Supplemental Materials for "Sensitivity Analysis using Approximate Moment Condition Models"

Timothy B. Armstrong* Yale University

Michal Kolesár† Princeton University

February 15, 2019

Appendix D gives the construction of the optimal sensitivity path under ℓ_∞ bounds. Appendix E gives a construction of a submodel satisfying Appendix C.1, verifies the conditions in Appendix C in the misspecified IV model, and collects auxiliary results used in Appendix C.

D Optimal sensitivity under ℓ_∞ bounds

In this case, Equation (22) can be written as

$$\min_k k'\Sigma k/2 \qquad \text{s.t.} \quad \|B'k\|_\infty \leq \overline{B}/M \qquad \text{and} \quad H = -k'\Gamma. \tag{S1}$$

Using the linear transformation T defined in Equation (24), the Lagrangian for this problem can be written as

$$\min_\kappa \kappa' S\kappa/2 + \sum_{i\in I}(\lambda_{+,i}(\kappa_i - b) - \lambda_{-,i}(\kappa_i + b)) + \mu'(H' + G'\kappa).$$

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

The first-order conditions are given by

$$\begin{aligned} i \in I^C: &\qquad e_i'S\kappa + e_i'G\mu = 0 \\ i \in I: &\qquad e_i'S\kappa + e_i'G\mu + \lambda_{+,i} - \lambda_{-,i} = 0 \end{aligned}$$

*email: timothy.armstrong@yale.edu
†email: mkolesar@princeton.edu

The complementary slackness conditions are $\lambda_{+,i}(\kappa_i - b) \geq 0$ and $\lambda_{-,i}(\kappa_i + b) \leq 0$ for $i \in I$, and the feasibility constraints are $\lambda_{+,i}, \lambda_{-,i} \geq 0$ for $i \in I$, and $H' = -G'\kappa$.

Let $\mathcal{A}$ denote the subset of I for which $|\kappa_i| = b$, and let $\mathcal{A}^C = \{1, \ldots, d_g\} \setminus \mathcal{A}$. If $i \in \mathcal{A}$, then either $\kappa_i = b$, so that $\lambda_{-,i} = 0$, and hence $e_i' S\kappa + e_i' G\mu = -\lambda_{+,i} \leq 0$, or else $\kappa_i = -b$, so that $\lambda_{+,i} = 0$, and hence $e_i' S\kappa + e_i' G\mu = \lambda_{-,i} \geq 0$, so that

$$\operatorname{sign}(\kappa_i)(e_i' S\kappa + e_i' G\mu) \leq 0, \qquad i \in \mathcal{A}.$$

If $i \in \mathcal{A}^C$, whether $i \in I$, we have

$$e_i' S\kappa + e_i' G\mu = 0 \qquad i \in \mathcal{A}^C.$$

Combining this with the feasibility constraint, we can write the conditions compactly as

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

This implies

$$\begin{aligned} \mu &= (G'_{\mathcal{A}^C} S^{-1}_{\mathcal{A}^C\mathcal{A}} G_{\mathcal{A}^C})^{-1}(H' + (G'_{\mathcal{A}} - G'_{\mathcal{A}^C} S^{-1}_{\mathcal{A}^C\mathcal{A}^C} S_{\mathcal{A}^C\mathcal{A}})\kappa_{\mathcal{A}}), \\ \kappa_{\mathcal{A}^C} &= -S^{-1}_{\mathcal{A}^C\mathcal{A}^C} G_{\mathcal{A}^C}\mu - S^{-1}_{\mathcal{A}^C\mathcal{A}^C} S_{\mathcal{A}^C\mathcal{A}}\kappa_{\mathcal{A}}. \end{aligned}$$

