Supplement to "Sensitivity analysis using approximate moment condition models"

(*Quantitative Economics*, Vol. 12, No. 1, January 2021, 77–108)

Timothy B. Armstrong
Department of Economics, Yale University

Michal Kolesár
Department of Economics, Princeton University

Appendix A contains the details of calculations for particular sets $\mathcal{C}$ discussed in Section 4.3.4 and Remark 3.1. Appendix B develops a specification test of the null $H_0: c \in \mathcal{C}$. Appendix C contains the asymptotic coverage and efficiency results discussed in Section 4. Appendix D discusses extensions allowing for global misspecification. Appendix E gives a construction of a submodel satisfying Assumption C.1, verifies the conditions in Appendix C in the misspecified IV model, and collects auxiliary results used in Appendix C.

Appendix A: Details of calculations

This Appendix contains the details of calculations for particular sets $\mathcal{C}$ discussed in Section 4.3.4 and Remark 3.1.

A.1 *Cressie–Read divergences*

Consider the problem in equation (15) under constraints of the form $\{c: c\Sigma^{-1}c \leq M^2\}$. The Lagrangian for this problem can be written as

$$2H\theta + \lambda_1\big(\delta^2/4 - (c - \Gamma\theta)'\Sigma^{-1}(c - \Gamma\theta)\big) + \lambda_2\big(M^2 - c'\Sigma^{-1}c\big)$$

(we multiply the objective function by 2 so that its optimized value equals $\omega(\delta)$). The first-order conditions imply that at optimum, $c = \frac{\lambda_1}{\lambda_1+\lambda_2}\Gamma\theta$, and $\theta = \frac{\lambda_1+\lambda_2}{\lambda_1\lambda_2}(\Gamma'\Sigma^{-1}\Gamma)^{-1}H'$. Plugging these expressions into the constraints yields $M^2 = H(\Gamma'\Sigma^{-1}\Gamma)^{-1}H'/\lambda_2^2$ and $\delta^2/4 = H(\Gamma'\Sigma^{-1}\Gamma)^{-1}H'/\lambda_1^2$. Since $H(\Gamma'\Sigma^{-1}\Gamma)^{-1}H' = k_{\text{LS},0}'\Sigma k_{\text{LS},0}$, solving for λ_1 and λ_2, and plugging into the expression for θ yields

$$\theta = \frac{\delta/2 + M}{\sqrt{k_{\text{LS},0}'\Sigma k_{\text{LS},0}}} \cdot \big(\Gamma'\Sigma^{-1}\Gamma\big)^{-1}H'.$$

Timothy B. Armstrong: timothy.armstrong@yale.edu
Michal Kolesár: mkolesar@princeton.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE1609

Thus, $\omega(\delta) = 2H\theta = (\delta + 2M)\sqrt{k_{\text{LS},0}'\Sigma k_{\text{LS},0}}$. With this form of ω, the bound in equation (18) becomes

$$\kappa_*(H, \Gamma, \Sigma, \mathcal{C}) = \frac{(1-\alpha)(z_{1-\alpha} + M) + \phi(z_{1-\alpha})}{\text{cv}_\alpha(M)}.$$

This efficiency equals at least $\min\{\kappa_{*,\alpha}^L, 1-\alpha\}$, where $\kappa_{*,\alpha}^L = ((1-\alpha)z_{1-\alpha} + \phi(z_{1-\alpha}))/z_{1-\alpha/2}$ denotes the efficiency given in equation (19) when $\mathcal{C}$ is a linear subspace. To show this, observe that $\text{cv}_\alpha'(M) \leq 1$ for all $M \geq 0$. Therefore, the derivative of

$$(1-\alpha)(z_{1-\alpha} + M) + \phi(z_{1-\alpha}) - \min\bigl\{1-\alpha, \kappa_{*,\alpha}^L\bigr\}\,\text{cv}_\alpha(M)$$

with respect to M, given by $1-\alpha-\min\{1-\alpha, \kappa_{*,\alpha}^L\}\,\text{cv}_\alpha'(M)$, is always nonnegative. Since the expression in the above display equals $(\kappa_{*,\alpha}^L - \min\{\kappa_{*,\alpha}^L, 1-\alpha\})z_{1-\alpha/2} \geq 0$ at $M = 0$, it follows that it is always nonnegative. Rearranging it then yields $\kappa_*(H, \Gamma, \Sigma, \mathcal{C}) \geq \min\{\kappa_{*,\alpha}^L, 1-\alpha\}$. Furthermore, it follows from equation (S29) below that the efficiency of one-sided CIs at $c = 0$ is given by $\kappa_*^{\text{OCI},\beta} = 1$.

A.2 ℓ_p *bounds*

We now consider the form of the optimal sensitivity when the set $\mathcal{C} = \mathcal{C}(M)$ takes the form in equation (12), and $\|\cdot\|$ corresponds to an ℓ_p norm, as discussed in Remark 3.1. First, we explain the connection with penalized estimation. Since $c = B\gamma$, one can write the approximately linear model in equation (13) as

$$Y = -\Gamma\theta + B\gamma + \Sigma^{1/2}\varepsilon,$$

which one can think of as a regression model with correlated errors, design matrix $(-\Gamma, B)$, and coefficient vector $(\theta', \gamma')'$. With this interpretation, it is clear that if the number of regressors $d_\theta + d_\gamma$ is greater than the number of observations d_g, the constraint on the norm of γ is necessary to make the model informative. Using the observation from Remark 3.1 that $\overline{\text{bias}}_{\mathcal{C}(1)}(k) = \|B'k\|_{p'}$, it follows that the optimization problem in equation (10) under $\mathcal{C} = \mathcal{C}(1)$ is equivalent to

$$\min_k k'\Sigma k \quad \text{s.t. } H = -k'\Gamma \quad \text{and} \quad \left\|B'k\right\|_{p'} \leq \overline{B}. \tag{S1}$$

We now specialize the results to the cases $p = 2$ and $p = \infty$. We discuss the case $p = 1$ in a working paper version of this paper (Armstrong and Kolesár (2020)).

A.2.1 $p = 2$ In this case, the Lagrangian form of (S1) becomes

$$\min_k k'\bigl(\Sigma + \lambda BB'\bigr)k \quad \text{s.t. } H = -k'\Gamma,$$

with the Lagrange multiplier λ giving the relative weight on bias. Optimizing this objective is isomorphic to deriving the minimum variance unbiased estimator of $H\theta$ in a regression model with design matrix $-\Gamma$ and variance $\Sigma + \lambda BB'$, so the Gauss–Markov theorem implies that the optimal sensitivities are $k_\lambda' = -H(\Gamma' W_\lambda \Gamma)^{-1}\Gamma' W_\lambda$ where $W_\lambda = [\Sigma + \lambda BB']^{-1}$.

A.2.2 $p = \infty$ Write the Lagrangian form of (S1) as

$$\min_k k'\Sigma k/2 + \lambda \left\| B'k \right\|_1 \quad \text{s.t. } H = -k'\Gamma. \tag{S2}$$

It will be convenient to transform the problem so that the ℓ_1 constraint only involves d_γ elements of k. Let

$$T = \begin{pmatrix} B'_\perp \\ (B'B)^{-1} B' \end{pmatrix}, \qquad T^{-1} = \begin{pmatrix} B_\perp & B \end{pmatrix}, \tag{S3}$$

where $B_\perp$ is an orthonormal matrix that's orthogonal to B. Then, since $TB = (0, I_{d_\gamma})'$, the above minimization problem is equivalent to the problem

$$\min_\kappa \kappa' S \kappa/2 + \lambda \sum_{i \in I} |\kappa_i| \quad \text{s.t. } H' = -G'\kappa,$$

where $\kappa = T'^{-1}k$, $S = T\Sigma T'$, $G = T\Gamma$, and $I = \{d_g - d_\gamma, \dots, d_g\}$ indexes the last d_γ elements of κ.

To minimize the above display and give the solution path as λ varies, we use arguments similar to those in Theorem 2 of Rosset and Zhu (2007). For $i \in I$, write $\kappa_i = \kappa_{+,i} - \kappa_{-,i}$, where $\kappa_{+,i} = \max\{\kappa_i, 0\}$ and $\kappa_{-,i} = -\min\{\kappa_i, 0\}$. We minimize the objective function in the preceding display over $\{\kappa_{+,i}, \kappa_{-,i}, \kappa_j \colon i \in I, j \notin I\}$ subject to the constraints $\kappa_{+,i} \geq 0$ and $\kappa_{-,i} \geq 0$. Let μ denote a vector of Lagrange multipliers on the restriction $-H' = G'\kappa$. Then the Lagrangian can be written as

$$\kappa' S \kappa/2 + \lambda \sum_{i \in I} (\kappa_{+,i} + \kappa_{-,i}) + \mu' \left(H' + G'\kappa \right) - \sum_{i \in I} (\lambda_{+,i}\kappa_{+,i} + \lambda_{-,i}\kappa_{-,i}).$$

The first-order conditions are given by

$$e_i' S\kappa + e_i' G\mu = 0, \quad i \in I^C, \tag{S4}$$

$$e_i' S\kappa + e_i' G\mu + \lambda = \lambda_{+,i}, \quad i \in I, \tag{S5}$$

$$-\left(e_i' S\kappa + e_i' G\mu\right) + \lambda = \lambda_{-,i}, \quad i \in I. \tag{S6}$$

The complementary slackness conditions are given by $\lambda_{+,i}\kappa_{+,i} = 0$ and $\lambda_{-,i}\kappa_{-,i} = 0$ for $i \in I$, and the feasibility constraints are $\lambda_{+,i} \geq 0$, $\lambda_{-,i} \geq 0$ for $i \in I$ and $-H' = G'\kappa$.

Let $\mathcal{A}^C = \{i \colon i \in I, \kappa_i = 0\}$, and let $\mathcal{A} = \{i \colon i \notin \mathcal{A}^C\}$ denote the set of active constraints. Let s denote a vector of length $|\mathcal{A}|$ with elements $s_i = \operatorname{sign}(\kappa_i)$ if $i \in I$ and $s_i = 0$ otherwise.

The slackness and feasibility conditions imply that if for $i \in I$, $\kappa_i > 0$, then $\lambda_{+,i} = 0$, and if $\kappa_i < 0$ or $\lambda_{-,i} = 0$. It therefore follows from (S5) and (S6) that $e_i' S\kappa + e_i' G\mu = -\operatorname{sign}(\kappa_i)\lambda = -s_i\lambda$. We can combine this condition with (S4) and write

$$e_i' S\kappa + e_i' G\mu = -s_i\lambda, \quad i \in \mathcal{A}. \tag{S7}$$

On the other hand, if $i \in \mathcal{A}^C$, then since $\lambda_{+,i}$ and $\lambda_{-,i}$ are nonnegative, it follows from (S5) and (S6) that

$$\left| e_i' S\kappa + e_i' G\mu \right| \leq \lambda = \left| e_j' S\kappa + e_j' G\mu \right|, \quad i \in \mathcal{A}^C, j \in \mathcal{A}. \tag{S8}$$

Let $\kappa_{\mathcal{A}}$ denote the subset of κ corresponding to the active moments, $G_{\mathcal{A}}$ denote the corresponding rows of G, and $S_{\mathcal{A}\mathcal{A}}$ the corresponding submatrix of S. Then we can write the condition (S7) together with the feasibility constraint $G'\kappa = -H'$ compactly as

$$\begin{pmatrix} 0 & G'_{\mathcal{A}} \\ G_{\mathcal{A}} & S_{\mathcal{A}\mathcal{A}} \end{pmatrix} \begin{pmatrix} \mu \\ \kappa_{\mathcal{A}} \end{pmatrix} = \begin{pmatrix} -H' \\ -s\lambda \end{pmatrix}.$$

Using the block matrix inverse formula, this implies

$$\begin{aligned} \mu &= \left(G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} G_{\mathcal{A}}\right)^{-1} \left(H' - G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} s\lambda\right), \\ \kappa_{\mathcal{A}} &= -S^{-1}_{\mathcal{A}\mathcal{A}} G_{\mathcal{A}} \mu - S^{-1}_{\mathcal{A}\mathcal{A}} s\lambda \\ &= S^{-1}_{\mathcal{A}\mathcal{A}} G_{\mathcal{A}} \left(G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} G_{\mathcal{A}}\right)^{-1} \left(G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} s\lambda - H'\right) - S^{-1}_{\mathcal{A}\mathcal{A}} s\lambda. \end{aligned}$$

Consequently, if we are in a region in where the solution path is differentiable with respect to λ, we have

$$\frac{\partial \kappa_{\mathcal{A}}}{\partial \lambda} = S^{-1}_{\mathcal{A}\mathcal{A}} G_{\mathcal{A}} \left(G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} G_{\mathcal{A}}\right)^{-1} G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} s - S^{-1}_{\mathcal{A}\mathcal{A}} s. \tag{S9}$$

The differentiability of path is violated if either (a) the constraint (S8) is violated for some $i \in \mathcal{A}^C$ if $\kappa(\lambda)$ keeps moving in the same direction, and we add i to $\mathcal{A}$ at a point at which (S8) holds with equality; or else (b) the sensitivity $\kappa_i(\lambda)$ for some $i \in \mathcal{A}$ reaches zero. In this case, drop i from $\mathcal{A}$. In either case, we need to re-calculate the direction (S9) using the new definition of $\mathcal{A}$.

Based on the arguments above and the fact that $\kappa(0) = -S^{-1}G(G'S^{-1}G)^{-1}H'$, we can derive the following algorithm, similar to the LAR-LASSO algorithm, to generate the path of optimal sensitivities $\kappa(\lambda)$:

1. Initialize $\lambda = 0$, $\mathcal{A} = \{1, \dots, d_g\}$, $\mu = (G'S^{-1}G)^{-1}H'$, $\kappa = -S^{-1}G\mu$. Let s be a vector of length d_g with elements $s_i = \mathbb{I}\{i \in I\}\operatorname{sign}(\kappa_i)$, and calculate initial directions as $\mu_\Delta = -(G'S^{-1}G)^{-1}G'S^{-1}s$, $\kappa_\Delta = -S^{-1}(G\mu_\Delta + s)$

2. While $(|\mathcal{A}| > \max\{d_g - d_\gamma, d_\theta\})$:

   (a) Set step size to $d = \min\{d_1, d_2\}$, where

   $$\begin{aligned} d_1 &= \min\{d > 0 \colon \kappa_i + d\kappa_{\Delta,i} = 0, i \in \mathcal{A} \cap \mathcal{I}\}, \\ d_2 &= \min\left\{d > 0 \colon \left|e_i'(S\kappa + G\mu) + d e_i'(S\kappa_\Delta + G\mu_\Delta)\right| = \lambda + d, i \in \mathcal{A}^C\right\}. \end{aligned}$$

   Take step of size d: $\kappa \mapsto \kappa + d\kappa_\Delta$, $\mu \mapsto \mu + d\mu_\Delta$, and $\lambda \mapsto \lambda + d$.

   (b) If $d = d_1$, drop $\operatorname{argmin}(d_1)$ from $\mathcal{A}$, and if $d = d_2$, then add $\operatorname{argmin}(d_2)$ to $\mathcal{A}$. Let s be a vector of length d_g with elements $s_i = -\mathbb{I}\{i \in I\}\operatorname{sign}(e_i'S\kappa + e_i'G\mu)$, and calculate new directions as

   $$\mu_\Delta = -\left(G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} G_{\mathcal{A}}\right)^{-1} G'_{\mathcal{A}} S^{-1}_{\mathcal{A}\mathcal{A}} s_{\mathcal{A}},$$

$$(\kappa_\Delta)_{\mathcal{A}} = -S_{\mathcal{A}\mathcal{A}}^{-1}(G_{\mathcal{A}}\mu_\Delta + s_{\mathcal{A}}),$$
$$(\kappa_\Delta)_{\mathcal{A}^C} = 0.$$

The solution path $k(\lambda)$ is then obtained as $k(\lambda) = T'\kappa(\lambda)$.

Finally, we show that in the limit $M \to \infty$, the optimal sensitivity corresponds to a method of moments estimator based on the most informative set of d_θ moments, with the remaining $d_g - d_\theta$ moments dropped. The optimal sensitivity as $M \to \infty$ obtains by solving (S2) as $\lambda \to \infty$. If B corresponds to columns of the identity matrix, then this is equivalent to minimizing $\|k_I\|_1$ subject to $H = -k'\Gamma$. This can be written as a linear program $\min k_{I,+} + k_{I,-i}$ st $-H' = \Gamma'(k_+ - k_-)$, $k_+, k_- \geq 0$. The minimization problem is done on a d_θ-dimensional hyperplane, and solution must occur at a boundary point of the feasible set, where only d_θ variables are nonzero. So the optimal k has d_θ nonzero elements.

Appendix B: Specification test

One can test the null hypothesis of correct specification (i.e., the null hypothesis that $c = 0$) using the J statistic

$$J = n \min_\theta \hat{g}(\theta)'\hat{\Sigma}^{-1}\hat{g}(\theta) = n\hat{g}(\hat{\theta})'\hat{\Sigma}^{-1}\hat{g}(\hat{\theta}),$$

where $\hat{\theta} = \operatorname{argmin}_\theta \hat{g}(\theta)'\hat{\Sigma}^{-1}\hat{g}(\theta)$. Alternatively, letting $\hat{\Sigma}^{-1/2}$ denote the symmetric square root of $\hat{\Sigma}^{-1}$, one can project $\hat{\Sigma}^{-1/2}\hat{g}(\tilde{\theta})$, where $\tilde{\theta}$ is some consistent estimate, onto the complement of the space spanned by $\hat{\Sigma}^{-1/2}\hat{\Gamma}$,

$$S = n\hat{g}(\tilde{\theta})'\hat{\Sigma}^{-1/2}\hat{R}\hat{\Sigma}^{-1/2}\hat{g}(\tilde{\theta}),$$

where $\hat{R} = I - \hat{\Sigma}^{-1/2}\hat{\Gamma}(\hat{\Gamma}'\hat{\Sigma}^{-1}\hat{\Gamma})^{-1}\hat{\Gamma}'\hat{\Sigma}^{-1/2}$. If the model is correctly specified, so that $c = 0$, S and J are asymptotically equivalent (Newey and McFadden (1994, p. 2231)), and distributed $\chi^2_{d_g - d_\theta}$.

Under local misspecification, the J statistic has a noncentral χ^2 distribution, with noncentrality parameter depending on c (Newey (1985)), and the asymptotic equivalence of J and S still holds. In this section, we use this observation to form a test of the null hypothesis $H_0\colon c \in \mathcal{C}$. When $\mathcal{C}$ takes the form in equation (12) for some norm $\|\cdot\|$, inverting these tests gives a lower CI for M. We begin with a lemma deriving the asymptotic distribution of S and J under local misspecification.

Lemma B.1. *Suppose that equations* (1), (2), *and* (3) *hold, and that $\hat{\theta}$ and $\tilde{\theta}$ satisfy, for some K and $K'_{\text{opt}} = -(\Gamma'\Sigma^{-1}\Gamma)^{-1}\Gamma'\Sigma^{-1}$,*

$$\sqrt{n}(\hat{\theta} - \theta_0) = K'_{\text{opt}}\sqrt{n}\hat{g}(\theta_0), \quad \textit{and} \quad \sqrt{n}(\tilde{\theta} - \theta_0) = K'\sqrt{n}\hat{g}(\theta_0).$$

Suppose that $\hat{\Sigma}$ and $\hat{\Gamma}$ are consistent estimates of Σ and Γ, and that Σ and Γ are full rank. Then $S = J + o_P(1)$ and S and J converge in distribution to a noncentral chi-square distribution with $d_g - d_\theta$ degrees of freedom and noncentrality parameter $c'\Sigma^{-1/2}R\Sigma^{-1/2}c$ where $R = I - \Sigma^{-1/2}\Gamma(\Gamma'\Sigma^{-1}\Gamma)^{-1}\Gamma\Sigma^{-1/2}$.

Proof. By equations (1), (2), and (3), $\sqrt{n}\hat{g}(\tilde{\theta}) = (I + \Gamma K')\Sigma^{1/2}(\Sigma^{-1/2}c + Z_n) + o_P(1)$ where $Z_n = \Sigma^{-1/2}[\sqrt{n}\hat{g}(\theta_0) - c] \xrightarrow{d} \mathcal{N}(0, I_{d_g})$, so that

$$S = \big(\Sigma^{-1/2}c + Z_n\big)'\Sigma^{1/2}\big(\Sigma^{-1/2} + \Sigma^{-1/2}\Gamma K'\big)' R\big(\Sigma^{-1/2} + \Sigma^{-1/2}\Gamma K'\big)\Sigma^{1/2}\big(\Sigma^{-1/2}c + Z_n\big) + o_P(1)$$
$$= \big(\Sigma^{-1/2}c + Z_n\big)' R\big(\Sigma^{-1/2}c + Z_n\big) + o_P(1) \xrightarrow{d} \big(\Sigma^{-1/2}c + Z\big)' R\big(\Sigma^{-1/2}c + Z\big),$$

where $Z \sim \mathcal{N}(0, I_{d_g})$ and we use the fact that $R(I + \Sigma^{-1/2}\Gamma K'\Sigma^{1/2}) = R$. Similarly,

$$\sqrt{n}\hat{g}(\hat{\theta}) = \big(I - \Gamma\big(\Gamma'\Sigma^{-1}\Gamma\big)\Gamma'\Sigma^{-1}\big)\big(c + \Sigma^{1/2}Z_n\big) + o_P(1) = \Sigma^{1/2}R\big(\Sigma^{-1/2}c + Z_n\big) + o_P(1),$$

so that $J = (\Sigma^{-1/2}c + Z_n)'R(\Sigma^{-1/2}c + Z_n) + o_P(1) = S + o_P(1)$. To prove the second claim, decompose $R = P_1P_1'$, where $P_1 \in \mathbb{R}^{d_\theta \times (d_g - d_\theta)}$ corresponds to the eigenvectors associated with nonzero eigenvalues of R. Then

$$\big(\Sigma^{-1/2}c + Z\big)' R\big(\Sigma^{-1/2}c + Z\big) = \big(P_1'\Sigma^{-1/2}c + P_1'Z\big)'\big(P_1'\Sigma^{-1/2}c + P_1'Z\big).$$

Since $P_1'Z \sim \mathcal{N}(0, I_{d_g - d_\theta})$, it follows that the random variable in the preceding display has a noncentral χ^2 distribution with $d_g - d_\theta$ degrees of freedom and noncentrality parameter $c'\Sigma^{-1/2}R\Sigma^{-1/2}c$. □

Lemma B.1 can be interpreted in using the limiting experiment described in Section 4.1. In particular, the asymptotic distribution of the S and J statistics is given by the distribution of the statistic $Y'\Sigma^{-1/2}R\Sigma^{-1/2}Y$ in the limiting experiment.

