}.$$

By Lemma 8, the optimal objective value of the primal is equal to that of the dual, so $\hat{\theta}$ achieves the optimum for the primal, and we argued above that the primal constraints are satisfied at $\hat{\theta}$ when γ solves the dual. We have thus verified a solution to the primal with B binding and $\gamma_{n,B} = \gamma$ whenever $\gamma \in \widehat{V}$.

Finally, recall from the proof of Lemma 5 that if $W_{n,B}$ is invertible and there is a solution to the primal with B binding, then the Kuhn-Tucker conditions hold with $M_B \gamma = (W_{n,B}')^{-1} e_1$ and the other entries of γ equal to zero, so by the sufficiency of the Kuhn-Tucker conditions, γ solves the primal whenever B is binding in the dual. It follows that $\{Y_n$ such that B is binding in the primal$\} = \{Y_n$ such that $\gamma \in \widehat{V}\}$. Observing that when $\gamma \in \widehat{V}$, $S_{n,\gamma} = S_{n,B}$, it is then immediate from Lemmas 6 and 9 that the definition of $\mathcal{V}^{lo}$ and $\mathcal{V}^{up}$ in equations (14) and (15) coincides with that in equations (22) and (23). □

Proof of Lemma 10 Uniqueness and non-degeneracy of the solution $\hat{\theta}$ implies that $|B| = p + 1$. To see that this is the case, note that if $|B| < p + 1$ then there exists a nonzero vector v such that $W_{n,B} v = 0$. If $e_1' v = 0$ then for α sufficiently small $\hat{\theta} + \alpha \cdot v$ is also a solution to the primal problem, contradicting our assumption of uniqueness. If instead $e_1' v \neq 0$, then for sufficiently small $\alpha > 0$, $\hat{\theta} - \alpha \cdot sign\,(e_1' v)\, v$ also satisfies the constraints of the primal problem and attains a smaller value of the objective, contradicting the optimality of $\hat{\theta}$. Likewise, if $|B| > p + 1$, since W_n has $p + 1$ columns the rows of $W_{n,B}$ cannot be linearly independent, violating our assumption of non-degeneracy. Thus, our assumptions imply that $W_{n,B}$ must be a full-rank $(p + 1) \times (p + 1)$ matrix.

We next show that there must be $p+1$ strictly positive multipliers. Note that from the complementary slackness conditions, $\gamma_j = 0$ for $j \notin B$, so there can be at most $p + 1$ strictly positive multipliers. Let $\hat{\gamma}$ be a solution to the dual problem (21). By (21) in Section 10.4 of Schrijver (1986), non-degeneracy of the primal problem implies that for v in an open neighborhood of zero,

$$\begin{array}{c} \min_\theta e_1'\theta \\ \text{subject to } (Y_n + v) - W_n\theta \leq 0 \end{array} = e_1'\hat{\theta} + \hat{\gamma}' v, \tag{28}$$

so $\hat{\gamma}$ gives the marginal change in the objective for small changes in Y_n.

Uniqueness of $\hat{\theta}$ implies that for $\hat{\gamma}_B$ the elements of $\hat{\gamma}$ corresponding to B, $\hat{\gamma}_B > 0$. To see that this is the case, suppose not. Then there exists $\hat{j} \in B$ with $\hat{\gamma}_{\hat{j}} = 0$. In this case, note that for $e_{\hat{j}}$ the vector with a one in entry $\hat{j}$ and zeros everywhere else, $\hat{\gamma}' e_{\hat{j}} = 0$. We know that for α sufficiently small, there continues to be a unique solution with only constraints B binding after we perturb Y_n by $\alpha \cdot e_{\hat{j}}$, and thus that we can write

$$\hat{\theta}\left(\alpha \cdot e_{\hat{j}}\right) = \begin{array}{c} \arg\min_\theta e_1'\theta \\ \text{subject to } \left(Y_n + \alpha \cdot e_{\hat{j}}\right) - W_n\theta \leq 0 \end{array} = W_{n,B}^{-1}\left(Y_{n,B} + \alpha \cdot M_B e_{\hat{j}}\right),$$

for M_B the selection matrix that selects rows in B. Further, by (28) we know that $e_1'\hat{\theta}\left(\alpha \cdot e_{\hat{j}}\right) = \hat{\theta}(0) = \hat{\eta}$, so this perturbation does not affect the objective. Let us define $\tilde{\theta}(\alpha) = \hat{\theta} + \alpha \cdot W_{n,B}^{-1} M_B e_{\hat{j}}$. Note that $e_1'\tilde{\theta}(\alpha) = \hat{\eta}$, while

$$Y_n - W_n\tilde{\theta}(\alpha) = Y_n - W_n\hat{\theta} - \alpha W_n W_{n,B}^{-1} M_B e_j.$$

However, for all $\alpha \geq 0$

$$M_B\left(Y_n - W_n\tilde{\theta}(\alpha)\right) = Y_{n,B} - W_{n,B}\hat{\theta} - \alpha W_{n,B} W_{n,B}^{-1} M_B e_j = Y_{n,B} - W_{n,B}\hat{\theta} - \alpha M_B e_j \leq 0.$$

Since the other rows of $Y_n - W_n\hat{\theta}$ are not binding, they remain nonbinding for α sufficiently small. Thus, there exists $\alpha^* > 0$ such that $Y_n - W_n\tilde{\theta}(\alpha^*) \leq 0$ and $e_1'\tilde{\theta}(\alpha^*) = \hat{\eta}$. There is thus another solution to the primal problem, which contradicts our assumption of uniqueness. $\square$

Proof of Proposition 6 Monotonicity of the conditional distribution in $\gamma'\mu_n$ implies that the test has conditional size α given $\gamma \in \widehat{V}$ and $S_{n,\gamma} = s$ for almost every s. For this section only, we make the dependence of $\mathcal{V}^{lo}$ and $\mathcal{V}^{up}$ on γ explicit, writing $\mathcal{V}^{lo}(s,\gamma)$ and $\mathcal{V}^{up}(s,\gamma)$. Note that for all $V \in \mathcal{V}$, Lemma 11 implies

$$\mathcal{V}^{lo}(S_{n,\gamma_j},\gamma_j) = \mathcal{V}^{lo}(S_{n,\gamma_k},\gamma_k)\forall\gamma_j,\gamma_k \in V$$

$$\mathcal{V}^{up}(S_{n,\gamma_j},\gamma_j) = \mathcal{V}^{up}(S_{n,\gamma_k},\gamma_k)\forall\gamma_j,\gamma_k \in V,$$

and in particular, however the value $\hat{\gamma}$ is selected from $\widehat{V}$,

$$Pr_{\mu_n}\left\{c_{\alpha,C}\left(\gamma,\mathcal{V}^{lo}(S_{n,\gamma},\gamma),\mathcal{V}^{up}(S_{n,\gamma},\gamma),\Sigma\right)=c_{\alpha,C}\left(\hat{\gamma},\mathcal{V}^{lo}(S_{n,\hat{\gamma}},\hat{\gamma}),\mathcal{V}^{up}(S_{n,\hat{\gamma}},\hat{\gamma}),\Sigma\right)|\gamma\in\widehat{V}\right\}=1.$$

Lemma 9, the monotonicity of the conditional distribution in $\gamma'\mu_n$, Assumption 1, and the fact (argued in the proof of Proposition 3) that $Pr_{\mu_n}\left\{\mathcal{V}^{lo}(S_{n,\gamma},\gamma)<\mathcal{V}^{up}(S_{n,\gamma},\gamma)\right\}=1$ imply that for almost every s in the support of S_n, given $\gamma\in\widehat{V}$ and $S_n=s$,

$$\sup_{\mu_n\in\mathcal{M}_0} Pr_{\mu_n}\left\{\hat{\eta}>c_{\alpha,C}\left(\gamma,\mathcal{V}^{lo}(S_{n,\gamma},\gamma),\mathcal{V}^{up}(S_{n,\gamma},\gamma),\Sigma\right)|\gamma\in\widehat{V},S_n=s\right\}=\alpha,$$

from which it follows that

$$\sup_{\mu_n\in\mathcal{M}_0} Pr_{\mu_n}\left\{\hat{\eta}>c_{\alpha,C}\left(\hat{\gamma},\mathcal{V}^{lo}(S_{n,\hat{\gamma}},\hat{\gamma}),\mathcal{V}^{up}(S_{n,\hat{\gamma}},\hat{\gamma}),\Sigma\right)|\gamma\in\widehat{V},S_n=s\right\}=\alpha,$$

and thus that

$$\sup_{\mu_n\in\mathcal{M}_0} E_{\mu_n}\left[\phi_C|\tilde{\gamma}\in\widehat{V},S_n=s\right]=E_0\left[\phi_C|\tilde{\gamma}\in\widehat{V},S_n=s\right]=\alpha.$$

For the first equality we have used the fact that the sup is achieved at $\mu=0$, which again follows monotonicity of the conditional distribution. The law of iterated expectations then immediately implies the first result in the proposition,

$$\sup_{\mu_n\in\mathcal{M}_0} E_{\mu_n}\left[\phi_C|\tilde{\gamma}\in\widehat{V}\right]=E_0\left[\phi_C|\tilde{\gamma}\in\widehat{V}\right]=\alpha.$$

To obtain the second part of the proposition, note that by Lemma 11 the events $\widehat{V}=V_j$, $j\in\{1,...,m\}$ are disjoint, and their union occurs with probability one. Thus,

$$E_{\mu_n}[\phi_C]=\sum_{j=1}^{m} Pr_{\mu_n}\left\{\widehat{V}=V_j\right\}E_{\mu_n}\left[\phi_C|\widehat{V}=V_j\right].$$

By Lemma 11, however,

$$E_{\mu_n}\left[\phi_C|\widehat{V}=V_j\right]=E_{\mu_n}\left[\phi_C|\tilde{\gamma}\in\widehat{V}\right]\ \forall\tilde{\gamma}\in\widehat{V}.$$

Thus, the argument above implies that

$$\sup_{\mu_n \in \mathcal{M}_0} E\left[\phi_C\right] = E_0\left[\phi_C\right] = \alpha.$$

□

Proof of Proposition 7 Size control conditional on $\hat{\eta} \leq c_{\kappa,LF}(X_n, \Sigma)$ and $\gamma \in \widehat{V}$ holds by the same argument as the proof of Proposition 6, replacing $\mathcal{V}^{up}$ with $\mathcal{V}^{up,H}$ as in the text.

To prove unconditional size control, note that

$$E_{\mu_n}\left[\phi_H\right] =$$

$$E_{\mu_n}\left[\phi_H | \hat{\eta} \leq c_{\kappa,LF}(X_n,\Sigma)\right] Pr_{\mu_n}\left\{\hat{\eta} \leq c_{\kappa,LF}\right\} + E_{\mu_n}\left[\phi_H | \hat{\eta} > c_{\kappa,LF}(X_n,\Sigma)\right] Pr_{\mu_n}\left\{\hat{\eta} > c_{\kappa,LF}\right\}.$$

From the first part of the proposition and the law of iterated expectations we know that $E_{\mu_n}\left[\phi_H | \hat{\eta} \leq c_{\kappa,LF}(X_n,\Sigma)\right]$ is bounded above by $\frac{\alpha-\kappa}{1-\kappa}$ while by the construction of the hybrid test we know that $E_{\mu_n}\left[\phi_H | \hat{\eta} > c_{\kappa,LF}(X_n,\Sigma)\right] = 1$. Thus, we see that for $\mu_n \in H_0$,

$$E_{\mu_n}\left[\phi_H\right] \leq \frac{\alpha-\kappa}{1-\kappa} Pr_{\mu_n}\left\{\hat{\eta} \leq c_{\kappa,LF}\right\} + 1 - Pr_{\mu_n}\left\{\hat{\eta} \leq c_{\kappa,LF}\right\}.$$

This expression is decreasing in $Pr_{\mu_n}\left\{\hat{\eta} \leq c_{\kappa,LF}\right\}$, so to obtain an upper bound we need to make $Pr_{\mu_n}\left\{\hat{\eta} \leq c_{\kappa,LF}\right\}$ as small as possible. By Proposition 1 we know $Pr_{\mu_n}\left\{\hat{\eta} \leq c_{\kappa,LF}\right\} \geq 1-\kappa$ under the null, which yields

$$\sup_{\mu_n \in H_0} E_{\mu_n}\left[\phi_H\right] \leq \frac{\alpha-\kappa}{1-\kappa}(1-\kappa) + \kappa = \alpha.$$

Note, further, that both of the bounds we used above are tightest at $\mu_n = 0$, and both bind in this case provided $\hat{\eta}$ is continuously distributed. However, Assumption 1 implies that $\hat{\eta}$ is continuously distributed, so $E_0\left[\phi_H\right] = \alpha$. □

Proof of Lemma 11 Each element of $\widehat{V}$ is also a vertex of the feasible set

$$F = \left\{\gamma : \gamma \geq 0,\, W_n'\gamma = e_1\right\}.$$

Again denoting the vertices of the feasible set by V_F, we thus see that $\widehat{V}$ has support equal to a subset of the power set of V_F. Note, however, that if we consider two values $\gamma_1, \gamma_2 \in V_F$, then since Y_n is normally distributed,

$$Pr\left\{\gamma_1' Y_n = \gamma_2' Y_n\right\} \in \{0, 1\}. \tag{29}$$

Thus, a given set of optimal vertices V in the dual problem (21) either always or never arise together. From this, and the finiteness of the power set of V_F, it follows that there exists a finite set

$$\mathcal{V} = \{V_1, V_2, ..., V_m\}$$

such that $V_j \neq V_k$ for $j \neq k$, $Pr\left\{\widehat{V} \in \mathcal{V}\right\} = 1$, and $Pr\left\{\widehat{V} = V_j\right\} > 0$ for all j, which establishes the first part of the result.

To complete the proof, note that the restriction that each V_j must arise with positive probability together with (29) implies that $V_j \cap V_k = \emptyset$ for all $j \neq k$. To see that this is the case, suppose there exists an element $\gamma \in V_j \cap V_k$. The restrictions that $Pr\left\{\widehat{V} = V_j\right\} > 0$ and $Pr\left\{\widehat{V} = V_k\right\} > 0$ together with (29) imply that

$$Pr\left\{\gamma' Y_n = \gamma_j' Y_n = \gamma_k' Y_n \ \forall (\gamma_j, \gamma_k) \in V_j \times V_k\right\} = 1.$$

However, this is inconsistent with the restriction that $Pr\left\{\widehat{V} = V_j\right\} > 0$ and $V_j \neq V_k$. Thus, we see that $V_j \cap V_k = \emptyset$. $\square$

D Asymptotics

In Sections 4 and 5 of the main text, we derived finite-sample results in the normal model (7), which we motivated in Section 3 as an asymptotic approximation. In this section, we show that these finite sample results translate to asymptotic validity of our proposed tests over a large class of data generating processes. In particular, we establish uniform asymptotic validity of least favorable and least favorable projection tests under minimal conditions. We likewise establish the uniform asymptotic validity of conditional and hybrid tests over classes of data generating processes implying different μ_n values, but these results impose more stringent conditions on X_n and Σ. Specifically, our conditions for these results imply that the dual linear program (21) has a unique solution with probability tending to one, which in turn implies that the primal problem (10) has a non-degenerate solution with probability tending to one.[35]

We conduct our analysis conditional on a sequence of values for the instruments, $\{Z_i\} = \{Z_i\}_{i=1}^{\infty}$, and assume that conditional on $\{Z_i\}_{i=1}^{\infty}$ the data are independent but potentially not identically distributed

$$D_i \perp D_{i'} | \{Z_j\}_{j=1}^{\infty} \text{ for all } i \neq i'.$$

We further assume that for some common conditional distribution $P_{D|Z}$,

$$D_i|Z_i = z \sim P_{D|Z}(z),$$

where the conditional distribution belongs to a family $\mathcal{P}_{D|Z}$ of conditional distributions, $P_{D|Z} \in \mathcal{P}_{D|Z}$. We explore conditions on $\mathcal{P}_{D|Z}$ under which the procedures we suggest are uniformly asymptotically valid.

