Supplemental Materials for "Robust Empirical Bayes Confidence Intervals"

Timothy B. Armstrong*
University of Southern California

Michal Kolesár†
Princeton University

Mikkel Plagborg-Møller‡
Princeton University

May 14, 2022

This supplement is organized as follows. Supplemental Appendix D gives proofs of the formal results in the main text and details on Assumption C.5. Supplemental Appendix E gives details on the simulations. Supplemental Appendix F discusses the power of tests based on our empirical Bayes confidence intervals (EBCIs), and Supplemental Appendix G works through examples of the general shrinkage estimators in Section 6.1.

Appendix D Theoretical details and proofs

Supplemental Appendix D.1 gives technical details on Assumption C.5. The remainder of this Supplemental Appendix provides the proofs of all results in the main paper and in this supplement.

D.1 Primitive conditions for Assumption C.5

To verify Assumption C.5, we will typically have to define θ_i to be scaled by a rate of convergence. Let $\tilde{Y}_i$ be an estimator of a parameter $\vartheta_{i,n}$ with rate of convergence κ_n and asymptotic variance estimate $\hat{\sigma}_i^2$. Suppose that

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \sup_{t\in\mathbb{R}} \left| P\left(\frac{\kappa_n(\tilde{Y}_i - \vartheta_{i,n})}{\hat{\sigma}_i} \leq t \right) - \Phi(t) \right| = 0. \tag{S1}$$

*email: timothy.armstrong@usc.edu
†email: mkolesar@princeton.edu
‡email: mikkelpm@princeton.edu

Then Assumption C.5 holds with $\theta_i = \kappa_n \vartheta_{i,n}$ and $Y_i = \kappa_n \tilde{Y}_i$. Consider an affine estimator $\hat{\vartheta}_i = a_i/\kappa_n + w_i \tilde{Y}_i = (a_i + w_i Y_i)/\kappa_n$ with standard error $\widetilde{\mathrm{se}}_i = w_i \hat{\sigma}_i/\kappa_n$. The corresponding affine estimator of θ_i is $\hat{\theta}_i = \kappa_n \hat{\vartheta}_i = a_i + w_i Y_i$ with standard error $\mathrm{se}_i = \kappa_n \cdot \widetilde{\mathrm{se}}_i = w_i \hat{\sigma}_i$. Then $\vartheta_{i,n} \in \{\hat{\vartheta}_i \pm \widetilde{\mathrm{se}}_i \cdot \hat{\chi}_i\}$ iff. $\theta_i \in \{\hat{\theta}_i \pm \mathrm{se}_i \cdot \hat{\chi}_i\}$. Thus, Theorem C.2 guarantees average coverage of the intervals $\{\hat{\vartheta}_i \pm \widetilde{\mathrm{se}}_i \cdot \hat{\chi}_i\}$ for $\vartheta_{i,n}$. Note that, in order for the moments of θ_i to converge to a non-degenerate constant, we will need to consider triangular arrays $\vartheta_{i,n}$ that converge to zero at a κ_n rate.

As an example, we now verify Assumption C.5 for the linear fixed effects panel data model

$$W_{it} = \vartheta_{i,n} + X_{it}'\beta + u_{it}, \quad i = 1, \ldots, n, t = 1, \ldots, T_i,$$

where X_{it} are covariates in the fixed effects regression.[1] We assume that the T_is increase at the same rate so that, letting $\bar{T} = \frac{1}{n}\sum_{i=1}^n T_i$, we can apply the approach described above with $\kappa_n = \sqrt{\bar{T}}$ to verify Assumption C.5 with $\theta_i = \sqrt{\bar{T}}\vartheta_{i,n}$. We consider the fixed effects estimate of $\vartheta_{i,n}$ formed by regressing W_{it} on X_{it} and indicator variables for each individual i, along with the heteroskedasticity robust variance estimate from this regression. To give the formulas for these estimates, we first define some notation. Let $\bar{W}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} W_{it}$, $\bar{X}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} X_{it}$, $\ddot{X}_{it} = X_{it} - \bar{X}_i$, $\ddot{W}_{it} = W_{it} - \bar{W}_i$, $\bar{u}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} u_{it}$ and $\bar{T} = \frac{1}{n}\sum_{i=1}^n T_i$. Letting $\hat{Q}_{XX} = \frac{1}{n\bar{T}}\sum_{i=1}^n \sum_{t=1}^{T_i} \ddot{X}_{it}\ddot{X}_{it}'$, the fixed effect estimate of β is given by $\hat{\beta} = \hat{Q}_{XX}^{-1}\sum_{i=1}^n \sum_{t=1}^{T_i} \ddot{X}_{it} W_{it}/(n\bar{T})$, and the fixed effect estimate of $\vartheta_{i,n}$ is given by

$$\tilde{Y}_i = \bar{W}_i - \bar{X}_i'\hat{\beta} = \sum_{j=1}^n \sum_{t=1}^{T_j} \left(\mathrm{I}\{i = j\}\frac{1}{T_i} - \frac{1}{n\bar{T}}\bar{X}_i'\hat{Q}_{XX}^{-1}\ddot{X}_{it} \right) W_{it}. \tag{S2}$$

We assume that the T_is grow at the same rate, so that all $\tilde{Y}_i$'s converge at the same rate $1/\sqrt{\bar{T}}$. An estimate of the variance of $\sqrt{\bar{T}}(\tilde{Y}_i - \vartheta_{i,n})$ that is robust to heteroskedasticity in u_{it} is given by

$$\hat{\sigma}_i^2 = \bar{T}\sum_{j=1}^n \sum_{t=1}^{T_i} \left(\mathrm{I}\{i = j\}\frac{1}{T_i} - \frac{1}{n\bar{T}}\bar{X}_i'\hat{Q}_{XX}^{-1}\ddot{X}_{jt} \right)^2 \hat{u}_{jt}^2, \tag{S3}$$

where $\hat{u}_{it} = W_{it} - X_{it}'\hat{\beta} - \tilde{Y}_i$.

We consider "large n large T" asymptotics in which the T_i's are implicitly indexed by n. We make the following assumptions about the T_i's and the distribution $\tilde{P} = \tilde{P}^{(n)}$ of $\{X_{it}, u_{it}\}_{i=1,\ldots,n,\, t=1,\ldots,T_i}$.

Assumption D.1. *For some constants $\gamma > 0$ and $K > 0$,*

[1] We note that, despite the similarity in notation, we do not make any assumption about the relation between the individual level prediction variables X_i used in the individual level predictive regression and the covariates X_{it} used in the fixed effects regression.

1. *u_{it} is mean zero and independent across i and t with $1/K \leq E_{\tilde{P}} u_{it}^2$ and $E_{\tilde{P}}|u_{it}|^{2+\gamma} \leq K$.*

2. *$|X_{it}| \leq K$ for all i, t.*

3. *$n \to \infty$ and $\min_{1 \leq i \leq n} T_i \to \infty$ and $T_i/T_j \leq K$ for all $i, j \leq n$.*

4. *Under $\tilde{P}$, $\sqrt{n\bar{T}}(\hat{\beta} - \beta) = \mathcal{O}(1)$ and the minimum eigenvalue of $\hat{Q}_{XX}$ is greater than $1/K$ with probability approaching one as $n \to \infty$.*

Assumption D.1 is meant to give a simple set of sufficient conditions, and it could be modified for other settings, so long as large n and T asymptotics allow for valid inference on the individual fixed effects. For example, one could relax the independence assumption on the u_{it}'s and modify the standard errors to take into account dependence, so long as one puts enough structure on the dependence that consistent variance estimation is possible as n and T increase. The assumption of bounded covariates is made for simplicity, and could be relaxed, at the possible expense of strengthening the moment condition on u_{it}. The convergence rate assumption on $\hat{\beta}$ follows from standard arguments under appropriate conditions on u_{it} and X_{it} (see, e.g., Stock and Watson, 2008).

Theorem D.1. *Consider the fixed effects setting given above, and suppose Assumption D.1 holds. Then Assumption C.5 holds with $\theta_i = \sqrt{\bar{T}}\vartheta_{i,n}$, $Y_i = \sqrt{\bar{T}}\tilde{Y}_i$ where $\tilde{Y}_i$ is the fixed effects estimator defined in Eq. (S2), and $\hat{\sigma}_i^2$ is the variance estimate defined in Eq. (S3).*

To prove Theorem D.1, we first prove a series of lemmas.