The quantiles of a noncentral chi-square distribution are increasing in the noncentrality parameter (Sun, Baricz, and Zhou (2010)). Thus, to test the null hypothesis $H_0\colon c \in \mathcal{C}$, the appropriate critical value for tests based on the J or S statistic is based on a noncentral chi-squared distribution, with noncentrality parameter

$$\bar{\lambda} = \sup_{c \in \mathcal{C}} c'\Sigma^{-1/2}R\Sigma^{-1/2}c.$$

If $\mathcal{C} = \{B\gamma\colon \|\gamma\|_p \le M\}$, then this becomes

$$\bar{\lambda} = \sup_{\|t\|_p \le M} t'B'\Sigma^{-1/2}R\Sigma^{-1/2}Bt = \sup_{\|t\|_p \le 1} M^2\big\|R\Sigma^{-1/2}Bt\big\|_2^2 = M^2\|A\|_{p,2}^2,$$

where the second equality uses the fact that R is idempotent, $A = R\Sigma^{-1/2}B$, and $\|A\|_{p,q} = \max_{\|x\|_p \le 1}\|Ax\|_q$ is the (p, q) operator norm. For $p = 2$, the operator norm has a closed form, which gives $\bar{\lambda} = M \max \text{eig}(B'\Sigma^{-1/2}R\Sigma^{-1/2}B)$.

Appendix C: Asymptotic coverage and efficiency

This Appendix contains the asymptotic coverage and efficiency results discussed in Section 4. In particular, we prove Theorem 4.1. In order to allow for stronger statements, we state upper and lower bounds separately. Theorem 4.1 then follows by combining these results. Theorem 4.1 focuses on two-sided CIs in the case where $\mathcal{C}$ is centrosymmetric, in addition to being convex. In this Appendix, we also prove analogous results for one-sided CIs, and we generalize these results to the case where $\mathcal{C}$ is a convex but asymmetric set. When $\mathcal{C}$ is convex but asymmetric, the negative results about the scope for improvement when c is close to zero no longer hold. Therefore, we consider the general problem of optimizing quantiles of excess length over a set $\mathcal{D} \subseteq \mathcal{C}$, which may be a strict subset of $\mathcal{C}$.

The remainder of this Appendix is organized as follows. Appendix C.1 presents notation and definitions, as well as an overview of the results. Appendix C.2 contains results on least favorable submodels as well as a two-point testing lemma used in later proofs. We then use this to obtain efficiency bounds for one-sided CIs in Appendix C.3, and for two-sided CIs in Appendix C.4. Appendix C.5 shows that our CIs achieve (or, for two-sided CIs, nearly achieve) these bounds. Appendix C.6 shows how Theorem 4.1 follows from these results, and also gives a one-sided version of this theorem. Primitive conditions for the misspecified linear IV model, as well as a general construction of a least favorable submodel satisfying the assumptions used in this section, are given in the supplemental Appendix.

C.1 *Setup*

While our focus is on parameter spaces that place restrictions on c, we will also allow for local restrictions on θ in some results. This allows us to bound the scope for "directing power" at particular values of θ. Formally, for some parameter θ^*, we consider the local parameter space that restricts $(\sqrt{n}(\theta-\theta^*)', c')'$ to some set $\mathcal{F} \subseteq \mathbb{R}^{d_\theta+d_g}$. The unrestricted case considered throughout most of the main text corresponds to $\mathcal{F} = \mathbb{R}^{d_\theta} \times \mathcal{C}$ (in which case θ^* does not affect the definition of the parameter space). We also allow for additional restrictions on θ by placing it in some set Θ_n. Finally, we use $\mathcal{P}$ to denote the set of distributions P over which we require coverage.

With this notation, the set of values of θ that are consistent with the model under P (i.e., the identified set under P) is

$$\Theta_I(P) = \Theta_I(P; \mathcal{F}, \Theta_n) = \left\{\theta \in \Theta_n : \sqrt{n}\big((\theta-\theta^*)', g_P(\theta)'\big)' \in \mathcal{F}\right\},$$

and the set of pairs (θ, P) over which coverage is required is given by

$$\mathcal{S}_n = \left\{(\theta, P) \in \Theta_n \times \mathcal{P} : \theta \in \Theta_I(P)\right\} = \left\{(\theta, P) \in \Theta_n \times \mathcal{P} : \sqrt{n}\big((\theta-\theta^*)', g_P(\theta)'\big)' \in \mathcal{F}\right\},$$

which reduces to the definition in Theorem 4.1 when $\mathcal{F} = \mathbb{R}^{d_\theta} \times \mathcal{C}$. The coverage requirement for a CI $\mathcal{I}_n$ is then given in equation (20) with this definition of $\mathcal{S}_n$. To compare one-sided CIs $[\hat{c}, \infty)$, we will consider the β quantile of excess length. Rather than restricting ourselves to the minimax criterion, we consider worst-case excess length over a potentially smaller parameter space $\mathcal{G}$, which may place additional restrictions on θ and c. Let

$$q_{\beta,n}(\hat{c}; \mathcal{P}, \mathcal{G}, \Theta_n) = \sup_{P\in\mathcal{P}} \sup_{\theta\in\Theta_I(P;\mathcal{G},\Theta_n)} q_{P,\beta}\big(h(\theta) - \hat{c}\big),$$

where $q_{P,\beta}$ denotes the β quantile under P. We will also consider bounds on $q_{P,\beta}(h(\theta)-\hat{c})$ at a single P, which corresponds to the optimistic case of optimizing length at a single distribution. For two-sided CIs, we will consider expected length.

Our efficiency bounds can be thought of as applying the bounds in Armstrong and Kolesár (2018) to a local asymptotic setting, which corresponds to the limiting model in equation (13) with $\Gamma = \Gamma_{\theta^*,P_0}$, $\Sigma = \Sigma_{\theta^*,P_0}$ and $H = H_{\theta^*}$. The between class modulus of continuity for this model is

$$\omega(\delta; \mathcal{F}, \mathcal{G}, H, \Gamma, \Sigma) = \sup H(s_1 - s_0) \quad \text{s.t. } \big(s_0', c_0'\big)' \in \mathcal{F}, \big(s_1', c_1'\big)' \in \mathcal{G},$$
$$\big[(c_1 - c_0) - \Gamma(s_1 - s_0)\big]'\Sigma^{-1}\big[(c_1 - c_0) - \Gamma(s_1 - s_0)\big] \le \delta^2. \tag{S10}$$

We use the notation $\omega(\delta)$ and $\omega(\delta; \mathcal{F}, \mathcal{G})$ when the context is clear. In the case where $\mathcal{G} = \mathcal{F} = \mathbb{R}^{d_\theta} \times \mathcal{C}$ and $\mathcal{C}$ is centrosymmetric, the solution satisfies $s_1 = -s_0$ and $c_1 = -c_0$, which gives the same optimization problem as in equation (15), with the objective multiplied by two (this matches the definition of $\omega(\cdot)$ used to define κ_* in the main text).

For one-sided CIs, we show that, for any CI satisfying the coverage condition in equation (20) for a rich enough class $\mathcal{P}$, we will have

$$\liminf_{n\to\infty} \sqrt{n} q_{\beta,n}(\hat{c}; \mathcal{P}, \mathcal{G}, \Theta_n) \ge \omega(\delta_\beta; \mathcal{F}, \mathcal{G}, H, \Gamma, \Sigma), \tag{S11}$$

where $\delta_\beta = z_{1-\alpha} + z_\beta$, where z_τ denotes the τ quantile of the $\mathcal{N}(0,1)$ distribution. For bounds on excess length at a single P_0 with $E_{P_0} g(w_i, \theta^*) = 0$, we obtain this bound with $\mathcal{G} = \{0\}$:

$$\liminf_{n\to\infty} \sqrt{n} q_{P_0,\beta}\big(h\big(\theta^*\big) - \hat{c}\big) \ge \omega\big(\delta_\beta; \mathcal{F}, \{0\}, H, \Gamma, \Sigma\big). \tag{S12}$$

These results can be thought of as a local asymptotic version of Theorem 3.1 in Armstrong and Kolesár (2018) applied to our setting.

For two-sided CIs, we show that, if a CI $\mathcal{I}_n = \{\hat{h} \pm \hat{\chi}\}$ satisfies the coverage condition in equation (20) for a rich enough class $\mathcal{P}$ then, for any P_0 with $E_{P_0} g(w_i, \theta^*) = 0$, expected length satisfies

$$\begin{aligned} &\liminf_{T\to\infty} \liminf_{n\to\infty} E_{P_0} \min\{\sqrt{n} 2\hat{\chi}, T\} \\ &\quad \ge (1-\alpha) E\big[\omega\big(z_{1-\alpha} - Z; \{0\}, \mathcal{F}, H, \Gamma, \Sigma\big) \\ &\qquad + \omega\big(z_{1-\alpha} - Z; \mathcal{F}, \{0\}, H, \Gamma, \Sigma\big) \mid Z \le z_{1-\alpha}\big], \end{aligned} \tag{S13}$$

where $Z \sim \mathcal{N}(0,1)$. The above bound uses truncated expected length to avoid technical issues with convergence of moments when achieving the bound (note however that this bound immediately implies the same bound on excess length without truncation). Our results constrain the CI to take the form of an interval. We conjecture that the bound applies to arbitrary confidence sets (with length defined as Lebesgue measure) under additional regularity conditions.

Here, "rich enough" means that $\mathcal{P}$ contains a least favorable submodel. Appendix C.2 begins the derivation of our efficiency results by giving conditions on this submodel. In Appendix E.1, we construct a submodel satisfying these conditions under mild conditions.

C.2 *Least favorable submodel*

Let P_0 be a distribution with $E_{P_0} g(w_i, \theta^*) = 0$ (i.e., the model holds for this data-generating process with $\theta = \theta^*$ and $c = 0$), and consider a parametric submodel P_t indexed by $t \in \mathbb{R}^{d_g}$ (i.e., the dimension of t is the same as the dimension of the values of $g(w_i, \theta)$) with P_t equal to P_0 at $t = 0$. We assume that $\{w_i\}_{i=1}^n$ are i.i.d. under P_t. Let $\pi_t(w_i)$ denote the density of a single observation with respect to its distribution under P_0, so that $E_{P_t} f(w_i) = E_{P_0} f(w_i)\pi_t(w_i)$ for any function f. We expect that the least favorable submodel for this problem will be the one that makes estimating $E_P g(W_i, \theta^*)$ most difficult. This corresponds to any subfamily with score function $g(w_i, \theta^*)$. We also place additional conditions on this submodel, given in the following assumption.

Assumption C.1. *The data are i.i.d. under P_t for all t in a neighborhood of zero, and the density $\pi_t(w_i)$ for a single observation is quadratic mean differentiable at $t = 0$ with score function $g(w_i, \theta^*)$, where $E_{P_0} g(w_i, \theta^*) = 0$. In addition, the function $(t', \theta')' \mapsto E_{p_t} g(w_i, \theta)$ is continuously differentiable at $(0', \theta^{*\prime})'$ with*

$$\left[\frac{d}{d\big(t', \theta'\big)} E_{p_t} g(w_i, \theta) \right]_{t=0,\, \theta=\theta^*} = (\Sigma, \Gamma), \tag{S14}$$

where Σ and Γ are full rank.

To understand Assumption C.1, note that Problem 12.17 in Lehmann and Romano (2005) gives the Jacobian with respect to t as Σ in the case where $g(w_i, \theta^*)$ is bounded, and the Jacobian with respect to θ is equal to Γ by definition. Assumption C.1 requires the slightly stronger condition that $E_{p_t} g(w_i, \theta)$ is continuously differentiable with respect to $(t', \theta')'$ for t close to 0 and θ close to θ^*. This is needed to apply the implicit function theorem in the derivations that follow. In the supplemental materials, we give a construction of a quadratic mean differentiable family satisfying this condition, without requiring boundedness of $g(w_i, \theta^*)$ (Lemma E.1 in Appendix E.1).

The bounds in Armstrong and Kolesár (2018) are obtained by bounding the power of a two-point test (simple null and simple alternative) where the null and alternative are given by the points that achieve the modulus. To obtain analogous results in our setting, we use a bound on the power of a two-point test in a least favorable submodel.

Consider sequences of local parameter values $(\theta'_{0,n}, c'_{0,n})'$ and $(\theta'_{1,n}, c'_{1,n})'$ where, for some s_0, c_0, s_1, and c_1,

$$\theta_{d,n} = \theta^* + \big(s_d + o(1)\big)/\sqrt{n}, \qquad c_{d,n} = c_d + o(1) \quad d \in \{0, 1\}. \tag{S15}$$

Consider a sequence of tests of $(\theta'_{0,n}, c'_{0,n})'$ vs $(\theta'_{1,n}, c'_{1,n})'$. Formally, for any $(\theta', c')'$, let

$$\mathcal{P}_n(\theta, c) = \big\{P \in \mathcal{P} : E_P g(w_i, \theta) = c/\sqrt{n}\big\} \tag{S16}$$

be the set of probability distributions in $\mathcal{P}$ that are consistent with the parameter values $(\theta', c')'$. We derive a bound on the asymptotic minimax power of a level α test of

$$H_{0,n} : P \in \mathcal{P}_n(\theta_{0,n}, c_{0,n}) \quad \text{vs} \quad H_{1,n} : P \in \mathcal{P}_n(\theta_{1,n}, c_{1,n}), \tag{S17}$$

as well as a bound on the power of a test of $H_{0,n}$ at P_0. Let Φ be the standard normal cdf and let

$$\overline{\beta}(s_0, c_0, s_1, c_1) = \Phi\left(\sqrt{[c_1 - c_0 - \Gamma(s_1 - s_0)]'\Sigma^{-1}[c_1 - c_0 - \Gamma(s_1 - s_0)]} - z_{1-\alpha}\right).$$

Lemma C.1. *Consider a sequence of tests ϕ_n satisfying* $\limsup_n \sup_{P\in\mathcal{P}_n(\theta_{0,n},c_{0,n})} E_P\phi_n \leq \alpha$. *Then, provided the class of distributions $\mathcal{P}$ contains a family P_t that satisfies Assumption C.1, we have*

$$\limsup_n E_{P_0}\phi_n \leq \overline{\beta}(s_0, c_0, 0, 0) \quad \textit{and} \quad \limsup_n \inf_{P\in\mathcal{P}_n(\theta_{1,n},c_{1,n})} E_P\phi_n \leq \overline{\beta}(s_0, c_0, s_1, c_1).$$

Lemma C.1 says that the asymptotic minimax power of any test of $H_{0,n}$ vs $H_{1,n}$ is bounded by $\overline{\beta}(s_0, c_0, s_1, c_1)$. Furthermore, if we take $s_1 = 0$ and $c_1 = 0$, then this bound is achieved at P_0. Note that, in keeping with the analogy with the linear model in equation (13), $\overline{\beta}(s_0, c_0, s_1, c_1)$ is the power of the optimal (Neyman–Pearson) test of the simple null (s_0', c_0') versus the simple alternative (s_1', c_1') in the model in equation (13).

Proof of Lemma C.1. The proof involves two steps. First, we use the implicit function theorem to find sequences $t_{0,n}$ and $t_{1,n}$ such that $P_{t_{0,n}}$ satisfies $H_{0,n}$ and $P_{t_{1,n}}$ satisfies $H_{1,n}$. Next, we apply a standard result on testing in quadratic mean differentiable families to obtain the limiting power of the optimal test of $P_{t_{0,n}}$ versus $P_{t_{1,n}}$, which gives an upper bound on the limiting minimax power of any test of $H_{0,n}$ versus $H_{1,n}$.

Let $f(t, \theta, a) = E_{P_t} g(w_i, \theta) - a$ so that $(\theta', c')'$ is consistent with P_t iff $f(t, \theta, c/\sqrt{n}) = 0$. Under Assumption C.1, it follows from the implicit function theorem that there exists a function $r(\theta, a)$ such that, for θ in a neighborhood of θ^* and a in a neighborhood of zero,

$$E_{P_{r(\theta,a)}} g(w_i, \theta) - a = f\big(r(\theta, a), \theta, a\big) = 0.$$

Thus, letting $t_{0,n} = r(\theta_{0,n}, c_{0,n}/\sqrt{n})$ and $t_{1,n} = r(\theta_{1,n}, c_{1,n}/\sqrt{n})$, $P_{t_{0,n}}$ satisfies $H_{0,n}$ and $P_{t_{1,n}}$ satisfies $H_{1,n}$. Furthermore,

$$\left[\frac{d}{d(\theta', a')} r(\theta, a)\right]_{(\theta', a') = (\theta^*, 0)} = -\Sigma^{-1}(\Gamma, -I_{d_g})$$

so that

$$r(\theta, a) = \Sigma^{-1} a - \Sigma^{-1}\Gamma(\theta - \theta^*) + o\big(\|\theta - \theta^*\| + \|a\|\big).$$

Thus, letting $t_{0,\infty} = \Sigma^{-1} c_0 - \Sigma^{-1}\Gamma s_0$, we have

$$\begin{aligned} t_{0,n} = r(\theta_{0,n}, c_{0,n}/\sqrt{n}) &= \Sigma^{-1} c_{0,n}/\sqrt{n} - \Sigma^{-1}\Gamma(\theta_{0,n} - \theta^*) + o\big(\|\theta_{0,n} - \theta^*\| + \|c_{0,n}\|/\sqrt{n}\big) \\ &= \Sigma^{-1} c_0/\sqrt{n} - \Sigma^{-1}\Gamma s_0/\sqrt{n} + o(1/\sqrt{n}) = t_{0,\infty}/\sqrt{n} + o(1/\sqrt{n}). \end{aligned}$$

Similarly, $t_{1,n} = t_{1,\infty}/\sqrt{n} + o(1/\sqrt{n})$ where $t_{1,\infty} = \Sigma^{-1} c_1 - \Sigma^{-1}\Gamma s_1$.

Since the information matrix for this submodel evaluated at $t = 0$ is Σ, it follows from the arguments in Example 12.3.12 in Lehmann and Romano (2005), extended to the case where the null and alternative are both drifting sequences (rather than just the alternative), that the limit of the power of the Neyman–Pearson test of $P_{t_{0,n}}$ vs $P_{t_{1,n}}$ is

$$\Phi\Big(\sqrt{[t_{1,\infty}-t_{0,\infty}]'\Sigma[t_{1,\infty}-t_{0,\infty}]}-z_{1-\alpha}\Big)=\overline{\beta}(s_0,c_0,s_1,c_1).$$

This gives the required bound on minimax power over $H_{1,n}$. To obtain the bound on power at P_0, note that, for $\theta_{1,n}=\theta^*$ and $c_{1,n}=0$, $t_{0,n}=0$, the bound also corresponds to the power of a test that is optimal for $P_{t_{0,n}}$ versus P_0. □

C.3 *One-sided CIs*

We prove the following efficiency bound for one-sided CIs.