We first assume that the average conditional variance of Y_i given Z_i converges uniformly to some limit which may depend on $P_{D|Z}$, and that this limit is uniformly bounded over $\mathcal{P}_{D|Z}$.

Assumption 2 *For some* $\Sigma\left(P_{D|Z}\right)$,

$$\lim_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \left\| \frac{1}{n}\sum Var_{P_{D|Z}}\left(Y_i|Z_i\right) - \Sigma\left(P_{D|Z}\right) \right\| \to 0. \tag{30}$$

Further, for all $P_{D|Z} \in \mathcal{P}_{D|Z}$,

$$\Sigma\left(P_{D|Z}\right) \in \Lambda = \left\{ \Sigma : 1/\bar{\lambda} \leq \min_j \Sigma_{jj} \leq \max_j \Sigma_{jj} \leq \bar{\lambda} \right\}$$

where $\bar{\lambda}$ *is a finite constant.*

To justify this assumption, note that for an iid sample from P, if the conditional distribution of $D_i|Z_i$ is $P_{D|Z}$, the strong law of large numbers implies that for almost every sequence $\{Z_i\}_{i=1}^{\infty}$,

$$\frac{1}{n}\sum Var_{P_{D|Z}}\left(Y_i|Z_i\right) \to E_P\left[Var_{P_{D|Z}}\left(Y_i|Z_i\right)\right],$$

[35] Note, however, that uniqueness of the dual solution holds automatically if Σ has full rank, and can be ensured by adding full-rank, mean-zero noise to Y_n. Moreover, since our results are uniform in μ_n, they allow that the "population" version of (21), with $Y_n = \mu_n$, may have a non-unique solution as in one of our simulation specifications.

provided the right hand side exists and is finite. Thus, the convergence in (30) holds pointwise under minimal conditions, and we merely strengthen it to hold uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$. The second part of the assumption then requires that the average conditional variance of each of the moments be bounded above and below, which is again a mild condition. We do not require the matrix $\Sigma\left(P_{D|Z}\right)$ to have full rank, which is important since it allows us to accommodate moment equalities represented as pairs of moment inequalities.

We next suppose that we have a uniformly consistent estimator of the variance $\Sigma\left(P_{D|Z}\right)$. We discuss primitive conditions for this assumption in Section D.3 below, but for the moment take the existence of suitable estimator $\widehat{\Sigma}$ as given.

Assumption 3 *We have an estimator $\widehat{\Sigma}$ for the average conditional variance $\Sigma\left(P_{D|Z}\right)$ which is uniformly consistent in the sense that for all $\varepsilon > 0$,*

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

We further assume that the scaled sample average Y_n is uniformly asymptotically normal once recentered around μ_n. To state this assumption we use the fact that uniform convergence in distribution is equivalent to uniform convergence in bounded Lipschitz metric (see Theorem 1.12.4 of Van der Vaart & Wellner (1996)).

Assumption 4 *For BL_1 the class of Lipschitz functions which are bounded in absolute value by one and have Lipschitz constant bounded by one, and $\xi_{P_{D|Z}} \sim N\left(0, \Sigma\left(P_{D|Z}\right)\right)$,*

$$\lim_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{f\in BL_1} \left| E_{P_{D|Z}}\left[f\left(Y_n - \mu_n\right)\right] - E\left[f\left(\xi_{P_{D|Z}}\right)\right]\right| = 0.$$

Under Assumption 2, Assumption 4 holds whenever the average conditional distribution of $Y_i - \mu_i$ given Z_i is uniformly integrable over $P_{D|Z} \in \mathcal{P}_{D|Z}$.

Lemma 12 *Under Assumption 2, if for all $\varepsilon > 0$,*

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \frac{1}{n}\sum_i E_{P_{D|Z}}\left[\|Y_i - \mu_i\| 1\left\{\|Y_i - \mu_i\| > \varepsilon\sqrt{n}\right\} |Z_i\right] = 0,$$

then Assumption 4 holds.

D.1 Uniform Validity of Least Favorable Tests

Assumptions 2-4 imply the uniform asymptotic validity of feasible least favorable and least favorable projection tests which replace Σ by the estimator $\widehat{\Sigma}$ in all expressions. To formally state this result, it is helpful to define $\mathcal{P}^0_{D|Z}$ as the class of conditional distributions consistent with our conditional moment restriction,

$$\mathcal{P}^0_{D|Z} = \left\{ P_{D|Z} \in \mathcal{P}_{D|Z} : \exists \delta \ s.t. \ E_{P_{D|Z}}\left[Y_i - X_i\delta | Z_i\right] \leq 0 \text{ for all } i \right\}.$$

Proposition 8 *Under Assumptions 2-4, the least favorable projection test is uniformly asymptotically valid*

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}^0_{D|Z}} Pr_{P_{D|Z}}\left\{ \hat{\eta} > c_{\alpha,LF(\delta)}\left(\widehat{\Sigma}\right)\right\} \leq \alpha.$$

The least favorable test is likewise uniformly valid once the critical value is increased by an arbitrarily small amount. In particular, for any $\varepsilon > 0$

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}^0_{D|Z}} Pr_{P_{D|Z}}\left\{ \hat{\eta} > c_{\alpha,LF}\left(X_n, \widehat{\Sigma}\right) + \varepsilon\right\} \leq \alpha.$$

We adjust the critical value in the least favorable test by ε to accommodate the possibility that the distribution of $\widehat{\eta}$ may become degenerate asymptotically. D. Andrews & Shi (2013) termed this an infinitesimal uniformity factor. We next discuss assumptions which rule out such degeneracy, and so ensure asymptotic validity of least favorable tests with $\varepsilon = 0$.

Continuity of the Limit Distribution We next consider assumptions which ensure a continuous limiting distribution for $\hat{\eta}$. These assumptions restrict the behavior of X_n and $\Sigma\left(P_{D|Z}\right)$ but, critically, impose no restrictions on μ_n, and so allow any combination of binding and non-binding moments.

We first assume that X_n, appropriately scaled, converges to some limit as $n \to \infty$.

Assumption 5 $X_n^* = \frac{1}{\sqrt{n}} X_n \to X$ *for a constant matrix* X.

As with Assumption 2, if the data are drawn iid from some distribution P with $E_P\left[X_i\right]$ finite, then the strong law of large numbers implies that this assumption holds for almost every $\{Z_i\}_{i=1}^{\infty}$ if we take $X = E_P\left[X_i\right]$.

Our next assumption concerns the vertices $V_F(X, \Sigma)$ of the feasible region

$$F(X, \Sigma) = \left\{ \gamma : \gamma \geq 0, W'\gamma = e_1 \right\}$$

in the dual problem, where as in (12) $W_j = \left[\begin{array}{cc} \sqrt{\Sigma_{jj}} & X_j \end{array} \right]$.

Assumption 6 *For all $P_{D|Z} \in \mathcal{P}_{D|Z}$, $\Sigma\left(P_{D|Z}\right) \in \mathcal{S}$ where $\mathcal{S} \subseteq \Lambda$ is a compact set of matrices. Moreover, for some finite J,*

$$V_F(X, \Sigma) = \left\{ \gamma^1(X, \Sigma), ..., \gamma^J(X, \Sigma) \right\}$$

where each $\gamma^j(X, \Sigma)$ is unique and continuous in both arguments on $B(X) \times \mathcal{S}$ for $B(X)$ an open neighborhood of X.

This assumption requires that the vertices $V_F(X, \Sigma)$ of the feasible region be continuous at the limiting pair (X, Σ). This will generally fail if the columns of W are multi-collinear, since in this case some of the constraints in $W'\gamma = e_1$ are redundant, and the dimension of the feasible region $F(X, \Sigma)$ changes discontinuously in (X, Σ). This assumption thus implies an asymptotic rank condition, requiring that the different elements of the nuisance parameter vector δ have distinguishable effects on the vector of moments, and can in this sense be understood as an identification condition on δ.

Our final condition restricts the relationship between the variance matrix Σ and the vertices $V(X, \Sigma)$.

Assumption 7 *For all $\Sigma \in \mathcal{S}$ and all $\gamma_1, \gamma_2 \in V_F(X, \Sigma)$ with $\gamma_1 \neq \gamma_2$,*

1. $1/\bar{\lambda} \leq \gamma_1' \Sigma \gamma_1$
2. $(\gamma_1 - \gamma_2)' \Sigma (\gamma_1 - \gamma_2) \geq \frac{1}{\bar{\lambda}}$.

To interpret this assumption, recall that

$$\hat{\eta} = \max_{\gamma \in V_F\left(X_n, \widehat{\Sigma}\right)} \gamma' Y_n,$$

where the asymptotic variance of Y_n is Σ. Thus, $\hat{\eta}$ is a (data-dependent) linear combination of the elements of Y_n. The first part of Assumption 7, $1/\bar{\lambda} \leq \gamma_1' \Sigma \gamma_1$, bounds the asymptotic variance of these linear combinations away from zero, and can be interpreted as an asymptotic analog of Assumption 1 in the main text. The second part of Assumption 7, $(\gamma_1 - \gamma_2)' \Sigma (\gamma_1 - \gamma_2) \geq \frac{1}{\lambda}$, ensures that $\gamma_1' Y_n$ and $\gamma_2' Y_n$ are not perfectly correlated asymptotically.

Both conditions hold automatically if we bound the minimal eigenvalue of Σ away from zero. As noted above, however, we do not wish to rule out moment equalities represented as pairs of inequalities, and so do not impose this condition. More broadly, this assumption implies the existence of a unique solution in the dual problem (21), and thus non-degeneracy of the primal solution, with probability going to one. While this does not require uniqueness in the primal problem (see Corollary 1 in Tijssen & Sierksma (1998)), it rules out the sort of exact primal degeneracy which Appendix A shows can be accommodated in the normal model.

It is worth contrasting Assumption 7 with conditions used elsewhere in the literature on subvector inference. Gafarov (2019), Cho & Russell (2019), and Flynn (2019) all impose versions of the linear independence constraint qualification, which requires that the Jacobian of the binding moments have full rank in a population problem.[36] This rules out degenerate solutions. The linear programs studied in these papers differ from ours, in that they aim to minimize or maximize a parameter of interest subject to moment constraints in the population, while we aim to minimize η subject to constraints in the sample. Assumption 7 then rules out degenerate solutions to our primal problem in-sample. The distinction between the sample and population problems is important, however, since Assumption 7 imposes no restrictions on μ_n, and as we note above can be made to hold mechanically by adding full-rank normal noise to the moments.

With these conditions, we obtain asymptotic validity of ϕ_{LF} with $\varepsilon = 0$,

Corollary 1 *Under Assumptions 2-7, the least favorable test is uniformly valid without an increase in the critical value,*

$$\limsup_{n \to \infty} \sup_{P_{D|Z} \in \mathcal{P}^0_{D|Z}} Pr_{P_{D|Z}} \left\{ \hat{\eta} > c_{\alpha,LF}\left(X_n, \widehat{\Sigma}\right) \right\} \leq \alpha.$$

[36] See Kaido et al. (2019) on the role of constraint qualifications for inference for partially identified models.

D.2 Uniform Validity of Conditional and Hybrid Tests

We next turn to the asymptotic properties of conditional and hybrid tests. Note that the feasible conditional test based on the estimated variance $\widehat{\Sigma}$ can be written as

$$\phi_C = 1\left\{\hat{\gamma}'Y_n > c_{\alpha,C}\left(\hat{\gamma}, \mathcal{V}^{lo}\left(S_{n,\hat{\gamma}}\right), \mathcal{V}^{up}\left(S_{n,\hat{\gamma}}\right), \widehat{\Sigma}\right)\right\},$$

where

$$\hat{\gamma} \in \arg\max_{\gamma \in V_F\left(X_n, \widehat{\Sigma}\right)} \gamma' Y_n.$$

We make the following additional assumption, which ensures that the vertices of the feasible set $V_F(X, \Sigma)$ are either zero or nonzero on a neighborhood of (X, Σ).

Assumption 8 *For all* $\Sigma \in \mathcal{S}$ *and all* $\gamma(X, \Sigma) \in V_F(X, \Sigma)$, $1\{\gamma_j(X, \Sigma) = 0\}$ *is constant on* $B(X) \times B(\Sigma)$ *for all* j*.*

Recall that $\hat{\gamma}$ can be interpreted as the vector of Lagrange multipliers in the primal problem (12). This condition requires that when we consider the set of potential Lagrange multipliers $V_F(X, \Sigma)$, the elements do not switch from zero to nonzero at (X, Σ). Critically, since the realized multiplier $\hat{\gamma}$ is also determined by Y_n, this still allows the distribution of the realized $\hat{\gamma}$ to vary depending on μ_n, which remains unrestricted.

To prove our asymptotic results, we use a modified version of the conditional test which never rejects if $\hat{\eta} < -C$ for C a large positive constant. We do this for technical reasons, since when μ_n diverges to $-\infty$, both $\hat{\eta}$ and our conditional critical values may likewise diverge, and size control for the unmodified test ϕ_C requires that we control the relative rates of divergence. At the same time, this modification is reasonable on substantive grounds, since when $\hat{\eta}$ is very small it is clear from the data that the moments hold, and rejections of the null in this case reflect extreme realizations of the conditional critical values.

Proposition 9 *Under Assumptions 2-8, the modified conditional test*

$$\phi_C^* = \phi_C 1\{\hat{\eta} \geq -C\}$$

is uniformly asymptotically valid,

$$\limsup_{n\to\infty} \sup_{P_{D|Z} \in \mathcal{P}_{D|Z}^0} Pr_{P_{D|Z}}\{\phi_C^* = 1\} \leq \alpha.$$

Analogously, we can write the feasible hybrid test as

$$\phi_H = 1\left\{\hat{\gamma}' Y_n > c_{\frac{\alpha-\kappa}{1-\kappa},H}\left(\hat{\gamma}, \mathcal{V}^{lo}\left(S_{n,\hat{\gamma}}\right), \mathcal{V}^{up}\left(S_{n,\hat{\gamma}}\right), \widehat{\Sigma}\right)\right\}.$$

Once we modify the test to never reject if $\hat{\eta} < -C$, asymptotic validity follows under the same conditions.

Corollary 2 *Under Assumptions 2-8, the modified hybrid test*

$$\phi_H^* = \phi_H 1\left\{\hat{\eta} \geq -C\right\}$$

is uniformly asymptotically valid

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}^0_{D|Z}} Pr_{P_{D|Z}}\left\{\phi_H^* = 1\right\} \leq \alpha.$$

D.3 Asymptotic Variance Estimation

Our asymptotic results have thus far taken as given the existence of a uniformly consistent estimator $\widehat{\Sigma}$ for the conditional variance $\Sigma\left(P_{D|Z}\right)$. Here, we establish the uniform consistency of a particular estimator under mild conditions.

Following Abadie et al. (2014), we consider the nearest-neighbor variance estimator

$$\widehat{\Sigma} = \frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - Y_{\ell_Z(i)}\right)\left(Y_i - Y_{\ell_Z(i)}\right)' \tag{31}$$

where for Ξ_n a positive-definite matrix,

$$\ell_Z(i) = \text{argmin}_{j\in\{1,\ldots,n\},j\neq i}\left(Z_i - Z_j\right)'\Xi_n\left(Z_i - Z_j\right)$$

selects the index for the observation j with Z_j as close as possible to Z_i in distance defined by Ξ_n. One natural choice of Ξ_n is the inverse of the sample variance, $\Xi_n = \widehat{Var}\left(Z_i\right)^{-1}$, provided the sample variance has full rank. For ease of exposition we assume that Z_i has at least one continuously distributed dimension, so that $\ell_Z(i)$ is unique for all i. If instead Z_i is entirely discrete, one can estimate $\widehat{\Sigma}$ using the average of the sample conditional variances.