Lemma D.1. *For any $\eta > 0$, $\max_{1 \leq i \leq n} \tilde{P}\left(\sqrt{\bar{T}}|\tilde{Y}_i - \vartheta_{i,n} - \bar{u}_i| > \eta\right) \to 0$.*

Proof. The result is immediate from Assumption D.1 since $\tilde{Y}_i - \vartheta_{i,n} - \bar{u}_i = \bar{X}_{it}'(\beta - \hat{\beta})$. ☐

Lemma D.2. *For any $\eta > 0$, $\max_{1 \leq i \leq n} \tilde{P}\left(\left|\frac{1}{T_i}\sum_{t=1}^{T_i}(\hat{u}_{it}^2 - u_{it}^2)\right| > \eta\right) \to 0$. Furthermore, if $A_{it,n}$ is a triangular array of random variables that are bounded almost surely uniformly in n and i, t, then, for any $\eta > 0$, there exists C such that $\max_{1 \leq i \leq n} \tilde{P}\left(\left|\frac{1}{T_i}\sum_{t=1}^{T_i} A_{it,n}\hat{u}_{it}^2\right| > C\right) < \eta$ and $\tilde{P}\left(\left|\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i} A_{it,n}\hat{u}_{it}^2\right| > C\right) < \eta$ for large enough n.*

Proof. Some algebra shows that $\hat{u}_{it} = \ddot{X}_{it}'(\beta - \hat{\beta}) + u_{it} - \bar{u}_i$. Thus,

$$\hat{u}_{it}^2 = u_{it}^2 + (\beta - \hat{\beta})'\ddot{X}_{it}\ddot{X}_{it}'(\beta - \hat{\beta}) + \bar{u}_i^2 + 2u_{it}\ddot{X}_{it}'(\beta - \hat{\beta}) - 2\bar{u}_i\ddot{X}_{it}'(\beta - \hat{\beta}) - 2\bar{u}_i u_{it}. \quad \text{(S4)}$$

It follows that $\left|\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i} A_{it,n}\hat{u}_{it}^2\right|$ is bounded by $\max_{i,t,n}|A_{it,n}|$ times

$$\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i} u_{it}^2 + (\beta - \hat{\beta})'\hat{Q}_{XX}(\beta - \hat{\beta}) + \frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i} \bar{u}_i^2$$

$$+\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i}2|u_{it}|\cdot|\ddot{X}_{it}'(\beta-\hat{\beta})|-\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i}2|\bar{u}_i||\ddot{X}_{it}'(\beta-\hat{\beta})|-\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i}2|\bar{u}_iu_{it}|.$$

The second term converges in probability to zero by the assumptions on X_{it} and $\hat{\beta}$. The remaining terms are bounded by a constant times $\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(u_{it}^2+\bar{u}_i^2+|u_{it}|+|\bar{u}_i|+|\bar{u}_iu_{it}|)$. By Jensen's inequality, we have $\bar{u}_i^2\leq\frac{1}{T_i}\sum_{i=1}^{T_i}u_{it}^2$, $|\bar{u}_i|\leq\frac{1}{T_i}\sum_{i=1}^{T_i}|u_{it}|$ and

$$\sum_{t=1}^{T_i}|\bar{u}_i||u_{it}|=|\bar{u}_i|\sum_{t=1}^{T_i}|u_{it}|\leq\frac{1}{T_i}\left[\sum_{t=1}^{T_i}|u_{it}|\right]^2\leq T_i\frac{1}{T_i}\sum_{t=1}^{T_i}u_{it}^2=\sum_{t=1}^{T_i}u_{it}^2.$$

This gives a bound of a constant times $\frac{1}{n\bar{T}}\sum_{i=1}^{n}\sum_{t=1}^{T_i}(u_{it}^2+|u_{it}|)$. The last statement in the lemma then follows by Markov's inequality. The second statement in the lemma follows from similar arguments.

For the first statement in the lemma, it follows from (S4) that $\frac{1}{T_i}\sum_{t=1}^{T_i}(u_{it}^2-\hat{u}_{it}^2)$ is equal to

$$(\hat{\beta}-\beta)'\left(\frac{1}{T_i}\sum_{t=1}^{T_i}\ddot{X}_{it}\ddot{X}_{it}'\right)(\hat{\beta}-\beta)-\bar{u}_i^2+2\frac{1}{T_i}\sum_{t=1}^{T_i}u_{it}\ddot{X}_{it}'(\beta-\hat{\beta})-2\frac{\bar{u}_i}{T_i}\sum_{t=1}^{T_i}\ddot{X}_{it}'(\beta-\hat{\beta}).$$

The first term is bounded by a constant that does not depend on i times $|\hat{\beta}-\beta|^2$ (the squared Euclidean norm), which converges in probability to 0 by assumption. The second term has expectation bounded by $\bar{T}^{-1}$ times a constant that does not depend on i. From the bounds on the support of X_{it} and the first moment of u_{it} it follows that the last two terms are bounded by $|\hat{\beta}-\beta|$ times a constant that does not depend on i. This gives the first statement of the lemma. □

Lemma D.3. *Let* $\sigma_i^2=\frac{\bar{T}}{T_i^2}\sum_{t=1}^{T_i}E_{\tilde{P}}u_{it}^2$. *For any* $\eta>0$, $\max_{1\leq i\leq n}\tilde{P}\left(|\hat{\sigma}_i^2-\sigma_i^2|>\eta\right)\to 0$.

Proof. We have $\hat{\sigma}_i^2=I+II+III$ where $I=\frac{\bar{T}}{T_i^2}\sum_{t=1}^{T_i}\hat{u}_{it}^2$,

$$II=\frac{1}{n^2\bar{T}}\sum_{j=1}^{n}\sum_{t=1}^{T_i}\bar{X}_i'\hat{Q}_{XX}^{-1}\ddot{X}_{jt}\ddot{X}_{jt}'\hat{Q}_{XX}^{-1}\bar{X}_i\hat{u}_{jt}^2=\frac{1}{n}\bar{X}_i'\hat{Q}_{XX}^{-1}\hat{Q}_{XXu}\hat{Q}_{XX}^{-1}\bar{X}_i,$$

where $\hat{Q}_{XXu}=\frac{1}{n\bar{T}}\sum_{j=1}^{n}\sum_{t=1}^{T_i}\ddot{X}_{jt}\ddot{X}_{jt}'\hat{u}_{it}^2$, and

$$III=-2\frac{1}{nT_i}\sum_{t=1}^{T_i}\bar{X}_i'\hat{Q}_{XX}^{-1}\ddot{X}_{it}\hat{u}_{it}^2=-2\frac{1}{n}\bar{X}_i'\hat{Q}_{XX}^{-1}\hat{Q}_{Xu,i},$$

where $\hat{Q}_{Xu,i}=\frac{1}{T_i}\sum_{i=1}^{n}\ddot{X}_{it}\hat{u}_{it}^2$. By Lemma D.2 and the condition on the minimum eigenvalue of $\hat{Q}_{XX}$, it follows that $\max_{1\le i\le n}\tilde{P}(|II+III|>\eta/3)\to 0$. It also follows from Lemma D.2 that $\max_{1\le i\le n}\tilde{P}\left(\left|I-\frac{\bar{T}}{T_i^2}\sum_{t=1}^{T_i}u_{it}^2\right|>\eta/3\right)\to 0$. It now suffices to show that $\max_{1\le i\le n}\tilde{P}\left(\left|\frac{\bar{T}}{T_i^2}\sum_{t=1}^{T_i}\left(u_{it}^2-E_{\tilde{P}}u_{it}^2\right)\right|>\eta/3\right)\to 0$. By von Bahr and Esseen (1965, Theorem 3),

$$E_{\tilde{P}}\left|\frac{\bar{T}}{T_i^2}\sum_{t=1}^{T_i}\left(u_{it}^2-E_{\tilde{P}}u_{it}^2\right)\right|^{1+\gamma/2}\le 2(\bar{T}/T_i^2)^{1+\gamma/2}\sum_{t=1}^{T_i}E_{\tilde{P}}\left|u_{it}^2-E_{\tilde{P}}u_{it}^2\right|^{1+\gamma/2},$$

which is bounded by a constant times $\bar{T}^{-\gamma/2}$ by the moment bound on u_{it} and the bound on T_i/T_j. The result now follows from Markov's inequality. □

Let $\tilde{Z}_i=\sqrt{\bar{T}}\bar{u}_i/\sigma_i$, $R_{1,i}=\sqrt{\bar{T}_i}(\tilde{Y}_i-\vartheta_{i,n}-\bar{u}_i)/\sigma_i$ and $R_{2,i}=\hat{\sigma}_i-\sigma_i$. We have

$$\frac{\sqrt{\bar{T}}(\tilde{Y}_i-\vartheta_{i,n})}{\hat{\sigma}_i}=\left(\tilde{Z}_i+R_{1,i}\right)\frac{\sigma_i}{\sigma_i+R_{2,i}}=\tilde{Z}_i-\tilde{Z}_i\frac{R_{2,i}}{\sigma_i+R_{2,i}}+R_{1,i}\frac{\sigma_i}{\sigma_i+R_{2,i}}.$$

It follows from the Lyapounov Central Limit Theorem (applied to Z_{i_n} for arbitrary sequences $i_n\le n$) that $\lim_{n\to\infty}\max_{1\le i\le n}\sup_{t\in\mathbb{R}}\left|P\left(\tilde{Z}_i\le t\right)-\Phi(t)\right|=0$. The conclusion of Theorem D.1 then follows so long as $\max_{1\le i\le n}P\left(\left|\tilde{Z}_i\frac{R_{2,i}}{\sigma_i+R_{2,i}}\right|+\left|R_{1,i}\frac{\sigma_i}{\sigma_i+R_{2,i}}\right|>\eta\right)\to 0$ for any $\eta>0$. But this follows by Lemmas D.1 and D.3 and the fact that σ_i is bounded from above and from below away from zero by the moment assumptions on u_{it}.

D.2 Proof of Lemma 4.1

We first show that the non-coverage probability is weakly decreasing in $w_{EB,i}$. Let $\Gamma(m)$ denote the space of probability measures on $\mathbb{R}$ with second moment bounded above by $m>0$. Abbreviating $z_{1-\alpha/2}$ by z, let $\tilde{\rho}(w)=\rho(1/w-1,z/\sqrt{w})$ denote the maximal undercoverage when $w_{EB,i}=w$. By definition of ρ,

$$\tilde{\rho}(w)=\sup_{F\in\Gamma(1/w-1)}E_{b\sim F}\left[P(|b-Z|>z/\sqrt{w}\mid b)\right]=\sup_{F\in\Gamma(1/w-1)}P_{b\sim F}\left(\sqrt{w}|b-Z|>z\right),\tag{S5}$$

where Z denotes a $N(0,1)$ variable that is independent of b.