Theorem C.1. *Let $\mathcal{P}$ be a class of distributions that contains a submodel P_t satisfying Assumption C.1. Let $\Theta_n(C)=\{\theta \mid \|\theta-\theta^*\|\le C/\sqrt{n}\}$ for some constant C, and let $\mathcal{F}$ be given. Let $[\hat{c},\infty)$ be a sequence of CIs such that, for all C, the coverage condition in equation* (20) *holds with $\Theta_n=\Theta_n(C)$. Let $\mathcal{G}\subseteq\mathcal{F}$ be a set such that the limiting modulus ω is well-defined and continuous for all δ. Then the asymptotic lower bounds* (S11) *and* (S12) *hold.*

Proof. Consider a sequence of simple null and alternative values of θ and c that satisfy (S15) for some s_0, c_0, s_1, c_1, with $(\sqrt{n}(\theta_{0,n}-\theta^*)', c'_{0,n})'\in\mathcal{F}$ and $(\sqrt{n}(\theta_{1,n}-\theta^*)', c'_{1,n})'\in\mathcal{G}$, for each n. Note that

$$\lim_{n\to\infty}\sqrt{n}\big[h(\theta_{1,n})-h(\theta_{0,n})\big]=H(s_1-s_0).$$

Consider the testing problem $H_{0,n}: P\in\mathcal{P}_n(\theta_{0,n},c_{0,n})$ vs $H_{1,n}: P\in\mathcal{P}_n(\theta_{1,n},c_{1,n})$ defined in (S16) and (S17). Suppose that

$$q_{\beta,n}(\hat{c};\mathcal{P},\mathcal{G},\Theta_n)<h(\theta_{1,n})-h(\theta_{0,n}). \tag{S18}$$

Let ϕ_n denote the test that rejects when $h(\theta_{0,n})\notin[\hat{c},\infty)$. Since, for any $P\in\mathcal{P}_n(\theta_{1,n},c_{1,n})$, we have $q_{P,\beta}(h(\theta_{1,n})-\hat{c})\le q_{\beta,n}(\hat{c};\mathcal{P},\mathcal{G},\Theta_n)$ by construction, it follows that, for all $P\in\mathcal{P}_n(\theta_{1,n},c_{1,n})$,

$$E_P\phi_n=P\big(h(\theta_{1,n})-\hat{c}<h(\theta_{1,n})-h(\theta_{0,n})\big)\ge P\big(h(\theta_{1,n})-\hat{c}\le q_{P,\beta}\big(h(\theta_{1,n})-\hat{c}\big)\big)\ge\beta,$$

where the last step follows from properties of quantiles (Lemma 21.1 in van der Vaart (1998)). The coverage requirement in equation (20) implies that the test ϕ_n that rejects when $h(\theta_{0,n})\notin[\hat{c},\infty)$ has asymptotic level α for $H_{0,n}$. Thus, by Lemma C.1, we must have $\beta\le\overline{\beta}(s_0,c_0,s_1,c_1)$ if (S18) holds infinitely often.

It follows that, if $\overline{\beta}(s_0,c_0,s_1,c_1)<\beta$, we must have

$$\liminf_{n\to\infty}\sqrt{n}q_{\beta,n}(\hat{c};\mathcal{P},\mathcal{G},\Theta_n)\ge H(s_1-s_0)$$

since otherwise, equation (S18) would hold infinitely often. Since the sequences and limiting $(s'_0,c'_0)\in\mathcal{F}$ and $(s'_1,c'_1)\in\mathcal{G}$ were arbitrary, the above bound holds for any

$(s_0', c_0') \in \mathcal{F}$ and $(s_1', c_1') \in \mathcal{G}$ with $\overline{\beta}(s_0, c_0, s_1, c_1) \leq \beta - \eta$, where $\eta > 0$ is arbitrary. The maximum of the right-hand side over s_0, c_0, s_1, c_1 in this set is equal to $\omega(\delta_{\beta-\eta}; \mathcal{F}, \mathcal{G}, H, \Gamma, \Sigma)$ by definition, so taking $\eta \to 0$ gives the result. □

C.4 *Two-sided CIs*

We prove the following efficiency bound for two-sided CIs.

Theorem C.2. *Suppose that, for all C, $\{\hat{h} \pm \hat{\chi}\}$ satisfies the local coverage condition in equation* (20) *with $\Theta_n = \Theta_n(C) = \{\theta \mid \|\theta - \theta^*\| \leq C/\sqrt{n}\}$, where $\mathcal{P}$ contains a submodel P_t satisfying Assumption* C.1. *Suppose also that $0_{d_\theta + d_g} \in \mathcal{F}$ and a minimizer $(s_\vartheta', c_\vartheta')'$ of $(c - \Gamma s)'\Sigma^{-1}(c - \Gamma s)$ subject to $Hs = \vartheta$ and $(s', c')' \in \mathcal{F}$ exists for all $\vartheta \in \mathbb{R}$. Then the asymptotic lower bound* (S13) *holds.*

In the case where $\mathcal{F} = \mathbb{R}^{d_\theta} \times \mathcal{C}$, which is the focus of the main text, a sufficient condition for the existence of the minimizer $(s_\vartheta', c_\vartheta')'$ is that $\mathcal{C}$ is compact, H is not equal to the zero vector and Γ is full rank.

Proof. For each $\vartheta \in \mathbb{R}$, let $\tilde{\theta}_{\vartheta,n} = \theta^* + s_\vartheta/\sqrt{n}$, and let $\phi_{\vartheta,n} = I(h(\tilde{\theta}_{\vartheta,n}) \notin \{\hat{h} \pm \hat{\chi}\})$ be the test that rejects when $h(\tilde{\theta}_{\vartheta,n})$ is not in the CI. When the constant C defining $\Theta_n = \Theta_n(C)$ is large enough, the asymptotic coverage condition in equation (20) implies that $\phi_{\vartheta,n}$ is an asymptotic level α test for $H_{0,n} : P \in \mathcal{P}_n(\tilde{\theta}_{\vartheta,n}, c_\vartheta)$ defined in (S16) and (S17). Thus, by Lemma C.1,

$$\limsup_{n\to\infty} E_{P_0}\phi_{\vartheta,n} \leq \Phi(\delta_\vartheta - z_{1-\alpha}) \quad \text{where } \delta_\vartheta = \sqrt{(c_\vartheta - \Gamma s_\vartheta)'\Sigma^{-1}(c_\vartheta - \Gamma s_\vartheta)}. \tag{S19}$$

We apply this bound to a grid of values of ϑ. Let $\mathcal{E}_n(m)$ denote the grid centered at zero with length $2m$ and meshwidth $1/m$:

$$\mathcal{E}_n(m) = \{j/m : j \in \mathbb{Z}, |j| \leq m^2\}.$$

Let

$$\widetilde{\mathcal{E}}_n(m) = \{\sqrt{n}[h(\tilde{\theta}_{\vartheta,n}) - h(\theta^*)] : \vartheta \in \mathcal{E}_n(m)\}.$$

Note that $h(\tilde{\theta}_{\vartheta,n}) = h(\theta^*) + (1 + o(1))Hs_\vartheta/\sqrt{n} = h(\theta^*) + (1 + o(1))\vartheta/\sqrt{n}$. Thus, letting $a_1, \ldots, a_{2m^2+1}$ denote the ordered elements in $\mathcal{E}_n(m)$ and $\tilde{a}_1, \ldots, \tilde{a}_{m^2+1}$ the ordered elements in $\widetilde{\mathcal{E}}_n$, we have $\tilde{a}_j \to a_j$ for each j as $n \to \infty$.

Let $\mathcal{N}(n, m)$ be the number of elements $\tilde{a}_j$ in $\widetilde{\mathcal{E}}_n$ such that $h(\theta^*) + \tilde{a}_j/\sqrt{n} = h(\tilde{\theta}_{a_j,n}) \in \{\hat{h} \pm \hat{\chi}\}$. Then

$$E_{P_0}\mathcal{N}(n, m) = \sum_{j=1}^{2m^2+1} E_{P_0} I(h(\tilde{\theta}_{a_j,n}) \in \{\hat{h} \pm \hat{\chi}\}) = \sum_{j=1}^{2m^2+1} [1 - E_{P_0}\phi_{a_j,n}].$$

It follows from (S19) that (assuming the constant C that defines $\Theta_n(C)$ is large enough),

$$\liminf_{n\to\infty} E_{P_0}\mathcal{N}(n,m) \geq \sum_{j=1}^{2m^2+1}\left[1-\Phi(\delta_{a_j}-z_{1-\alpha})\right] = \sum_{j=1}^{2m^2+1}\Phi(z_{1-\alpha}-\delta_{a_j}).$$

Note that $2\hat{\chi} \geq n^{-1/2}[\mathcal{N}(n,m)-1]\cdot\min_{1\leq j\leq 2m^2}(\tilde{a}_{j+1}-\tilde{a}_j) = n^{-1/2}[\mathcal{N}(n,m)-1]\cdot m^{-1}\cdot(1+\varepsilon_n)$ where $\varepsilon_n = \min_{1\leq j\leq 2m^2}(\tilde{a}_{j+1}-\tilde{a}_j)/m^{-1}-1$ is a nonrandom sequence converging to zero. This, combined with the above display, gives

$$\liminf_{n\to\infty} E_{P_0}\min\{2n^{1/2}\hat{\chi}, T\} \geq \left[m^{-1}\sum_{j=1}^{2m^2+1}\Phi(z_{1-\alpha}-\delta_{a_j}) - m^{-1}\right]$$

for any $T > 2m$. We have

$$m^{-1}\sum_{j=1}^{2m^2+1}\Phi(z_{1-\alpha}-\delta_{a_j}) = m^{-1}\sum_{j=1}^{2m^2+1}\int I(\delta_{a_j}\leq z_{1-\alpha}-z)\,d\Phi(z). \tag{S20}$$

Following the proof of Theorem 3.2 in Armstrong and Kolesár (2018), note that, for $\vartheta \geq 0$, $t \geq 0$, we have $\delta_\vartheta \leq t$ iff $\vartheta \leq \omega(t;\{0\},\mathcal{F})$. Indeed, note that $\omega(\delta_\vartheta;\{0\},\mathcal{F}) \geq Hs_\vartheta = \vartheta$ by feasibility of 0 and s_ϑ, c_ϑ for this modulus problem. Since the modulus is increasing, this means that, if $\delta_\vartheta \leq t$, we must have $\vartheta \leq \omega(t;\{0\},\mathcal{F})$. Now suppose $\vartheta \leq \omega(t;\{0\},\mathcal{F})$. Then $Hs_{\omega(t;\{0\},\mathcal{F})} \geq \vartheta$, so, for some $\lambda \in [0,1]$, $(s'_\lambda, c'_\lambda) = \lambda(s'_{\omega(t;\{0\},\mathcal{F})}, c'_{\omega(t;\{0\},\mathcal{F})})$ satisfies $Hs_\lambda = \vartheta$, which means that $\delta_\vartheta \leq \sqrt{(c_\lambda - \Gamma s_\lambda)'\Sigma^{-1}(c_\lambda-\Gamma s_\lambda)} \leq t$ as claimed.

Thus, the part of the expression in (S20) corresponding to terms in the sum with $a_j \geq 0$ is given by

$$\begin{aligned} m^{-1}&\sum_{j=1}^{2m^2+1}\int I\big(0\leq a_j \leq \omega\big(z_{1-\alpha}-z;\{0\},\mathcal{F}\big)\big)\,d\Phi(z) \\ &\geq \int_{z\leq z_{1-\alpha}} \min\big\{\omega\big(z_{1-\alpha}-z;\{0\},\mathcal{F}\big)-1/m, m\big\}\,d\Phi(z). \end{aligned}$$

By the dominated convergence theorem, this converges to $\int_{z\leq z_{1-\alpha}}\omega(z_{1-\alpha}-z;\{0\},\mathcal{F})\,d\Phi(z)$ as $m\to\infty$. Similarly, for $\vartheta < 0$, $t \geq 0$, we have $\delta_\vartheta \leq t$ iff. $-\vartheta \leq \omega(t;\mathcal{F},\{0\})$, so that an analogous argument shows that, for arbitrary $\varepsilon > 0$, there exists m such that $\int_{z\leq z_{1-\alpha}}\omega(z_{1-\alpha}-z;\mathcal{F},\{0\})\,d\Phi(z)-\varepsilon$ is an asymptotic lower bound for the part of the expression (S20) that corresponds to terms in the sum with $a_j < 0$. Thus, for any $\varepsilon > 0$, there exist constants C and T such that, if the coverage condition in equation (20) holds with $\Theta_n = \Theta_n(C)$,

$$\liminf_{n\to\infty} E_{P_0}\min\{n^{1/2}2\hat{\chi}, T\} \geq \int_{z\leq z_{1-\alpha}}\big[\omega\big(z_{1-\alpha}-z;\{0\},\mathcal{F}\big)+\omega\big(z_{1-\alpha}-z;\mathcal{F},\{0\}\big)\big]\,d\Phi(z)-2\varepsilon.$$

This gives the result. □

C.5 *Achieving the bound*

This section gives formal results showing that the CIs proposed in the main text are asymptotically valid, and that, if the sensitivities are chosen optimally, they achieve the efficiency bound in Theorem C.1 in the one-sided case, and nearly achieve the bound in Theorem C.2 in the two-sided case (where "nearly" means up to the sharp efficiency bound κ_* in the limiting model, given in equation (18), in the case where $\mathcal{C}$ is centrosymmetric).

We specialize to the case considered in the main text where we require coverage without local restrictions on θ. In the notation of Appendices C.3 and C.4, this corresponds to $\mathcal{F} = \mathbb{R}^{d_\theta} \times \mathcal{C}$ for a convex (but possibly asymmetric) set $\mathcal{C}$.

In the main text, we focused on the case where $\mathcal{C}$ is centrosymmetric. To allow for general convex $\mathcal{C}$, we use estimators that are asymptotically affine, rather than linear. We focus on one-step estimators, which take the form

$$\hat{h} = h(\hat{\theta}_{\text{initial}}) + \hat{k}' g(\hat{\theta}_{\text{initial}}) + \hat{a}/\sqrt{n}$$

for some vector $\hat{k}$ and some scalar $\hat{a}$. We continue to require the condition

$$\hat{H} = -\hat{k}'\hat{\Gamma}, \tag{S21}$$

where $\hat{\Gamma}$ is an estimator of Γ satisfying conditions to be given below.

To deal with asymmetric $\mathcal{C}$, and to state results involving worst-case quantiles of excess length over different sets, it will be helpful to separately define worst-case upper and lower bias. For a set $\mathcal{C} \in \mathbb{R}^{d_g}$, let

$$\overline{\text{bias}}_{\mathcal{C}}(k, a) = \sup_{c \in \mathcal{C}} k'c + a, \qquad \underline{\text{bias}}_{\mathcal{C}}(k, a) = \inf_{c \in \mathcal{C}} k'c + a.$$

A one-sided asymptotic $1-\alpha$ CI is given by $[\hat{c}, \infty)$ where

$$\begin{aligned}
\hat{c} &= \hat{h} - \overline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a})/\sqrt{n} - z_{1-\alpha}\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}/\sqrt{n} \\
&= h(\hat{\theta}_{\text{initial}}) + \hat{k}' g(\hat{\theta}_{\text{initial}}) + \hat{a}/\sqrt{n} - \overline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a})/\sqrt{n} - z_{1-\alpha}\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}/\sqrt{n} \\
&= h(\hat{\theta}_{\text{initial}}) + \hat{k}' g(\hat{\theta}_{\text{initial}}) - \overline{\text{bias}}_{\mathcal{C}}(\hat{k}, 0)/\sqrt{n} - z_{1-\alpha}\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}/\sqrt{n},
\end{aligned}$$

and $\hat{\Sigma}$ is an estimate of Σ. Thus, the intercept term $\hat{a}$ does not matter for the one-sided CI and can be taken to be zero in this case. For two-sided CIs, however, the choice of $\hat{a}$ matters, and we assume that $\hat{a}$ is chosen so that the estimator is centered:

$$\overline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a}) = \sup_{c \in \mathcal{C}} \hat{k}'c + \hat{a} = -\left(\inf_{c \in \mathcal{C}} \hat{k}'c + \hat{a}\right) = -\underline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a}). \tag{S22}$$

A two-sided asymptotic $1-\alpha$ CI is then given by $\hat{h} \pm \hat{\chi}$ where

$$\hat{\chi} = \text{cv}_\alpha\left(\overline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a})/\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}\right)\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}/\sqrt{n}, \quad \text{where } \text{cv}_\alpha(t) \text{ is the } 1-\alpha \text{ quantile of } |\mathcal{N}(t,1)|.$$

For both forms of CIs, we first state a result for general $\hat{k}$, $\hat{a}$, and then specialize to optimal weights. For the one-sided case, we consider CIs that optimize worst-case length over $(\sqrt{n}(\theta-\theta^*)', c')'$ in some set $\mathcal{G}$, subject to coverage over $\mathcal{F}=\mathbb{R}^{d_\theta}\times\mathcal{C}$. In principle, this allows for confidence sets that "direct power" not only at particular values of c but also at particular values of θ. However, Lemma E.2 in Appendix E.3.1 shows that the optimal weights for this problem are the same as the optimal weights when $\mathcal{G}$ is replaced by $\mathbb{R}^{d_\theta}\times\mathcal{D}(\mathcal{G})$, where $\mathcal{D}(\mathcal{G})=\{c\colon \text{there exists } s \text{ s.t. } (s',c')'\in\mathcal{G}\}$. Thus, it is without loss of generality to consider weights that optimize worst-case excess length over $c\in\mathcal{D}$ subject to coverage over $c\in\mathcal{C}$ where $\mathcal{D}\subseteq\mathcal{C}$ is a compact convex set.

The optimal weights take the form $\hat{k}=k(\delta_\beta,\hat{H},\hat{\Gamma},\hat{\Sigma})$ where

$$k(\delta,H,\Gamma,\Sigma)'=\frac{\left(\left(c^*_{1,\delta}-c^*_{0,\delta}\right)-\Gamma\left(s^*_{1,\delta}-s^*_{0,\delta}\right)\right)'\Sigma^{-1}}{\left(\left(c^*_{1,\delta}-c^*_{0,\delta}\right)-\Gamma\left(s^*_{1,\delta}-s^*_{0,\delta}\right)\right)'\Sigma^{-1}\Gamma H'/HH'} \tag{S23}$$

and $c_{0,\delta}$, $s_{0,\delta}$, $c_{1,\delta}$, $s_{1,\delta}$ solve the between class modulus problem (S10) with $\mathcal{F}=\mathbb{R}^{d_\theta}\times\mathcal{C}$ and $\mathcal{G}=\mathbb{R}^{d_\theta}\times\mathcal{D}$. For a two-sided CI of the form given above, the optimal weights take this form with $\mathcal{D}=\mathcal{C}$, δ minimizing $\hat{\chi}$, and with $\hat{a}$ chosen to center the CI so that (S22) holds. We note that, in the case where $\mathcal{D}=\mathcal{C}$ and $\mathcal{C}$ is centrosymmetric, $s^*_{1,\delta}=s^*_{0,\delta}$ and $c^*_{1,\delta}=c^*_{0,\delta}$, and (S10) reduces to two times the optimization problem in equation (15). The weights $\hat{k}$ then take the form given in equation (14) in the main text, and, since $\mathcal{C}$ is centrosymmetric, $\hat{a}=0$, which gives the two-sided CI proposed in the main text.

For our general result showing coverage for possibly suboptimal weights $\hat{k}$, $\hat{a}$, we make the following assumptions. In the following, for a set $\mathcal{A}_n$, random variables $A_{n,\theta,P}$ and $B_{n,\theta,P}$ and a sequence a_n, we say $A_{n,\theta,P}=B_{n,\theta,P}+o_P(a_n)$ uniformly over (θ,P) in $\mathcal{A}_n$ if, for all $\varepsilon>0$, $\sup_{(\theta,P)\in\mathcal{A}_n}P(a_n^{-1}\|A_{n,\theta,P}-B_{n,\theta,P}\|>\varepsilon)\to 0$. We say $A_{n,\theta,P}=B_{n,\theta,P}+\mathcal{O}_P(a_n)$ uniformly over (θ,P) in a set $\mathcal{A}_n$ if $\lim_{C\to\infty}\limsup_{n\to\infty}\sup_{(\theta,P)\in\mathcal{A}_n}P(a_n^{-1}\|A_{n,\theta,P}-B_{n,\theta,P}\|>C)=0$. In the following, the set $\mathcal{S}_n$ defined in Appendix C.1 over which coverage is required is defined with $\mathcal{F}=\mathbb{R}^{d_\theta}\times\mathcal{C}$.

Assumption C.2. *The set $\mathcal{C}$ is compact or takes the form $\widetilde{\mathcal{C}}\times\mathbb{R}^{d_{g_2}}$ where $d_{g_1}+d_{g_2}=d_g$ and $\widetilde{\mathcal{C}}$ is a compact subset of $\mathbb{R}^{d_{g_1}}$. In addition, $\hat{\theta}_{\text{initial}}-\theta=\mathcal{O}_P(1/\sqrt{n})$, $\hat{g}(\hat{\theta}_{\text{initial}})-\hat{g}(\theta)=\Gamma_{\theta,P}(\hat{\theta}_{\text{initial}}-\theta)+o_P(1/\sqrt{n})$ and $h(\hat{\theta}_{\text{initial}})-h(\theta)=H_\theta(\hat{\theta}_{\text{initial}}-\theta)+o_P(1/\sqrt{n})$ uniformly over $(\theta,P)\in\mathcal{S}_n$.*

Assumption C.3. *$\hat{g}(\theta)-g_P(\theta)=\mathcal{O}(1/\sqrt{n})$ uniformly over $(\theta,P)\in\mathcal{S}_n$. Furthermore, for a collection of matrices $\Sigma_{\theta,P}$ such that $k'_{\theta,P}\Sigma_{\theta,P}k_{\theta,P}$ is bounded away from zero and infinity,*

$$\sup_{t\in\mathbb{R}}\sup_{(\theta,P)\in\mathcal{S}_n}\left|P\left(\frac{\sqrt{n}k'_{\theta,P}\left(\hat{g}(\theta)-g_P(\theta)\right)}{\sqrt{k'_{\theta,P}\Sigma_{\theta,P}k_{\theta,P}}}\le t\right)-\Phi(t)\right|\to 0.$$

Assumption C.4. *$\hat{k}-k_{\theta,P}=o_P(1)$ uniformly over $(\theta,P)\in\mathcal{S}_n$, and similarly for $\hat{a}$, $\hat{\Gamma}$, $\hat{H}$ and $\hat{\Sigma}$. Furthermore, $k_{\theta,P}$, $a_{\theta,P}$, $\Gamma_{\theta,P}$, H_θ and $\Sigma_{\theta,P}$ are bounded uniformly over $(\theta,P)\in\mathcal{S}_n$. In the case where $\mathcal{C}=\widetilde{\mathcal{C}}\times\mathbb{R}^{d_{g_2}}$, assume that the last d_{g_2} elements of $\hat{k}$ are zero with probability one for all $P\in\mathcal{P}$.*

THEOREM C.3. *Suppose that Assumptions C.2, C.3, and C.4 hold and let $\hat{c}$ be defined above with $\hat{k}$, $\hat{\Gamma}$, and $\hat{H}$ satisfying* (S21). *Then*

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{S}_n} P\big(h(\theta)\in[\hat{c},\infty)\big)\geq 1-\alpha,$$

and

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{\theta\in\Theta_I(P;\mathbb{R}^{d_\theta}\times\mathcal{D},\Theta_n)}\Big\{\sqrt{n}q_{\beta,P}\big(h(\theta)-\hat{c}\big) \\ -\Big[\overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P},0)-\underline{\text{bias}}_{\mathcal{D}}(k_{\theta,P},0)+(z_{1-\alpha}+z_\beta)\sqrt{k_{\theta,P}'\Sigma_{\theta,P}k_{\theta,P}}\Big]\Big\}\leq 0.$$

PROOF. If $\mathcal{C}=\widetilde{\mathcal{C}}\times\mathbb{R}^{d_{g_2}}$ with $\widetilde{\mathcal{C}}$ compact, the theorem can equivalently be stated as holding with $\hat{k}$ redefined to be the vector in $\mathbb{R}^{d_{g_1}}$ that contains the first d_{g_1} elements of the original sensitivity $\hat{k}$, and with other objects redefined similarly. Therefore, it suffices to consider the case where $\mathcal{C}$ is compact.