The intuition for the estimator $\widehat{\Sigma}$ is straightforward. Provided the conditional mean and variance of Y_i given Z_i are continuous in Z_i, if $Z_{\ell_Z(i)}$ is close to Z_i it will have nearly the same mean and variance. Hence, the variance of $Y_i - Y_{\ell_Z(i)}$ will be approximately twice the variance of Y_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 $\widehat{\Sigma}$ will converge to the average conditional variance $\Sigma\left(P_{D|Z}\right)$. The next assumption formalizes the conditions needed for this argument.

Assumption 9 *For $\lambda_{\max}(A)$ the maximal eigenvalue of a matrix A, the following conditions hold*

1. $\{Z_i\}_{i=1}^{\infty} \subset \mathcal{Z}^{\infty}$ *for $\mathcal{Z}$ a compact set*
2. $\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \frac{1}{n}\sum E_{P_{D|Z}}\left[\|Y_i\|^4 |Z_i\right]$ *is finite*
3. $\mu_{P_{D|Z}}(z) = E_{P_{D|Z}}[Y_i|Z_i = z]$ *is Lipschitz in z with Lipschitz constant uniformly bounded over $P_{D|Z} \in \mathcal{P}_{D|Z}$, and is uniformly bounded over $P_{D|Z} \in \mathcal{P}_{D|Z}$*
4. $V_{P_{D|Z}}(z) = E_{P_{D|Z}}[Y_iY_i'|Z_i = z]$ *is Lipschitz in z with Lipschitz constant uniformly bounded over $P_{D|Z} \in \mathcal{P}_{D|Z}$*
5. $\sup_{P_{D|Z}\in\mathcal{P}_{D|Z}} \sup_{z\in\mathcal{Z}} \lambda_{\max}\left(Var_{P_{D|Z}}(Y_i|Z_i = z)\right)$ *is finite*
6. $\Xi_n \to \Xi$ *for a positive-definite limit Ξ*

Assumption 9(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 9(2) and (5) restrict the variance and second moment of Y_i, and are satisfied under a wide range of data generating processes. Assumption 9(3) and (4) impose Lipschitz continuity on the mean and second moment of Y_i, consistent with the heuristic argument given above. Finally, 9(6) requires only that Ξ_n converge to a positive-definite limit.

Proposition 10 *Under Assumptions 2 and 9, for $\widehat{\Sigma}$ as defined in (31), and all $\varepsilon > 0$*

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

so Assumption 3 holds.

E Proofs for Asymptotic Results

This section collects the proofs for the asymptotic results stated in Section D, along with the statements and proofs of some auxiliary results. Section E.1 proves Proposition 8, while Section E.2 proves Proposition 9, and Section E.3 proves Proposition 10.

Proof of Lemma 12 Towards contradiction, suppose the conclusion of the lemma fails. Then there exists a sequence of distributions and sample sizes $(P_{D|Z,m}, n_m)$ and a constant $\varepsilon > 0$ such that

$$\liminf_{m\to\infty} \sup_{f\in BL_1} \left| E_{P_{D|Z,m}}\left[f\left(Y_{n_m} - \mu_{n_m}\right)\right] - E\left[f\left(\xi_{P_{D|Z,m}}\right)\right]\right| > \varepsilon. \tag{32}$$

Since the set Λ specified in Assumption 2 is compact, there exists a subsequence of distributions and sample sizes $(P_{D|Z,l}, n_l)$ along which $\Sigma\left(P_{D|Z,l}\right) \to \Sigma$ for $\Sigma \in \Lambda$. Under this subsequence, however, the Linderberg Feller Central Limit Theorem (see e.g. Proposition 2.27 in Van der Vaart 1998), along with the assumptions of the lemma, implies that

$$Y_{n_l} - \mu_{n_l} \to_d N\left(0, \Sigma\right),$$

and thus that

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

This contradicts (32), completing the proof. □

E.1 Proof of Validity For Least Favorable Tests

As a preliminary step, we show that for test statistics $R\left(\xi, \Sigma\right)$ which are (a) constant outside compact sets of values ξ and (b) bounded Lipschitz in both arguments, the critical value function is likewise bounded Lipschitz. To prove this statement, we use the metric

$$d\left(\Sigma_1, \Sigma_2\right) = \left\|\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\right\| + \left\|\Sigma_1 - \Sigma_2\right\|$$

for $\|A\|$ the Euclidean norm if A is a vector, and the operator norm if A is a matrix.

Lemma 13 *Suppose* $R(\xi,\Sigma)$ *is (a) constant in* ξ *when* $\max_j \left\{|\xi_j| / \sqrt{\Sigma_{jj}}\right\} > C$ *for some constant* C *and (b) bounded Lipschitz in both arguments for* $\Sigma \in \Lambda$ *with Lipschitz constant* K. *Then for* $c_\alpha(\Sigma)$ *the* $1-\alpha$ *quantile of* $R(\xi,\Sigma)$ *under* $\xi \sim N(0,\Sigma)$, $c_\alpha(\Sigma)$ *is bounded Lipschitz with a constant that depends only on* C, K, *and* $\bar{\lambda}$.

Proof of Lemma 13 That $c_\alpha(\Sigma)$ is bounded follows immediately from boundedness of $R(\xi,\Sigma)$. Next, note that we can write $R(\xi,\Sigma) = R\left(\Sigma^{\frac{1}{2}}\zeta, \Sigma\right)$ for $\zeta \sim N(0,I)$. Since $R\left(\Sigma^{\frac{1}{2}}\zeta, \Sigma\right)$ is constant for ζ outside a compact set $\mathcal{C}$, it suffices to limit attention to $(\zeta,\Sigma) \in \mathcal{C} \times \Lambda$. Note further that for any pair Σ_1, $\Sigma_2 \in \Lambda$ and any $\zeta \in \mathcal{C}$,

$$R\left(\Sigma_1^{\frac{1}{2}}\zeta, \Sigma_1\right) - R\left(\Sigma_2^{\frac{1}{2}}\zeta, \Sigma_2\right) \le K\|\Sigma_1 - \Sigma_2\| + K\left\|\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\right\| + K\left\|\left(\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\right)\zeta\right\|$$

$$\le K\|\Sigma_1 - \Sigma_2\| + K\left\|\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\right\| + K\left\|\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\right\|\|\zeta\|$$

$$\le K\|\Sigma_1 - \Sigma_2\|(1 + \|\zeta\|) + K\left\|\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\right\| + K\left\|\Sigma_1^{\frac{1}{2}} - \Sigma_2^{\frac{1}{2}}\right\|\|\zeta\|$$

$$\le K(1 + \|\mathcal{C}\|)\, d(\Sigma_1, \Sigma_2)$$

for $\|\mathcal{C}\| = \sup_{\zeta\in\mathcal{C}}\|\zeta\|$, where the second line follows from the definition of the operator norm, the third line adds a weakly positive term to the RHS, and the final line uses the definition of the metric and takes a supremum.

Thus, we see that

$$1 - \alpha = Pr\left\{R(\xi_1, \Sigma_1) \le c_\alpha(\Sigma_1)\right\}$$

$$\le Pr\left\{R(\xi_2, \Sigma_2) \le c_\alpha(\Sigma_1) + K(1 + \|\mathcal{C}\|)\, d(\Sigma_1, \Sigma_2)\right\},$$

and hence that $c_\alpha(\Sigma_2) \le c_\alpha(\Sigma_1) + K(1 + \|\mathcal{C}\|)\, d(\Sigma_1, \Sigma_2)$. Repeating the argument in the other direction, we obtain that

$$|c_\alpha(\Sigma_1) - c_\alpha(\Sigma_2)| \le K(1 + \|\mathcal{C}\|)\, d(\Sigma_1, \Sigma_2),$$

and hence that $c_\alpha(\Sigma)$ is Lipschitz in Σ, as we aimed to show. □

Lemma 13 applies only to test statistics that are (a) globally Lipschitz and (b) constant for ξ large. Our next result builds on this lemma to establish asymptotic validity for tests based on a much wider range of statistics.

Assumption 10 *For all constants C, $R(\xi,\Sigma)$ is bounded Lipschitz in (ξ,Σ) for*

$$\left\{(\xi,\Sigma)\,:\,\Sigma\in\Lambda,\,\max_j\left\{|\xi_j|\,/\sqrt{\Sigma_{jj}}\right\}\leq C\right\}$$

with Lipschitz constant $K(C)$.

Lemma 14 *Under Assumptions 2- 4, for any $\varepsilon>0$ and any sequence of test statistics R_n satisfying Assumption 10 for a common $K(C)$, and corresponding critical values $c_{\alpha,n}\left(\widehat{\Sigma}\right)$,*

$$\lim_{n\to\infty}\sup_{P_{D|Z}\in\mathcal{P}_{D|Z}}Pr_{P_{D|Z}}\left\{R_n\left(Y_n-\mu_n,\widehat{\Sigma}\right)\geq c_{\alpha,n}\left(\widehat{\Sigma}\right)+\varepsilon\right\}\leq\alpha.$$

Proof of Lemma 14 For constants (C_1,C_2) with $0<C_1<C_2$ let us define $\varsigma(\xi,\Sigma)=\max_j\left\{|\xi_j|\,/\sqrt{\Sigma_{jj}}\right\}$ and

$$\psi(R,\xi,\Sigma,C_1,C_2)=\left(1\left\{\varsigma(\xi,\Sigma)<C_1\right\}+\frac{C_2-\varsigma(\xi,\Sigma)}{C_2-C_1}1\left\{C_1\leq\varsigma(\xi,\Sigma)<C_2\right\}\right)R(\xi,\Sigma).$$

$\psi(R,\xi,\Sigma,C_1,C_2)$ is equal to $R(\xi,\Sigma)$ when $\varsigma(\xi,\Sigma)$ is small, and continuously censors to zero when $\varsigma(\xi,\Sigma)$ is large. Note that for any (C_1,C_2), the assumptions of the lemma and the fact that products of bounded Lipschitz functions are bounded Lipschitz imply that $\psi(R,\xi,\Sigma,C_1,C_2)$ is bounded Lipschitz in (ξ,Σ) for ξ unrestricted and $\Sigma\in\Lambda$. By Lemma 13, if we define $c_{\alpha,n}(\Sigma,C_1,C_2)$ as the $1-\alpha$ quantile of $\psi(R_n,\xi,\Sigma,C_1,C_2)$ under $\xi\sim N(0,\Sigma)$, we see that $c_{\alpha,n}(\Sigma,C_1,C_2)$, and thus the difference

$$\psi(R_n,\xi,\Sigma,C_1,C_2)-c_{\alpha,n}(\Sigma,C_1,C_2)$$

is bounded Lipschitz as well.

Towards contradiction, suppose the conclusion to the lemma fails. Then there exists a sequence of distributions $\left\{P_{D|Z,m}\right\}\subset\mathcal{P}_{D|Z}$, sample sizes n_m, and a constant $\nu>0$ such that

$$\liminf_{m\to\infty}Pr_{P_{D|Z,m}}\left\{R_{n_m}\left(Y_{n_m}-\mu_{n_m},\widehat{\Sigma}\right)>c_{\alpha,n_m}\left(\widehat{\Sigma}\right)+\varepsilon\right\}\geq\alpha+\nu.$$

Let us choose $C_1 > 0$ such that

$$\sup_{\Sigma \in \Lambda} Pr_{\Sigma} \left\{ \varsigma(\xi, \Sigma) \geq \frac{1}{2} C_1 \right\} < \frac{\nu}{4}.$$

Since $R_{n_m}(\xi, \Sigma)$ and $\psi(R_{n_m}, \xi, \Sigma, C_1, C_2)$ are equal when $\varsigma(\xi, \Sigma) \leq C_1$, we see that

$$c_{\alpha+\nu/4, n_m}(\Sigma, C_1, C_2) \leq c_{\alpha, n_m}(\Sigma) \leq c_{\alpha-\nu/4, n_m}(\Sigma, C_1, C_2).$$

Assumptions 2-4 imply that

$$\limsup_{m\to\infty} Pr_{P_{D|Z,m}} \left\{ \varsigma\left(Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}\right) > C_1 \right\} < \frac{\nu}{4}. \tag{33}$$

To see that this is the case, note that since the set of matrices Λ is compact, for any sequence of distributions and sample sizes $(P_{D|Z,s}, n_s)$ there exists a subsequence $(P_{D|Z,s_t}, n_{s_t})$ such that $\Sigma(P_{D|Z,s_t}) \to \Sigma$ for some $\Sigma \in \Lambda$. Under this subsequence, $Y_{n_{s_t}} - \mu_{n_{s_t}} \to_d N(0, \Sigma)$, $\widehat{\Sigma} \to_p \Sigma$, and

$$\limsup_{t\to\infty} Pr_{P_{D|Z,s_t}} \left\{ \varsigma\left(Y_{n_{s_t}} - \mu_{n_{s_t}}, \widehat{\Sigma}\right) > C_1 \right\} < \frac{\nu}{4}$$

by the continuous mapping theorem and the portmanteau Lemma (see Lemma 2.2 of Van der Vaart (2000)). Since such a subsequence can be extracted for any sequence, the claim follows.

Since $R_{n_m}\left(Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}\right)$ and $\psi\left(R_{n_m}, Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}, C_1, C_2\right)$ are equal for $\varsigma\left(Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}\right) \leq C_1$, this implies that

$$\limsup_{m\to\infty} Pr_{P_{D|Z,m}} \left\{ R_{n_m}\left(Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}\right) \neq \psi\left(R_{n_m}, Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}, C_1, C_2\right) \right\} < \frac{\nu}{4}.$$

Thus,

$$\limsup_{m\to\infty} Pr_{P_{D|Z,m}} \left\{ \psi\left(R_{n_m}, Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}, C_1, C_2\right) > c_{\alpha, n_m}\left(\widehat{\Sigma}\right) + \varepsilon \right\} \geq \alpha + \frac{3}{4}\nu.$$

Since we have shown that $c_{\alpha, n_m}(\Sigma) \geq c_{\alpha+\nu/4, n_m}(\Sigma, C_1, C_2)$, this implies that

$$\limsup_{m\to\infty} Pr_{P_{D|Z,m}} \left\{ \psi\left(R_{n_m}, Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}, C_1, C_2\right) > c_{\alpha+\nu/4, n_m}\left(\widehat{\Sigma}, C_1, C_2\right) + \varepsilon \right\} \geq \alpha + \frac{3}{4}\nu.$$

Let

$$\mathcal{T}_m = \psi\left(R_{n_m}, Y_{n_m} - \mu_{n_m}, \widehat{\Sigma}, C_1, C_2\right) - c_{\alpha+\nu/4,n_m}\left(\widehat{\Sigma}, C_1, C_2\right)$$

and

$$\mathcal{T}_{m,\infty} = \psi\left(R_{n_m}, \xi, \Sigma, C_1, C_2\right) - c_{\alpha+\nu/4,n_m}\left(\Sigma, C_1, C_2\right),$$

for $\xi \sim N\left(0, \Sigma\left(P_{D|Z,m}\right)\right)$. The difference between $\mathcal{T}_m$ and $\mathcal{T}_{m,\infty}$ is that the former uses the finite-sample distribution of $Y_{n_m} - \mu_{n_m}$ and $\widehat{\Sigma}$ while the latter uses the asymptotic normal distribution for ξ and the exact value of Σ. Our arguments above show that, viewed as a function of $\left(Y_n - \mu_n, \widehat{\Sigma}\right)$, $\mathcal{T}_m$ is bounded Lipschitz. Since compositions of bounded Lipschitz functions are bounded Lipschitz, Assumptions 3 and 4 imply that

$$\lim_{m\to\infty} \sup_{f\in BL_1} \left|E\left[f\left(\mathcal{T}_m\right)\right] - E\left[f\left(\mathcal{T}_{m,\infty}\right)\right]\right| = 0. \tag{34}$$

Since $\mathcal{T}_m$ is a sequence of bounded variables, by Prohorov's theorem there exists a subsequence m_s and a random variable $\mathcal{T}$ such that $\mathcal{T}_{m_s} \to_d \mathcal{T}$. By (34) and the Portmanteau lemma (see Lemma 2.2 of Van der Vaart (2000)), however, we also have $\mathcal{T}_{m,\infty} \to_d \mathcal{T}$. From the Portmanteau lemma, it follows that

$$\alpha+\frac{3}{4}\nu \leq \limsup_{s\to\infty} Pr\left\{\mathcal{T}_{m_s} \geq \varepsilon\right\} \leq Pr\left\{\mathcal{T} \geq \varepsilon\right\} \leq Pr\left\{\mathcal{T} > 0\right\} \leq \liminf_{s\to\infty} Pr\left\{\mathcal{T}_{m_s,\infty} > 0\right\}.$$

However, $Pr\left\{\mathcal{T}_{m_s,\infty} > 0\right\} \leq \alpha + \frac{\nu}{4}$ for all m by the definition of the quantile function. Thus, since $\nu > 0$ we have arrived at a contradiction. □

Lemma 15 *Provided*

$$\inf_{\delta} \max_{j} \left\{X_{n,j}\delta\right\} \neq -\infty,$$

the statistic

$$\min_{\delta} S\left(\xi - X_n\delta, \Sigma\right) = \min_{\delta} \max_{j} \left\{\left(\xi_j - X_{n,j}\delta\right)/\sqrt{\Sigma_{jj}}\right\}$$

satisfies Assumption 10 with Lipschitz constants independent of X_n.