Consider any w_0,w_1 such that $0<w_0\le w_1<1$. Let $F_1^*\in\Gamma(1/w_1-1)$ denote the least-favorable distribution—i.e., the distribution that achieves the supremum (S5)—when $w=w_1$. (Proposition B.1 implies that the supremum is in fact attained at a particular discrete distribution.) Let $\tilde{F}_0$ denote the distribution of the linear combination

$$\sqrt{\frac{w_1}{w_0}}b - \sqrt{\frac{w_1 - w_0}{w_0}}Z$$

when $b \sim F_1^*$ and $Z \sim N(0,1)$ are independent. Note that the second moment of this distribution is $\frac{w_1}{w_0} \cdot \frac{1-w_1}{w_1} + \frac{w_1-w_0}{w_0} = \frac{1-w_0}{w_0}$, so $\tilde{F}_0 \in \Gamma(1/w_0 - 1)$. Thus, if we let $\tilde{Z}$ denote another $N(0,1)$ variable that is independent of (b, Z), then

$$\begin{aligned}
\tilde{\rho}(w_0) &\geq P_{b\sim\tilde{F}_0}\left(\sqrt{w_0}|b - Z| > z\right) = P_{b\sim F_1^*}\left(\sqrt{w_0}\left|\sqrt{\frac{w_1}{w_0}}b - \sqrt{\frac{w_1-w_0}{w_0}}\tilde{Z} - Z\right| > z\right) \\
&= P_{b\sim F_1^*}\left(\left|\sqrt{w_1}b - \underbrace{(\sqrt{w_1 - w_0}\tilde{Z} + \sqrt{w_0}Z)}_{\sim N(0,w_1)}\right| > z\right) = P_{b\sim F_1^*}\left(\sqrt{w_1}|b - Z| > z\right) = \tilde{\rho}(w_1).
\end{aligned}$$

Next, we derive the limit of the non-coverage probability as $w_{EB,i} \to 0$. It follows from Proposition B.1 that

$$\rho(t, \chi) = \sup_{0\leq\lambda\leq1} (1-\lambda)r(0,\chi) + \lambda r((t/\lambda)^{1/2}, \chi).$$

Note that $r(0, z/\sqrt{w}) \to 0$ as $w \to 0$. Thus,

$$\lim_{w\to0} \tilde{\rho}(w) = \lim_{w\to0} \rho\left(1/w - 1, z/\sqrt{w}\right) = \lim_{w\to0} \sup_{0\leq\lambda\leq1} \lambda r\left(\lambda^{-1/2}(1/w-1)^{1/2}, zw^{-1/2}\right),$$

provided the latter limit exists. We will first show that the supremum above is bounded below by an expression that tends to $1/\max\{z^2, 1\}$. Then we will show that the supremum is bounded above by an expression that tends to $1/z^2$ (and the supremum is obviously also bounded above by 1).

Let $\varepsilon(w) \geq 0$ be any function of w such that $\varepsilon(w) \to 0$ and $\varepsilon(w)(1/w-1)^{1/2} \to \infty$ as $w \to 0$. Let $\tilde{z} = \max\{z, 1\}$. Note first that, by setting $\lambda = (\tilde{z}(1-w)^{-1/2} + \varepsilon(w))^{-2} \in [0,1]$,

$$\sup_{0\leq\lambda\leq1} \lambda r\left(\lambda^{-1/2}(1/w-1)^{1/2}, zw^{-1/2}\right) \geq \frac{r\left((\tilde{z}(1-w)^{-1/2} + \varepsilon(w))(1/w-1)^{1/2}, zw^{-1/2}\right)}{(\tilde{z}(1-w)^{-1/2} + \varepsilon(w))^2} \to \frac{1}{\tilde{z}^2}$$

as $w \to 0$, since $r(b, \chi) \to 1$ when $(b - \chi) \to \infty$, and

$$\begin{aligned}
(\tilde{z}(1-w)^{-1/2} + \varepsilon(w))(1/w-1)^{1/2} - zw^{-1/2} &\geq (z(1-w)^{-1/2} + \varepsilon(w))(1/w-1)^{1/2} - zw^{-1/2} \\
&= \varepsilon(w)(1/w-1)^{1/2} \to \infty.
\end{aligned}$$

Second,

$$\sup_{0\leq\lambda\leq 1} \lambda r\left(\lambda^{-1/2}(1/w-1)^{1/2}, zw^{-1/2}\right) \\ \leq \Phi\left(-zw^{-1/2}\right) + \sup_{0\leq\lambda\leq 1} \lambda\Phi\left(\lambda^{-1/2}(1/w-1)^{1/2} - zw^{-1/2}\right).$$

The first term above tends to 0 as $w \to 0$. The second term equals

$$\max\Bigg\{ \sup_{0\leq\lambda\leq(z-\varepsilon(w))^{-2}} \lambda\Phi\left(\lambda^{-1/2}(1/w-1)^{1/2} - zw^{-1/2}\right), \\ \sup_{(z-\varepsilon(w))^{-2}<\lambda\leq 1} \lambda\Phi\left(\lambda^{-1/2}(1/w-1)^{1/2} - zw^{-1/2}\right) \Bigg\},$$

where the first argument is bounded above by $\sup_{0\leq\lambda\leq(z-\varepsilon(w))^{-2}} \lambda = (z-\varepsilon(w))^{-2} \to \frac{1}{z^2}$. The second argument tends to 0 as $w \to 0$, since

$$\lambda^{-1/2}(1/w-1)^{1/2} - zw^{-1/2} \leq (\lambda^{-1/2} - z)(1/w-1)^{1/2} \leq -\varepsilon(w)(1/w-1)^{1/2}$$

for all $\lambda > (z-\varepsilon(w))^{-2}$, and the far right-hand side above tends to $-\infty$ as $w \to 0$.

D.3 Proof of Proposition B.1

Since $r(b,\chi)$ is symmetric in b, Eq. (5) is equivalent to maximizing $E_F[r_0(t,\chi)]$ over distributions F of t with $E_F[t] = m_2$. Let $\bar{r}(t,\chi)$ denote the least concave majorant of $r_0(t,\chi)$. We first show that $\rho(m_2,\chi) = \bar{r}(m_2,\chi)$.

Observe that $\rho(m_2,\chi) \leq \bar{\rho}(m_2,\chi)$, where $\bar{\rho}(m_2,\chi)$ denotes the value of the problem

$$\bar{\rho}(m_2,\chi) = \sup_F E_F[\bar{r}(t,\chi)] \quad \text{s.t.} \quad E_F[t] = m_2.$$

Furthermore, since $\bar{r}$ is concave, by Jensen's inequality, the optimal solution F^* to this problem puts point mass on m_2, so that $\bar{\rho}(m_2,\chi) = \bar{r}(m_2,\chi)$, and hence $\rho(m_2,\chi) \leq \bar{r}(m_2,\chi)$.

Next, we show that the reverse inequality holds, $\rho(m_2,\chi) \geq \bar{r}(m_2,\chi)$. By Corollary 17.1.4 on page 157 in Rockafellar (1970), the majorant can be written as

$$\bar{r}(t,\chi) = \sup\{\lambda r_0(x_1,\chi) + (1-\lambda) r_0(x_2,\chi) \colon \lambda x_1 + (1-\lambda)x_2 = t,\ 0 \leq x_1 \leq x_2, \lambda \in [0,1]\}, \quad \text{(S6)}$$

which corresponds to the problem in Eq. (5), with the distribution F constrained to be a discrete distribution with two support points. Since imposing this additional constraint on

F must weakly decrease the value of the solution, it follows that $\rho(m_2, \chi) \geq \bar{r}(m_2, \chi)$. Thus, $\rho(m_2, \chi) = \bar{r}(m_2, \chi)$. The proposition then follows by Lemma D.5 below.

Lemma D.4. *Let $r_0(t, \chi) = r(\sqrt{t}, \chi)$. If $\chi \leq \sqrt{3}$, then r_0 is concave in t. If $\chi > \sqrt{3}$, then its second derivative is positive for t small enough, negative for t large enough, and crosses zero exactly once, at some $t_1 \in [\chi^2 - 3, (\chi - 1/\chi)^2]$.*

Proof. Letting ϕ denote the standard normal density, the first and second derivative of $r_0(t) = r_0(t, \chi)$ are given by

$$\begin{aligned} r_0'(t) &= \frac{1}{2\sqrt{t}} \left[\phi(\sqrt{t} - \chi) - \phi(\sqrt{t} + \chi) \right] \geq 0, \\ r_0''(t) &= \frac{\phi(\chi - \sqrt{t})(\chi\sqrt{t} - t - 1) + \phi(\chi + \sqrt{t})(\chi\sqrt{t} + t + 1)}{4t^{3/2}} \\ &= \frac{\phi(\chi + \sqrt{t})}{4t^{3/2}} \left[e^{2\chi\sqrt{t}}(\chi\sqrt{t} - t - 1) + (\chi\sqrt{t} + t + 1) \right] = \frac{\phi(\chi + \sqrt{t})}{4t^{3/2}} f(\sqrt{t}), \end{aligned}$$

where the last line uses $\phi(a + b)e^{-2ab} = \phi(a - b)$, and

$$f(u) = (\chi u + u^2 + 1) - e^{2\chi u}(u^2 - \chi u + 1).$$

Thus, the sign of $r_0''(t)$ corresponds to that of $f(\sqrt{t})$, with $r_0''(t) = 0$ if and only if $f(\sqrt{t}) = 0$. Observe $f(0) = 0$, and $f(u) < 0$ is negative for u large enough, since the term $-u^2 e^{2\chi u}$ dominates. Furthermore,

$$\begin{aligned} f'(u) &= 2u + \chi - e^{2\chi u}(2\chi(u^2 - \chi u + 1) + 2u - \chi) & f'(0) &= 0 \\ f''(u) &= e^{2\chi u}(4\chi^3 u - 4\chi^2 u^2 - 8\chi u - 2) + 2 & f''(0) &= 0 \\ f^{(3)}(u) &= 4\chi e^{2\chi u}(2\chi^3 u + \chi^2(1 - 2u^2) - 6\chi u - 3) & f^{(3)}(0) &= 4\chi(\chi^2 - 3). \end{aligned}$$

Therefore for $u > 0$ small enough, $f(u)$, and hence $r_0''(u^2)$ is positive if $\chi^2 \geq 3$, and negative otherwise.