Note that

$$\begin{aligned}
&\sqrt{n}\big(\hat{h}-h(\theta)\big)\\
&\quad=H_\theta\sqrt{n}(\hat{\theta}_{\text{initial}}-\theta)+\hat{k}\sqrt{n}\hat{g}(\theta)+\hat{k}\sqrt{n}\big(\hat{g}(\hat{\theta}_{\text{initial}})-\hat{g}(\theta)\big)+\hat{a}+o_P(1)\\
&\quad=H_\theta\sqrt{n}(\hat{\theta}_{\text{initial}}-\theta)+\hat{k}\sqrt{n}\big(\hat{g}(\theta)-g_P(\theta)\big)+\hat{k}'c+\hat{k}\sqrt{n}\Gamma_{\theta,P}(\hat{\theta}_{\text{initial}}-\theta)+\hat{a}+o_P(1)\\
&\quad=\big(H_\theta+k_{\theta,P}'\Gamma_{\theta,P}\big)\sqrt{n}(\hat{\theta}_{\text{initial}}-\theta)+k_{\theta,P}'c+a_{\theta,P}+k_{\theta,P}'\sqrt{n}\big(\hat{g}(\theta)-g_P(\theta)\big)+o_P(1),
\end{aligned}$$

where $c=\sqrt{n}g_P(\theta)$ and the $o_P(1)$ terms are uniform over $(\theta,P)\in\mathcal{S}_n$ (the last equality uses the fact that $\mathcal{C}$ is compact). By Assumption C.4 and (S21), $H_\theta+k_{\theta,P}'\Gamma_{\theta,P}=0$ so this implies

$$\sqrt{n}\big(\hat{h}-h(\theta)\big)=k_{\theta,P}'c+a_{\theta,P}+k_{\theta,P}'\sqrt{n}\big(\hat{g}(\theta)-g_P(\theta)\big)+o_P(1) \tag{S24}$$

uniformly over $(\theta,P)\in\mathcal{S}_n$. By compactness of $\mathcal{C}$ and Assumption C.4, we also have

$$\overline{\text{bias}}_{\mathcal{C}}(\hat{k},\hat{a})=\overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P},a_{\theta,P})+o_P(1),\quad \hat{k}'\hat{\Sigma}\hat{k}=k_{\theta,P}'\Sigma_{\theta,P}k_{\theta,P}+o_P(1)$$

uniformly over $(\theta,P)\in\mathcal{S}_n$. Thus,

$$\begin{aligned}
\sqrt{n}\big(\hat{c}-h(\theta)\big)&=\sqrt{n}\big(\hat{h}-h(\theta)\big)-\overline{\text{bias}}_{\mathcal{C}}(\hat{k},\hat{a})-z_{1-\alpha}\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}\\
&=k_{\theta,P}'c+a_{\theta,P}+k_{\theta,P}'\sqrt{n}\big(\hat{g}(\theta)-g_P(\theta)\big)-\overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P},a_{\theta,P})\\
&\quad-z_{1-\alpha}\sqrt{k_{\theta,P}'\Sigma_{\theta,P}k_{\theta,P}}+o_P(1)
\end{aligned} \tag{S25}$$

uniformly over $(\theta,P)\in\mathcal{S}_n$. Since $k_{\theta,P}'c+a_{\theta,P}-\overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P},a_{\theta,P})\leq 0$ by definition, the first part of the theorem (coverage) now follows from Assumption C.3. For the last part of the theorem, note that, using the above display and the fact that $k_{\theta,P}'c+a_{\theta,P}\geq$

$\underline{\text{bias}}_{\mathcal{D}}(k_{\theta,P}, a_{\theta,P})$ for any (θ, P) with $c = \sqrt{n}E_P g(w_i, \theta) \in \mathcal{D}$, it follows that $\sqrt{n}(h(\theta) - c)$ is less than or equal to

$$\overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P}, a_{\theta,P}) - \underline{\text{bias}}_{\mathcal{D}}(k_{\theta,P}, a_{\theta,P}) + z_{1-\alpha}\sqrt{k'_{\theta,P}\Sigma_{\theta,P}k_{\theta,P}} + k'_{\theta,P}\sqrt{n}\big(\hat{g}(\theta) - g_P(\theta)\big) + o_P(1)$$

uniformly over (θ, P) with $\sqrt{n}E_P g(w_i, \theta) \in \mathcal{D}$. This, along with Assumption C.3, gives the last part of the theorem. □

THEOREM C.4. *Suppose that Assumptions C.2, C.3, and C.4 hold and let $\hat{h}$ and $\hat{\chi}$ be defined above with $\hat{k}$, $\hat{a}$, $\hat{\Gamma}$, and $\hat{H}$ satisfying (S21) and (S22). Then*

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{S}_n} P\big(h(\theta) \in \{\hat{h} \pm \hat{\chi}\}\big) \geq 1 - \alpha.$$

In addition, we have

$$\sqrt{n}\hat{\chi} - \text{cv}_\alpha\left(\frac{\overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P}, a_{\theta,P})}{\sqrt{k'_{\theta,P}\Sigma_{\theta,P}k_{\theta,P}}}\right)\sqrt{k'_{\theta,P}\Sigma_{\theta,P}k_{\theta,P}} \xrightarrow{p} 0$$

uniformly over $(\theta, P) \in \mathcal{S}_n$.

PROOF. As with Theorem C.3, it suffices to consider the case where $\mathcal{C}$ is compact. Let (θ_n, P_n) be a sequence in $\mathcal{S}_n$ and let $c_n = \sqrt{n}g_{P_n}(\theta_n)$. Let $b_n = k'_{\theta_n,P_n}c_n + a_{\theta_n,P_n}$, $\text{sd}_n = \sqrt{k'_{\theta_n,P_n}\Sigma_{\theta_n,P_n}k_{\theta_n,P_n}}$ and $\overline{b}_n = \overline{\text{bias}}_{\mathcal{C}}(k_{\theta_n,P_n}, a_{\theta_n,P_n})$. Note that, by (S22), $\overline{\text{bias}}_{\mathcal{C}}(k_{\theta_n,P_n}, a_{\theta_n,P_n}) = -\underline{\text{bias}}_{\mathcal{C}}(k_{\theta_n,P_n}, a_{\theta_n,P_n})$ when Assumption C.4 holds. It therefore follows that $-\overline{b}_n \leq b_n \leq \overline{b}_n$.

Let $Z_n = \sqrt{n}k'_{\theta_n,P_n}(\hat{g}(\theta_n) - g_{P_n}(\theta_n))/\text{sd}_n$. Note that Z_n converges in distribution (under P_n) to a $\mathcal{N}(0, 1)$ random variable by Assumption C.3. By (S24),

$$\sqrt{n}\big(\hat{h} - h(\theta_n)\big) = b_n + \text{sd}_n Z_n + o_{P_n}(1).$$

Using the fact that sd_n is bounded away from zero and $\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}/\text{sd}_n$ converges in probability to one under P_n, it also follows that

$$\sqrt{n}\big(\hat{h} - h(\theta_n)\big)/\sqrt{\hat{k}'\hat{\Sigma}\hat{k}} = b_n/\text{sd}_n + Z_n + o_{P_n}(1).$$

Also, by Assumption C.4, we have, for a large enough constant K,

$$\left|\text{cv}_\alpha\left(\frac{\overline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a})}{\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}}\right) - \text{cv}_\alpha\left(\frac{\overline{b}_n}{\text{sd}_n}\right)\right| \leq K\big\{\big[\overline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a}) - \overline{b}_n\big] + \big[\sqrt{\hat{k}'\hat{\Sigma}\hat{k}} - \text{sd}_n\big]\big\} \xrightarrow{p} 0.$$

This, along with the fact that $\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}/\text{sd}_n$ converges in probability to one under P_n, gives the second part of the theorem. Furthermore, it follows from the above display that

$$P_n\big(h(\theta_n) > \hat{h} + \hat{\chi}\big) = P_n\left(\frac{\sqrt{n}\big(\hat{h} - h(\theta_n)\big)}{\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}} < -\text{cv}_\alpha\big(\overline{\text{bias}}_{\mathcal{C}}(\hat{k}, \hat{a})/\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}\big)\right)$$

$$= P_n\big(b_n/\mathrm{sd}_n + Z_n < -\mathrm{cv}_\alpha(\overline{b}_n/\mathrm{sd}_n) + o_{P_n}(1)\big)$$
$$= \Phi\big(-b_n/\mathrm{sd}_n - \mathrm{cv}_\alpha(\overline{b}_n/\mathrm{sd}_n)\big) + o(1).$$

Similarly,

$$P_n\big(h(\theta_n) < \hat{h} - \hat{\chi}\big) = P_n\left(\frac{\sqrt{n}\big(\hat{h} - h(\theta_n)\big)}{\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}} > \mathrm{cv}_\alpha\Big(\overline{\mathrm{bias}}_{\mathcal{C}}(\hat{k}, \hat{a})/\sqrt{\hat{k}'\hat{\Sigma}\hat{k}}\Big)\right)$$
$$= P_n\big(b_n/\mathrm{sd}_n + Z_n > \mathrm{cv}_\alpha(\overline{b}_n/\mathrm{sd}_n) + o_{P_n}(1)\big)$$
$$= 1 - \Phi\big(-b_n/\mathrm{sd}_n + \mathrm{cv}_\alpha(\overline{b}_n/\mathrm{sd}_n)\big) + o(1).$$

Thus, the probability of the CI not covering is given, up to $o(1)$, by

$$1 - \Phi\big(-b_n/\mathrm{sd}_n + \mathrm{cv}_\alpha(\overline{b}_n/\mathrm{sd}_n)\big) + \Phi\big(-b_n/\mathrm{sd}_n - \mathrm{cv}_\alpha(\overline{b}_n/\mathrm{sd}_n)\big).$$

This is the probability that the absolute value of a $\mathcal{N}(b_n/\mathrm{sd}_n, 1)$ variable is greater than $\mathrm{cv}_\alpha(\overline{b}_n/\mathrm{sd}_n)$, which is less than $1-\alpha$ since $|b_n| \le \overline{b}_n$. □

We now specialize to the case where the optimal weights are used. We make a uniform consistency assumption on $\hat{\Gamma}$, $\hat{H}$, and $\hat{\Sigma}$, as well as assumptions on the rank of H, Γ, and Σ. The latter are standard regularity conditions for the correctly specified ($\mathcal{C} = \{0\}$) case.

Assumption C.5. *The estimators $\hat{\Gamma}$, $\hat{H}$, and $\hat{\Sigma}$ are full rank with probability one and satisfy $\hat{\Gamma} - \Gamma_{\theta,P} = o_P(1)$, $\hat{H} - H_\theta = o_P(1)$ and $\hat{\Sigma} - \Sigma_{\theta,P} = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$.*

Assumption C.6. *There exists a compact set $\mathcal{B}$ that contains the set $\{(H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P}) : \theta \in \Theta_n, P \in \mathcal{P}\}$ for all n, such that (i) in the case where $\mathcal{C}$ is compact, $H \neq 0$ and Γ and Σ are full rank for any $(H, \Gamma, \Sigma) \in \mathcal{B}$ or (ii) in the case where $\mathcal{C} = \widetilde{\mathcal{C}} \times \mathbb{R}^{d_{g_2}}$ with $\widetilde{\mathcal{C}}$ compact, the same holds for the submatrices corresponding to the first d_{g_1} moments.*

Using these assumptions, we can verify that Assumption C.4 holds with weights $k_{\theta,P}$ that achieve the efficiency bound in Theorem C.1 and nearly achieve the efficiency bound in Theorem C.2. This gives the following results.

Theorem C.5. *Suppose that Assumptions C.2, C.3, C.5, and C.6 hold and let $\hat{c}$ be defined above with $\hat{k} = k(\delta_\beta, \hat{H}, \hat{\Gamma}, \hat{\Sigma})$. Then*

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{S}_n} P\big(h(\theta) \in [\hat{c}, \infty)\big) \ge 1 - \alpha$$

and

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sup_{\theta\in\Theta_I(P;\mathbb{R}^{d_\theta}\times\mathcal{D},\Theta_n)} \big[\sqrt{n} q_{\beta,P}\big(h(\theta) - \hat{c}\big) - \omega\big(\delta_\beta; \mathbb{R}^{d_\theta} \times \mathcal{C}, \mathbb{R}^{d_\theta} \times \mathcal{D}, H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P}\big)\big]$$
$$\le 0.$$

PROOF. In the case where $\mathcal{C}$ is compact, it follows from Lemma E.6 in Appendix E.3.2, $k(\delta, H, \Gamma, \Sigma)$ is continuous on $\{\delta\} \times \mathcal{B}$. Since $\mathcal{B}$ is compact, this means that $k(\delta, H, \Gamma, \Sigma)$ is uniformly continuous. Thus, Assumption C.5 implies that $\hat{k}$ satisfies Assumption C.4 with $k_{\theta,P} = k(\delta, H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P})$. Furthermore, $\hat{k}$ satisfies (S21) by assumption. By properties of the modulus (equation (24) in Armstrong and Kolesár (2018)),

$$\overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P}, 0) - \underline{\text{bias}}_{\mathcal{D}}(k_{\theta,P}, 0) + (z_{1-\alpha} + z_\beta)\sqrt{k_{\theta,P}' \Sigma_{\theta,P} k_{\theta,P}}$$
$$= \omega\big(\delta_\beta; \mathbb{R}^{d_\theta} \times \mathcal{C}, \mathbb{R}^{d_\theta} \times \mathcal{D}, H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P}\big)$$

for this $k_{\theta,P}$. Applying Theorem C.3 gives the result.

In the case where $\mathcal{C} = \widetilde{\mathcal{C}} \times \mathbb{R}^{d_{g_2}}$ with $\widetilde{\mathcal{C}}$ compact, the last d_{g_2} elements of $\hat{k}$ are equal to zero as required by Assumption C.4, and the first d_{g_1} elements are the same as the weights computed from the modulus problem with the last d_{g_2} components thrown away and H, Γ and Σ redefined to be the submatrices corresponding to the first d_{g_1} elements of the moments. Thus, the same arguments apply in this case. □

For two-sided CIs, we consider weights $\hat{k} = k(\delta^*(\hat{H}, \hat{\Gamma}, \hat{\Sigma}), \hat{H}, \hat{\Gamma}, \hat{\Sigma})$ given by (S23) with $\mathcal{G} = \mathcal{F} = \mathbb{R}^{d_\theta} \times \mathcal{C}$, where δ^* may depend on the data through $\hat{H}$, $\hat{\Gamma}$, and $\hat{\Sigma}$. If δ^* is chosen to optimize CI length, it will be given by $\delta_\chi(\hat{H}, \hat{\Gamma}, \hat{\Sigma})$ where

$$\delta_\chi(H, \Gamma, \Sigma) = \operatorname*{argmin}_{\delta} \text{cv}_\alpha\left(\frac{\omega(\delta)}{2\omega'(\delta)} - \frac{\delta}{2}\right)\omega'(\delta), \tag{S26}$$

where $\omega(\delta) = \omega(\delta; \mathbb{R}^{d_\theta} \times \mathcal{C}, \mathbb{R}^{d_\theta} \times \mathcal{C}, H, \Gamma, \Sigma)$ is the single class modulus (see Section 3.4 in Armstrong and Kolesár (2018)).

We make a continuity assumption on δ^*.

ASSUMPTION C.7. *δ^* is a continuous function of its arguments on the set $\mathcal{B}$ given in Assumption* C.6.

THEOREM C.6. *Suppose that Assumptions* C.2, C.3, C.5, C.6, *and* C.7 *hold and let $\hat{h}$ be defined above with $\hat{k} = k(\delta^*(\hat{H}, \hat{\Gamma}, \hat{\Sigma}), \hat{H}, \hat{\Gamma}, \hat{\Sigma})$. Then the conclusion of Theorem* C.4 *holds. If, in addition, $\delta^* = \delta_\chi(\hat{H}, \hat{\Gamma}, \hat{\Sigma})$ for δ_χ the CI length optimizing choice of δ given in* (S26), *then the half-length $\hat{\chi}$ satisfies $\sqrt{n}\hat{\chi} = \chi(\theta, P) + o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$, where*

$$\chi(\theta, P) = \min_{\delta} \text{cv}_\alpha\left(\frac{\omega(\delta)}{2\omega'(\delta)} - \frac{\delta}{2}\right)\omega'(\delta), \quad \omega(\delta) = \omega\big(\delta; \mathbb{R}^{d_\theta} \times \mathcal{C}, \mathbb{R}^{d_\theta} \times \mathcal{C}, H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P}\big).$$

PROOF. The result follows from using the same arguments as in the proof of Theorem C.5, along with continuity of δ^*, to verify Assumption C.4. The form of the limiting half-length for the optimal weights follows from properties of the modulus (see Section 3.4 in Armstrong and Kolesár (2018)). □

C.6 *Centrosymmetric case*

Theorem 4.1 in Section 4 gives a bound for two-sided CIs in the case where $\mathcal{C}$ is centrosymmetric. This follows from applying Theorems C.6 and C.2 in the centrosymmetric case. In particular, comparing the asymptotic length in Theorem C.6 to the bound in Theorem C.2 and using the fact that $\omega(\delta; \mathbb{R}^{d_\theta} \times \mathcal{C}, \{0\}, H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P}) = \omega(\delta; \{0\}, \mathbb{R}^{d_\theta} \times \mathcal{C}, H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P}) = \frac{1}{2}\omega(2\delta; \mathbb{R}^{d_\theta} \times \mathcal{C}, \mathbb{R}^{d_\theta} \times \mathcal{C}, H_\theta, \Gamma_{\theta,P}, \Sigma_{\theta,P})$ when $\mathcal{C}$ is centrosymmetric gives the bound $\kappa_*(H_\theta, \Gamma_{\theta,P_0}, \Sigma_{\theta,P_0}, \mathcal{C})$ from the statement of Theorem 4.1. This corresponds to the bound in Corollary 3.3 of Armstrong and Kolesár (2018). The universal lower bound for κ_* follows from the following result.

Theorem C.7. *For any H, Γ, Σ, and $\mathcal{C}$, the efficiency κ_* given in equation* (18) *is lower bounded by*

$$\left(z_{1-\alpha}(1-\alpha) - \tilde{z}_\alpha \Phi(\tilde{z}_\alpha) + \phi(z_{1-\alpha}) - \phi(\tilde{z}_\alpha)\right)/z_{1-\alpha/2},$$

where $\tilde{z}_\alpha = z_{1-\alpha} - z_{1-\alpha/2}$ and Φ and ϕ denote the standard normal cdf, and pdf respectively. The lower bound is sharp in the sense that it holds with equality if $\omega(\delta) = K_0 \min\{\delta, 2z_{1-\alpha/2}\}$, for some constant K_0.