Consequently, if we're in a region in where the solution path is differentiable with respect to b, we have

$$\begin{aligned} \frac{\partial}{\partial b}\kappa_{\mathcal{A}} &= \operatorname{sign}(\kappa_{\mathcal{A}}) \\ \frac{\partial}{\partial b}\mu &= (G'_{\mathcal{A}^C} S^{-1}_{\mathcal{A}^C\mathcal{A}} G_{\mathcal{A}^C})^{-1}(G'_{\mathcal{A}} - G'_{\mathcal{A}^C} S^{-1}_{\mathcal{A}^C\mathcal{A}^C} S_{\mathcal{A}^C\mathcal{A}}) \operatorname{sign}(\kappa_{\mathcal{A}}) \\ \frac{\partial}{\partial b}\kappa_{\mathcal{A}^C} &= -S^{-1}_{\mathcal{A}^C\mathcal{A}^C} G_{\mathcal{A}^C} \frac{\partial}{\partial b}\mu - S^{-1}_{\mathcal{A}^C\mathcal{A}^C} S_{\mathcal{A}^C\mathcal{A}} \operatorname{sign}(\kappa_{\mathcal{A}}). \end{aligned}$$

The differentiability of path is violated if either (a) the constraint $|\kappa_i| \leq b$ is violated for some $i \in \mathcal{A}^C \cap I$ if $\kappa(b)$ keeps moving in the same direction, and we add i to $\mathcal{A}$ at a point at which this constraint holds with equality; or else (b) the derivative $e_i'(S\kappa + G\mu)$ 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 directions in the preceding display using the new definition of $\mathcal{A}$.

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

1. Initialize $\mu = (G'S^{-1}G)^{-1}H'$, $\kappa = -S^{-1}G\mu$, $b = \|\kappa_I\|_\infty$, and $\mathcal{A} = \operatorname{argmax}_{i\in I}|\kappa_i|$.

2. While $(|\mathcal{A}| < d_g - d_\theta + 1)$:

   (a) Calculate directions:

$$\begin{aligned}(\kappa_\Delta)_\mathcal{A} &= \operatorname{sign}(\kappa_\mathcal{A}) \\ \mu_\Delta &= (G'_{\mathcal{A}^C}S^{-1}_{\mathcal{A}^C\mathcal{A}}G_{\mathcal{A}^C})^{-1}(G'_\mathcal{A} - G'_{\mathcal{A}^C}S^{-1}_{\mathcal{A}^C\mathcal{A}^C}S_{\mathcal{A}^C\mathcal{A}})\operatorname{sign}(\kappa_\mathcal{A}) \\ (\kappa_\Delta)_{\mathcal{A}^C} &= -S^{-1}_{\mathcal{A}^C\mathcal{A}^C}G_{\mathcal{A}^C}\mu_\Delta - S^{-1}_{\mathcal{A}^C\mathcal{A}^C}S_{\mathcal{A}^C\mathcal{A}}\operatorname{sign}(\kappa_\mathcal{A}).\end{aligned}$$

   (b) Set step size to $d = \min\{d_1, d_2\}$, where[1]

$$\begin{aligned}d_1 &= \min\{d > 0 \colon e_i'(S\kappa + G\mu) - de_i'(S\kappa_\Delta + G\mu_\Delta) = 0, i \in \mathcal{A}\} \\ d_2 &= \min\{d > 0 \colon |\kappa_i - d\kappa_{\Delta,i}| = b - d, i \in \mathcal{A}_C \cap \mathcal{I}\}\end{aligned}$$

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

   (c) If $d = d_1$, drop $\operatorname{argmin}(d_1)$ from $\mathcal{A}$, and if $d = d_2$, then add $\operatorname{argmin}(d_2)$ to $\mathcal{A}$.

The solution path $k(\overline{B})$ is then obtained as $k(\overline{B}) = T'\kappa(bM)$.

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 $\varepsilon > 0$,*

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

Let

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

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.

[1] If $|\mathcal{A}| = |I|$, set $d = M$.

Proof. Quadratic mean differentiability follows from Problem 12.6 in Lehmann and Romano (2005), so we just need to show that (35) 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 (35) 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,

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

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),

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

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

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

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

Now consider the derivative of

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

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

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

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

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

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

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

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

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

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

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

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* $\Sigma_{\theta,P} = var_P((y_i - x_i'\theta)z_i)$ *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-Schwartz, the first condition implies $E_P|v_1v_2v_3v_4|^{1+\varepsilon/4} < 1/\varepsilon$ 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 $P \in \mathcal{P}$.