Now suppose that $f(u_0) = 0$ for some $u_0 > 0$, so that

$$\chi u_0 + u_0^2 + 1 = e^{2\chi u_0}(u_0^2 - \chi u_0 + 1) \tag{S7}$$

Since $\chi u + u^2 + 1$ is strictly positive, it must be the case that $u_0^2 - \chi u_0 + 1 > 0$. Multiplying and dividing the expression for $f'(u)$ above by $u_0^2 - \chi u_0 + 1$ and plugging in the identity in

Eq. (S7) and simplifying the expression yields

$$\begin{aligned} f'(u_0) &= \frac{(u_0^2 - \chi u_0 + 1)(2u_0 + \chi) - (\chi u_0 + u_0^2 + 1)(2\chi(u_0^2 - \chi u_0 + 1) + 2u_0 - \chi)}{u_0^2 - \chi u_0 + 1} \\ &= \frac{2u_0^2\chi(\chi^2 - 3 - u_0^2)}{u_0^2 - \chi u_0 + 1}. \end{aligned} \tag{S8}$$

Suppose $\chi^2 < 3$. Then $f'(u_0) < 0$ at all positive roots u_0 by Eq. (S8). But if $\chi^2 < 3$, then $f(u)$ is initially negative, so by continuity it must be that $f'(u_1) \geq 0$ at the first positive root u_1. Therefore, if $\chi^2 \leq 3$, f, and hence r_0'', cannot have any positive roots. Thus, if $\chi^2 \leq 3$, r_0 is concave as claimed.

Now suppose that $\chi^2 \geq 3$, so that $f(u)$ is initially positive. By continuity, this implies that $f'(u_1) \leq 0$ at its first positive root u_1. By Eq. (S8), this implies $u_1 \geq \sqrt{\chi^2 - 3}$. As a result, again by Eq. (S8), $f(u_i) \leq 0$ for all remaining positive roots. But since by continuity, the signs of f' must alternate at the roots of f, this implies that f has at most a single positive root. Since f is initially positive, and negative for large enough u, it follows that it has a single positive root $u_1 \geq \sqrt{\chi^2 - 3}$. Finally, to obtain an upper bound for $t_1 = u_1^2$, observe that if $f(u_1) = 0$, then, by Taylor expansion of the exponential function,

$$1 + \frac{2\chi u_1}{\chi u_1 + u_1^2 + 1} = e^{2\chi u_1} \geq 1 + 2\chi u_1 + 2(\chi u_1)^2,$$

which implies that $1 \geq (1 + \chi u_1)(\chi u_1 + u_1^2 + 1)$, so that $u_1 \leq \chi - 1/\chi$. $\square$

Lemma D.5. *The problem in Eq. (S6) can be written as*

$$\bar{r}(t, \chi) = \sup_{u \geq t} \{(1 - t/u) r_0(0, \chi) + \frac{t}{u} r_0(u, \chi)\}. \tag{S9}$$

Let $t_0 = 0$ if $\chi \leq \sqrt{3}$, and otherwise let $t_0 > 0$ denote the solution to $r_0(0, \chi) - r_0(u, \chi) + u\frac{\partial}{\partial u} r_0(u, \chi) = 0$. This solution is unique, and the optimal u solving Eq. (S9) satisfies $u = t$ for $t > t_0$ and $u = t_0$ otherwise.

Proof. If in the optimization problem in Eq. (S6), the constraint on x_2 binds, or either constraint on λ binds, then the optimum is achieved at $r_0(t) = r_0(t, \chi)$, with $x_1 = t$ and $\lambda = 1$ and x_2 arbitrary; $x_2 = t$ and $\lambda = 0$ and x_1 arbitrary; or else $x_1 = x_2$ and λ arbitrary. In any of these cases $\bar{r}$ takes the form in Eq. (S9) as claimed. If, on the other hand, these constraints do not bind, then $x_2 > t > x_1$, and substituting $\lambda = (x_2 - t)/(x_2 - x_1)$ into the objective function yields the first-order conditions

$$r_0(x_2) - (x_2 - x_1)r_0'(x_1) - r_0(x_1) = \mu\frac{(x_2 - x_1)^2}{(x_2 - t)}, \tag{S10}$$

$$r_0(x_2) + (x_1 - x_2)r_0'(x_2) - r_0(x_1) = 0, \tag{S11}$$

where $\mu \geq 0$ is the Lagrange multiplier on the constraint that $x_1 \geq 0$. Subtracting Eq. (S11) from Eq. (S10) and applying the fundamental theorem of calculus then yields

$$\mu\frac{x_2 - x_1}{(x_2 - t)} = r_0'(x_2) - r_0'(x_1) = \int_{x_1}^{x_2} r_0''(t)\,dt > 0, \tag{S12}$$

which implies that $\mu > 0$. Here the last inequality follows because by Taylor's theorem, Eq. (S11) implies that $\int_{x_1}^{x_2} r_0''(t)(t - x_1)\,dt = 0$. Since r_0'' is positive for $t \leq t_1$ and negative for $t \geq t_1$ by Lemma D.4, it follows that $x_1 \leq t_1 \leq x_2$, and hence that

$$0 = \int_{x_1}^{t_1} r_0''(t)(t - x_1)\,dt + \int_{t_1}^{x_2} r_0''(t)(t - x_1)\,dt$$
$$< (t_1 - x_1)\int_{x_1}^{t_1} r_0''(t)\,dt + (t_1 - x_1)\int_{t_1}^{x_2} r_0''(t)\,dt = (t_1 - x_1)\int_{x_1}^{x_2} r_0''(t)\,dt.$$

Finally Eq. (S12) implies that $\mu > 0$, so that $x_1 = 0$ at the optimum. Consequently, the problem in Eq. (S6) takes the form in Eq. (S9) as claimed.

To show the second part of Lemma D.5, note that by Lemma D.4, if $\chi \leq \sqrt{3}$, r_0 is concave, so that we can put $u = t$ in Eq. (S9). Otherwise, let $\mu \geq 0$ denote the Lagrange multiplier associated with the constraint $u \geq t$ in the optimization problem in Eq. (S9). The first-order condition is then given by

$$r_0(0) - r_0(u) + ur_0'(u) = \frac{-\mu u^2}{t}.$$

Let $f(u) = r_0(0) - r_0(u) + ur_0'(u)$. Since $f'(u) = ur_0''(u)$, it follows from Lemma D.4 that $f(u)$ is increasing for $u \leq t_1$ and decreasing for $u \geq t_1$. Since $f(0) = 0$ and $\lim_{u\to\infty} f(u) < r_0(0) - 1 < 0$, it follows that $f(u)$ has exactly one positive zero, at some $t_0 > t_1$. Thus, if $t < t_0$, $u = t_0$ is the unique solution to the first-order condition. If $t > t_0$, $u = t$ is the unique solution. □

D.4 Proof of Proposition B.2

Since $r(b, \chi)$ is symmetric in b, letting $t = b^2$, we can equivalently write the optimization problem as

$$\rho(m_2, \kappa, \chi) = \sup_F E_F[r_0(t, \chi)] \quad \text{s.t.} \quad E_F[t] = m_2,\ E_F[t^2] = \kappa m_2^2, \tag{S13}$$

where $r_0(t, \chi) = r(\sqrt{t}, \chi)$, and the supremum is over all distributions supported on the positive part of the real line. The dual of this problem is

$$\min_{\lambda_0, \lambda_1, \lambda_2} \lambda_0 + \lambda_1 m_2 + \lambda_2 \kappa m_2^2 \qquad \text{s.t.} \quad \lambda_0 + \lambda_1 t + \lambda_2 t^2 \geq r_0(t), \quad 0 \leq t < \infty,$$

where λ_0 the Lagrange multiplier associated with the implicit constraint that $E_F[1] = 1$, and $r_0(t) = r_0(t, \chi)$. So long as $\kappa > 1$ and $m_2 > 0$, so that the moments $(m_2, \kappa m_2^2)$ lie in the interior of the space of possible moments of F, by the duality theorem in Smith (1995), the duality gap is zero, and if F^* and $\lambda^* = (\lambda_0^*, \lambda_1^*, \lambda_2^*)$ are optimal solutions to the primal and dual problems, then F^* has mass points only at those t with $\lambda_0^* + \lambda_1^* t + \lambda_2^* t^2 = r(\sqrt{t}, \chi)$.

Define t_0 as in Lemma D.5. First, we claim that if $m_2 \geq t_0$, then $\rho(m_2, \kappa, \chi) = \rho(m_2, \chi)$, the value of the objective function in Proposition B.1. The reason that adding the constraint $E_F[t^2] = \kappa m_2^2$ does not change the optimum is that it follows from the proof of Proposition B.1 that the distribution achieving the rejection probability $\rho(m_2, \chi)$ is a point mass on m_2. Consider adding another support point $x_2 = \sqrt{n}$ with probability $\kappa m_2^2/n$, with the remaining probability on the support point m_2. Then, as $n \to \infty$, the mean of this distribution converges to m_2, and its second moment converges to κm_2^2, so that the constraints in Eq. (S13) are satisfied, while the rejection probability converges to $\rho(m_2, \chi)$. Since imposing the additional constraint $E_F[t^2] = \kappa m_2^2$ cannot increase optimum, the claim follows.

Suppose that $m_2 < t_0$. At optimum, the majorant $g(x) = \lambda_0 + \lambda_1 t + \lambda_2 t^2$ in the dual constraint must satisfy $g(x_0) = r_0(x_0)$ for at least one $x_0 > 0$. Otherwise, if the constraint never binds, we could lower the value of the objective function by decreasing λ_0; furthermore, $x_0 = 0$ cannot be the unique point at which the constraint binds, since by the duality theorem, this would imply that the distribution that puts point mass on 0 maximizes the primal, which cannot be the case.