Proof of Lemma 15 Note, first, that for any fixed δ the statistic

$$\tilde{S}\left(\xi, X_n, \Sigma; \delta\right) = \max_{j} \left\{\left(\xi_j - X_{n,j}\delta\right)/\sqrt{\Sigma_{jj}}\right\}$$

is Lipschitz in (ξ, Σ) for $\Sigma \in \Lambda$ and ξ such that $\max_j \left\{ |\xi_j| / \sqrt{\Sigma_{jj}} \right\} \leq C$, with a Lipschitz constant that does not depend on δ or X_n. Since the minimum of a collection of functions with a common Lipschitz constant is Lipschitz with the same constant, this implies that

$$\tilde{S}(\xi, X_n, \Sigma) = \min_{\delta} \tilde{S}(\xi, X_n, \Sigma; \delta)$$

is Lipschitz with the same constant.

To see that the statistic is bounded observe that the assumption that $\inf_\delta \max_j \{X_{n,j}\delta\} \neq -\infty$ implies that

$$\tilde{S}(\xi, X_n, \Sigma) \geq \min_j \left\{ \xi_j / \sqrt{\Sigma_{jj}} \right\},$$

since otherwise the span of X_n must contain a strictly negative vector, and hence $\inf_\delta \max_j \{X_{n,j}\delta\} = -\infty$. On the other hand, by construction

$$\tilde{S}(\xi, X_n, \Sigma) \leq \tilde{S}(\xi, X_n, \Sigma; 0) = \max_j \left\{ \xi_j / \sqrt{\Sigma_{jj}} \right\}.$$

Thus, we see that for $\max_j \left\{ |\xi_j| / \sqrt{\Sigma_{jj}} \right\} \leq C$ for any constant C, $\tilde{S}(\xi, X_n, \Sigma)$ is bounded between $-C$ and C. $\square$

We next build on these preliminary results to prove uniform size control for the least favorable test.

Proof of Proposition 8 If X_n is such that $\inf_\delta \max_j \{X_{n,j}\delta\} = -\infty$ then $\hat{\eta} = -\infty$ with probability one, and our tests never reject. For the remainder of the proof we thus assume that $\inf_\delta \max_j \{X_{n,j}\delta\} \neq -\infty$.

For the least favorable projection test, note that this test rejects if and only if

$$S\left(Y_n - X_n\delta, X_n, \widehat{\Sigma}\right) > c_{\alpha, H_0(\delta)}\left(\widehat{\Sigma}\right)$$

for all δ. Note that under the null, there exists a value δ^* such that $\mu_n - X_n\delta^* \leq 0$. Hence,

$$1\left\{S\left(Y_n - X_n\delta^*, X_n, \widehat{\Sigma}\right) > c_{\alpha, H_0(\delta)}\left(\widehat{\Sigma}\right)\right\} \leq 1\left\{S\left(Y_n - \mu_n, X_n, \widehat{\Sigma}\right) > c_{\alpha, H_0(\delta)}\left(\widehat{\Sigma}\right)\right\}.$$

Note, however, that $S\left(Y_n - \mu_n, X_n, \widehat{\Sigma}\right)$ is the (scaled) maximum of a finite number of normal random variables with nonzero variance, and is Lipschitz in $(Y_n - \mu_n, \Sigma)$ for

$\Sigma \in \Lambda$ and $Y_n - \mu_n$ bounded. Lemma 14 thus implies that for any $\varepsilon > 0$,

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}^0_{D|Z}} Pr_{P_{D|Z}}\left\{S\left(Y_n-\mu_n, X_n, \widehat{\Sigma}\right) > c_{\alpha,H_0(\delta)}\left(\widehat{\Sigma}\right)+\varepsilon\right\} \leq \alpha. \tag{35}$$

Moreover, for $\xi \sim N(0,\Sigma)$, $S(\xi, X_n, \Sigma)$ is continuously distributed with density bounded uniformly over $\Sigma \in \Lambda$ (see e.g. Theorem 3 of Chernozhukov et al. (2015)). Thus, since (35) holds for all $\varepsilon > 0$, it follows that is also holds for $\varepsilon = 0$.

To establish size control for least favorable tests, we note that since the test statistic is monotonically increasing in Y_n, the fact that $\mu_n \leq 0$ under the null implies that

$$1\left\{\hat{\eta} > c_\alpha\left(X_n, \widehat{\Sigma}\right)+\varepsilon\right\} \leq 1\left\{\min_\delta S\left(Y_n-\mu_n-X_n\delta, X_n, \widehat{\Sigma}\right) > c_\alpha\left(X_n, \widehat{\Sigma}\right)+\varepsilon\right\}.$$

Thus, if we can prove that the right hand side has asymptotic rejection probability less than or equal to α under the null, the left hand side must as well. Since Lemma 15 shows that $\min_\delta S(\xi - X_n\delta, X_n, \Sigma)$ satisfies the conditions of Lemma 14 with Lipschitz constants that do not depend on X_n, Lemma 14 immediately implies that

$$\limsup_{n\to\infty} \sup_{P_{D|Z}\in\mathcal{P}^0_{D|Z}} Pr_{P_{D|Z}}\left\{\min_\delta S\left(Y_n-\mu_n, X_n, \widehat{\Sigma}\right) \geq c_{\alpha,n}\left(\widehat{\Sigma}\right)+\varepsilon\right\} \leq \alpha,$$

as we aimed to show. $\square$

Proof of Corollary 1 As in the proof of Lemma 14, let us assume the result fails. Then there exists a sequence of distributions $\left\{P_{D|Z,m}\right\} \subset \mathcal{P}_{D|Z}$, sample sizes n_m, and a constant $\nu > 0$ such that, for $\tilde{S}$ defined as in the proof of Lemma 15,

$$\liminf_{m\to\infty} Pr_{P_{D|Z,m}}\left\{\tilde{S}\left(Y_{n_m}-\mu_{n_m}, X_{n_m}, \widehat{\Sigma}, X_{n_m}\right) > c_{\alpha,LF}\left(\widehat{\Sigma}, X_{n_m}\right)\right\} \geq \alpha+\nu.$$

Let us choose $C_1 > 0$ such that

$$\sup_{\Sigma\in\Lambda} Pr_\Sigma\left\{\varsigma(\xi,\Sigma) \geq \frac{1}{2}C_1\right\} < \frac{\nu}{4},$$

where we again define $\varsigma(\xi,\Sigma) = \max_j\left\{|\xi_j|/\sqrt{\Sigma_{jj}}\right\}$.

As argued in the proof of Lemma 14, this implies that (for ψ as defined in that proof)

$$\limsup_{m\to\infty} Pr_{P_{D|Z},m}\left\{\tilde{S}\left(Y_{n_m}-\mu_{n_m},\widehat{\Sigma}\right)\neq\psi\left(\tilde{S},Y_{n_m}-\mu_{n_m},\widehat{\Sigma},C_1,C_2\right)\right\}<\frac{\nu}{4}.$$

and

$$\limsup_{m\to\infty} Pr_{P_{D|Z,m}}\left\{\psi\left(\tilde{S},Y_{n_m}-\mu_{n_m},\widehat{\Sigma},C_1,C_2\right)>c_{\alpha,LF}\left(\widehat{\Sigma},X_{n_m}\right)\right\}\geq\alpha+\frac{3}{4}\nu.$$

and thus that

$$\limsup_{m\to\infty} Pr_{P_{D|Z,m}}\left\{\psi\left(\tilde{S},Y_{n_m}-\mu_{n_m},\widehat{\Sigma},C_1,C_2\right)>c_{\alpha+\nu/4,n_m}\left(\widehat{\Sigma},C_1,C_2\right)\right\}\geq\alpha+\frac{3}{4}\nu,$$

for $c_{\alpha+\nu/4,n_m}\left(\widehat{\Sigma},C_1,C_2\right)$ the $1-\alpha-\nu/4$ quantile of $\tilde{S}\left(\xi,X_{n_m},\Sigma\right)$ for $\xi\sim N(0,\Sigma)$.

Since the set Λ is compact, we can extract a further subsequence n_s along which $\Sigma\left(P_{D|Z,n_s}\right)\to\Sigma$.[37] We see, however, that along this subsequence the continuous mapping theorem implies

$$\tilde{S}\left(Y_{n_s}-\mu_{n_s},X_{n_s},\widehat{\Sigma}\right)\to_d\max_{\gamma\in V_F(X,\Sigma)}\gamma'\xi,$$

and $c_{\alpha+\nu/4,n_s}\left(\widehat{\Sigma},C_1,C_2\right)\to_p c_{\alpha+\nu/4,n_s}\left(\Sigma,C_1,C_2\right)$, where we have used the continuity of $\tilde{S}\left(\xi,X,\Sigma\right)$ implied by Lemma 19 below, as well as the continuity of the critical value implied by Lemma 13.

The proof of Lemma 20 below then implies that

$$\tilde{S}\left(Y_{n_s}-\mu_{n_s},X_{n_s},\widehat{\Sigma}\right)-c_{\alpha+\nu/4,n_s}\left(\widehat{\Sigma},C_1,C_2\right)\tag{36}$$

converges in distribution to a continuous random variable. Note, however, that the total variation distance between (36) and

$$\mathcal{T}_s=\psi\left(\tilde{S},Y_{n_s}-\mu_{n_s},\widehat{\Sigma},C_1,C_2\right)-c_{\alpha+\nu/4,n_s}\left(\widehat{\Sigma},C_1,C_2\right)$$

is bounded above by $\nu/4$ asymptotically by the argument following (33) in the proof

[37] We write (s,n_s) rather than (m_s,n_{m_s}) for readability.

of Lemma 14. If we define

$$\mathcal{T}_{s,\infty} = \psi\left(\tilde{S}, \xi, \Sigma, C_1, C_2\right) - c_{\alpha+\nu/4,n_s}\left(\Sigma, C_1, C_2\right),$$

then as in the proof of Lemma 14 we know that

$$\lim_{s\to\infty} \sup_{f\in BL_1} \left|E\left[f\left(\mathcal{T}_s\right)\right] - E\left[f\left(\mathcal{T}_{s,\infty}\right)\right]\right| = 0.$$

As in the proof of Lemma 14, by Prohorov's Theorem we know there exists a further subsequence s_t along with $\mathcal{T}_{s_t} \to \mathcal{T}$ for a random variable $\mathcal{T}$. Moreover, we know that $\mathcal{T}_{s_t,\infty}$ converges to the same limit, and thus that by the Portmanteau lemma

$$\alpha + \frac{3}{4}\nu \leq \limsup_{t\to\infty} Pr\left\{\mathcal{T}_{s_t} \geq 0\right\} \leq Pr\left\{\mathcal{T} \geq 0\right\}$$

and

$$Pr\left\{\mathcal{T} > 0\right\} \leq \liminf_{t\to\infty} Pr\left\{\mathcal{T}_{s_t,\infty} > 0\right\} \leq \alpha + \nu/4$$

by the definition of the critical value. Thus, we see that $Pr\left\{\mathcal{T} = 0\right\} \geq \frac{\nu}{2}$. However, we have argued that for large t, $\mathcal{T}_{s_t}$ is within total variation distance $\frac{\nu}{4}$ of a sequence of random variables that converge in distribution to a continuous limit, which implies that $Pr\left\{\mathcal{T} = 0\right\} \leq \frac{\nu}{4}$. Thus, we have reached a contradiction. □

E.2 Proof of Validity for Conditional and Hybrid Tests

We next turn to the proof of Proposition 9. Let us define

$$T\left(Y_n, X_n, \widehat{\Sigma}\right) = \hat{\gamma}' Y_n - c_{\alpha,C}\left(\hat{\gamma}, \mathcal{V}^{lo}\left(S_{n,\hat{\gamma}}\right), \mathcal{V}^{up}\left(S_{n,\hat{\gamma}}\right), \widehat{\Sigma}\right) \tag{37}$$

for

$$\hat{\gamma} = \operatorname{argmax}_{\gamma\in V_F\left(X_n,\widehat{\Sigma}\right)} \gamma' Y_n.$$

Note that $\hat{\eta}$ exceeds the conditional critical value if and only if $T\left(Y_n, X_n, \widehat{\Sigma}\right)$ is strictly positive. As in the last section, we begin by proving several auxiliary lemmas.

Lemma 16 *For all* $\tilde{\delta} \in \mathbb{R}^p$,

$$T\left(Y_n, X_n, \widehat{\Sigma}\right) = T\left(Y_n + X_n^*\tilde{\delta}, X_n^*, \widehat{\Sigma}\right),$$

where again $X_n^* = \frac{1}{\sqrt{n}} X_n$.

Proof of Lemma 16 Recall that the feasible region $F(X,\Sigma)$ is the set of points $\gamma \geq 0$ such that $\sqrt{diag(\Sigma)}'\gamma = 1$ and $X'\gamma = 0$. It follows that $F(X_n,\Sigma) = F(X_n^*,\Sigma)$, and hence that the set of vertices $V_F\left(X_n,\widehat{\Sigma}\right) = V_F\left(X_n^*,\widehat{\Sigma}\right)$. From this we see immediately that $T\left(Y_n,X_n,\widehat{\Sigma}\right) = T\left(Y_n,X_n^*,\widehat{\Sigma}\right)$. Since $\gamma' X_n^* = 0$, we also see that $\hat{\gamma}$ calculated with Y_n is the same as $\hat{\gamma}$ calculated with $Y_n + X_n^*\tilde{\delta}$, and

$$\hat{\gamma}' Y_n = \hat{\gamma}'\left(Y_n + X_n^*\tilde{\delta}\right).$$

Likewise, for all $\hat{\gamma},\gamma \in V_F\left(X_n^*,\widehat{\Sigma}\right)$,

$$\gamma' S_{n,\hat{\gamma}} = \gamma' Y_n - \frac{\gamma'\widehat{\Sigma}\hat{\gamma}}{\hat{\gamma}'\widehat{\Sigma}\hat{\gamma}}\hat{\gamma}' Y_n = \gamma'\left(Y_n + X_n^*\tilde{\delta}\right) - \frac{\gamma'\widehat{\Sigma}\hat{\gamma}}{\hat{\gamma}'\widehat{\Sigma}\hat{\gamma}}\hat{\gamma}'\left(Y_n + X_n^*\tilde{\delta}\right).$$

Thus, $\gamma' S_{n,\hat{\gamma}}$ calculated with Y_n is equal to $\gamma' S_{n,\hat{\gamma}}$ calculated with $Y_n + X_n^*\tilde{\delta}$. From (22) and (23), it is thus clear that $V^{lo}(s)$ and $V^{up}(s)$ are the same when calculated with Y_n as with $Y_n + X_n^*\tilde{\delta}$. This suffices to establish the result. $\square$

Lemma 17 *Under Assumptions 6 and 7, for all* μ^* *with* $\mu_j^* \in [-\infty, 0]$ *for all* j*,*

$$\hat{\gamma}(\xi, X, \Sigma) = argmax_{\gamma \in V_F(X,\Sigma)} \gamma'(\xi + \mu^*),$$

$\hat{\gamma}(\xi, X, \Sigma)$ *is almost surely continuous at* (ξ, X, Σ) *for* $\xi \sim N(0,\Sigma)$ *and* (X,Σ) *non-stochastic, where we define* $0 \cdot \infty = 0$.