Proof. Since $\mathrm{cv}_\alpha(b) \leq b + z_{1-\alpha/2}$, the denominator in equation (18) is upper bounded by

$$\begin{aligned} &\min_\delta 2\,\mathrm{cv}_\alpha\left(\frac{\omega(\delta)}{2\omega'(\delta)} - \frac{\delta}{2}\right)\omega'(\delta) \\ &\leq 2\,\mathrm{cv}_\alpha\left(\frac{\omega(2z_{1-\alpha/2})}{2\omega'(2z_{1-\alpha/2})} - z_{1-\alpha/2}\right)\omega'(2z_{1-\alpha/2}) \leq \omega(2z_{1-\alpha/2}). \end{aligned} \tag{S27}$$

On the other hand, the numerator in equation (18) can be decomposed as

$$\begin{aligned} &(1-\alpha)E\left[\omega\left(2(z_{1-\alpha} - Z)\right) \mid Z \leq z_{1-\alpha}\right] \\ &= E\left[\omega\left(2(z_{1-\alpha} - Z)\right)\mathbb{I}\{Z \leq z_{1-\alpha} - z_{1-\alpha/2}\}\right] \\ &\quad + E\left[\omega\left(2(z_{1-\alpha} - Z)\right)\mathbb{I}\{z_{1-\alpha} - z_{1-\alpha/2} \leq Z \leq z_{1-\alpha}\}\right]. \end{aligned}$$

Since the modulus $\omega(\delta)$ is nondecreasing, the first summand is lower bounded by

$$E\left[\omega(2z_{1-\alpha/2})\mathbb{I}\{Z \leq z_{1-\alpha} - z_{1-\alpha/2}\}\right] = \omega(2z_{1-\alpha/2})\Phi(z_{1-\alpha} - z_{1-\alpha/2}).$$

Since the modulus $\omega(\delta)$ is concave, $\omega(2(z_{1-\alpha} - Z)) \geq (z_{1-\alpha} - Z)/z_{1-\alpha/2} \cdot \omega(2z_{1-\alpha/2})$, so that the second summand is lower bounded by

$$\begin{aligned} &\frac{\omega(2z_{1-\alpha/2})}{z_{1-\alpha/2}} E\left[(z_{1-\alpha} - Z)\mathbb{I}\{z_{1-\alpha} - z_{1-\alpha/2} \leq Z \leq z_{1-\alpha}\}\right] \\ &= \frac{\omega(2z_{1-\alpha/2})}{z_{1-\alpha/2}}\left(z_{1-\alpha}\left(1 - \alpha - \Phi(z_{1-\alpha} - z_{1-\alpha/2})\right) + \phi(z_{1-\alpha}) - \phi(z_{1-\alpha} - z_{1-\alpha/2})\right), \end{aligned}$$

where the equality follows by the formula for the expectation of a truncated normal random variable. Combining the two preceding displays then yields

$$(1-\alpha)E\big[\omega\big(2(z_{1-\alpha}-Z)\big)\mid Z\le z_{1-\alpha}\big] \\ \ge \omega(2z_{1-\alpha/2})\frac{z_{1-\alpha}(1-\alpha)-\tilde{z}_\alpha\Phi(\tilde{z}_\alpha)+\phi(z_{1-\alpha})-\phi(\tilde{z}_\alpha)}{z_{1-\alpha/2}}, \tag{S28}$$

where $\tilde{z}_\alpha = z_{1-\alpha}-z_{1-\alpha/2}$. Combining this with the bound in (S27) then yields the result. The sharpness of the bound for the case $\omega(\delta)=K_0\min\{\delta, 2z_{1-\alpha/2}\}$ follows from by noting that in this case, both (S27) and (S28) hold as equalities. □

For the one-sided case, we obtain the following bound.

Theorem C.8. *Consider the setting of Theorem C.5, with $\mathcal{C}$ centrosymmetric. Then the weights $\hat{k}=\hat{k}(\delta_\beta,\hat{H},\hat{\Gamma},\hat{\Sigma})$ with $\mathcal{D}=\mathcal{C}$ are identical to the weights $\hat{k}(\delta_{\tilde{\beta}},\hat{H},\hat{\Gamma},\hat{\Sigma})$ computed with $\mathcal{D}=\{0\}$, but with $\tilde{\beta}=\Phi((z_\beta-z_{1-\alpha})/2)$. Furthermore, letting $\hat{c}_{\text{minimax}}$ denote the lower endpoint of the CI computed with these weights ($\hat{k}(\delta_\beta,\hat{H},\hat{\Gamma},\hat{\Sigma})$ with $\mathcal{D}=\mathcal{C}$), we have*

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{\theta\in\Theta_I(P;\mathbb{R}^{d_\theta}\times\{0\},\Theta_n)}\Big\{\sqrt{n}q_{\beta,P}\big(h(\theta)-\hat{c}_{\text{minimax}}\big)-\frac{1}{2}\big[\omega_{\theta,P}(\delta_\beta)+\delta_\beta\omega'_{\theta,P}(\delta_\beta)\big]\Big\}\le 0,$$

where $\omega_{\theta,P}(\delta)=\omega(\delta;\mathbb{R}^{d_\theta}\times\mathcal{C},\mathbb{R}^{d_\theta}\times\mathcal{C},H_\theta,\Gamma_{\theta,P},\Sigma_{\theta,P})$. For $\hat{c}$ computed instead with $\mathcal{D}=\{0\}$, we obtain

$$\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{\theta\in\Theta_I(P;\mathbb{R}^{d_\theta}\times\{0\},\Theta_n)}\Big\{\sqrt{n}q_{\beta,P}\big(h(\theta)-\hat{c}\big)-\frac{1}{2}\omega_{\theta,P}(2\delta_\beta)\Big\}\le 0.$$

Proof. The first statement follows from Corollary 3.2 in Armstrong and Kolesár (2018). The second statement follows from applying Theorem C.3 as in the proof of Theorem C.5, noting that $\underline{\text{bias}}_{\{0\}}(k_{\theta,P},0)=0$, and using arguments from the proof of Corollary 3.2 in Armstrong and Kolesár (2018). The last statement follows from Theorem C.5 and the fact that $\omega(\delta;\mathbb{R}^{d_\theta}\times\mathcal{C},\mathbb{R}^{d_\theta}\times\{0\},H_\theta,\Gamma_{\theta,P},\Sigma_{\theta,P})=\frac{1}{2}\omega(2\delta;\mathbb{R}^{d_\theta}\times\mathcal{C},\mathbb{R}^{d_\theta}\times\mathcal{C},H_\theta,\Gamma_{\theta,P},\Sigma_{\theta,P})$. □

Thus, directing power toward the correctly specified case yields the same one-sided CI once one changes the quantile over which one optimizes excess length. If one does attempt to direct power, the scope for doing so is bounded by a factor of

$$\kappa_*^{\text{OCI},\beta}(H_\theta,\Gamma_{\theta,P_0},\Sigma_{\theta,P_0},\mathcal{C})=\frac{\omega_{\theta,P}(2\delta_\beta)}{\omega_{\theta,P}(\delta_\beta)+\delta_\beta\omega'_{\theta,P}(\delta_\beta)}. \tag{S29}$$

This gives a bound for the one-sided case analogous to the bound κ_* in equation (18) for two-sided CIs.

A consistent estimate of these bounds can be obtained by plugging in $\omega(\delta;\mathbb{R}^{d_\theta}\times\mathcal{C},\mathbb{R}^{d_\theta}\times\mathcal{C},\hat{H},\hat{\Gamma},\hat{\Sigma})$ for $\omega_{\theta,P}(\delta)=\omega(\delta;\mathbb{R}^{d_\theta}\times\mathcal{C},\mathbb{R}^{d_\theta}\times\mathcal{C},H_\theta,\Gamma_{\theta,P},\Sigma_{\theta,P})$. Table 2 reports estimates of this bound under different forms of misspecification in the empirical application in Section 6.

Appendix D: Global misspecification

We now describe two approaches to the construction of CIs that are robust to global misspecification. The first approach is generally applicable, and, for a one-sides CIs, yields CIs that are asymptotically equivalent under local misspecification to the CIs proposed in the main text. The second approach also exhibits this equivalence property for two-sided CIs, but the regularity conditions it imposes may not be satisfied in all applications.

Before describing the procedures in Appendices D.1 and D.2 below, let us briefly describe the setup. Under global misspecification, the true parameter θ_0 satisfies

$$g_P(\theta_0) = E_P g(w_i, \theta_0) = \tilde{c}, \quad \tilde{c} \in \widetilde{\mathcal{C}}, \tag{S30}$$

where $\widetilde{\mathcal{C}}$ is fixed with the sample size n. To accommodate both local and global misspecification with the same notation, we consider a sequence $\widetilde{\mathcal{C}} = \widetilde{\mathcal{C}}_n$ of sets. Under global misspecification $\widetilde{\mathcal{C}}_n$ is fixed with n, whereas, under local misspecification, $\widetilde{\mathcal{C}}_n = \mathcal{C}/\sqrt{n} = \{c/\sqrt{n} : c \in \mathcal{C}\}$ where $\mathcal{C}$ is fixed with n. The rest of the setup is the same as the formal setup in Section 4.2 and Appendix C: we are interested in a CI for $h(\theta)$ that satisfies the coverage condition in equation (20), where $\mathcal{S}_n = \{(\theta, P) \in \Theta_n \times \mathcal{P} : g_P(\theta) \in \widetilde{\mathcal{C}}\}$ denotes the set of pairs (θ, P) such that θ is in the identified set under P.

For concreteness, we focus on GMM estimators. We treat the weighting matrix W as given, and construct CIs that are asymptotically equivalent to the CIs given in Section 2.1 with sensitivity $k' = -H(\Gamma' W \Gamma)^{-1}\Gamma' W$. To make these CIs optimal under local misspecification, we need to choose the weighting matrix W so that this sensitivity is optimal under local misspecification. This can be done by computing the optimal sensitivity $\hat{k}$ under local misspecification using first stage estimates, following our implementation in equation (20), and then computing an equivalent GMM weighting matrix as described in Remark 3.2.

D.1 *CIs based on recentering the moments*

Let $\mathcal{I}_{\tilde{c}}$ be a family of CIs, indexed by $\tilde{c} \in \widetilde{\mathcal{C}}$, such that, for each $\tilde{c}$, $\mathcal{I}_{\tilde{c}}$ is asymptotically valid for the GMM model defined by the moment function $\theta \mapsto g(w_i, \theta) - \tilde{c}$. We consider the CI $\mathcal{I} = \bigcup_{\tilde{c} \in \widetilde{\mathcal{C}}} \mathcal{I}_{\tilde{c}}$. Since this CI contains a CI based on the moment conditions $\theta \mapsto \hat{g}(\theta) - \tilde{c}_0$, where $\tilde{c}_0 = E_P g(w_i, \theta_0)$ is the true value of $\tilde{c}$, it will have correct asymptotic coverage under standard conditions. (While we omit a formal statement, we note that this follows by showing that $\mathcal{I}_{c_n}$ has correct coverage under a drifting sequence of moment functions $\tilde{g}_n(w_i, \theta) = g(w_i, \theta) - \tilde{c}_n$ for sequences $\tilde{c}_n \in \widetilde{\mathcal{C}}$. This follows from the usual arguments for asymptotic coverage of GMM estimators under correct specification.) As we discuss in Appendix D.1.1 below, this CI can be computed using constrained optimization, where the objective function will typically be smooth so long as $\theta \mapsto g(w_i, \theta)$ is smooth. However, since the problem involves minimization of a GMM objective function, it will typically not be convex.

We now show that this approach can be used to construct a one-sided CI that is asymptotically equivalent under local misspecification to the CI proposed in the main text. We consider CIs based on the GMM estimator $\hat{\theta}_{W,\tilde{c}} = \arg\min_\theta [\hat{g}(\theta) - \tilde{c}]' W [\hat{g}(\theta) - \tilde{c}]$ based on the moment function $g(w_i, \theta) - \tilde{c}$. Let $\hat{\Gamma}_\theta$ denote an estimate of the Jacobian matrix $\Gamma_{\theta,P} = \frac{d}{dt'} E_P g(w_i, t)|_{t=\theta}$. Under local misspecification, the identified set for θ shrinks toward a point, so that the same estimate $\hat{\Sigma}$ is consistent for $\Sigma_{\theta,P}$ uniformly over the identified set. This is no longer the case under global misspecification, and we instead need to use a class of estimates $\hat{\Sigma}_\theta$ for $\Sigma_{\theta,P}$ indexed by θ. In the i.i.d. case, we can take $\hat{\Sigma}_\theta = \frac{1}{n}\sum_{i=1}^n g(w_i, \theta) g(w_i, \theta)'$, in which case the estimate of $\hat{\Sigma}$ in equation (20) corresponds to $\hat{\Sigma} = \hat{\Sigma}_{\hat{\theta}_{\text{initial}}}$. We allow for the possibility that W is data dependent, in which case we will assume that it converges to some W_P, which may depend on the underlying distribution P. Let $\hat{k}'_\theta = -H_\theta(\hat{\Gamma}'_\theta W \hat{\Gamma}_\theta)^{-1}\hat{\Gamma}'_\theta W$ denote an estimate of the sensitivity $k'_{\theta,P} = -H_\theta(\Gamma'_{\theta,P} W_P \Gamma_{\theta,P})^{-1}\Gamma'_{\theta,P} W$. This gives the standard error $\text{se}_{\hat{\theta}_{W,\tilde{c}}}$ where $\text{se}_\theta = \sqrt{\hat{k}'_\theta \hat{\Sigma}_\theta \hat{k}_\theta / n}$. The nominal $1-\alpha$ one-sided CI based on the given $\tilde{c}$ is then $[h(\hat{\theta}_{W,\tilde{c}}) - z_{1-\alpha}\,\text{se}_{W,\tilde{c}}, \infty)$. The lower endpoint of a one-sided CI that is asymptotically valid for the set $\widetilde{\mathcal{C}}$ is then given by $[\hat{c}_{\text{glob}}, \infty)$ where

$$\hat{c}_{\text{glob}} = \inf_{\tilde{c}\in\widetilde{\mathcal{C}}} \big[h(\hat{\theta}_{W,\tilde{c}}) - z_{1-\alpha}\,\text{se}_{\hat{\theta}_{W,\tilde{c}}}\big].$$

The CI proposed in the main text takes the form $[\hat{c}_{\text{loc}}, \infty)$ where

$$\hat{c}_{\text{loc}} = h(\hat{\theta}_{\text{initial}}) + \hat{k}'\hat{g}(\hat{\theta}_{\text{initial}}) - \sup_{c\in\mathcal{C}} \hat{k}'c/\sqrt{n} - z_{1-\alpha}\sqrt{\hat{k}\hat{\Sigma}\hat{k}}/\sqrt{n}.$$

We assume that the sensitivity $\hat{k}$ corresponds to the same GMM weighting matrix, so that Assumptions C.2, C.3, and C.4 hold with $k_{\theta,P}$, $\Gamma_{\theta,P}$, H_θ and $\Sigma_{\theta,P}$ given above. In addition, we use some further regularity conditions for the estimator $\hat{\theta}_{W,\tilde{c}}$.

Assumption D.1. *The estimator $\hat{\theta}_{W,\tilde{c}}$ satisfies*

$$\sup_{c\in\mathcal{C}} \big|h(\hat{\theta}_{W,c/\sqrt{n}}) - h(\theta) - k'_{\theta,P}\big[\hat{g}(\theta) - c/\sqrt{n}\big]\big| = o_P(1/\sqrt{n})$$

and

$$\sup_{c\in\mathcal{C}} \big|\hat{k}'_{\hat{\theta}_{W,c/\sqrt{n}}} \hat{\Sigma}_{\hat{\theta}_{W,c/\sqrt{n}}} \hat{k}_{\hat{\theta}_{W,c/\sqrt{n}}} - k'_{\theta,P}\Sigma_{\theta,P} k_{\theta,P}\big| = o_P(1)$$

uniformly over $(\theta, P) \in \mathcal{S}_n$, where $\widetilde{\mathcal{C}} = \widetilde{\mathcal{C}}_n = \mathcal{C}/\sqrt{n}$.

Assumption D.1 imposes an influence function representation for the GMM estimator $\hat{\theta}_{W,c}$, and a uniform consistency condition for the asymptotic variance estimator. We verify this assumption under primitive conditions in the linear IV setting in Appendix E.2.3. Note that, while the CI is robust to global misspecification, Assumption D.1 imposes conditions that are required only under local misspecification.

Theorem D.1. *Let $\hat{c}_{\text{loc}}$ and $\hat{c}_{\text{glob}}$ be given above with $\widetilde{\mathcal{C}} = \mathcal{C}/\sqrt{n}$, where $\mathcal{C}$ is a compact set. Suppose that Assumptions C.2, C.3, and C.4 hold for $\hat{c}_{\text{loc}}$, and Assumption D.1 holds for*

$\hat{c}_{\text{glob}}$, *with the same* $k_{\theta,P}$ *and* $\Sigma_{\theta,P}$, *and assume that, for each* $P \in \mathcal{P}$, *there exists* $\theta \in \Theta_n$ *such that* $(\theta, P) \in \mathcal{S}_n$ *(i.e., the identified set for* θ *under each* $P \in \mathcal{P}$ *is nonempty). Then* $\sqrt{n}(\hat{c}_{\text{loc}} - \hat{c}_{\text{glob}})$ *converges in probability to zero uniformly over* $P \in \mathcal{P}$.

Proof. Under these assumptions,

$$\begin{aligned}\hat{c}_{\text{glob}} &= \inf_{c \in \mathcal{C}} h(\theta) + k_{\theta,P}'\hat{g}(\theta) - k_{\theta,P}'c/\sqrt{n} - z_{1-\alpha}\sqrt{k_{\theta,P}'\Sigma_{\theta,P}k_{\theta,P}}/\sqrt{n} + o_P(1/\sqrt{n}) \\ &= h(\theta) + k_{\theta,P}'\hat{g}(\theta) - \overline{\text{bias}}_{\mathcal{C}}(k_{\theta,P}, 0)/\sqrt{n} - z_{1-\alpha}\sqrt{k_{\theta,P}'\Sigma_{\theta,P}k_{\theta,P}}/\sqrt{n} + o_P(1/\sqrt{n})\end{aligned}$$

uniformly over $(\theta, P) \in \mathcal{S}_n$. The result follows by noting that this matches the asymptotic expression for $\hat{c}_{\text{loc}}$ given in equation (S25) in the proof of Theorem C.3. □

The CI $[\hat{c}_{\text{glob}}, \infty)$ is valid under global misspecification under standard conditions, and Theorem D.1 shows that this CI is asymptotically equivalent to the CI $[\hat{c}_{\text{loc}}, \infty)$ under local misspecification.

D.1.1 *Computation* The problem of computing this CI can be written as a nested optimization problem, in which one minimizes the GMM objective function for a given $\tilde{c}$, and then minimizes the lower endpoint of the CI over $\tilde{c}$. Alternatively, in the spirit of recent papers on MPEC (e.g., Dubé, Fox, and Su (2012)), one can write this as an optimization problem over $\theta, \tilde{c}$ subject to the constraint that θ minimizes the GMM objective under $\tilde{c}$:

$$\min_{\theta,\tilde{c}} h(\theta) - z_{1-\alpha}\,\text{se}_\theta \quad \text{s.t. } \theta = \arg\min_t \big[\hat{g}(t) - \tilde{c}\big]'W\big[\hat{g}(t) - \tilde{c}\big], \tilde{c} \in \widetilde{\mathcal{C}}.$$

In the case where $\hat{g}(\theta)$ is smooth, one may also relax the constraint that θ minimizes the GMM objective by instead imposing only the first-order conditions:

$$\min_{\theta,\tilde{c}} h(\theta) - z_{1-\alpha}\,\text{se}_\theta \quad \text{s.t. } \hat{\Gamma}_\theta'W\big[\hat{g}(\theta) - \tilde{c}\big] = 0, \tilde{c} \in \widetilde{\mathcal{C}}.$$

Since a constraint is relaxed, this can only make the resulting CI more conservative.

D.2 *CIs based on misspecification-robust standard errors*

In some cases, it may be possible to construct estimates $\hat{B}$ of the worst-case asymptotic bias of the estimator $\hat{h}$ that are asymptotically normal. In such cases, one can construct CIs that are valid under global misspecification by using misspecification-robust standard errors.

Let $\theta_c^* = \text{argmin}_\theta(g_P(\theta) - c)'W(g_P(\theta) - c)$, so that under the model (S30), the true parameter is given by $\theta_0 = \theta_{\tilde{c}}^*$, and the pseudo-true parameter, the estimand of the GMM estimator $\hat{\theta}$ with weighting matrix W, is given by θ_0^*. Note that θ_c^* depends on P, but we leave this dependence implicit for clarity of notation. Hall and Inoue (2003) showed that under regularity conditions, if $\widetilde{\mathcal{C}}$ is fixed with n,

$$\sqrt{n}\big(\hat{h} - h(\theta_0^*)\big) \xrightarrow{d} \mathcal{N}(0, \Omega_h),$$

where, unlike in the correctly specified case, the asymptotic variance Ω_h generally depends on the weighting matrix W. Hall and Inoue (2003) also showed how to construct consistent estimates of Ω_h. Let $B_{\tilde{\mathcal{C}}} = \max_{\tilde{c}\in\tilde{\mathcal{C}}} |h(\theta_0^*) - h(\theta_{\tilde{c}}^*)|$ denote the worst-case bias. Suppose that we have available an estimator $\hat{B}$ of $B_{\tilde{\mathcal{C}}}$, such that $(\hat{h}, \hat{B})$ are jointly asymptotically normal. Then, by the delta method, for some asymptotic variances Ω_+ and Ω_-, it holds that

$$\frac{\sqrt{n}\big(\hat{h} - \hat{B} - \big(h(\theta_0^*) - B_{\tilde{\mathcal{C}}}\big)\big)}{\Omega_-^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1),$$
$$\frac{\sqrt{n}\big(\hat{h} + \hat{B} - \big(h(\theta_0^*) + B_{\tilde{\mathcal{C}}}\big)\big)}{\Omega_+^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{S31}$$

Note that $B_{\tilde{\mathcal{C}}}$, Ω_+, and Ω_- may depend on P, although we leave this implicit in the notation.