At such x_0, we must also have $g'(x_0) = r_0'(x_0)$, otherwise the constraint would be locally violated. Using this fact together with the equality $g(x_0) = r_0(x_0)$, we therefore have that $\lambda_0 = r_0(x_0) - \lambda_1 x_0 - \lambda_2 x_0^2$ and $\lambda_1 = r_0'(x_0) - 2\lambda_2 x_0$, so that the dual problem may be written as

$$\min_{x_0>0,\lambda_2} r_0(x_0) + r_0'(x_0)(m_2 - x_0) + \lambda_2((x_0 - m_2)^2 + (\kappa - 1)m_2^2)$$
$$\text{s.t.} \quad r_0(x_0) + r_0'(x_0)(x - x_0) + \lambda_2(x - x_0)^2 \geq r_0(x). \quad \text{(S14)}$$

Since $\kappa > 1$, the objective is increasing in λ_2. Therefore, given x_0, the optimal value of λ_2 is as small as possible while still satisfying the constraint,

$$\lambda_2 = \sup_{x>0} \delta(x; x_0), \qquad \delta(x; x_0) = \frac{r_0(x) - r_0(x_0) - r_0'(x_0)(x - x_0)}{(x - x_0)^2}.$$

Next, we claim that the dual constraint cannot bind for $x_0 > t_0$. Observe that $\lambda_2 \geq 0$, otherwise the constraint would be violated for t large enough. However, setting $\lambda_2 = 0$ still satisfies the constraint. This is because the function $h(x) = r_0(x_0) + r_0'(x_0)(x - x_0) - r_0(x)$ is minimized at $x = x_0$, with its value equal to 0. To see this, note that its derivative equals zero if $r_0'(x_0) = r'(x)$. By Lemma D.4, $r_0'(t)$ is increasing for $t \leq t_0$ and decreasing for $t > t_0$. Therefore, if $r_0'(x_0) < r_0'(0)$, $h'(x) = 0$ has a unique solution, $x = x_0$. If $r_0'(x_0) > r_0'(0)$, there is another solution at some $x_1 \in [0, t_0]$. However, $h''(x_1) = -r_0''(x_1) < 0$, so $h(x)$ achieves a local maximum here. Since $h(0) > 0$ by arguments in the proof of Lemma D.4, it follows that the maximum of $h(x)$ occurs at $x = x_0$, and equals 0. However, Eq. (S14) cannot be maximized at $(x_0, 0)$, since by Proposition B.1, setting $(x_2, \lambda_2) = (t_0, 0)$ achieves a lower value of the objective function, which proves the claim.

Therefore, Eq. (S14) can be written as

$$\min_{0<x_0\leq t_0} r_0(x_0) + r_0'(x_0)(m_2 - x_0) + ((x_0 - m_2)^2 + (\kappa - 1)m_2^2) \sup_{x\geq 0} \delta(x; x_0),$$

To finish the proof of the proposition, it remains to show that δ cannot be maximized at $x > t_0$. This follows from observing that the dual constraint in Eq. (S14) binds at any x that maximizes δ. However, by the claim above, the constraint cannot bind for $x > t_0$.

D.5 Proof of Theorem C.1

To prove this theorem, we begin with some lemmas.

Lemma D.6. *Under Assumption C.1, we have, for any deterministic $\chi_1, \ldots, \chi_n$, and any*

$\mathcal{X} \in \mathcal{A}$ *with* $N_{\mathcal{X},n} \to \infty$*,*

$$\lim_{n\to\infty} \frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \tilde{P}\left(|Z_i| > \chi_i\right) - \frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} r(b_{i,n}, \chi_i) = 0.$$

Furthermore, if $Z_i - \tilde{b}_i$ *is independent over* i *under* $\tilde{P}$*, then*

$$\frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \mathrm{I}\{|Z_i| > \chi_i\} - \frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} r(b_{i,n}, \chi_i) = o_{\tilde{P}}(1).$$

Proof. For any $\varepsilon > 0$, $\frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \mathrm{I}\{|Z_i| > \chi_i\}$ is bounded from above by

$$\frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \mathrm{I}\{|Z_i - \tilde{b}_i + b_{i,n}| > \chi_i - \varepsilon\} + \frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \mathrm{I}\{|\tilde{b}_i - b_{i,n}| \geq \varepsilon\}.$$

The expectation under $\tilde{P}$ of the second term converges to zero by Assumption C.1. The expectation under $\tilde{P}$ of the first term is $\frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \tilde{r}_{i,n}(b_{i,n}, \chi_i - \varepsilon)$ where $\tilde{r}_{i,n}(b,\chi) = \tilde{P}(Z_i - \tilde{b}_i < -\chi - b) + 1 - \tilde{P}(Z_i - \tilde{b}_i \leq \chi - b)$. Note that $r_{i,n}(b,\chi)$ converges to $r(b,\chi)$ uniformly over b, χ under Assumption C.1, using the fact that the convergence in Assumption C.1 is uniform in t by Lemma 2.11 in van der Vaart (1998), and the fact that $\tilde{P}(Z_i - \tilde{b}_i < -\chi - b) = \lim_{t\uparrow -\chi - b} P(Z_i - \tilde{b}_i \leq t)$. It follows that the expectation of the above display under $\tilde{P}$ is bounded by $\frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \tilde{r}(b_{i,n}, \chi_i - \varepsilon) + o(1)$. If $Z_i - \tilde{b}_i$ is independent over i, the variance of each term in the above display converges to zero, so that the above display equals $\frac{1}{N_{\mathcal{X},n}} \sum_{i\in\mathcal{I}_{\mathcal{X},n}} \tilde{r}(b_{i,n}, \chi_i - \varepsilon) + o_{\tilde{P}}(1)$. Taking $\varepsilon \to 0$ and noting that $r(b,\chi)$ is uniformly continuous in both arguments, and using an analogous argument with a lower bound, gives the result. □

Lemma D.7. $\rho_g(\chi; m)$ *is continuous in* χ*. Furthermore, for any* m^* *in the interior of the set of values of* $\int g(b)\, dF(b)$*, where* F *ranges over all probability measures on* $\mathbb{R}$*,* $\rho_g(\chi; m)$ *is continuous with respect to* m *at* m^**.*

Proof. To show continuity with respect to χ, note that

$$|\rho_g(\chi; m) - \rho_g(\tilde{\chi}; m)| \leq \sup_F \left| \int [r(b,\chi) - r(b,\tilde{\chi})]\, dF(b) \right| \quad \text{s.t.} \int g(b)\, dF(b) = m,$$

where we use the fact that the difference between suprema of two functions over the same constraint set is bounded by the supremum of the absolute difference of the two functions. The above display is bounded by $\sup_b |r(b,\chi) - r(b,\tilde{\chi})|$, which is bounded by a constant times $|\tilde{\chi} - \chi|$ by uniform continuity of the standard normal CDF.

To show continuity with respect to m, note that, by Lemma D.8 below, the conditions for the Duality Theorem in Smith (1995, p. 812) hold for m in a small enough neighborhood of m^*, so that

$$\rho_g(\chi; m) = \inf_{\lambda_0,\lambda} \lambda_0 + \lambda' m \quad \text{s.t.} \quad \lambda_0 + \lambda' g(b) \geq r(b, \chi) \text{ for all } b \in \mathbb{R}$$

and the above optimization problem has a finite solution. Thus, for m in this neighborhood of m^*, $\rho_g(\chi; m)$ is the infimum of a collection of affine functions of m, which implies that it is concave function of m (Boyd and Vandenberghe, 2004, p. 81). By concavity, $\rho_g(\chi; m)$ is also continuous as a function of m in this neighborhood of m^*. □

Lemma D.8. *Suppose that μ is in the interior of the set of values of $\int g(b)\, dF(b)$ as F ranges over all probability measures with respect to the Borel sigma algebra, where $g : \mathbb{R} \to \mathbb{R}^p$. Then $(1, \mu')'$ is in the interior of the set of values of $\int (1, g(b)')'\, dF(b)$ as F ranges over all measures with respect to the Borel sigma algebra.*

Proof. Let μ be in the interior of the set of values of $\int g(b)\, dF(b)$ as F ranges over all probability measures with respect to the Borel sigma algebra. We need to show that, for any $a, \tilde{\mu}$ with $(a, \tilde{\mu}')'$ close enough to $(1, \mu')$, there exists a measure F such that $\int (1, g(b)')' dF(b) = (a, \tilde{\mu}')'$. To this end, note that, $\tilde{\mu}/a$ can be made arbitrarily close to μ by making $(a, \tilde{\mu}')'$ close to $(1, \mu')$. Thus, for $(a, \tilde{\mu}')'$ close enough to $(1, \mu')$, there exists a probability measure $\tilde{F}$ with $\int g(b)\, d\tilde{F}(b) = \tilde{\mu}/a$. Let F be the measure defined by $F(A) = a\tilde{F}(A)$ for any measurable set A. Then $\int (1, g(b)')' dF(b) = a \int (1, g(b)')' d\tilde{F}(b) = (a, \tilde{\mu})$. This completes the proof. □

Lemma D.9. *Let M be a compact subset of the interior of the set of values of $\int g(b)\, dF(b)$, where F ranges over all measures on $\mathbb{R}$ with the Borel σ-algebra. Suppose $\lim_{b\to\infty} g_j(b) = \lim_{b\to-\infty} g_j(b) = \infty$ and that $\inf_b g_j(b) \geq 0$ for some j. Then $\lim_{\chi\to\infty} \sup_{m\in M} \rho_g(\chi; m) = 0$ and $\rho_g(\chi; m)$ is uniformly continuous with respect to $(\chi, m')'$ on the set $[0, \infty) \times M$.*