Proof of Lemma 17 To prove this result, note first that Assumption 7 implies that for any pair γ_1, $\gamma_2 \in V_F(X,\Sigma)$, $(\gamma_1 - \gamma_2)'\xi$ has a non-degenerate normal distribution. By Assumption 6, the same also holds on a neighborhood of (X,Σ). This implies, however, that on a neighborhood of (X,Σ), $\hat{\gamma}(\xi, X, \Sigma)$ is unique with probability one. Almost everywhere continuity of $\hat{\gamma}(\xi, X, \Sigma)$ then follows from Assumption 6. $\square$

Lemma 18 *Under Assumptions 6 and 7, the conditional critical value*

$$c_{\alpha,C}\left(\hat{\gamma}(\xi, X, \Sigma), \mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}(\xi,X,\Sigma)}\right), \mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}(\xi,X,\Sigma)}\right), \Sigma\right)$$

is almost surely continuous at (ξ, X, Σ) *when computed with*

$$\tilde{S}_{n,\hat{\gamma}(\xi,X,\Sigma)} = \left(I - \frac{\Sigma\hat{\gamma}(\xi,X,\Sigma)\,\hat{\gamma}(\xi,X,\Sigma)'}{\hat{\gamma}(\xi,X,\Sigma)'\,\Sigma\hat{\gamma}(\xi,X,\Sigma)} \right)(\xi + \mu^*)$$

when $\xi \sim N(0,\Sigma)$ *and* μ^* *is as in Lemma 17.*

Proof of Lemma 18 For brevity of notation we abbreviate $\hat{\gamma}(\xi, X, \Sigma)$ by $\hat{\gamma}$. To prove the result, recall that

$$c_{\alpha,C}\left(\gamma, \mathcal{V}^{lo}(S_{n,\gamma}), \mathcal{V}^{up}(S_{n,\gamma}), \Sigma\right)$$

$$= \sqrt{\gamma'\Sigma\gamma}\cdot\Phi^{-1}\left((1-\alpha)\,\Phi\left(\frac{\mathcal{V}^{up}(S_{n,\gamma})}{\sqrt{\gamma'\Sigma\gamma}}\right) + \alpha\Phi\left(\frac{\mathcal{V}^{lo}(S_{n,\gamma})}{\sqrt{\gamma'\Sigma\gamma}}\right)\right).$$

That $\sqrt{\hat{\gamma}'\Sigma\hat{\gamma}}$ and $1/\sqrt{\hat{\gamma}'\Sigma\hat{\gamma}}$ are almost everywhere continuous follows from Assumption 6 and Lemma 17.

Note, next, that provided $\gamma'\Sigma\gamma$ is nonzero,

$$\Phi^{-1}\left((1-\alpha)\,\Phi\left(\frac{\mathcal{V}^{up}}{\sqrt{\gamma'\Sigma\gamma}}\right) + \alpha\Phi\left(\frac{\mathcal{V}^{lo}}{\sqrt{\gamma'\Sigma\gamma}}\right)\right)$$

is continuous in $\left(\mathcal{V}^{lo}, \mathcal{V}^{up}\right)$ on $(\mathbb{R}\cup\{-\infty,\infty\})^2$. This is obvious when at least one of $\left(\mathcal{V}^{lo}, \mathcal{V}^{up}\right)$ is finite. When $\mathcal{V}^{lo} \to -\infty$ and $\mathcal{V}^{up} \to \infty$,

$$\Phi^{-1}\left((1-\alpha)\,\Phi\left(\frac{\mathcal{V}^{up}}{\sqrt{\gamma'\Sigma\gamma}}\right) + \alpha\Phi\left(\frac{\mathcal{V}^{lo}}{\sqrt{\gamma'\Sigma\gamma}}\right)\right) \to \Phi^{-1}(1-\alpha) = \Phi^{-1}\left((1-\alpha)\,\Phi(\infty) + \alpha\Phi(-\infty)\right),$$

while when both $\mathcal{V}^{lo}, \mathcal{V}^{up} \to -\infty$,

$$\Phi^{-1}\left((1-\alpha)\,\Phi\left(\frac{\mathcal{V}^{up}}{\sqrt{\gamma'\Sigma\gamma}}\right) + \alpha\Phi\left(\frac{\mathcal{V}^{lo}}{\sqrt{\gamma'\Sigma\gamma}}\right)\right) \to -\infty = \Phi^{-1}\left((1-\alpha)\,\Phi(-\infty) + \alpha\Phi(-\infty)\right)$$

and when both $\mathcal{V}^{lo}, \mathcal{V}^{up} \to \infty$,

$$\Phi^{-1}\left((1-\alpha)\,\Phi\left(\frac{\mathcal{V}^{up}}{\sqrt{\gamma'\Sigma\gamma}}\right) + \alpha\Phi\left(\frac{\mathcal{V}^{lo}}{\sqrt{\gamma'\Sigma\gamma}}\right)\right) \to \infty = \Phi^{-1}\left((1-\alpha)\,\Phi(\infty) + \alpha\Phi(\infty)\right).$$

To complete the argument, it suffices to show that $\left(\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right), \mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right)\right)$ are continuous at almost every (ξ, X, Σ). To see that this is the case, recall that $\hat{\gamma}$ is almost everywhere continuous by Lemma 17. Note, next, that for a given $\hat{\gamma}$,

$$\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right) = \min\left\{c : c = \max_{\gamma \in V_F(X,\Sigma)} \gamma'\left(\tilde{S}_{n,\hat{\gamma}} + \frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)\right\}$$
$$= \min\left\{c : 0 = \max_{\gamma \in V_F(X,\Sigma)} \hat{a}_\gamma + \hat{b}_\gamma c\right\}$$

for $\hat{a}_\gamma = \gamma'\tilde{S}_{n,\hat{\gamma}}$ and $\hat{b}_\gamma = \frac{\gamma'\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}} - 1$. Note that $\hat{a}_{\hat{\gamma}} = \hat{b}_{\hat{\gamma}} = 0$, so $0 \leq \max_{\gamma \in V_F(X,\Sigma)} \hat{a}_\gamma + \hat{b}_\gamma c$ for all c. Moreover, for $c = \hat{\gamma}'Y_n$ the max is attained at $\hat{\gamma}$ by construction. Hence, the set over which we are minimizing is non-empty.

Intuitively, if we plot $\hat{a}_\gamma + \hat{b}_\gamma c$ as a function of c, each $\gamma \in V_F(X, \Sigma)$ defines a line, and we are interested in the set of values c such that zero lies on the upper envelope of this collection of lines. As this characterization suggests, to find the lower bound $\mathcal{V}^{lo}$ it suffices to limit attention to $\gamma \in V_F(X, \Sigma)$ with $\hat{b}_\gamma \leq 0$.

For given $\hat{\gamma}$, $\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ thus is equal to either $-\infty$ or the largest solution to

$$c = \gamma'\left(\tilde{S}_{n,\hat{\gamma}} + \frac{\Sigma\hat{\gamma}}{\hat{\gamma}'\Sigma\hat{\gamma}}c\right)$$

for γ in $V_F(X, \Sigma)$ with $\gamma'\Sigma\hat{\gamma} < \hat{\gamma}'\Sigma\hat{\gamma}$. Among γ with $\gamma'\Sigma\hat{\gamma} \neq \hat{\gamma}'\Sigma\hat{\gamma}$, this largest solution is well-defined and continuous. Matters are more delicate for γ with $\gamma'\Sigma\hat{\gamma} = \hat{\gamma}'\Sigma\hat{\gamma}$: in this case we may have discontinuities in Σ, but only if $\hat{\gamma}'\tilde{S}_{n,\hat{\gamma}} = \gamma'\tilde{S}_{n,\hat{\gamma}}$. However, $\hat{\gamma}'\tilde{S}_{n,\hat{\gamma}} = \gamma'\tilde{S}_{n,\hat{\gamma}}$ with positive probability if and only if $\gamma'\xi - \hat{\gamma}'\xi = 0$ with positive probability, which for $\gamma \neq \hat{\gamma}$ is ruled out by Assumption 7. Hence, we see that $\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ is almost everywhere continuous in the limit problem, as desired. The analogous argument applies for $\mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right)$, so overall we obtain that the critical value function is almost everywhere continuous, as we wanted to show. □

Lemma 19 *Under Assumptions 6-8, for μ^* such that $\mu_j^* \in [-\infty, 0]$ for all j,*

$$\max_{\gamma \in V_F(X,\Sigma)} \gamma'(\xi + \mu^*)$$

is almost everywhere continuous at (ξ, X, Σ) for $\xi \sim N(0, \Sigma)$ and (X, Σ) constant.

Proof of Lemma 19 To see that this is the case, note, first, that almost everywhere continuity of $\hat{\gamma}'\xi$ is immediate from Lemma 17. Thus, what remains is to show almost everywhere continuity of $\hat{\gamma}'\mu^* = \sum \hat{\gamma}_j \mu_j^*$. For those elements μ_j that are finite, almost everywhere continuity of $\hat{\gamma}_j \mu_j^*$ is again immediate from Lemma 17. To complete the proof we need only to show that $\hat{\gamma}_j \mu_j^*$ is almost everywhere continuous when $\mu_j^* = -\infty$. However, this follows from Assumption 8, which ensures that for every $\gamma(X, \Sigma) \in V_F(X, \Sigma)$, $\gamma_j(X, \Sigma)\mu_j^*$ is constant on a neighborhood of (X, Σ) when $\mu_j^* = -\infty$. □

Lemma 20 *Under Assumptions 6 and 7, for μ^* such that $\mu_j^* \in [-\infty, 0]$ for all j, if $\max_{\gamma \in V_F(X,\Sigma)} \gamma'\mu^*$ is finite then $T(\xi + \mu^*, X, \Sigma)$ as defined in (37) is finite with probability one and continuously distributed for $\xi \sim N(0, \Sigma)$ and (X, Σ) constant.*

Proof of Lemma 20 We first prove finiteness. In particular, note that since ξ is finite with probability one and $V_F(X, \Sigma)$ is a finite set, finiteness of $\max_{\gamma \in V_F(X,\Sigma)} \gamma'\mu^*$ implies finiteness of $\hat{\eta} = \max_{\gamma \in V_F(X,\Sigma)} \gamma'(\xi + \mu^*)$. Recall from the proof of Lemma 18 that the conditional critical value is infinite only if $V^{lo}(s) = V^{up}(s) = \infty$ or $V^{lo}(s) = V^{up}(s) = -\infty$. Since $\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right) \leq \hat{\eta} \leq \mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right)$, however, this implies that $\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ is not equal to ∞ and $\mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ is not equal to $-\infty$, and thus that $c_{\alpha,C}\left(\hat{\gamma}, \mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right), \mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right), \Sigma\right)$ is finite. Hence, $T(\xi + \mu^*, X, \Sigma)$ is finite.

To complete the proof, note that for fixed γ, $\gamma'Y_n$ is continuously distributed and independent of $\tilde{S}_{n,\gamma}$, and thus of $\left(\mathcal{V}^{lo}\left(\tilde{S}_{n,\gamma}\right), \mathcal{V}^{up}\left(\tilde{S}_{n,\gamma}\right)\right)$. In particular,

$$Pr\left\{\gamma'Y_n = \mathcal{V}^{lo}\left(\tilde{S}_{n,\gamma}\right)\right\} = 0.$$

Since $V_F(X, \Sigma)$ is finite, it follows that $Pr\left\{\hat{\eta} = \mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right)\right\} = 0$, and thus that $\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right) < \mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ with probability one. Recall that $\hat{\eta}$ lies between $\mathcal{V}^{lo}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ and $\mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ with probability one, and conditional on $\hat{\gamma}$ and $\tilde{S}_{n,\hat{\gamma}}$ follows a truncated normal distribution with untruncated variance $\hat{\gamma}'\Sigma\hat{\gamma} > 0$. Hence $T(\xi + \mu^*, X, \Sigma)$ is continuously distributed conditional on $\hat{\gamma}$ and $\tilde{S}_{n,\hat{\gamma}}$ for almost every $\hat{\gamma}$ and $\tilde{S}_{n,\hat{\gamma}}$. It follows that $T(\xi + \mu^*, X, \Sigma)$ is continuously distributed unconditionally as well. □

Proof of Proposition 9 If X_n is such that $\inf_\delta \max_j \{X_{n,j}\delta\} = -\infty$ then $\hat{\eta} = -\infty$ with probability one, and our tests never reject. For the remainder of the proof we thus assume that $\inf_\delta \max_j \{X_{n,j}\delta\} \neq -\infty$.

As in D. Andrews et al. (2019), note that uniform asymptotic size control is equivalent to asymptotic size control under all sequences of distributions $P_{D|Z,n} \in \mathcal{P}^0_{D|Z}$.

Towards contradiction, assume the test ϕ_C^* fails to control asymptotic size. Then there exists a sequence of distributions $P_{D|Z,n_m}$, a sequence of sample sizes n_m, and a value $\nu > 0$ such that

$$\liminf_{m\to\infty} Pr_{P_{D|Z,n_m}} \{\phi_C^* = 1\} > \alpha + \nu.$$

By the compactness of $\mathcal{S}$, for any such sequence, there exists a subsequence $n_{m,1}$ along which $\Sigma\left(P_{D|Z,n_{m,1}}\right) \to \Sigma \in \mathcal{S}$. For each n, since $P_{D|Z,n} \in \mathcal{P}^0_{D|Z}$ we know there exists a δ_n such that $\mu_n - X_n^*\delta_n \leq 0$. Thus, there exists a further subsequence $n_{m,2}$ along which $\mu_{n_{m,2},1} - X^*_{n_{m,2},1}\delta_{n_{m,2}} \to \mu_1^*$ for $\mu_1^* \in [-\infty, 0]$ for $\mu_{n_{m,2},1}$ the first component of $\mu_{n_{m,2}}$. Passing to further such subsequences, we see that there exists a subsequence $n_{m,k+1}$ such that $\Sigma\left(P_{D|Z,n_{m,k+1}}\right) \to \Sigma$ and

$$\mu_{n_{m,k}} - X^*_{n_{m,k}}\delta_{n_{m,k}} \to \mu^*$$

where $\mu_j^* \in [-\infty, 0]$ for all j. For simplicity of notation, for the remainder of the proof we assume that this property holds for the initial pair (m, n_m), so $\Sigma\left(P_{D|Z,n_m}\right) \to \Sigma$ and $\mu_{n_m} - X^*_{n_m}\delta_{n_m} \to \mu^*$.