If one has available estimators $\hat{\Omega}_+$ and $\hat{\Omega}_-$ that satisfy

$$\hat{\Omega}_+/\Omega_+ \xrightarrow{p} 1, \qquad \hat{\Omega}_-/\Omega_- \xrightarrow{p} 1, \tag{S32}$$

one can construct one-sided CIs as $[\hat{h} - \hat{B} - z_{1-\alpha}\hat{\Omega}_-^{1/2}/\sqrt{n}, \infty)$, and $(-\infty, \hat{h} + \hat{B} + z_{1-\alpha}\hat{\Omega}_+^{1/2}/\sqrt{n}]$. Two-sided CIs can be constructed as

$$\widetilde{CI} = \big[\hat{h} - \mathrm{cv}_\alpha\big(\sqrt{n}\hat{B}/\hat{\Omega}_-^{1/2}\big) \cdot \hat{\Omega}_-^{1/2}/\sqrt{n}, \hat{h} + \mathrm{cv}_\alpha\big(\sqrt{n}\hat{B}/\hat{\Omega}_+^{1/2}\big) \cdot \hat{\Omega}_+^{1/2}/\sqrt{n}\big].$$

The next result shows that these CIs are valid under global misspecification.

THEOREM D.2. *Suppose that the convergence in* (S31) *and* (S32) *holds uniformly over* $(\theta, P) \in \mathcal{S}_n = \{(\theta, P) \in \Theta_n \times \mathcal{P} : g_P(\theta) \in \tilde{\mathcal{C}}\}$, *with* $\tilde{\mathcal{C}}$ *is fixed. Then*

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{S}_n} P\big(h(\theta) \in [\hat{h} - \hat{B} - z_{1-\alpha}\hat{\Omega}_-^{1/2}/\sqrt{n}, \infty)\big) \geq 1 - \alpha.$$

Suppose, in addition, that $\hat{B}/B_{\tilde{\mathcal{C}}} \xrightarrow{p} 1$ *and* $|\Omega_+ - \Omega_-|/\sqrt{n}B_{\tilde{\mathcal{C}}} \to 0$ *uniformly over* $(\theta, P) \in \mathcal{S}_n$. *Then*

$$\liminf_{n\to\infty} \inf_{(\theta,P)\in\mathcal{S}_n} P\big(h(\theta) \in \widetilde{CI}\big) \geq 1 - \alpha.$$

The proof of this theorem is deferred to Appendix D.2.1. Under global misspecification, when $\sqrt{n}B_{\tilde{\mathcal{C}}} \to \infty$, the condition $|\Omega_+ - \Omega_-|/\sqrt{n}B_{\tilde{\mathcal{C}}} \to 0$ holds if Ω_+ and Ω_- are of the same order, which is typically the case. In this case, $\widetilde{CI}$ is asymptotically equivalent to the CI $[\hat{h} - \hat{B} - z_{1-\alpha}\hat{\Omega}_-/\sqrt{n}, \hat{h} + \hat{B} + z_{1-\alpha}\hat{\Omega}_+/\sqrt{n}]$. Since in large samples, the uncertainty about the endpoints of the identified set $[h(\theta_0^*) - B_{\tilde{\mathcal{C}}}, h(\theta_0^*) + B_{\tilde{\mathcal{C}}}]$ dominates by the uncertainty about the location of the endpoints, it suffices to use a one-sided critical value $z_{1-\alpha}$ (see Imbens and Manski (2004), for a discussion).

Under local misspecification, if the estimator $\hat{\theta}$ is asymptotically linear with sensitivity k, $\sqrt{n}B_{\tilde{\mathcal{C}}} = \overline{\mathrm{bias}}_{\sqrt{n}\tilde{\mathcal{C}}}(k)$ is bounded, so that the condition $|\Omega_+ - \Omega_-|/\sqrt{n}B_{\tilde{\mathcal{C}}} \to 0$ holds if $\Omega_+ - \Omega_- \to 0$. This is indeed the case if $\sqrt{n}\hat{B} = \overline{\text{bias}}_{\sqrt{n}\tilde{\mathcal{C}}}(k) + o_p(1)$, since then Ω_+ and Ω_- both equal $k'\Sigma k$. In this case, $\widetilde{CI}$ is asymptotically equivalent to the CI in equation (7). The CI thus automatically adapts to the misspecification magnitude.

Example D.1 (Linear IV model). To give an example of a setting in which the condition (S31) holds, consider the linear instrumental variables (IV) model from Section 5.1. In particular, suppose $h = H\theta$, and suppose that $g(\theta) = E[z_i(y_i - x_i'\theta)] \in \tilde{\mathcal{C}}$, with $\tilde{\mathcal{C}} = \{\sqrt{n}B\gamma\colon \|\gamma\| \leq M_n\}$, $B = E[z_i z_{Ii}']$, and for concreteness, suppose $\|\cdot\|$ corresponds to an ℓ_2 norm. If $M_n = M$ is fixed, this reduces to the local misspecification setup in the main text, but if $M_n = \sqrt{n}M$, the misspecification is global. Consider the 2SLS estimator $\hat{h} = \hat{k}' \sum_{i=1}^n z_i y_i$, with $\hat{k} = -H(\hat{\Gamma}'\hat{W}\hat{\Gamma})^{-1}\hat{\Gamma}'\hat{W}$, $\hat{\Gamma} = -n^{-1}\sum_{i=1}^n z_i x_i'$ and $\hat{W}^{-1} = \sum_{i=1}^n z_i z_i'$; let $\Gamma = -E[z_i x_i']$, $W = E[z_i z_i']^{-1}$, and $k = -H(\Gamma' W\Gamma)^{-1}\Gamma' W$.

Then $h(\theta_0^*) = H\theta + k'E[z_i z_{Ii}']\gamma/\sqrt{n}$, and $B_{\tilde{\mathcal{C}}} = \|k'E[z_i z_{Ii}']\| M_n/\sqrt{n}$. Consider the estimator $\hat{B} = \|\hat{k}'\frac{1}{n}\sum_{i=1}^n z_i z_{Ii}'\| M_n/\sqrt{n}$ of the worst-case bias, which is the same as the estimator $\overline{\text{bias}}_{\sqrt{n}\tilde{\mathcal{C}}}(\hat{k})/\sqrt{n}$ under local misspecification used in the main text. Since $\hat{B}$ and $\hat{h}$ depend on the data only through the sample means $S_n = n^{-1}(\sum_i z_i z_i', \sum_i x_i z_i', \sum_i x_i y_i')$, equation (S31) holds by the delta method, and consistent estimates $\hat{\Omega}_+$ and $\hat{\Omega}_-$ of Ω_+ and Ω_- can be constructed using a consistent estimator of the asymptotic variance of S_n, which yields the CI

$$\widetilde{CI} = \left[\hat{h} - \text{cv}_\alpha\left(\overline{\text{bias}}_{\sqrt{n}\tilde{\mathcal{C}}}(\hat{k})/\hat{\Omega}_-^{1/2}\right) \cdot \hat{\Omega}_-^{1/2}/\sqrt{n}, \hat{h} + \text{cv}_\alpha\left(\overline{\text{bias}}_{\sqrt{n}\tilde{\mathcal{C}}}(\hat{k})/\hat{\Omega}_+^{1/2}\right) \cdot \hat{\Omega}_+^{1/2}/\sqrt{n}\right].$$

Thus, relative to the CI described in the main text, $\widetilde{CI}$ differs only in that it uses variance estimates $\hat{\Omega}_+$ and $\hat{\Omega}_-$ that are valid under global misspecification. If $M_n = M$, so that misspecification is local, $\hat{\Omega}_+ = k'\Sigma k + o_p(1)$ and $\hat{\Omega}_+ = k'\Sigma k + o_p(1)$, and the CI is asymptotically equivalent to the CI described in the main text.

D.2.1 *Proof of Theorem* D.2

Proof. Let $Z_- = \sqrt{n}(\hat{h} - \hat{B} - (h(\theta_0^*) - B_{\tilde{\mathcal{C}}}))/\Omega_-^{1/2}$, and let $Z_+ = \sqrt{n}(\hat{h} + \hat{B} - (h(\theta_0^*) + B_{\tilde{\mathcal{C}}}))/\Omega_+^{1/2}$. Then

$$\begin{aligned}
&P\left(h(\theta) \geq \hat{h} - \hat{B} - z_{1-\alpha}\hat{\Omega}_-^{1/2}/\sqrt{n}\right) \\
&\quad = P\left(Z_- \leq \sqrt{n}\left(B_{\tilde{\mathcal{C}}} - \left(h\left(\theta_0^*\right) - h(\theta)\right)\right)/\Omega_-^{1/2} + z_{1-\alpha}\hat{\Omega}_-^{1/2}/\Omega_-^{1/2}\right) \\
&\quad \geq P\left(Z_- \leq z_{1-\alpha}\hat{\Omega}_-^{1/2}/\Omega_-^{1/2}\right) \geq 1 - \alpha + o(1),
\end{aligned}$$

where the equality follows by definition of Z_-, the first inequality follows since $B_{\tilde{\mathcal{C}}} \geq h(\theta_0^*) - h(\theta)$ by definition of $B_{\tilde{\mathcal{C}}}$, and the second inequality follows since $\hat{\Omega}_-^{1/2}/\Omega_-^{1/2} \xrightarrow{p} 1$, $Z_- \xrightarrow{d} \mathcal{N}(0,1)$, and since convergence in distribution to a continuous distribution implies uniform convergence of the cdfs (van der Vaart (1998, Lemma 2.11)). To show the result for the two-sided CI, let $b = h(\theta^*) - h(\theta)$ denote the asymptotic bias. Then

$$\begin{aligned}
P\left(h(\theta) \in \widetilde{CI}\right) &= P\left(\text{cv}_\alpha\left(\sqrt{n}\hat{B}/\hat{\Omega}_-^{1/2}\right) \cdot \hat{\Omega}_-^{1/2}/\sqrt{n} \geq \hat{h} - h(\theta) \geq -\text{cv}_\alpha\left(\sqrt{n}\hat{B}/\hat{\Omega}_+^{1/2}\right) \cdot \hat{\Omega}_+^{1/2}/\sqrt{n}\right) \\
&= P\left(Z_- + \sqrt{n}b/\Omega_-^{1/2} \leq A_-\right) + P\left(-Z_+ - \sqrt{n}b/\Omega_+^{1/2} \leq A_+\right) - 1,
\end{aligned}$$

where $A_- = \sqrt{n}(B_{\tilde{C}} - \hat{B})/\Omega_-^{1/2} + \text{cv}_\alpha(\sqrt{n}\hat{B}/\hat{\Omega}_-^{1/2}) \cdot \hat{\Omega}_-^{1/2}/\Omega_-^{1/2}$ and $A_+ = \text{cv}_\alpha(\sqrt{n}\hat{B}/\hat{\Omega}_+^{1/2}) \cdot \hat{\Omega}_+^{1/2}/\Omega_+^{1/2} + \sqrt{n}(B_{\tilde{C}} - \hat{B})/\Omega_+^{1/2}$. Now, since $\hat{\Omega}_-/\Omega_- \xrightarrow{p} 1$, applying first equation (S34) and next equation (S35) in Lemma D.1 below yields

$$A_- = \sqrt{n}(B_{\tilde{C}} - \hat{B})/\Omega_-^{1/2} + \text{cv}_\alpha\big(\sqrt{n}\hat{B}/\Omega_-^{1/2}\big) + o_p(1) = \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_-^{1/2}\big) + o_p(1),$$

where the $o_p(1)$ term is asymptotically negligible uniformly over $\mathcal{S}_n$. By similar argument, $A_+ = \text{cv}_\alpha(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}) + o_p(1)$. By equation (S31), it therefore follows that, up to a term that is asymptotically negligible uniformly over $\mathcal{S}_n$, $P(h(\theta) \in \widetilde{CI})$ equals

$$P\big(Z + \sqrt{n}b/\Omega_-^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_-^{1/2}\big)\big) + P\big(Z - \sqrt{n}b/\Omega_+^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big)\big) - 1, \quad \text{(S33)}$$

where Z denotes as standard normal random variable. Fix $\epsilon > 0$. To conclude the proof, we will show that for n large enough, this expression is bounded below by $1 - \alpha - \epsilon$.

Since equation (S33) is symmetric in Ω_+ and Ω_-, suppose without loss of generality that $\Omega_+ > \Omega_-$. By the assumption of the theorem, for n large enough and $\eta > 0$ specified below, $|\Omega_-^{1/2}/\Omega_+^{1/2} - 1| \le \eta\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}$.

We'll consider two cases, $\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2} > z_{1-\epsilon}$, and $\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2} \le z_{1-\epsilon}$. Suppose first $\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2} > z_{1-\epsilon}$. Then, if $b < 0$, equation (S33) is bounded below by $\Phi(\text{cv}_\alpha(\sqrt{n}B_{\tilde{C}}/\Omega_-^{1/2})) + 1 - \alpha - 1 \ge \Phi(\text{cv}_\alpha(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2})) - \alpha \ge 1 - \epsilon - \alpha$, where the last inequality follows since $\text{cv}_\alpha(t) \ge t$. If b is positive, it is bounded below by $1 - \alpha + \Phi(\text{cv}_\alpha(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2})) - 1 \ge 1 - \alpha - \epsilon$.

Next, suppose, $\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2} \le z_{1-\epsilon}$. Then $|\Omega_-^{1/2}/\Omega_+^{1/2} - 1| \le \eta z_{1-\epsilon}$, so that by equation (S35),

$$\big|\text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_-^{1/2}\big) - \sqrt{n}B_{\tilde{C}}/\Omega_-^{1/2} - \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big) + \sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big| \le \nu,$$

where $\nu = \eta z_{1-\epsilon}(z_{1-\alpha/2} - z_{1-\alpha})$. Therefore,

$$\begin{aligned}
&P\big(Z + \sqrt{n}b/\Omega_-^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_-^{1/2}\big)\big)\\
&\quad \ge P\big(Z + \sqrt{n}b/\Omega_-^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big) - \sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2} + \sqrt{n}B_{\tilde{C}}/\Omega_-^{1/2} - \nu\big)\\
&\quad = P\big(Z + \sqrt{n}b/\Omega_+^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big) + \sqrt{n}(B_{\tilde{C}} - b)\big(\Omega_-^{-1/2} - \Omega_+^{-1/2}\big) - \nu\big)\\
&\quad \ge P\big(Z + \sqrt{n}b/\Omega_+^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big) - \nu\big)\\
&\quad \ge P\big(Z + \sqrt{n}b/\Omega_+^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big)\big) + 1 - 2\Phi(\nu/2),
\end{aligned}$$

where the last equality follows since $\inf_x\{\Phi(x - \nu) - \Phi(x)\} = 1 - 2\Phi(\nu/2)$. It therefore follows that the coverage probability in equation (S33) is bounded below by

$$\begin{aligned}
&P\big(Z + \sqrt{n}b/\Omega_+^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big)\big) + P\big(Z - \sqrt{n}b/\Omega_+^{1/2} \le \text{cv}_\alpha\big(\sqrt{n}B_{\tilde{C}}/\Omega_+^{1/2}\big)\big)\\
&\quad - 1 + \big(1 - 2\Phi(\nu/2)\big) \ge 1 - \alpha + \big(1 - 2\Phi(\nu/2)\big),
\end{aligned}$$

where the inequality follows by definition of cv_α. Setting $\eta = 2z_{1/2+\epsilon/2}/(z_{1-\epsilon}(z_{1-\alpha/2} - z_{1-\alpha}))$ then implies that the right-hand side evaluates to $1 - \alpha - \epsilon$. Thus, equation (S33) is bounded below by $1 - \alpha - \epsilon$, concluding the proof. □

Lemma D.1. *The critical value* $\text{cv}_\alpha(t)$ *satisfies, for any* $a > 0$,

$$\sup_{b\geq 0}\left|\text{cv}_\alpha(ab)/a - \text{cv}_\alpha(b)\right| \leq z_{1-\alpha/2}\frac{|1-a|}{\max\{a,1\}}, \tag{S34}$$

and

$$\sup_{b\geq 0}\left|\text{cv}_\alpha(ab) - ab - \text{cv}_\alpha(b) + b\right| \leq (z_{1-\alpha/2} - z_{1-\alpha})\frac{|1-a|}{\max\{a,1\}}. \tag{S35}$$

Proof. Since the function cv_α is increasing and convex, with slope bounded by 1, for $b_1, b_2 \geq 0$

$$\text{cv}_\alpha(b_1 + b_2) \leq b_1 + \text{cv}_\alpha(b_2), \tag{S36}$$

and for $a \geq 1$ and $b \geq 0$,

$$\text{cv}_\alpha(ba)/a + \text{cv}_\alpha(0)(1 - 1/a) \geq \text{cv}_\alpha(b). \tag{S37}$$

Suppose $a \geq 1$. Then by equation (S36)

$$\text{cv}_\alpha(ab)/a - \text{cv}_\alpha(b) \leq \big(\text{cv}_\alpha(b) - b\big)(1/a - 1) \leq 0,$$

since $\text{cv}_\alpha(b) - b$ is bounded below by $z_{1-\alpha}$. On the other hand, by equation (S37), the left-hand side is greater than $-\text{cv}_\alpha(0)(1 - 1/a)$. If $a \leq 1$, the same argument with ab and b reversed then yields equation (S34).

To show equation (S35), suppose first that $a \geq 1$. By equation (S36), $\text{cv}_\alpha(ab) - ab - \text{cv}_\alpha(b) + b \leq 0$. On the other hand, by equation (S37),

$$\text{cv}_\alpha(ab) - ab - \text{cv}(b) + b \geq \big(\text{cv}_\alpha(ab) - ab - \text{cv}_\alpha(0)\big)(1 - 1/a) \geq (1 - 1/a)(z_{1-\alpha} - z_{1-\alpha/2}),$$

where the second inequality follows since $\text{cv}_\alpha(ab) - ab \geq z_{1-\alpha}$ and $\text{cv}_\alpha(0) = z_{1-\alpha/2}$. If $a < 1$, the same argument with ab and b reversed then yields equation (S35). □

Appendix E: Additional asymptotic results

E.1 *Construction of a submodel satisfying Assumption C.1*

We give here a construction of a submodel satisfying Assumption C.1 under mild conditions on the class $\mathcal{P}$. The construction follows Example 25.16 (p. 364) of van der Vaart (1998).

Lemma E.1. *Suppose that* $g(w_i, \theta)$ *is continuously differentiable almost surely in a neighborhood of* θ^* *where* $E_{P_0} g(w_i, \theta^*) = 0$, *and that, for some* $\varepsilon > 0$,

$$E_{P_0}\sup_{\|\theta-\theta^*\|\leq\varepsilon}\left|g(w_i,\theta)g(w_i,\theta)'\right| < \infty \quad \textit{and} \quad E_{P_0}\sup_{\|\theta-\theta^*\|\leq\varepsilon}\left\|\frac{d}{d\theta'}g(w_i,\theta)\right\| < \infty.$$

Let

$$\pi_t(w_i) = C(t)h\big(t'g\big(w_i,\theta^*\big)\big) \quad \textit{where } h(x) = 2\big[1 + \exp(-2x)\big]^{-1}$$

with $C(t)^{-1} = E_{P_0} h(t' g(w_i, \theta^*))$. *This submodel satisfies Assumption* C.1, *and the bounds on the moments in the above display hold with* P_0 *replaced by* P_t.

Proof. Quadratic mean differentiability follows from Problem 12.6 in Lehmann and Romano (2005), so we just need to show that equation (S14) holds, and that the derivative is continuous in a neighborhood of $(t', \theta')' = (0', \theta^{*\prime})'$. For this, it suffices to show that each partial derivative exists and is continuous as a function of $(t', \theta')'$ in a neighborhood of $(0', \theta^{*\prime})'$, and that the Jacobian matrix of partial derivatives takes the form in equation (S14) at $(t', \theta')' = (0', \theta^{*\prime})'$ (see Theorem 4.5.3 in Shurman (2016)).

To this end, we first show that $C(t)$ is continuously differentiable, and derive its derivative at 0. It can be checked that $h(x)$ is continuously differentiable, with $h(0) = h'(0) = 1$, and that $h(x)$ and $h'(x)$ are bounded. We have, for some constant K,

$$\left| \frac{d}{dt_j} h\big(t' g(w_i, \theta^*)\big) \right| = \left| h'\big(t' g(w_i, \theta^*)\big) g_j(w_i, \theta^*) \right| \leq K \left| g_j(w_i, \theta^*) \right|$$

so, since $E_{P_0} |g_j(w_i, \theta^*)| < \infty$, we have, by a corollary of the dominated convergence theorem (Corollary 5.9 in Bartle (1966)),

$$\frac{d}{dt_j} E_{P_0} h\big(t' g(w_i, \theta^*)\big) = E_{P_0} \frac{d}{dt_j} h\big(t' g(w_i, \theta^*)\big) = E_{P_0} h'\big(t' g(w_i, \theta^*)\big) g_j(w_i, \theta^*).$$

By boundedness of h' and the dominated convergence theorem, this is continuous in t. Thus, $C(t)$ is continuously differentiable in each argument, with

$$\frac{d}{dt_j} C(t) = -\big[E_{P_0} h\big(t' g(w_i, \theta^*)\big)\big]^{-2} E_{P_0} h'\big(t' g(w_i, \theta^*)\big) g_j(w_i, \theta^*)$$

which gives $[\frac{d}{dt_j} C(t)]_{t=0} = E_{P_0} g_j(w_i, \theta^*) = 0$.