Proof. The first claim (that $\lim_{\chi\to\infty} \sup_{m\in M} \rho_g(\chi; m) = 0$) follows by Markov's inequality and compactness of M. Given $\varepsilon > 0$, let $\overline{\chi}$ be large enough so that $\rho_g(\chi; m) < \varepsilon$ for all $\chi \in [\overline{\chi}, \infty)$ and all $m \in M$. By Lemma D.7, $\rho_g(\chi; m)$ is continuous on $[0, \overline{\chi} + 1] \times M$, so, since $[0, \overline{\chi} + 1] \times M$ is compact, it is uniformly continuous on this set. Thus, there exists δ such that, for any χ, m and $\tilde{\chi}, \tilde{m}$ with $\chi, \tilde{\chi} \leq \overline{\chi} + 1$ and $\|(\tilde{\chi}, \tilde{m}')' - (\chi, m')'\| \leq \delta$, we have $|\rho_g(\chi; m) - \rho_g(\tilde{\chi}; \tilde{m})| < \varepsilon$. If we also set $\delta < 1$, then, if either $\chi \geq \overline{\chi} + 1$ or $\tilde{\chi} \geq \overline{\chi} + 1$ we must have both $\chi \geq \overline{\chi}$ and $\tilde{\chi} \geq \overline{\chi}$, so that $\rho_g(\tilde{\chi}; \tilde{m}) < \varepsilon$ and $\rho_g(\chi; m) < \varepsilon$, which also implies $|\rho_g(\chi; m) - \rho_g(\tilde{\chi}; \tilde{m})| < \varepsilon$. This completes the proof. □

For any $\varepsilon > 0$, let

$$\overline{\rho}_g(\chi; m, \varepsilon) = \sup_{\tilde{m}\in B_\varepsilon(m)} \rho_g(\chi; \tilde{m}) \quad \text{and} \quad \underline{\rho}_g(\chi; m, \varepsilon) = \inf_{\tilde{m}\in B_\varepsilon(m)} \rho_g(\chi; \tilde{m}).$$

Lemma D.10. *Let M be a compact subset of the interior of the set of values of $\int g(b)\,dF(b)$, where F ranges over all measures on $\mathbb{R}$ with the Borel σ-algebra. Suppose $\lim_{b\to\infty} g_j(b) = \lim_{b\to-\infty} g_j(b) = \infty$ and $\inf_b g_j(b) \geq 0$ for some j. Then, for ε smaller than a constant that depends only on M, the functions $\overline{\rho}_g(\chi; m, \varepsilon)$ and $\underline{\rho}_g(\chi; m, \varepsilon)$ are continuous in χ. Furthermore, we have $\lim_{\varepsilon\to 0} \sup_{\chi\in[0,\infty), m\in M}[\overline{\rho}_g(\chi; m, \varepsilon) - \underline{\rho}_g(\chi; m, \varepsilon)] = 0$.*

Proof. For ε smaller than a constant that depends only on M, the set $\cup_{m\in M} B_\varepsilon(m)$ is contained in another compact subset of the interior of the set of values of $\int g(b)\,dF(b)$, where F ranges over all measures on $\mathbb{R}$ with the Borel σ-algebra. The result then follows from Lemma D.9, where, for the first claim, we use the fact that $|\overline{\rho}_g(\chi; m, \varepsilon) - \overline{\rho}_g(\tilde{\chi}; m, \varepsilon)| \leq \sup_{\tilde{m}\in B_\varepsilon(m)} |\rho_g(\chi; \tilde{m}) - \rho_g(\tilde{\chi}; \tilde{m})|$ and similarly for $\underline{\rho}_g$. $\square$

We now prove Theorem C.1. Given $\mathcal{X} \in \mathcal{A}$ and $\varepsilon > 0$, let $m_1, \ldots, m_J$ and $\mathcal{X}_1, \ldots, \mathcal{X}_J$ be as in Assumption C.3. Let $\underline{\chi}_j = \min\{\chi\colon \underline{\rho}_g(\chi; m_j, 2\varepsilon) \leq \alpha\}$. For $\hat{m}_i \in B_{2\varepsilon}(m_j)$, we have $\underline{\rho}_g(\chi; m_j, 2\varepsilon) \leq \rho_g(\chi; \hat{m}_i)$ for all χ, so that, using the fact that $\underline{\rho}_g(\chi; m_j, 2\varepsilon)$ and $\rho_g(\chi; \hat{m}_i)$ are weakly decreasing in χ, we have $\underline{\chi}_j \leq \hat{\chi}_i$. Thus, letting $\underline{\tilde{\chi}}^{(n)}$ denote the sequence with ith element equal to $\underline{\chi}_j$ when $\tilde{X}_i \in \mathcal{X}_j$, we have

$$\begin{aligned} ANC_n(\hat{\chi}^{(n)}; \mathcal{X}) &\leq \max_{1\leq j\leq J} ANC_n(\underline{\tilde{\chi}}^{(n)}; \mathcal{X}_j) \\ &\leq \max_{1\leq j\leq J}\left[\frac{1}{N_{\mathcal{X}_j,n}} \sum_{i\in\mathcal{I}_{\mathcal{X}_j,n}} \mathrm{I}\{\hat{m}_i \notin B_{2\varepsilon}(m_j)\} + \frac{1}{N_{\mathcal{X}_j,n}} \sum_{i\in\mathcal{I}_{\mathcal{X}_j,n}} \mathrm{I}\{|Z_i| > \underline{\chi}_j\}\right]. \end{aligned}$$

The first term is bounded by $\frac{1}{N_{\mathcal{X}_j,n}} \sum_{i\in\mathcal{I}_{\mathcal{X}_j,n}} \mathrm{I}\{\|\hat{m}_i - m(\tilde{X}_i)\| > \varepsilon\}$ since, for $i \in \mathcal{I}_{\mathcal{X}_j,n}$, we have $\|\hat{m}_i - m_j\| \leq \varepsilon + \|\hat{m}_i - m(\tilde{X}_i)\|$. This converges in probability (and expectation) to zero under $\tilde{P}$ by Assumption C.2. By Lemma D.6, the second term is equal to, letting $F_{j,n}$ denote the empirical distribution of the $b_{i,n}$'s for i with $x_i \in \mathcal{X}_j$,

$$\int r(b, \underline{\chi}_j)\,dF_{j,n}(b) + R_n \leq \overline{\rho}_g(\underline{\chi}_j; \mu_j, 2\varepsilon) + R_n$$

where R_n is a term such that $E_{\tilde{P}} R_n \to 0$ and such that, if $Z_i - \tilde{b}_i$ is independent over i under $\tilde{P}$, then R_n converges in probability to zero under $\tilde{P}$. The result will now follow if we can show that $\max_{1\leq j\leq J}[\overline{\rho}_g(\underline{\chi}_j; \mu_j, 2\varepsilon) - \alpha]$ can be made arbitrarily small by making ε small.

This holds by Lemma D.10 and the fact that $\underline{\rho}_g(\underline{\chi}_j; \mu_j, 2\varepsilon) \le \alpha$ by construction.

D.6 Proof of Theorem C.2

To prove Theorem C.2, we will verify the conditions of Theorem C.1 with $\mathcal{A}$ given in Assumption C.7, $m_j(\tilde{X}_i) = c(\gamma, \sigma_i)^{\ell_j} \mu_{0,\ell_j}$, $\tilde{b}_i = c(\hat{\gamma}, \hat{\sigma}_i)(\theta_i - \hat{X}_i'\hat{\delta})$ and $b_{i,n} = c(\gamma, \sigma_i)(\theta_i - \hat{X}_i'\delta)$ where $c(\gamma, \sigma) = \frac{w(\gamma,\sigma)-1}{w(\gamma,\sigma)\sigma}$. The first part of Assumption C.1 is immediate from Assumption C.5 since $Z_i - \tilde{b}_i = (Y_i - \theta_i)/\hat{\sigma}_i$. For the second part, we have

$$
\begin{aligned}
\tilde{b}_i - b_{i,n} &= c(\hat{\gamma}, \hat{\sigma}_i)(\theta_i - \hat{X}_i'\hat{\delta}) - c(\gamma, \sigma_i)(\theta_i - X_i'\delta) \\
&= [c(\hat{\gamma}, \hat{\sigma}_i) - c(\gamma, \sigma_i)](\theta_i - X_i'\delta) + c(\hat{\gamma}, \hat{\sigma}_i) \cdot [(\hat{X}_i - X_i)'\hat{\delta} - X_i'(\delta - \hat{\delta})].
\end{aligned}
$$

For $\|\theta_i\| + \|X_i\| \le C$, the above expression is bounded by

$$
[c(\hat{\gamma}, \hat{\sigma}_i) - c(\gamma, \sigma_i)] \cdot (\|\delta\| + 1) \cdot C + c(\hat{\gamma}, \hat{\sigma}_i) \left[\|\hat{\delta} - \delta\| \cdot C + \|\hat{X}_i - X_i\| \cdot (C + \|\hat{\delta} - \delta\|) \right].
$$

By uniform continuity of $c()$ on an open set containing $\{\gamma\} \times \mathcal{S}_1$, for every $\varepsilon > 0$ there exists $\eta > 0$ such that $\|(\hat{\sigma}_i - \sigma_i, \hat{\gamma} - \gamma, \hat{\delta}' - \delta', \hat{X}_i' - X_i')'\| \le \eta$ implies that the absolute value of the above display is less than ε. Thus, for any $\mathcal{X} \in \mathcal{A}$,

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{1}{N_{\mathcal{X},n}} \sum_{i \in \mathcal{I}_{\mathcal{X},n}} \tilde{P}(|\tilde{b}_i - b_{i,n}| \ge \varepsilon) \\
&\quad \le \lim_{n\to\infty} \frac{1}{N_{\mathcal{X},n}} \sum_{i \in \mathcal{I}_{\mathcal{X},n}} \tilde{P}(\|(\hat{\sigma}_i - \sigma_i, \hat{\gamma} - \gamma, \hat{\delta}' - \delta', \hat{X}_i' - X_i')'\| > \eta)\, \mathrm{I}\{\|\theta_i\| + \|X_i\| \le C\} \\
&\qquad + \limsup_{n\to\infty} \frac{1}{N_{\mathcal{X},n}} \sum_{i \in \mathcal{I}_{\mathcal{X},n}} \mathrm{I}\{\|\theta_i\| + \|X_i\| > C\}.
\end{aligned}
$$

The first limit is zero by Assumption C.6. The last limit converges to zero as $C \to \infty$ by the second part of Assumption C.7 and Markov's inequality. This completes the verification of Assumption C.5.