Lemma 16 implies that

$$T\left(Y_{n_m}, X_{n_m}, \widehat{\Sigma}\right) = T\left(Y_{n_m} - X^*_{n_m}\delta_{n_m}, X^*_{n_m}, \widehat{\Sigma}\right),$$

while Assumptions 2-5 imply that

$$\left(Y_{n_m} - X^*_{n_m}\delta_{n_m}, X^*_{n_m}, \widehat{\Sigma}\right) \to_d (\xi + \mu^*, X, \Sigma)$$

for $\xi \sim N(0, \Sigma)$. Together, Lemmas 18 and 19 imply that $T(\xi + \mu^*, X, \Sigma)$ is almost everywhere continuous with respect to the distribution of $(\xi + \mu^*, X, \Sigma)$, and thus, by the continuous mapping theorem, that

$$T\left(Y_{n_m}, X^*_{n_m}, \widehat{\Sigma}\right) \to_d T(\xi + \mu^*, X, \Sigma).$$

If $\max_{\gamma\in V_F(X,\Sigma)} \gamma'\mu^* = -\infty$, then $\hat{\eta} \to -\infty$. Hence, since the modified conditional test never rejects for $\hat{\eta} < -C$, this implies that $\lim_{m\to\infty} Pr\{\phi_C^* = 1\} = 0$, contradicting our assumption that size control fails. Thus, for the remainder of the argument we assume that $\max_{\gamma\in V_F(X,\Sigma)} \gamma'\mu^*$ is finite.[38] Under this assumption, Lemma 20 shows

[38]Recall that $\gamma \in V_F(X, \Sigma)$ implies that $\gamma \geq 0$, so we cannot have $\gamma'\mu^* = \infty$.

that $T(\xi+\mu^*,X,\Sigma)$ is continuously distributed. This implies that

$$\lim_{m\to\infty} Pr\left\{T\left(Y_{n_m},X^*_{n_m},\widehat{\Sigma}\right)>0\right\}\to Pr\left\{T(\xi+\mu^*,X,\Sigma)>0\right\},$$

and thus that

$$Pr\left\{T(\xi+\mu^*,X,\Sigma)>0\right\}\geq\alpha+\nu.$$

However, provided $\max_{\gamma\in V_F(X,\Sigma)}\gamma'\mu^*$ is finite, Proposition 6 shows that for $\mu^*\leq 0$

$$Pr\left\{T(\xi+\mu^*,X,\Sigma)>0\right\}\leq\alpha,$$

so we have reached a contradiction. □

Proof of Corollary 2 Note that the hybrid test is of nearly the same form as the conditional test, except that it uses the $\mathcal{V}^{up,H}(S_{n,\hat{\gamma}})=\min\left\{\mathcal{V}^{up}(S_{n,\hat{\gamma}}),c_{\kappa,LF}\left(X_n,\widehat{\Sigma}\right)\right\}$ instead of $\mathcal{V}^{up}(S_{n,\hat{\gamma}})$, and considers a different quantile of the conditional distribution. Building on the proof of Proposition 9, to prove asymptotic validity of ϕ^*_H it thus suffices to show that $\mathcal{V}^{up,H}(S_{n,\hat{\gamma}})$ is almost-everywhere continuous when computed using the set of limit distributions considered in that proof. However, we have already shown that $\mathcal{V}^{up}(S_{n,\hat{\gamma}})$ satisfies this property, so we need only show that $c_{\kappa,LF}(X,\Sigma)$ is continuous in (X,Σ).

Recall, however, that $c_{\kappa,LF}(X,\Sigma)$ is the $1-\kappa$ quantile of $\max_{\gamma\in V_F(X,\Sigma)}\gamma'\xi$ for $\xi\sim N(0,\Sigma)$. Lemma 19 shows that under our assumptions this max is almost everywhere continuous in (ξ,X,Σ), from which continuity of the $1-\kappa$ quantile follows immediately.

To complete the argument, recall that the proof of Lemma 18 shows that $\mathcal{V}^{up}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ is almost everywhere continuous in the limit problem, which together with the argument above shows that $\mathcal{V}^{up,H}\left(\tilde{S}_{n,\hat{\gamma}}\right)$ is almost everywhere continuous. Note that the hybrid test is unchanged if, rather than defining $c_{\frac{\alpha-\kappa}{1-\kappa},C}\left(\hat{\gamma},\mathcal{V}^{lo}(S_{n,\hat{\gamma}}),\mathcal{V}^{up,H}(S_{n,\hat{\gamma}}),\Sigma\right)$ to be $-\infty$ when $\mathcal{V}^{lo}(S_{n,\hat{\gamma}})>\mathcal{V}^{up,H}(S_{n,\hat{\gamma}})$, we instead define it to be $\mathcal{V}^{up,H}(S_{n,\hat{\gamma}})$.[39] With this modification, however, we see that $c_{\frac{\alpha-\kappa}{1-\kappa},C}\left(\hat{\gamma},\mathcal{V}^{lo}(S_{n,\hat{\gamma}}),\mathcal{V}^{up,H}(S_{n,\hat{\gamma}}),\Sigma\right)$ is almost-everywhere continuous in the limit problem by the same argument as in the proof of Lemma 18. Hence,

$$\hat{\eta}-c_{\frac{\alpha-\kappa}{1-\kappa},C}\left(\hat{\gamma},\mathcal{V}^{lo}(S_{n,\hat{\gamma}}),\mathcal{V}^{up,H}(S_{n,\hat{\gamma}}),\Sigma\right)$$

[39]Since $\mathcal{V}^{lo}(S_{n,\hat{\gamma}})>\mathcal{V}^{up,H}(S_{n,\hat{\gamma}})$ implies $\hat{\eta}>\mathcal{V}^{up,H}(S_{n,\hat{\gamma}})$.

is almost-everywhere continuous in the limit problem by the same arguments as in the proof of Proposition 9.

All that remains to show is that this quantity is continuously distributed. As argued in the proof of Lemma 20, however, if $\hat{\eta}$ is finite it is continuously distributed conditional on $\hat{\gamma}$ and $S_{n,\hat{\gamma}}$ for almost every $\hat{\gamma}$ and $S_{n,\hat{\gamma}}$. This implies that $\hat{\eta}$ is continuously distributed conditional on almost every realization of $c_{\frac{\alpha-\kappa}{1-\kappa},C}\left(\hat{\gamma}, \mathcal{V}^{lo}(S_{n,\hat{\gamma}}), \mathcal{V}^{up,H}(S_{n,\hat{\gamma}}), \Sigma\right)$, and so proves continuity. □

E.3 Proof of Variance Consistency

We first prove two auxiliary lemmas, which we then use to prove Proposition 10.

Lemma 21 *Under Assumption 9,*

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

uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$.

Proof of Lemma 21 Note that we can write

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_{\ell_Z(i)}Y'_{\ell_Z(i)} - V_{P_{D|Z}}\left(Z_i\right)\right) =$$

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

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 9(4),

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

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

for K the upper bound on the Lipschitz constant. Note, next, that since $\mathcal{Z}$ is compact by Assumption 9(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 $\frac{1}{n}\sum_{i=1}^{n}\left(V_{P_{D|Z}}\left(Z_{\ell_Z(i)}\right) - V_{P_{D|Z}}\left(Z_i\right)\right) \to 0$ uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$.

We next show that

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

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

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_{\ell_Z(i)}Y'_{\ell_Z(i)} - V_{P_{D|Z}}\left(Z_{\ell_Z(i)}\right)\right)_{jk} \mid \left\{Z_i\right\}_{i=1}^{\infty}\right)$$

$$\leq \mathcal{K}\left(\dim\left(Z_i\right)\right) Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_iY_i'\right)_{jk} \mid \left\{Z_i\right\}_{i=1}^{\infty}\right)$$

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

By Assumption 9(2) and Chebyshev's inequality, however, this implies that

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

uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$, which completes the proof. $\square$

Lemma 22 *Under Assumption 9,*

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

uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$.

Proof of Lemma 22 Note that we can write

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

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

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

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

By independence,

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

while the variance of the jkth element is

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

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

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

Assumption 9(5) thus implies that for some constant C,

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

which, together with Assumption 9(2) and the finiteness of the "kissing number"

$B\left(\dim\left(Z_i\right)\right)$ (see the proof of Lemma 21 above) implies that

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

and thus by Chebyshev's inequality that

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

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

To complete the proof, we need only show that

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

converges to zero uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$. Note, however, that by the triangle inequality and Assumption 9(3),

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

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

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

$$\leq \frac{K}{n}\sum_{i=1}^{n}\left\|\mu_{P_{D|Z}}\left(Z_i\right)\right\| \cdot \left\|Z_{\ell_Z(i)} - Z_i\right\| \leq \frac{KC}{n}\sum_{i=1}^{n}\left\|Z_{\ell_Z(i)} - Z_i\right\| \tag{38}$$

for K a Lipschitz constant and C a constant. As above, since $\mathcal{Z}$ is compact by Assumption 9(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 (38) converges to zero uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$. □

Proof of Proposition 10 Following proof of Lemma A.3 in Abadie et al. (2014), note that

$$\widehat{\Sigma} = \frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - Y_{\ell_Z(i)}\right)\left(Y_i - Y_{\ell_Z(i)}\right)'$$

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

Assumption 9(2) together with Chebyshev's inequality implies that

$$\frac{1}{2n}\sum_{i=1}^{n}\left(Y_i Y_i' - V_{P_{D|Z}}\left(Z_i\right)\right) \to_p 0$$

uniformly over $P_{D|Z} \in \mathcal{P}_{D|Z}$.

Since

$$Var\left(Y_i|Z_i\right) = V_{P_{D|Z}}\left(Z_i\right) - \mu_{P_{D|Z}}\left(Z_i\right)\mu_{P_{D|Z}}\left(Z_i\right)',$$

however, we see that

$$\frac{1}{n}\sum_i Var_{P_{D|Z}}\left(Y_i|Z_i\right) = \frac{1}{n}\sum_i V_{P_{D|Z}}\left(Z_i\right) - \frac{1}{n}\sum_i \mu_{P_{D|Z}}\left(Z_i\right)\mu_{P_{D|Z}}\left(Z_i\right)'.$$

Thus, to prove that

$$\widehat{\Sigma} - \frac{1}{n}\sum Var_{P_{D|Z}}\left(Y_i|Z_i\right) \to_p 0,$$

it suffices to prove that

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

and

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

where the first statement follows from Lemma 21 and the second from Lemma 22. Since

$$\frac{1}{n}\sum Var_{P_{D|Z}}\left(Y_i|Z_i\right) - \Sigma\left(P_{D|Z}\right) \to 0$$

uniformly over $P_{D|Z} \in P_{D|Z}$ by Assumption 2, however, the result follows by the triangle inequality. □

F Performance Without Nuisance Parameters

This appendix discusses the simulated performance of the procedures we consider in the simplified setting discussed in Section 5.1 of the paper. In particular, we assume that there are no nuisance parameters (and thus no matrix X_n), and that $Y_n \sim N(\mu_n, I)$, and want to test $H_0 : \mu_n \leq 0$. We simulate the power of the least favorable, conditional, and hybrid tests. A number of other tests have been studied in the setting without nuisance parameters, and for comparison we consider the test of Romano et al. (2014*a*) (henceforth RSW).[40] RSW include a simulation comparison of their test to that of D. Andrews & Barwick (2012), while Cox & Shi (2019) compare their test to both RSW and D. Andrews & Barwick (2012).

As noted in Section 5.1 of the paper the conditional test in this setting compares $\hat{\eta} = \max_j Y_{n,j}$ to a truncated normal critical value, truncated below at the second largest element of Y_n. The hybrid critical value considers the same test statistic but compares it to a truncated normal critical value which adds an upper truncation point equal to the level κ least-favorable critical value.

For our simulations, we consider either two, ten, or fifty moments, $k \in \{2, 10, 50\}$. When $k \in \{10, 50\}$ the parameter space is very large and we are unable to fully depict the power function. Instead, we focus on how the power varies in the first two elements of μ_n, while the remaining elements are held at a fixed value. In particular, we consider $(\mu_{n,1}, \mu_{n,2}) \in [-10, 10]^2$, while for $j > 2$ we set $\mu_{n,j} = \mu^*$ for a fixed value μ^*. Contours of the resulting power functions, based on 1000 simulations, are plotted in Figures 3-7. For visibility, we also include plots of the difference in power functions between the conditional and hybrid tests and the RSW test.

These simulations highlight a number of features discussed in the main text. Comparing the least favorable and conditional tests, we see that when the largest moment is substantially larger than the second largest, the conditional test has better power than does the least favorable test, particularly when the total number of moments is large. By contrast, when the two largest moments are approximately the same size the conditional test has poor power relative to the least favorable test. The hybrid test substantially improves on the conditional test in this case, while largely retaining the good performance of the conditional test in cases with many slack moments.

[40] Since this section considers a normal model with known variance, we consider a version of RSW based on the normal distribution, discussed in Supplement Section S.1.2 of that paper, rather than the bootstrap version they discuss in the main text.

Comparing the hybrid test and the test of RSW, we see that neither test dominates the other. The test of RSW has better power close to the diagonal, while the hybrid test has better power somewhat further from the diagonal.

From these results, we see that while the conditional test offers a substantial improvement over the least favorable test in some cases, the performance deterioration when the largest moment is not well-separated is a real problem. The hybrid testing approach largely corrects this weakness, and attains performance roughly comparable to the RSW approach, albeit with somewhat lower power close to the diagonal. Unlike the approach of RSW, however, the hybrid approach extends easily to settings with nuisance parameters, as we consider in the main text.

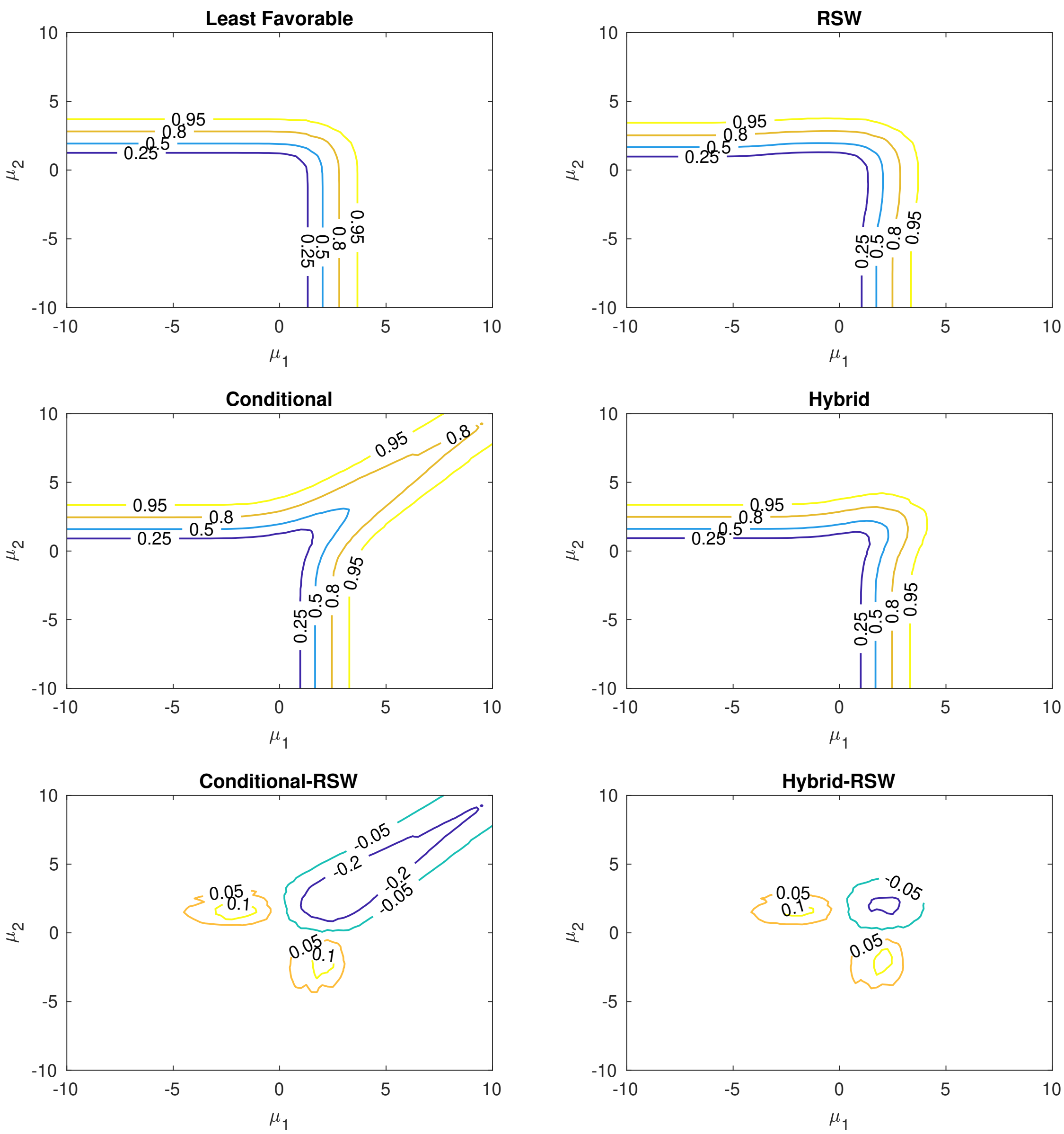


Figure 3: Power of tests with $k = 2$.