Now consider the derivative of

$$E_{P_t} g(w_i, \theta) = E_{P_0} g(w_i, \theta) \pi_t(w_i) = C(t) E_{P_0} g(w_i, \theta) h\big(t' g(w_i, \theta^*)\big)$$

with respect to elements of θ and t. We have, for each j, k,

$$\frac{d}{dt_j} g_k(w_i, \theta) h\big(t' g(w_i, \theta^*)\big) = g_k(w_i, \theta) h'\big(t' g(w_i, \theta^*)\big) g_j(w_i, \theta^*).$$

This is bounded by a constant times $|g_k(w_i, \theta) g_j(w_i, \theta^*)|$ by boundedness of h'. Also,

$$\frac{d}{d\theta_j} g_k(w_i, \theta) h\big(t' g(w_i, \theta^*)\big)$$

is bounded by a constant times $\frac{d}{d\theta_j} g_k(w_i, \theta)$ by boundedness of h. By the conditions of the lemma, the quantities in the above two displays are bounded uniformly over $(t', \theta')'$ in a neighborhood of $(\theta^{*\prime}, 0')'$ by a function with finite expectation under P_0. It follows that we can again apply Corollary 5.9 in Bartle (1966) to obtain the derivative of

$E_{P_0} g(w_i,\theta) h(t' g(w_i,\theta^*))$ with respect to each element of θ and t by differentiating under the expectation. Furthermore, the bounds above and continuous differentiability of $g(w_i,\theta)$ along with the dominated convergence theorem imply that the derivatives are continuous in $(t',\theta')'$.

Thus, $E_{P_t} g(w_i,\theta)$ is differentiable with respect to each argument of t and θ, with the partial derivatives continuous with respect to $(\theta',t')'$. It follows that $(t',\theta')' \mapsto E_{P_t} g(w_i,\theta)$ is differentiable at $t=0$, $\theta=\theta^*$. To calculate the Jacobian, note that

$$\frac{d}{dt'} E_{P_t} g(w_i,\theta) = C(t) E_{P_0} g(w_i,\theta) g\big(w_i,\theta^*\big)' h'\big(t' g\big(w_i,\theta^*\big)\big) + E_{P_0} g(w_i,\theta) h\big(t' g\big(w_i,\theta^*\big)\big) \frac{d}{dt'} C(t).$$

Evaluating this at $t=0$, $\theta=\theta^*$, the second term is equal to zero by calculations above, and the first term is given by $E_{P_0} g(w_i,\theta^*) g(w_i,\theta^*)$. For the derivative with respect to θ at $\theta=\theta^*$, $t=0$, this is equal to Γ_{θ^*,P_0} by definition. Thus, Assumption C.1 holds. Furthermore, the bounds on the moments of $g(w_i,\theta)$ hold with P_t replacing P_0 by boundedness of $\pi_t(w_i)$. □

E.2 *Example: Misspecified linear IV*

We verify our conditions in the misspecified linear IV model, defined by the equation

$$g_P(\theta) = E_P\big(y_i - x_i'\theta\big) z_i = c/\sqrt{n}, \quad c \in \mathcal{C},$$

where $\mathcal{C}$ is a compact convex set, y_i is a scalar valued random variable, x_i is a $\mathbb{R}^{d_\theta}$ valued random variable and z_i is a $\mathbb{R}^{d_g}$ valued random variable, with $d_g \geq d_\theta$. The derivative matrix and variance matrix are

$$\Gamma_{\theta,P} = \frac{d}{d\theta'} g_P(\theta) = -E_P z_i x_i', \qquad \Sigma_{\theta,P} = \operatorname{var}_P\big(\big(y_i - x_i'\theta\big) z_i\big).$$

Let $\Theta \subset \mathbb{R}^{d_\theta}$ be a compact set and let $h : \Theta \to \mathbb{R}$ be continuously differentiable with nonzero derivative at all $\theta \in \Theta$. Let ε be given and let $\mathcal{P}$ be a set of probability distributions P for $(x_i', z_i', y_i)'$. We make the following assumptions on $\mathcal{P}$.

Assumption E.1. *For all $P \in \mathcal{P}$, the following conditions hold*:

1. *For all j, $E_P|x_{i,j}|^{4+\varepsilon} < 1/\varepsilon$, $E_P|z_{i,j}|^{4+\varepsilon} < 1/\varepsilon$ and $E_P|y_i|^{4+\varepsilon} < 1/\varepsilon$.*
2. *The matrix $E_P z_i x_i'$ is full rank and $\|E_P z_i x_i' u\|/\|u\| > 1/\varepsilon$ for all $u \in \mathbb{R}^{d_g} \backslash \{0\}$ (i.e., the singular values of $E_P z_i x_i'$ are bounded away from zero).*
3. *The matrix $\Sigma_{\theta,P} = \operatorname{var}_P((y_i - x_i'\theta) z_i)$ satisfies $u'\Sigma_{\theta,P} u/\|u\|^2 > \varepsilon$ for all $u \in \mathbb{R}^{d_g} \backslash \{0\}$ and all θ such that there exists $c \in \mathcal{C}$ and $n \geq 1$ such that $E_P(y_i - x_i'\theta) z_i = c/\sqrt{n}$.*

Note that, applying Cauchy–Schwarz, the first condition implies $E_P|v_1 v_2 v_3 v_4|^{1+\varepsilon/4} < 1/\varepsilon$ for any v_1, v_2, v_3, v_4 where each v_k is an element of x_i, z_i or y_i. In particular, $z_i(y_i - x_i'\theta)$ has a bounded $2+\varepsilon/2$ moment uniformly over $\theta \in \Theta$ and $P \in \mathcal{P}$.

E.2.1 *Conditions for Theorems C.5 and C.6* We first verify the conditions of Appendix C.5. To verify the conditions of Theorems C.5 and C.6 (which show that the plug-in optimal weights $\hat{k} = k(\delta, \hat{H}, \hat{\Gamma}, \hat{\Sigma})$ lead to CIs that achieve or nearly achieve the efficiency bounds in Theorem C.1 and Theorem C.2), we must verify Assumptions C.2, C.3, C.5, and C.6.

Let

$$\hat{\theta}_{\text{initial}} = \left(\sum_{i=1}^{n} z_i x_i' W_n \sum_{i=1}^{n} x_i z_i'\right)^{-1} \sum_{i=1}^{n} z_i x_i' W_n \sum_{i=1}^{n} z_i y_i,$$

where $W_n = W_P + o_P(1)$ uniformly over $P \in \mathcal{P}$ and W_P is a positive definite matrix with $u' W_P u / \|u\|^2$ bounded away from zero uniformly over $P \in \mathcal{P}$. Let $\hat{H} = H_{\hat{\theta}}$ where H_θ is the derivative of h at θ. Let

$$\hat{\Gamma} = -\frac{1}{n}\sum_{i=1}^{n} z_i x_i', \qquad \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' \big(y_i - x_i' \hat{\theta}_{\text{initial}}\big)^2.$$

First, let us verify Assumption C.3. Indeed, it follows from a CLT for triangular arrays (Lemma E.7 with $v_i = u_n' [z_i(y_i - x_i'\theta) - E z_i(y_i - x_i'\theta)]$ with u_n an arbitrary sequence with $\|u_n\| = 1$ all n) that

$$\sup_{u \in \mathbb{R}^{d_g}} \sup_{t \in \mathbb{R}} \sup_{(\theta', c') \in \Theta \times \mathcal{C}} \sup_{P \in \mathcal{P}_n(\theta, c)} \left| P\left(\frac{\sqrt{n} u' \big(\hat{g}(\theta) - g_P(\theta)\big)}{\sqrt{u' \Sigma_{\theta,P} u}} \le t \right) - \Phi(t) \right| \to 0$$

(note that u can be taken to satisfy $\|u\| = 1$ without loss of generality, since the formula inside the probability statement is invariant to scaling). Note that this, along with compactness of $\mathcal{C}$, also implies that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i (y_i - x_i'\theta) = \sqrt{n}\hat{g}(\theta) = \mathcal{O}_P(1)$ uniformly over θ and P with $P \in \mathcal{P}(\theta, c)$ for some c.

For Assumption C.2, we have

$$\sqrt{n}(\hat{\theta}_{\text{initial}} - \theta) = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i' W_n \frac{1}{n}\sum_{i=1}^{n} x_i z_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} z_i x_i' W_n \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i \big(y_i - x_i'\theta\big).$$

Since $\frac{1}{n}\sum_{i=1}^{n} z_i x_i'$ converges in probability to $-\Gamma_{\theta,P}$ uniformly over P by Lemma E.8 and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i(y_i - x_i'\theta) = \mathcal{O}_P(1)$ uniformly over P by the verification of Assumption C.3 above, it follows that this display is $\mathcal{O}_P(1)$ uniformly over P and θ, as required. For the second part of the assumption, we have

$$\hat{g}(\hat{\theta}_{\text{initial}}) - g(\theta) = -\frac{1}{n}\sum_{i=1}^{n} z_i x_i' (\hat{\theta}_{\text{initial}} - \theta) = \Gamma_{\theta,P}(\hat{\theta}_{\text{initial}} - \theta) + (\hat{\Gamma} - \Gamma_{\theta,P})(\hat{\theta}_{\text{initial}} - \theta).$$

The last term is uniformly $o_P(1/\sqrt{n})$ as required since $(\hat{\theta}_{\text{initial}} - \theta) = \mathcal{O}_P(1/\sqrt{n})$ as shown above and $\hat{\Gamma} - \Gamma_{\theta,P}$ converges in probability to zero uniformly by an LLN for triangular arrays (Lemma E.8). For the last part of the assumption, we have, by the mean value theorem,

$$h(\hat{\theta}_{\text{initial}}) - h(\theta) = H_{\theta^*(\hat{\theta}_{\text{initial}})}(\hat{\theta}_{\text{initial}} - \theta) = H_\theta(\hat{\theta}_{\text{initial}} - \theta) + (H_{\theta^*(\hat{\theta}_{\text{initial}})} - H_\theta)(\hat{\theta}_{\text{initial}} - \theta),$$

where $\theta^*(\hat{\theta}_{\text{initial}}) - \theta$ converges uniformly in probability to zero. Since $\theta \mapsto H_\theta$ is uniformly continuous on θ (since it is continuous by assumption and Θ is compact), it follows that $H_{\theta^*(\hat{\theta}_{\text{initial}})} - H_\theta$ converges uniformly in probability to zero, which, along with the verification of the first part of the assumption above, gives the required result.

For Assumption C.5, the first two parts of the assumption (concerning uniform consistency of $\hat{\Gamma}$ and $\hat{H}$) follow from arguments above. For the last part (uniform consistency of $\hat{\Sigma}$), note that

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{n}\sum_{i=1}^{n} z_i z_i'(y_i - x_i'\hat{\theta}_{\text{initial}})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} z_i z_i'(y_i - x_i'\theta)^2 + \frac{1}{n}\sum_{i=1}^{n} z_i z_i'\big[(y_i - x_i'\hat{\theta}_{\text{initial}})^2 - (y_i - x_i'\theta)^2\big].\end{aligned}$$

The first term converges uniformly in probability to $\Sigma_{\theta,P}$ by an LLN for triangular arrays (Lemma E.8). The last term is equal to

$$\frac{1}{n}\sum_{i=1}^{n} z_i z_i'(x_i'\hat{\theta}_{\text{initial}} + x_i'\theta - 2y_i)x_i'(\hat{\theta}_{\text{initial}} - \theta).$$

This converges in probability to zero by an LLN for triangular arrays (Lemma E.8) and the moment bound in Assumption E.1(1)

Finally, Assumption C.6 follows by Assumption E.1(2), and the condition that the derivative is nonzero for all θ.

E.2.2 *Conditions for Theorems* C.1 *and* C.2 We now verify the conditions of the lower bounds, Theorems C.1 and C.2. Given $P_0 \in \mathcal{P}$ with $E_{P_0} g(w_i, \theta^*) = 0$, we need to show that a submodel P_t satisfying Assumption C.1 exists with $P_t \in \mathcal{P}$ for $\|t\|$ small enough. To verify this condition, we take $\mathcal{P}$ to be the set of all distributions satisfying Assumption E.1, and we assume that θ^* is in the interior of Θ.

Let P_t be the subfamily given in Lemma E.1. This satisfies Assumption C.1 by Lemma E.1 (the moment conditions needed for this lemma hold by Assumption E.1(1)), so we just need to check that $P_t \in \mathcal{P}$ for t small enough. For this, it suffices to show that $E_{P_t}|x_{i,j}|^{4+\varepsilon}$, $E_{P_t}|z_{i,j}|^{4+\varepsilon}$, $E_{P_t}|y_i|^{4+\varepsilon}$, $E_{P_t} z_i x_i'$, and $\text{var}_{P_t}(z_i(y_i - x_i'\theta))$ are continuous in t at $t = 0$, which holds by the dominated convergence theorem since the likelihood ratio $\pi_t(w_i)$ for this family is bounded and continuous with respect to t.

E.2.3 *Conditions for Appendix* D In Appendix D, we proposed a CI that is asymptotically valid under global misspecification and asymptotically equivalent to the CIs considered in the rest of the paper under local misspecification. Specializing to the present setting with misspecified IV, the CI is the union over $\tilde{c}$ of CIs that use the GMM estimator $\hat{\theta}_{W,\tilde{c}}$ based on the moment function $\theta \mapsto z_i(y_i - x_i'\theta) - \tilde{c}$. This estimator is given by $\theta_{W,\tilde{c}} = -(\hat{\Gamma}'W\hat{\Gamma})^{-1}\hat{\Gamma}'\hat{W}(\frac{1}{n}\sum_{i=1}^n z_i y_i - \tilde{c})$ where $\hat{\Gamma} = -\frac{1}{n}\sum_{i=1}^n z_i x_i'$ as defined above. We estimate $k_{\theta,P}' = -H_\theta(\Gamma_{\theta,P}'W_P\Gamma_{\theta,P})^{-1}\Gamma_{\theta,P}'W_P$ using $\hat{k}_\theta' = -H_\theta(\hat{\Gamma}'W\hat{\Gamma})^{-1}\Gamma'W_P$. We estimate $\Sigma_{\theta,P} = \text{var}_P(z_i(y_i - x_i'\theta))$ using $\hat{\Sigma}_\theta = \frac{1}{n}\sum_{i=1}^n z_i z_i'(y_i - x_i'\hat{\theta}_{W,\tilde{c}})^2 = \frac{1}{n}\sum_{i=1}^n z_i z_i'[y_i -$

$x_i'(\hat\Gamma'W\hat\Gamma)^{-1}\hat\Gamma'\hat W(\frac{1}{n}\sum_{i=1}^n z_i y_i - \tilde c)]^2$. In addition to Assumption E.1, we assume that the weighting matrix is given by a (possibly data dependent) sequence W_n such that $W_n - W_P = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$, where W_P is some family of limiting weighting matrices with $u'W_P u/\|u\|^2$ bounded away from zero and infinity uniformly over $P \in \mathcal{P}$. The population influence function weights are then given by $k_{\theta,P}' = H_\theta(\Gamma_{\theta,P}' W_{\theta,P}\Gamma_{\theta,P})^{-1}\Gamma_{\theta,P}'$.

To verify the asymptotic equivalence result (Assumption D.1), we need to verify Theorem D.1. To this end, first note that $\hat\Gamma - \Gamma_{\theta,P} = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$ by a law of large numbers (Lemma E.8). Thus, by the bounds on $E_P z_i x_i'$ in Assumption E.1, $\sup_{c\in\mathcal{C}} |\hat\theta_{W,c/\sqrt{n}} - \hat\theta_{W,0}| = \sup_{c\in\mathcal{C}} |(\hat\Gamma'W\hat\Gamma)^{-1}\hat\Gamma'Wc/\sqrt{n}| = \mathcal{O}_P(1/\sqrt{n})$ uniformly over $(\theta, P) \in \mathcal{S}_n$. Note that $\hat\theta_{W,0} = \hat\theta_{\text{initial}}$ where $\hat\theta_{\text{initial}}$ is defined in Appendix E.2.1 above, so it follows from arguments in that section that $\hat\theta_{W,0} - \theta = \mathcal{O}_P(1/\sqrt{n})$ uniformly over $(\theta, P) \in \mathcal{S}_n$. Thus, $\sup_{c\in\mathcal{C}} |\hat\theta_{W,c/\sqrt{n}} - \theta| = \mathcal{O}_P(1/\sqrt{n})$ uniformly over $(\theta, P) \in \mathcal{S}_n$. Similarly, we have $\sup_{c\in\mathcal{C}} |\hat\Sigma_{\theta_{W,c/\sqrt{n}}} - \hat\Sigma_{\theta_{W,c/\sqrt{n}}}| = o_P(1)$ and $\hat\Sigma_{\theta_{W,0}}$ corresponds to the estimate used in Appendix E.2.1 above, so that $\sup_{c\in\mathcal{C}} |\hat\Sigma_{\theta_{W,c/\sqrt{n}}} - \Sigma_{\theta,P}| = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$ by arguments in Appendix E.2.1.

The last part of Assumption D.1 will now follow if we can show that $\sup_{c\in\mathcal{C}}|\hat k_{\hat\theta_{W,c/\sqrt{n}}} - k_{\theta,P}| = o_P(1)$. Since we have already shown uniform consistency of $\hat\Gamma$, this will follow so long as $\sup_{c\in\mathcal{C}} |H_{\hat\theta_{W,c/\sqrt{n}}} - H_\theta| = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$. This follows by the fact that $\sup_{c\in\mathcal{C}} |\hat\theta_{W,c/\sqrt{n}} - \theta| = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$ along with uniform continuity of $\theta \mapsto H_\theta$ on Θ (since Θ is compact, continuity implies uniform continuity).

Finally, for the first display of Assumption D.1, note that, for some $\theta^*(c)$ on the line segment between θ and $\hat\theta_{W,c}$,

$$\begin{aligned}
&h(\hat\theta_{W,c/\sqrt{n}}) - h(\theta) - k_{\theta,P}'\big[\hat g(\theta) - c/\sqrt{n}\big] \\
&= H_{\theta^*(c)}(\hat\theta_{W,0} - \theta) - k_{\theta,P}'\hat g(\theta) + H_{\theta^*(c)}\big(\hat\Gamma'W\hat\Gamma\big)^{-1}\hat\Gamma'Wc/\sqrt{n} + k_{\theta,P}'c/\sqrt{n} \\
&= H_\theta(\hat\theta_{W,0} - \theta) - k_{\theta,P}'\hat g(\theta) + H_\theta\big(\hat\Gamma'W\hat\Gamma\big)^{-1}\hat\Gamma'Wc/\sqrt{n} + k_{\theta,P}'c/\sqrt{n} + R_{n,\theta,P}(c) \\
&= \big[-H_\theta\big(\hat\Gamma'W\hat\Gamma\big)^{-1}\hat\Gamma'W - k_{\theta,P}'\big]\left[\frac{1}{n}\sum_{i=1}^n z_i\big(y_i - x_i'\theta\big) - c/\sqrt{n}\right] + R_{n,\theta,P}(c),
\end{aligned}$$

where $\sup_{c\in\mathcal{C}} \sqrt{n}|R_{n,\theta,P}(c)| = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$. The first display of Assumption D.1 now follows from the fact that $\frac{1}{n}\sum_{i=1}^n z_i(y_i - x_i'\theta) - c/\sqrt{n} = \mathcal{O}_P(1/\sqrt{n})$ (by Lemma E.7) and $-H_\theta(\hat\Gamma'W\hat\Gamma)^{-1}\hat\Gamma'W - k_{\theta,P}' = o_P(1)$ uniformly over $(\theta, P) \in \mathcal{S}_n$.

E.3 *Auxiliary results*

This section contains auxiliary results used in Appendix C. Appendix E.3.1 shows that optimizing length over a set of the form $\mathcal{G} = \mathbb{R}^{d_\theta} \times \mathcal{D}$ is without loss of generality, as claimed in Appendix C.5. Appendix E.3.2 contains a result on the continuity of the optimal weights with respect to δ, Γ, Σ, and H. Appendix E.3.3 states a law of large numbers and central limit theorem for triangular arrays.

It will be convenient to state some of these results in the general setup of Donoho (1994), Low (1995), and Armstrong and Kolesár (2018). Using the notation in Armstrong and Kolesár (2018), the between class modulus problem is given by

$$\begin{aligned}\omega(\delta) &= \omega(\delta; \mathcal{F}, \mathcal{G}, L, K) \\ &= \sup L(g-f); \quad \text{s.t. } \|K(g-f)\| \leq \delta, f \in \mathcal{F}, g \in \mathcal{G}, \end{aligned} \tag{S38}$$

where $\mathcal{F}$ and $\mathcal{G}$ are convex sets with $\mathcal{G} \subseteq \mathcal{F}$, L is a linear functional and K is a linear operator from $\mathcal{F}$ to a Hilbert space with norm $\|\cdot\|$. In our case, this is given by equation (S10) in the main text, which fits into this setting with $(\theta', c')'$ playing the role of f, $\mathbb{R}^{d_\theta} \times \mathcal{C}$ playing the role of $\mathcal{F}$, K given by the transformation $(\theta', c')' \mapsto -\Gamma\theta + c$, and with the norm defined using the inner product $\langle x, y\rangle = x'\Sigma^{-1}y$. The linear functional L is given by $(\theta', c')' \mapsto H\theta$.

E.3.1 *Replacing* $\mathbb{R}^{d_\theta} \times \mathcal{D}$ *with a general set* $\mathcal{G}$ In Appendix C.5, we mentioned that directing power at sets that do not restrict θ is without loss of generality when we require coverage over a set that does not make local restrictions on θ. This holds by the following lemma (applied with $\mathcal{U} = \mathbb{R}^{d_\theta} \times \{0\}^{d_g}$).