We first verify the conditions of Section 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 $u'W_Pu/\|u\|^2$ 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

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

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

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

(note that u can be taken to satisfy $\|u\| = 1$ without loss of generality, since the formula inside of 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

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

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 a LLN for triangular arrays (Lemma E.8). For the last part of the assumption, we have, by the mean value theorem,

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

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

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

The first term converges uniformly in probability to $\Sigma_{\theta,P}$ by a 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 a 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.

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.3 Auxiliary results

This section contains auxiliary results used in Appendix C. Section 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 Section C.5. Section E.3.2 contains a result on the continuity of the optimal weights with respect to δ, Γ, Σ and H. Section 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

$$\omega(\delta) = \omega(\delta; \mathcal{F}, \mathcal{G}, L, K) = \sup L(g - f) \text{ s.t. } \|K(g - f)\| \leq \delta, f \in \mathcal{F}, g \in \mathcal{G}, \tag{S2}$$

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 (31) 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 Section 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*

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

where K is a linear operator, there is a solution f^*, g^* to the modulus problem (S2) for $\mathcal{F}$ and $\mathcal{G}$ with $g^* - f^* = \tilde{g}^* - \tilde{f}^*$. Furthermore, any solution to the modulus problem (S2) 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 (S2).

Lemma E.3. *For each δ, let (f^*_δ, g^*_δ) be a solution to the modulus problem (S2), 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*

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

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, $\omega(\|K\tilde{h}\|) \ge L\tilde{h} = [\omega(\delta_0)+\omega(\delta_1)]/2 \ge \min\{\omega(\delta_0),\omega(\delta_1)\}$. From this and the fact that ω is strictly increasing on an open interval containing δ_0 and δ_1, it follows that $\|K\tilde{h}\| \ge \min\{\delta_0,\delta_1\}$.

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,

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

Thus, $\|K(h_1-h_0)\|^2/4 \le \delta_1^2 - \min\{\delta_0^2,\delta_1^2\} - (\delta_1^2-\delta_0^2)/2 = |\delta_1^2-\delta_0^2|/2$ 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 (S2), 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*

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

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

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

Rearranging this gives

$$\|K(h-h^*)/2\|^2 = [\|Kh\|^2 + \|Kh^*\|^2]/2 - \|K(h+h^*)/2\|^2. \tag{S3}$$

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 (S2). Let* $\delta_0, L_0, K_0, \mathcal{F}_0, \mathcal{G}_0$ *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* $\delta, K, \mathcal{F}, \mathcal{G}$. *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 $\delta_n, \mathcal{F}_n, \mathcal{G}_n, L_n, K_n$ for large enough n. Thus, for large enough n,

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

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,

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

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

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

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,

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

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 (31) 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

$$\sup Hs \text{ s.t. } (c-\Gamma s)'\Sigma^{-1}(c-\Gamma s) \le \delta^2, c\in\mathcal{D}-\mathcal{C}, s\in\mathbb{R}^{d_\theta}. \tag{S4}$$

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 $\delta_0>0$, $H_0\neq 0$ 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 semi-definite 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 (S4). Let $\mathcal{A}$ denote the set of (δ,H,Γ,Σ) 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 Section 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 iid sequence of scalar valued random variables and let* $\mathcal{P}$ *be a set of probability distributions with* $E_P v_i^{2+\varepsilon} \leq 1/\varepsilon$*,* $1/\varepsilon \leq E_P v_i^2$ *and* $E_P v_i = 0$ *for all* $P \in \mathcal{P}$*. Then*

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

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 iid sequence of scalar valued random variables and let* $\mathcal{P}$ *be a set of probability distributions with* $E_P |v_i|^{1+\varepsilon} \leq 1/\varepsilon$ *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 $\sup_{P \in \mathcal{P}} E_P \left| \frac{1}{n}\sum_{i=1}^n v_i - E_P v_i \right|^{1+\min\{\varepsilon,2\}} \to 0$ follows from Theorem 3 in von Bahr and Esseen (1965). □

References

Armstrong, T. B. and Kolesár, M. (2018). Optimal inference in a class of regression models. *Econometrica*, 86(2):655–683.

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

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

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses*. Springer, New York, third edition.

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

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

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York, NY.

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