We now verify Assumption C.2. Given $\mathcal{X} \in \mathcal{A}$ and given $\varepsilon > 0$, we can partition $\mathcal{X}$ into sets $\mathcal{X}_1, \ldots, \mathcal{X}_J$ such that, for some $c_1, \ldots, c_J$, we have $|c(\gamma, \sigma_i)^{\ell_k} - c_j^{\ell_k}| < \varepsilon$ for all $k = 1, \ldots, p$ whenever $i \in \mathcal{I}_{\mathcal{X}_j,n}$ for some j. Thus, for each j and k,

$$
\frac{1}{N_{\mathcal{X}_j,n}} \sum_{i \in \mathcal{I}_{\mathcal{X}_j,n}} b_{i,n}^{\ell_k} - m_k(\tilde{X}_i) = \frac{1}{N_{\mathcal{X}_j,n}} \sum_{i \in \mathcal{I}_{\mathcal{X}_j,n}} c(\gamma, \sigma_i)^{\ell_k} \left[(\theta_i - X_i'\delta)^{\ell_k} - \mu_{0,\ell_k} \right]
$$

$$= c_j^{\ell_k} \cdot \frac{1}{N_{\mathcal{X}_j,n}} \sum_{i \in \mathcal{I}_{\mathcal{X}_j},n} \left[(\theta_i - X_i'\delta)^{\ell_k} - \mu_{0,\ell_k} \right]$$
$$+ \frac{1}{N_{\mathcal{X}_j,n}} \sum_{i \in \mathcal{I}_{\mathcal{X}_j},n} [c(\gamma, \sigma_i)^{\ell_k} - c_j^{\ell_k}] \left[(\theta_i - X_i'\delta)^{\ell_k} - \mu_{0,\ell_k} \right].$$

Under Assumption C.7, the first term converges to 0 and the second term is bounded up to an $o(1)$ term by ε times a constant that depends only on K. Since the absolute value of $\frac{1}{N_{\mathcal{X},n}} \sum_{i \in \mathcal{I}_{\mathcal{X}},n} b_{i,n}^{\ell_k} - m_k(\tilde{X}_i)$ is bounded by the maximum over j of the absolute value of the above display, and since ε can be chosen arbitrarily small, the first part of Assumption C.2 follows.

For the second part of Assumption C.2, we have $\hat{m}_{i,k} - m_k(\tilde{X}_i) = c(\gamma, \sigma_i)\hat{\mu}_{\ell_j} - c(\gamma, \sigma_i)^{\ell_j} \mu_{0,\ell_j}$. By uniform continuity of $(\tilde{\gamma}', \sigma, \mu_{\ell_1}, \ldots, \mu_{\ell_p})' \mapsto (c(\gamma, \sigma_i)^{\ell_1} \mu_{\ell_1}, \ldots, c(\gamma, \sigma_i)^{\ell_p} \mu_{\ell_p})'$ in an open set containing $\{\gamma\} \times \mathcal{S}_1 \times \{(\mu_{0,\ell_1}, \ldots, \mu_{0,\ell_p})'\}$, for any $\varepsilon > 0$, there exists $\eta > 0$ such that $\|(\hat{\gamma}' - \gamma', \hat{\sigma}_i - \sigma, \hat{\mu}_{\ell_1} - \mu_{0,\ell_1}, \ldots, \hat{\mu}_{\ell_p} - \mu_{0,\ell_p})\| < \eta$ implies $\|\hat{m}_{i,k} - m_k(\tilde{X}_i)\| < \varepsilon$. Thus,

$$\max_{1 \le i \le n} \tilde{P}(\|\hat{m}_i - m(\tilde{X}_i)\| \ge \varepsilon) \le \max_{1 \le i \le n} \tilde{P}(\|(\hat{\gamma}' - \gamma', \hat{\sigma}_i - \sigma, \hat{\mu}_{\ell_1} - \mu_{0,\ell_1}, \ldots, \hat{\mu}_{\ell_p} - \mu_{0,\ell_p})\| < \eta),$$

which converges to zero by Assumptions C.6 and C.7. This completes the verification of Assumption C.2.

Assumption C.3 follows immediately from compactness of the set $\mathcal{S}_1 \times \cdots \times \mathcal{S}_1$ and uniform continuity of $m()$ on this set. Assumption C.4 follows from Assumption C.7 and Lemma D.11 below. This completes the proof of Theorem C.2.

Lemma D.11. *Suppose that, as F ranges over all probability measures with respect to the Borel sigma algebra, $(\mu_{\ell_1}, \ldots, \mu_{\ell_p})'$ is interior to the set of values of $\int (b^{\ell_1}, \ldots, b^{\ell_p})' \, dF(b)$. Let $c \in \mathbb{R}$. Then, as F ranges over all probability measures with respect to the Borel sigma algebra, $(c^{\ell_1}\mu_{\ell_1}, \ldots, c^{\ell_p}\mu_{\ell_p})'$ is also in the interior of the set of values of $\int (b^{\ell_1}, \ldots, b^{\ell_p})' \, dF(b)$.*

Proof. We need to show that, for any vector r with $\|r\|$ small enough, there exists a probability measure F such that $\int (b^{\ell_1}, \ldots, b^{\ell_p})' \, dF(b) = (c^{\ell_1}\mu_{\ell_1} + r_1, \ldots, c^{\ell_p}\mu_{\ell_p} + r_p)'$. Let $\tilde{\mu}_{\ell_k} = \mu_{\ell_k} + r_k / c^{\ell_k}$. For $\|r\|$ small enough, there exists a probability measure $\tilde{F}$ with $\int b^{\ell_k} \, dF(b) = \tilde{\mu}_{\ell_k}$ for each k. Let F denote the probability measure of cB when B is a random variable distributed according to $\tilde{F}$. Then $\int b^{\ell_k} \, dF(b) = c^{\ell_k} \int b^{\ell_k} \, d\tilde{F} = c^{\ell_k}\tilde{\mu}_{\ell_k} = c^{\ell_k}\mu_{\ell_k} + r_k$ as required. $\square$

Appendix E Details for simulations

Supplemental Appendix E.1 gives details on the Monte Carlo designs in Section 4.4. Supplemental Appendix E.2 considers an additional Monte Carlo exercise calibrated to the empirical application in Section 7.

E.1 Details for panel data simulation designs

The simulation results reported in Section 4.4 consider the following six distributions for θ_i, each of which satisfies $\mathrm{var}(\theta_i) = \mu_2$:

1. Normal (kurtosis $\kappa = 3$): $\theta_i \sim N(0, \mu_2)$.

2. Scaled chi-squared ($\kappa = 15$): $\theta_i \sim \sqrt{\mu_2/2} \cdot \chi^2(1)$.

3. 2-point ($\kappa = 1/(0.9 \cdot 0.1) - 3 \approx 8.11$), with $\theta_i = 0$ w.p. 0.9 and $\theta_i = \mu_2/(0.9 \cdot 0.1)$ w.p. 0.1.

4. 3-point ($\kappa = 2$):
$$\theta_i \sim \begin{cases} -\sqrt{\mu_2/0.5} & \text{w.p. } 0.25, \\ 0 & \text{w.p. } 0.5, \\ \sqrt{\mu_2/0.5} & \text{w.p. } 0.25. \end{cases}$$

5. Least favorable for robust EBCI: The (asymptotically as $n, T \to \infty$) least favorable distribution for the robust EBCI that exploits only second moments, i.e.,
$$\theta_i \sim \begin{cases} -\sqrt{\mu_2/\min\{\frac{m_2}{t_0(m_2,\alpha)}, 1\}} & \text{w.p. } \frac{1}{2}\min\{\frac{m_2}{t_0(m_2,\alpha)}, 1\}, \\ 0 & \text{w.p. } 1 - \min\{\frac{m_2}{t_0(m_2,\alpha)}, 1\}, \\ \sqrt{\mu_2/\min\{\frac{m_2}{t_0(m_2,\alpha)}, 1\}} & \text{w.p. } \frac{1}{2}\min\{\frac{m_2}{t_0(m_2,\alpha)}, 1\}, \end{cases}$$
where $m_2 = 1/\mu_2$, and $t_0(m_2, \alpha)$ is the number defined in Proposition B.1 with $\chi = \mathrm{cva}_\alpha(m_2)$. The kurtosis $\kappa(\mu_2, \alpha) = 1/\min\{\frac{1/\mu_2}{t_0(1/\mu_2,\alpha)}, 1\}$ depends on μ_2 and α.

6. Least favorable for parametric EBCI: The (asymptotically) least favorable distribution for the parametric EBCI. This is the same distribution as above, except that now $t_0(m_2, \alpha)$ is the number defined in Proposition B.1 with $\chi = z_{1-\alpha/2}/\sqrt{\mu_2/(1+\mu_2)}$.