Lemma E.2. *Let $\mathcal{U}$ be a set with $0 \in \mathcal{U}$ such that $\mathcal{F} = \mathcal{F} - \mathcal{U}$ (i.e., $\mathcal{F}$ is invariant to adding elements in $\mathcal{U}$). Then, for any solution $\tilde{f}^*$, $\tilde{g}^*$ to the modulus problem*

$$\sup L(g-f) \quad \textit{s.t. } \|K(g-f)\| \leq \delta, f \in \mathcal{F}, g \in \mathcal{G} + \mathcal{U},$$

where K is a linear operator, there is a solution f^*, g^* to the modulus problem (S38) for $\mathcal{F}$ and $\mathcal{G}$ with $g^* - f^* = \tilde{g}^* - \tilde{f}^*$. Furthermore, any solution to the modulus problem (S38) for $\mathcal{F}$ and $\mathcal{G}$ is also a solution to the modulus problem for $\mathcal{F}$ and $\mathcal{G} + \mathcal{U}$.

Proof. Let $\tilde{f}$, $\tilde{g} + \tilde{u}$ be a solution to the modulus problem for $\mathcal{F}$ and $\mathcal{G} + \mathcal{U}$ with $\tilde{g} \in \mathcal{G}$ and $\tilde{u} \in \mathcal{U}$. Then $f = \tilde{f} - \tilde{u}$, and $g = \tilde{g}$ is feasible for $\mathcal{F}$ and $\mathcal{G}$ and achieves the same value of the objective function. Since it achieves the maximum for the objective function over the larger set $\mathcal{F} \times (\mathcal{G} + \mathcal{U})$ and is in $\mathcal{F} \times \mathcal{G}$, it must maximize the objective function over $\mathcal{F} \times \mathcal{G}$. Thus, f, g achieves the modulus for $\mathcal{F}$ and $\mathcal{G}$ and also for $\mathcal{F}$ and $\mathcal{G} + \mathcal{U}$. Since the modulus for $\mathcal{F}$ and $\mathcal{G}$ is the same as the modulus over $\mathcal{F}$ and the larger set $\mathcal{G} + \mathcal{U}$, it also follows that any solution to the former modulus problem is a solution to the latter modulus problem. □

E.3.2 *Continuity of optimal weights* We first give some lemmas under the general setup (S38).

Lemma E.3. *For each δ, let (f^*_δ, g^*_δ) be a solution to the modulus problem (S38), and let $h^*_\delta = g^*_\delta - f^*_\delta$. Let δ_0, δ_1 be given, and suppose that ω is strictly increasing on an open interval containing δ_0 and δ_1, and that a solution to the modulus problem exists for δ_0 and δ_1. Then $Kh^*_{\delta_0}$ and $Kh^*_{\delta_1}$ are defined uniquely (i.e., they do not depend on the particular solution (f^*_δ, g^*_δ)) and*

$$\|Kh^*_{\delta_0} - Kh^*_{\delta_1}\|^2 \leq 2|\delta_1^2 - \delta_0^2|.$$

PROOF. Let $f_0 = f^*_{\delta_0}$, $f_1 = f^*_{\delta_1}$ and similarly for g_0, g_1, h_0, and h_1. Let $\tilde{h} = (h_0 + h_1)/2$. Note that $\tilde{h} = \tilde{g} - \tilde{f}$ where $\tilde{g} = (g_0 + g_1)/2 \in \mathcal{G}$ and $\tilde{f} = (f_0 + f_1)/2 \in \mathcal{F}$ by convexity. Thus, $\omega(\|K\tilde{h}\|) \geq L\tilde{h} = [\omega(\delta_0) + \omega(\delta_1)]/2 \geq \min\{\omega(\delta_0), \omega(\delta_1)\}$. From this and the fact that ω is strictly increasing on an open interval containing δ_0 and δ_1, it follows that $\|K\tilde{h}\| \geq \min\{\delta_0, \delta_1\}$.

Note that $h_1 = \tilde{h} + (h_1 - h_0)/2$ and $\langle K\tilde{h}, K(h_1 - h_0)/2\rangle = \|Kh_1\|^2/4 - \|Kh_0\|^2/4 = (\delta_1^2 - \delta_0^2)/4$ (the last equality uses the fact that the constraint on $\|K(f-g)\|$ binds at any δ at which the modulus is strictly increasing). Thus,

$$\begin{aligned}\delta_1^2 = \|Kh_1\|^2 &= \|K\tilde{h}\|^2 + \left\|K(h_1 - h_0)/2\right\|^2 + \left(\delta_1^2 - \delta_0^2\right)/2 \\ &\geq \min\left\{\delta_0^2, \delta_1^2\right\} + \left\|K(h_1 - h_0)/2\right\|^2 + \left(\delta_1^2 - \delta_0^2\right)/2.\end{aligned}$$

Thus, $\|K(h_1 - h_0)\|^2/4 \leq \delta_1^2 - \min\{\delta_0^2, \delta_1^2\} - (\delta_1^2 - \delta_0^2)/2 = |\delta_1^2 - \delta_0^2|/2$ as claimed. The fact that $Kh^*_{\delta_0}$ is defined uniquely follows from applying the result with δ_1 and δ_0 both given by δ_0. □

LEMMA E.4. *For each δ, let (f^*_δ, g^*_δ) be a solution to the modulus problem* (S38), *and let $h^*_\delta = g^*_\delta - f^*_\delta$. Let δ_0 and $\varepsilon > 0$ be given, and suppose that ω is strictly increasing in a neighborhood of δ_0, and that the modulus is achieved at δ_0. Let $g \in \mathcal{G}$ and $f \in \mathcal{F}$ satisfy $L(g-f) > \omega(\delta_0) - \varepsilon$ with $\|K(g-f)\| \leq \delta_0$, and let $h = g - f$. Then*

$$\left\|K\left(h - h^*_{\delta_0}\right)\right\|^2 < 4\left[\delta_0^2 - \omega^{-1}\left(\omega(\delta_0) - \varepsilon\right)^2\right].$$

PROOF. Let $h^* = h^*_{\delta_0}$, $g^* = g^*_{\delta_0}$ and $f^* = f^*_{\delta_0}$. Using the fact that $\langle K(h + h^*)/2, K(h - h^*)/2\rangle = \|Kh\|^2/4 - \|Kh^*\|^2/4$, we have

$$\|Kh\|^2 = \left\|K\left(h + h^*\right)/2\right\|^2 + \left\|K\left(h - h^*\right)/2\right\|^2 + \|Kh\|^2/2 - \left\|Kh^*\right\|^2/2.$$

Rearranging this gives

$$\left\|K\left(h - h^*\right)/2\right\|^2 = \left[\|Kh\|^2 + \left\|Kh^*\right\|^2\right]/2 - \left\|K\left(h + h^*\right)/2\right\|^2. \tag{S39}$$

Let $\delta' = \omega^{-1}(\omega(\delta_0) - \varepsilon)$. Since $Lh > \omega(\delta')$ and $Lh^* = \omega(\delta_0)$, it follows that $L(h + h^*)/2 > [\omega(\delta') + \omega(\delta)]/2 \geq \omega(\delta')$. Since $(h + h^*)/2 = (g + g^*)/2 - (f + f^*)/2$ with $(g + g^*)/2 \in \mathcal{G}$ and $(f + f^*)/2 \in \mathcal{F}$, this means that $\|K(h + h^*)/2\| > \delta'$. Using this and the fact that $[\|Kh\|^2 + \|Kh^*\|^2]/2 \leq \delta_0^2$, it follows that $\|K(h - h^*)/2\|^2 \leq \delta_0^2 - \delta'^2$ as claimed. □

LEMMA E.5. *Let $h^*_{\delta,\mathcal{F},\mathcal{G},L,K} = g^*_{\delta,\mathcal{F},\mathcal{G},L,K} - f^*_{\delta,\mathcal{F},\mathcal{G},L,K}$ where $g^*_{\delta,\mathcal{F},\mathcal{G},L,K}, f^*_{\delta,\mathcal{F},\mathcal{G},L,K}$ is a solution to the modulus problem* (S38). *Let δ_0, L_0, K_0, $\mathcal{F}_0$, $\mathcal{G}_0$, and $\{\delta_n, L_n, K_n, \mathcal{F}_n, \mathcal{G}_n\}_{n=1}^\infty$ be given.*

Let $\mathcal{H}(\delta, K, \mathcal{F}, \mathcal{G}) = \{g - f : f \in \mathcal{F}, g \in \mathcal{G}, \|K(g-f)\| \leq \delta\}$ denote the feasible set of values of $g - f$ for the modulus problem for δ, K, $\mathcal{F}$, $\mathcal{G}$. Suppose that, for any $\varepsilon > 0$, we have, for large enough n, $\mathcal{H}(\delta_0 - \varepsilon, K_0, \mathcal{F}_0, \mathcal{G}_0) \subseteq \mathcal{H}(\delta_n, K_n, \mathcal{F}_n, \mathcal{G}_n) \subseteq \mathcal{H}(\delta_0 + \varepsilon, K_0, \mathcal{F}_0, \mathcal{G}_0)$. Suppose also that $L_n h - L_0 h \to 0$ and $\|(K_n - K_0)h\| \to 0$ uniformly over h in $\mathcal{H}(\delta_0 +$

$\varepsilon, K_0, \mathcal{F}_0, \mathcal{G}_0)$ for ε small enough. Suppose also that $\omega(\delta; \mathcal{F}_0, \mathcal{G}_0, L_0, K_0)$ is strictly increasing for δ in a neighborhood of δ_0. Then $\|K_n h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n} - K_0 h^*_{\delta_0,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0}\| \to 0$ and $L_n h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n} - L_0 h^*_{\delta_0,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0} \to 0$.

Proof. For any $\varepsilon > 0$, $g^*_{\delta_0-\varepsilon,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0}, f^*_{\delta_0-\varepsilon,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0}$ is feasible for the modulus problem under δ_n, $\mathcal{F}_n$, $\mathcal{G}_n$, L_n, K_n for large enough n. Thus, for large enough n,

$$\omega(\delta_0 - \varepsilon, \mathcal{F}_0, \mathcal{G}_0, L_0, K_0) = L h^*_{\delta_0-\varepsilon,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0} \le L_n h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n}.$$

Taking limits and using the fact that $(L_n - L) h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n} \to 0$, it follows that

$$\omega(\delta_0 - \varepsilon; \mathcal{F}_0, \mathcal{G}_0, L_0, K_0) - \varepsilon \le L h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n}$$

for large enough n. By continuity of the modulus in δ, for any $\eta > 0$ the left-hand side is strictly greater than $\omega(\delta_0 + \varepsilon; \mathcal{F}_0, \mathcal{G}_0, L_0, K_0) - \eta$ for ε small enough. Since $g^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n}$, $f^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n}$ is feasible for $\delta_0 + \varepsilon$, $\mathcal{F}_0$, $\mathcal{G}_0$, L_0, K_0 for n large enough, it follows from Lemma E.4 that

$$\begin{aligned}&\big\|K_0\big(h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n} - h^*_{\delta_0+\varepsilon,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0}\big)\big\| \\ &\quad < 4\big[(\delta_0+\varepsilon)^2 - \omega^{-1}\big(\omega(\delta_0+\varepsilon; \mathcal{F}_0, \mathcal{G}_0, L_0, K_0) - \eta; \mathcal{F}_0, \mathcal{G}_0, L_0, K_0\big)^2\big].\end{aligned}$$

By continuity of the modulus and inverse modulus, the right-hand side can be made arbitrarily close to zero by taking ε and η small. Thus,

$$\lim_{\varepsilon\downarrow 0}\limsup_{n}\big\|K_0\big(h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n} - h^*_{\delta_0+\varepsilon,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0}\big)\big\| = 0.$$

It then follows from Lemma E.3 that $\lim_{n\to\infty}\|K_0(h^*_{\delta_n,\mathcal{F}_n,\mathcal{G}_n,L_n,K_n} - h^*_{\delta_0,\mathcal{F}_0,\mathcal{G}_0,L_0,K_0})\| = 0$. The result then follows from the assumption that $\|(K_0 - K_n)h\| \to 0$ uniformly over $\mathcal{H}(\delta_0 + \varepsilon, K_0, \mathcal{F}_0, \mathcal{G}_0)$. □

We now specialize to our setting. Let $f^*_{\delta,H,\Gamma,\Sigma} = (s_0^{*\prime}, c_0^{*\prime})$ and $g^*_{\delta,H,\Gamma,\Sigma} = (s_1^{*\prime}, c_1^{*\prime})$ denote solutions to the modulus problem in equation (S10) with $\mathcal{F} = \mathbb{R}^{d_\theta} \times \mathcal{C}$ and $\mathcal{G} = \mathbb{R}^{d_\theta} \times \mathcal{D}$. Let $\omega(\delta; H, \Gamma, \Sigma) = \omega(\delta; \mathbb{R}^{d_\theta} \times \mathcal{C}, \mathbb{R}^{d_\theta} \times \mathcal{D}, H, \Gamma, \Sigma)$ denote the modulus. Let $h^*_{\delta,H,\Gamma,\Sigma} = f^*_{\delta,H,\Gamma,\Sigma} - g^*_{\delta,H,\Gamma,\Sigma}$ and let $K_{\Gamma,\Sigma} = \Sigma^{-1/2}(-\Gamma, I_{d_g\times d_g})$. Note that $h^*_{\delta,H,\Gamma,\Sigma,\mathcal{C}} = (s^{*\prime}, c^{*\prime})'$ where $(s^{*\prime}, c^{*\prime})'$ solves

$$\sup Hs \quad \text{s.t. } (c - \Gamma s)'\Sigma^{-1}(c - \Gamma s) \le \delta^2, c \in \mathcal{D} - \mathcal{C}, s \in \mathbb{R}^{d_\theta}. \tag{S40}$$

Furthermore, a solution does indeed exist so long as $\mathcal{C}$ and $\mathcal{D}$ are compact and Γ and Σ are full rank, since this implies that the constraint set is compact.

Let δ_0, H_0, Γ_0, and Σ_0 be such that $\delta_0 > 0$, $H_0 \ne 0$, and such that Γ_0 and Σ_0 are full rank. We wish to show that $K_{\Gamma,\Sigma} h^*_{\delta,H,\Gamma,\Sigma}$ is continuous as a function of δ, H, Γ, and Σ at $(\delta_0, H_0, \Gamma_0, \Sigma_0)$. To this end, let δ_n, H_n, Γ_n, and Σ_n be arbitrary sequences converging to δ_0, H_0, Γ_0, and Σ_0 (with Σ_n symmetric and positive semidefinite for each n). We will apply Lemma E.5. To verify the conditions of this lemma, first note that the modulus is strictly increasing by translation invariance (see Section C.2 in Armstrong and Kolesár (2018)). The conditions on uniform convergence of $(L_n - L)h$ and $(K_n - K)h$ follow since the constraint set for $h = g - f$ is compact. The condition on $\mathcal{H}(\delta, K, \mathcal{F}, \mathcal{G})$ follows because $(c - \Gamma s)'\Sigma^{-1}(c - \Gamma s)$ is continuous in Σ^{-1} and Γ uniformly over c and s in any compact set, and there exists a compact set that contains the constraint set for all n large enough. We record these results and some of their implications in a lemma.

Lemma E.6. *Let $\mathcal{C}$ and $\mathcal{D}$ be compact and let $c^*_{\delta,H,\Gamma,\Sigma}$, $s^*_{\delta,H,\Gamma,\Sigma}$ denote a solution to equation (S40). Let $\mathcal{A}$ denote the set of $(\delta, H, \Gamma, \Sigma)$ such that $\delta > 0$, $H \in \mathbb{R}^{d_\theta}\backslash\{0\}$, Γ is a full rank $d_g \times d_\theta$ matrix and Σ is a (strictly) positive definite $d_g \times d_g$ matrix. Then $\Sigma^{-1/2}(s^*_{\delta,H,\Gamma,\Sigma} - \Gamma c^*_{\delta,H,\Gamma,\Sigma})$ is defined uniquely for any $(\delta, H, \Gamma, \Sigma) \in \mathcal{A}$. Furthermore, the mappings $(\delta, H, \Gamma, \Sigma) \mapsto \Sigma^{-1/2}(s^*_{\delta,H,\Gamma,\Sigma} - \Gamma c^*_{\delta,H,\Gamma,\Sigma})$,*

$$k(\delta, H, \Gamma, \Sigma)' = \frac{(s^*_{\delta,H,\Gamma,\Sigma} - \Gamma c^*_{\delta,H,\Gamma,\Sigma})\Sigma^{-1}}{(s^*_{\delta,H,\Gamma,\Sigma} - \Gamma c^*_{\delta,H,\Gamma,\Sigma})\Sigma^{-1}\Gamma H / H'H} \quad \text{and} \quad \omega(\delta; H, \Gamma, \Sigma) = H s^*_{\delta,H,\Gamma,\Sigma}$$

are continuous functions on $\mathcal{A}$.

E.3.3 *CLT and LLN for triangular arrays* To verify the conditions of Appendix C.5, a CLT and LLN for triangular arrays (applied to the triangular arrays that arise from arbitrary sequences $P_n \in \mathcal{P}$) are useful. We state them here for convenience.

Lemma E.7. *Let $\varepsilon > 0$ be given. Let $\{v_i\}_{i=1}^n$ be an i.i.d. sequence of scalar valued random variables and let $\mathcal{P}$ be a set of probability distributions with $E_P v_i^{2+\varepsilon} \le 1/\varepsilon$, $1/\varepsilon \le E_P v_i^2$ and $E_P v_i = 0$ for all $P \in \mathcal{P}$. Then*

$$\sup_{P \in \mathcal{P}} \sup_{t \in \mathbb{R}} \left| P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n v_i / \sqrt{\mathrm{var}_P(v_i)} \le t \right) - \Phi(t) \right| \to 0.$$

Proof. The result is immediate from Lemma 11.4.1 in Lehmann and Romano (2005) applied to arbitrary sequences $P \in \mathcal{P}$ and the fact that convergence to a continuous cdf is always uniform over the point at which the cdf is evaluated (Lemma 2.11 in van der Vaart (1998)). □

Lemma E.8. *Let $\varepsilon > 0$ be given. Let $\{v_i\}_{i=1}^n$ be an i.i.d. sequence of scalar valued random variables and let $\mathcal{P}$ be a set of probability distributions with $E_P |v_i|^{1+\varepsilon} \le 1/\varepsilon$ for all $P \in \mathcal{P}$. Then $\frac{1}{n} \sum_{i=1}^n v_i - E_P v_i = o_P(1)$ uniformly over $P \in \mathcal{P}$.*

Proof. The stronger result $\sup_{P \in \mathcal{P}} E_P |\frac{1}{n} \sum_{i=1}^n v_i - E_P v_i|^{1+\min\{\varepsilon,2\}} \to 0$ follows from Theorem 3 in von Bahr and Esseen (1965). □

References

Armstrong, T. B. and M. Kolesár (2018), "Optimal inference in a class of regression models." *Econometrica*, 86 (2), 655–683. [8, 9, 13, 19, 20, 21, 34, 37]

Armstrong, T. B. and M. Kolesár (2020), "Sensitivity analysis using approximate moment condition models." arXiv:1808.07387v4. [2]

Bartle, R. G. (1966), *Elements of Integration*, first edition. John Wiley & Sons Inc, New York, NY. [29]

Donoho, D. L. (1994), "Statistical estimation and optimal recovery." *The Annals of Statistics*, 22 (1), 238–270. [34]

Dubé, J.-P., J. T. Fox, and C.-L. Su (2012), "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation." *Econometrica*, 80 (5), 2231–2267. [24]

Hall, A. R. and A. Inoue (2003), "The large sample behaviour of the generalized method of moments estimator in misspecified models." *Journal of Econometrics*, 114 (2), 361–394. [24, 25]

Imbens, G. W. and C. F. Manski (2004), "Confidence intervals for partially identified parameters." *Econometrica*, 72 (6), 1845–1857. [25]

Lehmann, E. L. and J. P. Romano (2005), *Testing Statistical Hypotheses*, third edition. Springer, New York, NY. [9, 10, 29, 37]

Low, M. G. (1995), "Bias-variance tradeoffs in functional estimation problems." *The Annals of Statistics*, 23 (3), 824–835. [34]

Newey, W. K. (1985), "Generalized method of moments specification testing." *Journal of Econometrics*, 29 (3), 229–256. [5]

Newey, W. K. and D. L. McFadden (1994), "Large sample estimation and hypothesis testing." In *Handbook of Econometrics*, Vol. 4 (D. L. McFadden and R. F. Engle, eds.), Chapter 36, 2111–2245, Elsevier. [5]

Rosset, S. and J. Zhu (2007), "Piecewise linear regularized solution paths." *The Annals of Statistics*, 35 (3), 1012–1030. [3]

Shurman, J. (2016), *Calculus and Analysis in Euclidean Space*, first edition. Springer, New York, NY. [29]

Sun, Y., Á. Baricz, and S. Zhou (2010), "On the monotonicity, log-concavity, and tight bounds of the generalized Marcum and Nuttall Q-functions." *IEEE Transactions on Information Theory*, 56 (3), 1166–1186. [6]

van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge University Press, New York, NY. [11, 26, 28, 37]

von Bahr, B. and C.-G. Esseen (1965), "Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$." *The Annals of Mathematical Statistics*, 36 (1), 299–303. [37]

Co-editor Andres Santos handled this manuscript.

Manuscript received 21 April, 2020; final version accepted 1 September, 2020; available online 16 September, 2020.