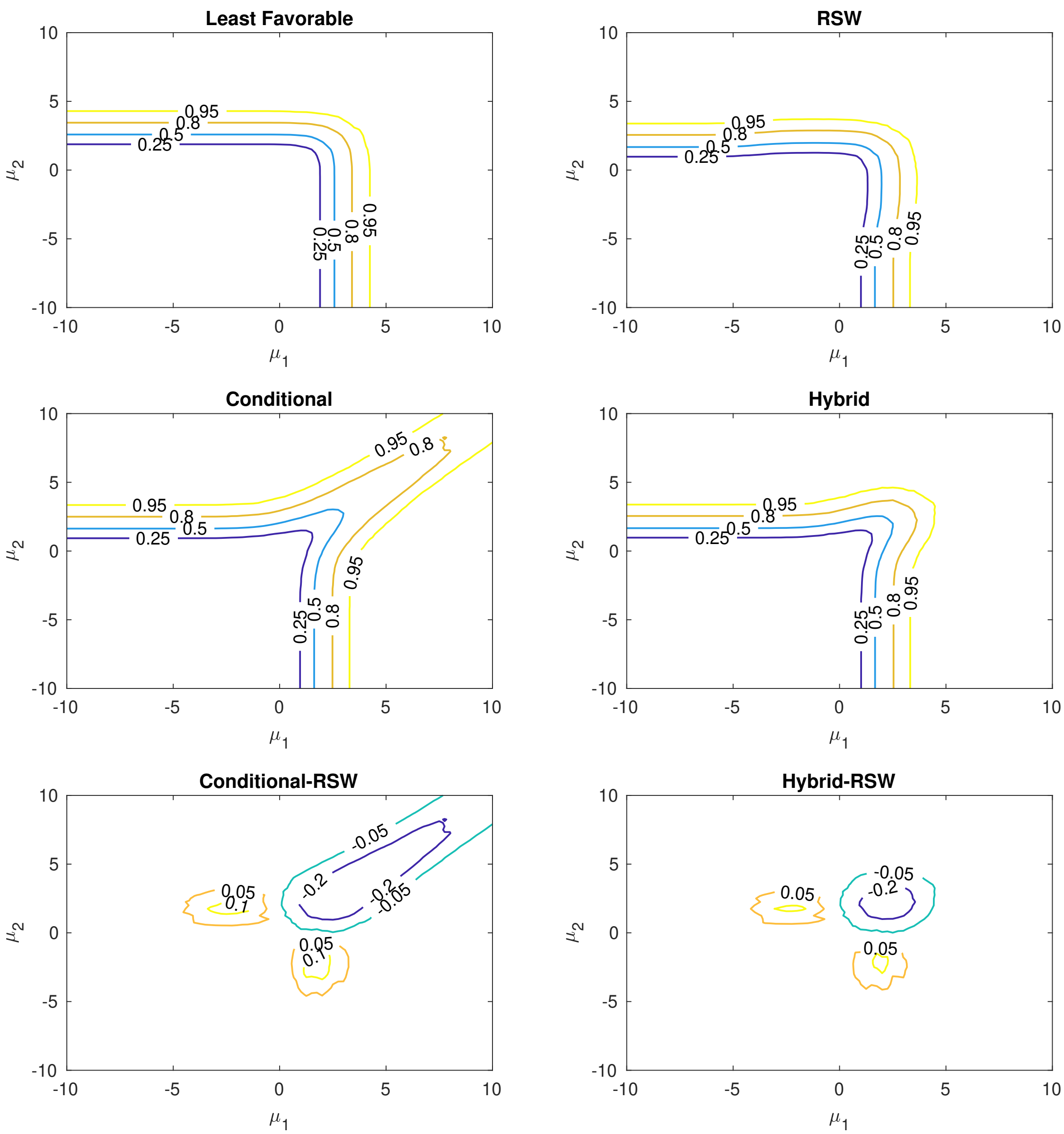


Figure 4: Power of tests with $k = 10$, $\mu^* = 0$.

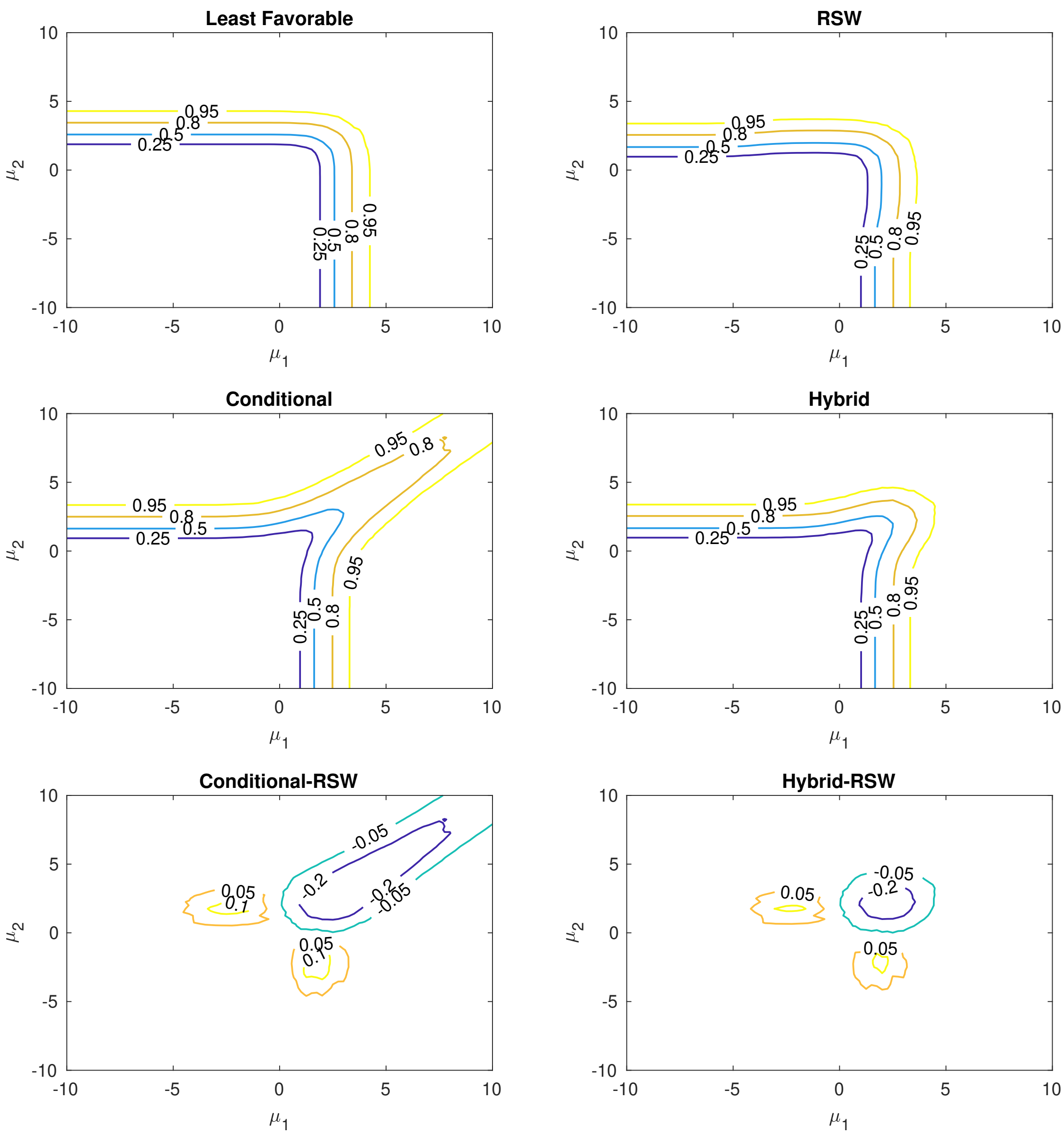


Figure 5: Power of tests with $k = 10$, $\mu^* = -10$.

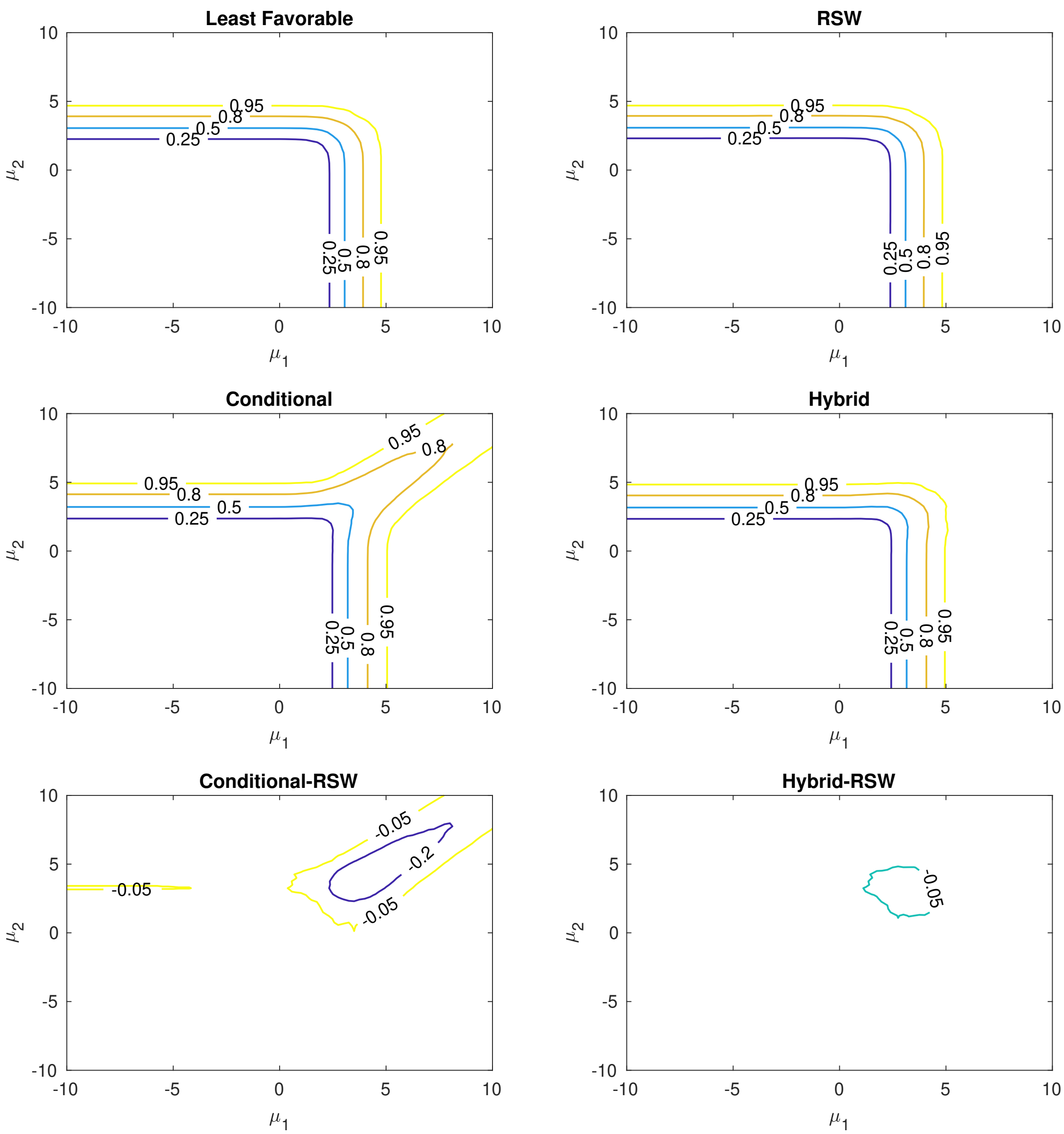


Figure 6: Power of tests with $k = 50$, $\mu^* = 0$.

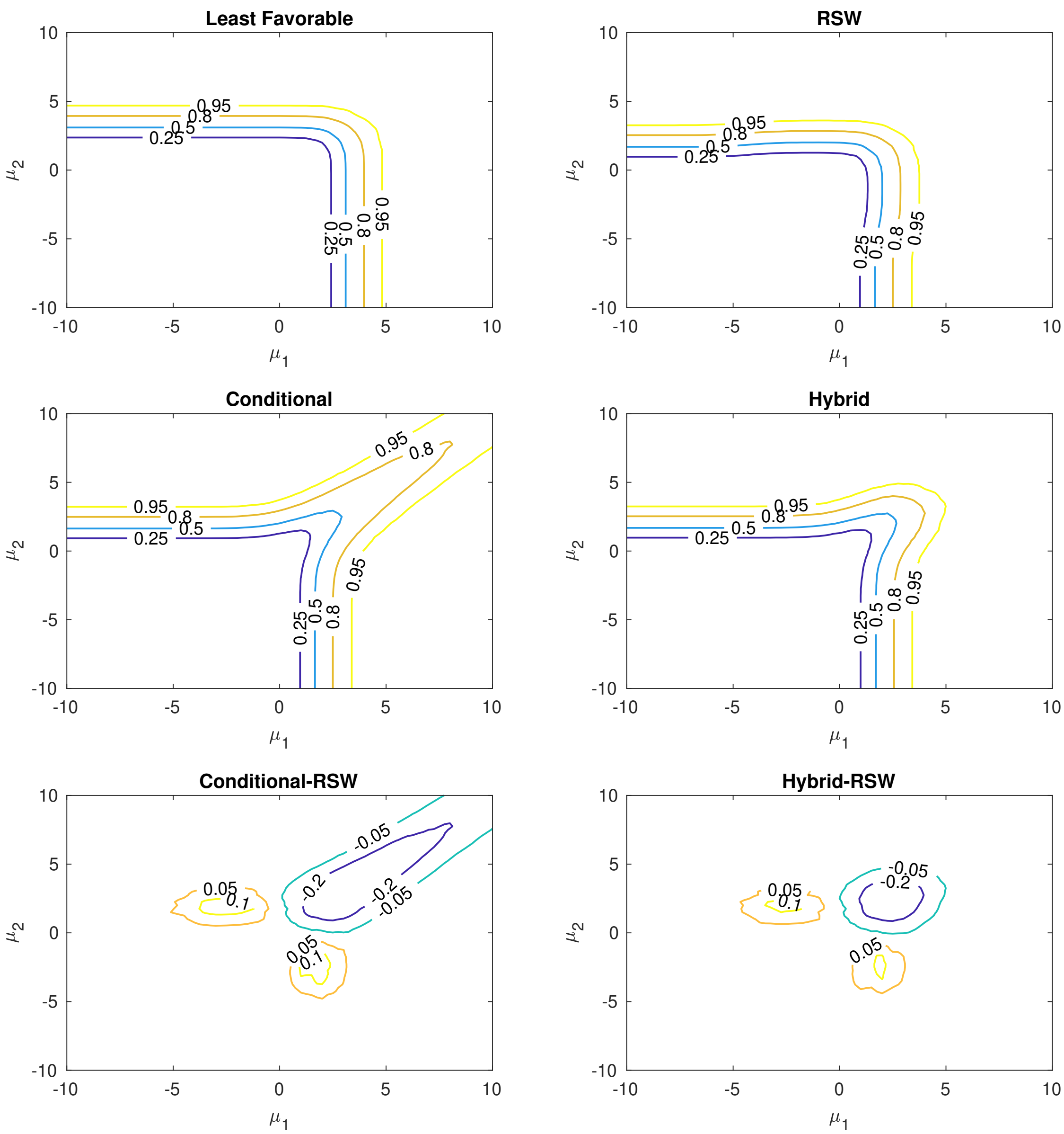


Figure 7: Power of tests with $k = 50$, $\mu^* = -10$.

G Simulation Appendix

G.1 The Simulated Model

G.1.1 Competition and Firm Decisions

We consider competition between F firms, who in each period decide which set of products to offer. As in Wollmann, the products (indexed by j) differ in their gross weight rating g_j, which can take on G possible values. The fixed cost of offering a product in the current period depends on whether it was offered in the previous period: if it was not previously marketed the costs are $\theta_c + \theta_g g_j$. If the product was previously marketed, the fixed costs scale down by a multiplicative factor β, so the cost of entering a previously marketed product is $\beta(\theta_c + \theta_g g_j)$.

Firm f estimates that marketing product j in period t will earn variable profits π^*_{jft}, and chooses to enter the product if and only if the expected profits exceed the fixed costs. Thus, if a firm offered product j in period $t-1$, then the firm chooses to offer j in period t iff

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

If the firm didn't offer the product j in period $t-1$, then it chooses to add product j iff

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

G.1.2 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 offered 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 offered 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 constuct 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.2 Calibration

G.2.1 Exogenous Parameter Values

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 the November 2018 version of Wollman (2018).[41] 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 of Wollman (2018), which gives the average GWR for different buyer types.

G.2.2 Simulating Data for Calibration

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 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.

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

G.2.3 Calibrating the Remaining Parameters

The parameter values to calibrate are $\{\mu_f\}, \sigma_\eta, \sigma_\epsilon, \sigma_\nu$.

Intuition for Calibration. 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 relatively 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 markups are fixed at 35%.

Technical Details for Calibration.

The calibration proceeds as follows:

1) We first calibrate $(\sigma_\eta, \sigma_\epsilon)$ 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 $(\sigma_\eta, \sigma_\epsilon)$, 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 Wollman (2018). 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 $(\sigma_\eta, \sigma_\epsilon)$ to match a measure of the variability of the number of products offered in Wollmann's data. In particular, Table I in Wollman (2018) 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 $\sigma_\eta = \sigma_\epsilon = 30$ because this 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. In particular, 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.[42] 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 3 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.2.4 Calibrated Parameters

Tables 3 and 4 show the calibrated values for the μ_f and variance parameters, respectively.

Table 3: 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

[42] 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, $Var(\pi_{jft}) = m^2p^2Var(Q_{jft})$ 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 4: Calibrated Variance Parameters

Parameter	Value
σ_η	30.00
σ_ϵ	30.00
σ_ν	57.96

G.3 Details of simulations in Section 7

G.3.1 Drawing from Independent Markets

Wollmann's original model 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. 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 Wollman (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 periods as burnout. For each simulated dataset, we then randomly subsample 500 periods from this chain.

G.3.2 Parameter Grids and Monte Carlo Draws

For all of our simulations, we conduct inference by discretizing the parameter space for the parameter of interest. For δ_g and the cost of the mean-weight truck, we use 1,001 gridpoints; for β, we use 100 gridpoints. The bounds for the grid depend on the specification, and are equal to the upper and lower bound of the x-axis shown in the rejection probabilty figures (Figures 1 and 2 and Appendix Figure 6).

To calculate the LFP critical values, we draw a fixed matrix Ξ of standard normal draws of size $M \times 10,000$, 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 $M \times 1000$.

G.3.3 Handling of Numerical Precision Errors

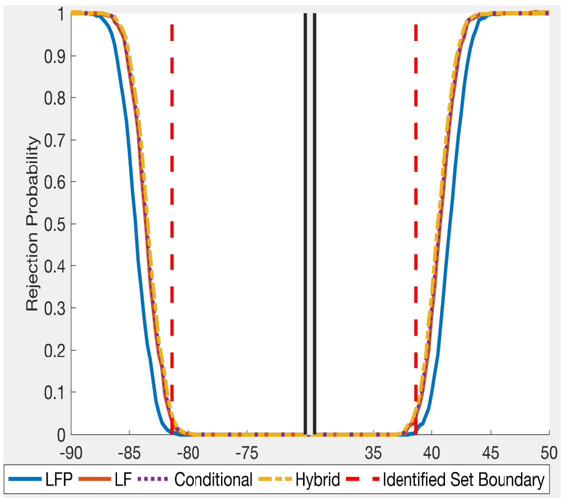
In simulating the draws for the LF approach, in certain very rare cases we encountered computational issues in which the linear program for one of the draws did not converge. In these cases, we treat the draw as if it were infinity, which pushes the estimated critical value slightly higher, and makes our estimate of the rejection probability slightly conservative. However, in all specifications this happens in no more than 0.01% of cases (of approximately 50 million simulations), and is thus unlikely to have any substantial impact on our results.

G.3.4 Additional Simulation Results

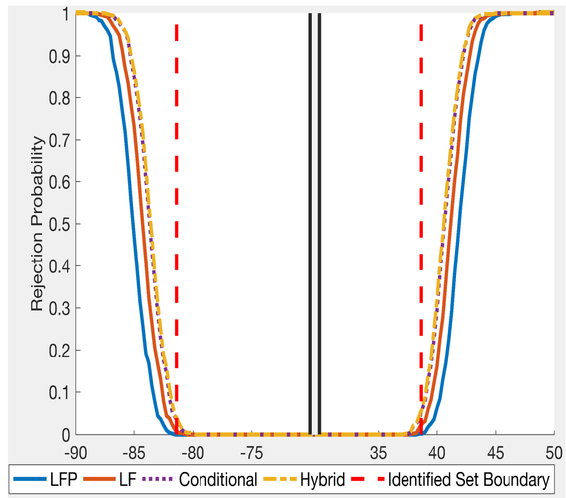
This appendix reports additional simulation results to complement the results reported in Section 7 of the main text. In particular, Figure 8 reports rejection probabilities for tests of hypotheses on δ_g, while Tables 5-7 report the 5th and 95th percentiles of the excess length distribution for the confidence sets we study.

Figure 8: Rejection probabilities for 5% tests of δ_g

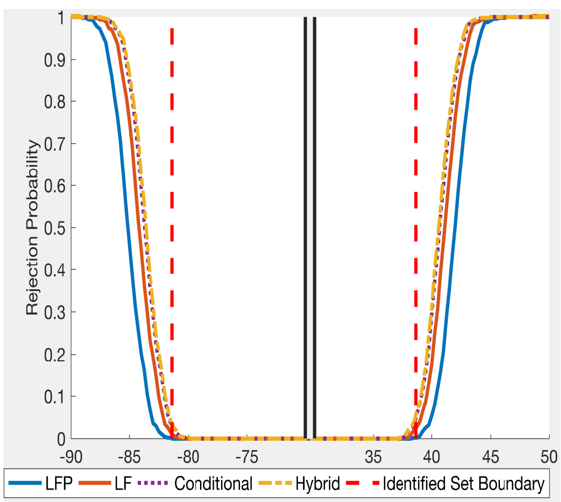
(a) 2 Parameters, 6 Moments



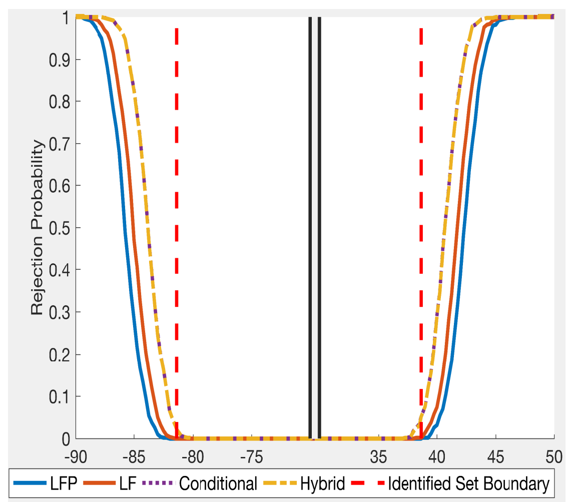
(b) 2 Parameters, 14 Moments



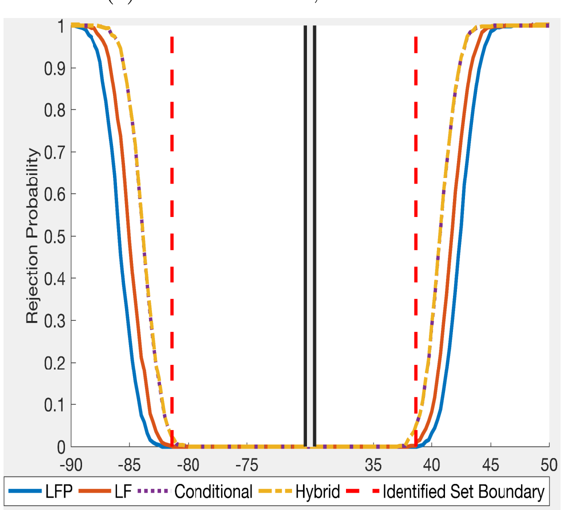
(c) 4 Parameters, 14 Moments



(d) 4 Parameters, 38 Moments



(e) 10 Parameters, 38 Moments



(f) 10 Parameters, 110 Moments

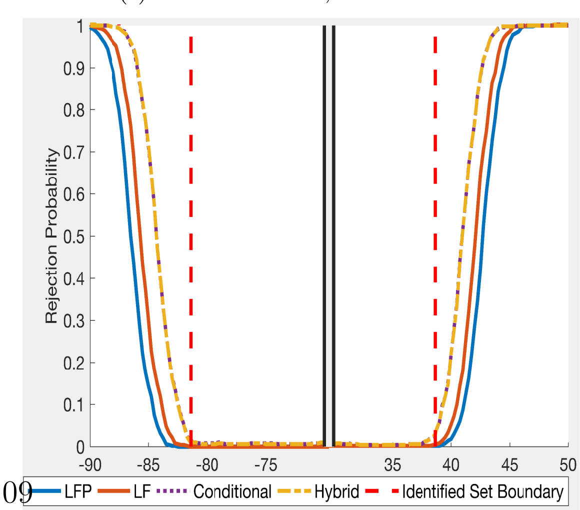


Table 5: Mean and Select Percentiles of Excess Length for Cost of Mean-Weight Truck

		Mean of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	5.38	4.05	4.02	3.89
2	14	12.79	10.61	10.74	8.72
4	14	7.59	5.94	4.23	4.25
4	38	18.79	16.03	14.33	11.13
10	38	12.76	10.24	4.87	4.79
10	110	25.55	22.27	17.74	14.03

		5th Percentile of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	3.38	2.06	2.12	1.80
2	14	7.60	5.44	4.97	3.34
4	14	5.51	3.88	2.09	2.09
4	38	15.02	11.67	7.53	3.96
10	38	10.34	7.82	2.43	2.43
10	110	22.45	18.89	11.58	7.19

		Median of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	5.32	3.99	4.07	3.75
2	14	12.75	10.48	10.49	8.54
4	14	7.56	5.91	4.07	4.37
4	38	19.08	16.33	14.68	11.60
10	38	12.70	10.20	4.71	4.71
10	110	25.61	22.36	17.91	14.34

		95th Percentile of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	7.45	6.16	6.02	6.02
2	14	17.99	15.99	17.97	15.85
4	14	9.78	8.07	6.32	6.48
4	38	22.07	19.77	20.05	17.61
10	38	15.22	12.63	6.98	7.31
10	110	28.43	25.58	23.11	19.70

Table 6: Mean and Select Percentiles of Excess Length for δ_g

		Mean of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	5.99	4.29	4.17	3.91
2	14	6.92	5.40	4.28	4.11
4	14	7.02	5.21	4.33	4.13
4	38	8.01	6.73	4.45	4.46
10	38	8.16	6.63	4.50	4.47
10	110	9.08	7.63	4.81	4.83

		5th Percentile of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	2.70	1.04	0.93	0.68
2	14	3.53	2.04	0.93	0.77
4	14	3.62	1.83	0.93	0.68
4	38	4.58	3.38	1.07	1.15
10	38	4.73	3.22	1.02	0.93
10	110	5.56	4.13	1.40	1.43

		Median of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	6.02	4.28	4.18	3.93
2	14	6.91	5.40	4.43	4.18
4	14	7	5.19	4.43	4.18
4	38	7.97	6.68	4.43	4.43
10	38	8.10	6.58	4.43	4.43
10	110	9.11	7.69	5.18	5.18

		95th Percentile of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
2	6	9.17	7.47	7.43	7.06
2	14	10.21	8.68	7.56	7.43
4	14	10.23	8.51	7.68	7.43
4	38	11.32	10	7.68	7.68
10	38	11.44	9.87	7.68	7.68
10	110	12.55	11.11	8.43	8.43

Table 7: Mean and Select Percentiles of Excess Length for β

		Mean of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
3	6	84.17^+	69.84^+	59.38^+	55.68^+
3	14	0.74	0.58	0.48^+	0.41
5	14	13.51^+	10.45^+	10.3^+	7.87^+
5	38	0.85	0.64	1.33^+	0.48
11	38	1.49	1.08	2.01^+	0.83
11	110	0.89	0.65	2.78^+	0.5

		5th Percentile of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
3	6	15.15	10.1	6.31	6.31
3	14	0.35	0.25	0.15	0.1
5	14	3.54	2.3	1.06	1.06
5	38	0.5	0.35	0.25	0.05
11	38	0.81	0.5	0.2	0.2
11	110	0.56	0.35	0.66	0.03

		Median of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
3	6	118.69	61.87	41.67	36.62
3	14	0.76	0.56	0.45	0.35
5	14	10.25	7.78	6.01	5.3
5	38	0.86	0.66	0.96	0.45
11	38	1.41	1.01	1.01	0.81
11	110	0.86	0.66	2.57	0.56

		95th Percentile of Excess Length Distribution			
#Parameters	#Moments	LFP	LF	Conditional	Hybrid
3	6	123.11^+	123.11^+	123.11^+	123.11^+
3	14	1.16	0.96	0.96	0.86
5	14	31.29^+	31.29^+	31.29^+	27.4
5	38	1.16	0.96	4.21^+	0.96
11	38	2.42	1.92	7.01^+	1.72
11	110	1.26	0.96	5.7^+	0.88

Note: For certain specifications and simulation draws, the rejection probability did not reach 1 at the edge of our grid for λ. In these cases, we truncate the excess length at the edge of the grid. A + denotes statistics that are affected by this truncation.

H Bisection Algorithm for Computing V^{lo} and V^{up}

When the conditions in step 2 in Section 6.3 do not hold, V^{lo} and V^{up} must be calculated by finding the minimum and maximum of the set

$$C = \left\{ c : \begin{array}{c} c = \max_{\tilde{\gamma}} \tilde{\gamma}' \left(s + \frac{\Sigma\gamma}{\gamma'\Sigma\gamma} c \right) \\ \text{subject to } \tilde{\gamma} \geq 0, \; W_n' \tilde{\gamma} = e_1 \end{array} \right\}$$

Recall that the set C is convex, and its endpoints, if they are finite, can therefore be calculated via bisection. We thus recommend the following procedure for calculating V^{up}. Begin by specifying a large value M, such that if $V^{up} > M$, for practical purposes we can consider $V^{up} = \infty$.[43] Then implement Algorithm 1 described in the box below.

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

Algorithm 1 Bisection Method for Calculating V^{up}

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

where we define the functions:

function LPVALUE(c)
  return

$$\max_{\tilde{\gamma}} \tilde{\gamma}' \left(s + \frac{\Sigma\gamma}{\gamma'\Sigma\gamma} c \right)$$
$$\text{subject to } \tilde{\gamma} \geq 0, \; W_n' \tilde{\gamma} = e_1$$

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

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.