E.2 Heteroskedastic design

We now provide average coverage and length results for a heteroskedastic simulation design. We base the design on the effect estimates and standard errors obtained in the empirical application in Section 7. Because we do not have access to the underlying data set, we treat the standard errors as known and impose exact conditional normality of the initial estimates. Let $(\hat{\theta}_i, \hat{\sigma}_i)$, $i = 1, \dots, n$, denote the $n = 595$ baseline shrinkage point estimates and associated standard errors from this application. Note for reference that $E_n[\hat{\theta}_i] = 0.0602$, and $E_n[(\hat{\theta}_i - \bar{\theta})^2] \cdot E_n[1/\hat{\sigma}_i^2] = 0.6698$, where E_n denotes the sample mean.

The simulation design imposes independence of θ_i and σ_i, consistent with the moment independence assumption required by our baseline EBCI procedure, see Remark 3.1. We calibrate the design to match one of three values for the signal-to-noise ratio $E[\varepsilon_i^2/\sigma_i^2] \in \{0.1, 0.5, 1\}$. Specifically, a simulation sample $(Y_i, \theta_i, \sigma_i)$, $i = 1, \dots, n$, is created as follows:

1. Sample $\tilde{\theta}_i$, $i = 1, \dots, n$, with replacement from the empirical distribution $\{\hat{\theta}_j\}_{j=1}^n$.
2. Sample σ_i, $i = 1, \dots, n$, with replacement from the empirical distribution $\{\hat{\sigma}_j\}_{j=1}^n$.
3. Compute $\theta_i = \bar{\theta} + \sqrt{m/c} \cdot (\tilde{\theta}_i - \bar{\theta})$, $i = 1, \dots, n$. Here m is the desired population value of $E[\varepsilon_i^2/\sigma_i^2]$ and $c = 0.6698$.
4. Draw $Y_i \overset{indep}{\sim} N(\theta_i, \sigma_i^2)$, $i = 1, \dots, n$.

The kurtosis of θ_i equals the sample kurtosis of $\hat{\theta}_i$, which is 3.0773. We use precision weights $\omega_i = \sigma_i^{-2}$ when computing the EBCIs, as in Section 7.

Table S1 shows that our baseline implementation of the 95% robust EBCI achieves average coverage above the nominal confidence level, regardless of the signal-to-noise ratio $E[\varepsilon_i^2/\sigma_i^2] \in \{0.1, 0.5, 1\}$. This contrasts with the feasible version of the parametric EBCI, which undercovers by 9.3 percentage points.

Appendix F Statistical power

The efficiency calculations in Figure 3 of Section 4.2 show that our EBCI is substantially shorter than the conventional confidence interval (CI) based on the unshrunk estimate Y_i if the signal-to-noise ratio is small enough. Here, we perform analogous calculations using the statistical power of tests based on a given CI as the measure of efficiency.

Consider testing $H_{0,i} : \theta_i = \theta_0$ for some null value θ_0 by rejecting when $\theta_0 \notin CI_i$, where CI_i is our robust EBCI. As with the efficiency calculations in Section 4.2, we consider efficiency under the baseline model in Eq. (9), and we consider the asymptotic setting in

Table S1: Monte Carlo simulation results: heteroskedastic design.

n	Robust, μ_2 only		Robust, μ_2 & κ		Parametric	
	Oracle	Baseline	Oracle	Baseline	Oracle	Baseline
Panel A: Average coverage (%), minimum across 3 DGPs						
595	98.9	96.0	96.1	96.0	94.3	85.7
Panel B: Relative average length, average across 3 DGPs						
595	1.56	1.51	1.00	1.48	0.89	0.86

Notes: Nominal average confidence level $1-\alpha = 95\%$. Top row: type of EBCI procedure. "Oracle": true μ_2 and κ (but not δ) known. "Baseline": $\hat{\mu}_2$ and $\hat{\kappa}$ estimates as in Section 3.2. For each DGP, "average coverage" and "average length" refer to averages across observations $i = 1, \dots, n$ and across 5,000 Monte Carlo repetitions. Average CI length is measured relative to the oracle robust EBCI that exploits μ_2 and κ.

which $\mu_{1,i} = X_i'\delta$, μ_2, σ_i^2 and $\kappa = 3$ can be treated as known. We compute the average power of this test (averaged over the baseline normal prior, conditional on X_i, σ_i), and we compare it to the average power of the conventional two-sided z-test based on the unshrunk estimate in the same setting. Since the distribution of θ_i is atomless, the average power is given by the rejection probability $P(\theta_0 \notin CI_i \mid X_i, \sigma_i)$. Let $d_i = (\mu_{1,i} - \theta_0)/\sigma_i$ denote the standardized average distance between the true parameter θ_i and the null θ_0. Under the baseline model in Eq. (9), the average power of a test based on the robust EBCI given in Eq. (12) with $\kappa = 3$ is thus given by

$$P\left(\theta_i \notin CI_i \mid X_i, \sigma_i\right) = P\left(\left|\frac{Y_i - \mu_{1,i}}{\sqrt{\sigma_i^2 + \mu_2}} + \frac{d_i \sigma_i}{w_{EB,i}\sqrt{\sigma_i^2 + \mu_2}}\right| > \frac{\mathrm{cva}_\alpha(\sigma_i^2/\mu_2, 3)}{\sqrt{1 + \mu_2/\sigma_i^2}} \middle| X_i, \sigma_i\right)$$
$$= r\left(\frac{d_i\sqrt{1 - w_{EB,i}}}{w_{EB,i}}, \mathrm{cva}_\alpha(1/w_{EB,i} - 1, 3)\sqrt{1 - w_{EB,i}}\right),$$

with r given in Eq. (4), and we use the fact that $Y_i - \mu_{1,i} \mid X_i, \sigma_i \sim \mathcal{N}(0, \sigma_i^2 + \mu_2)$ under Eq. (9). The two-sided z-test based on the unshrunk estimate Y_i rejects when $|Y_i - \theta_0| > z_{1-\alpha/2}\sigma_i$. By analogous reasoning, it follows that the average power of this test is given by

$$P\left(|Y_i - \theta_0| > z_{1-\alpha/2}\sigma_i \mid X_i, \sigma_i\right) = r\left(d_i\sqrt{1 - w_{EB,i}}, z_{1-\alpha/2}\sqrt{1 - w_{EB,i}}\right).$$

Both expressions depend only on d_i and the shrinkage $w_{EB,i}$ (or, equivalently, since $\mu_2/\sigma_i^2 = w_{EB,i}/(1 - w_{EB,i})$, the signal-to-noise ratio μ_2/σ_i^2).

Figure S1 computes the power of the robust EBCI-based test and the z-test as a function of the normalized distance $d_i = (\mu_{1,i} - \theta_0)/\sigma_i$ and the shrinkage $w_{EB,i}$ for $\alpha = 0.05$. The third panel shows the difference in power, with positive values indicating greater power for the EBCI-based test.

The graphs show that the EBCI-based test is more powerful than the z-test for a given shrinkage $w_{EB,i}$ (equivalently, given signal-to-noise ratio) when the normalized distance is large enough, while being less powerful when it is small enough. To get some intuition for this, note that the EBCI differs from the unshrunk CI in two ways: it is shorter, and it uses shrinkage to move the center of the CI toward the regression line $\mu_{1,i} = X_i'\delta$. Shortening the CI makes the EBCI more powerful than the test based on the unshrunk CI, but the effect of moving the center of the CI is ambiguous: it increases power when the regression line $\mu_{1,i}$ is far from the null θ_0, while decreasing power when $\mu_{1,i}$ is close to θ_0. On net, the graphs show that the EBCI-based test displays substantial gains in average power when the amount of shrinkage is large, even for small to moderate distances to the null.

Appendix G Applications of general shrinkage

Here we provide theoretical and numerical results for the soft thresholding EBCI and the Poisson EBCI, discussed in Examples 6.2 and 6.3 in Section 6.1.

G.1 Soft thresholding

The soft thresholding EBCI is obtained by calibrating the highest posterior density (HPD) set in the homoskedastic normal model with a baseline Laplace prior for θ_i. The HPD set $\mathcal{S}(Y_i;\chi)$ in Eq. (21) takes the form of an interval, and is available in closed form. In particular, it follows by direct calculation that the posterior density for θ is given by $p(\theta \mid Y_i) = e^{\bar{c}(Y_i) - \frac{1}{2\sigma^2}\theta^2 + Y_i\theta/\sigma^2 - |\theta|\sqrt{2/\mu_2}}$, where $\bar{c}(Y) = \frac{1}{2}\log(2/\pi\sigma^2) - \log(q(\sqrt{\sigma^2/\mu_2} - Y/\sigma\sqrt{2}) + q(\sqrt{\sigma^2/\mu_2} + Y/\sigma\sqrt{2}))$. Here $q(x) = 2e^{x^2}\Phi(-x\sqrt{2})$ is the scaled complementary error function. Consequently, $\mathcal{S}(Y;\chi)$ equals the intersection of the solution sets for two quadratic inequalities,

$$\mathcal{S}(Y;\chi) = \left\{\theta\colon \tfrac{\theta^2}{2\sigma^2} - \left(\tfrac{Y}{\sigma^2} - \sqrt{\tfrac{2}{\mu_2}}\right)\theta \leq \chi + \bar{c}(Y)\right\} \cap \left\{\theta\colon \tfrac{\theta^2}{2\sigma^2} - \left(\tfrac{Y}{\sigma^2} + \sqrt{\tfrac{2}{\mu_2}}\right)\theta \leq \chi + \bar{c}(Y)\right\}.$$

Since the quadratic term is positive in both inequalities, $\mathcal{S}(Y;\chi)$ is given by an intersection of two intervals, and is therefore itself an interval. The non-coverage function $\tilde{r}(\theta_i,\chi)$ in Eq. (18) is computed via numerical quadrature. The linear program in Eq. (19) is solved by discretizing the support for θ_i. In addition to computing a robust soft thresholding EBCI, we

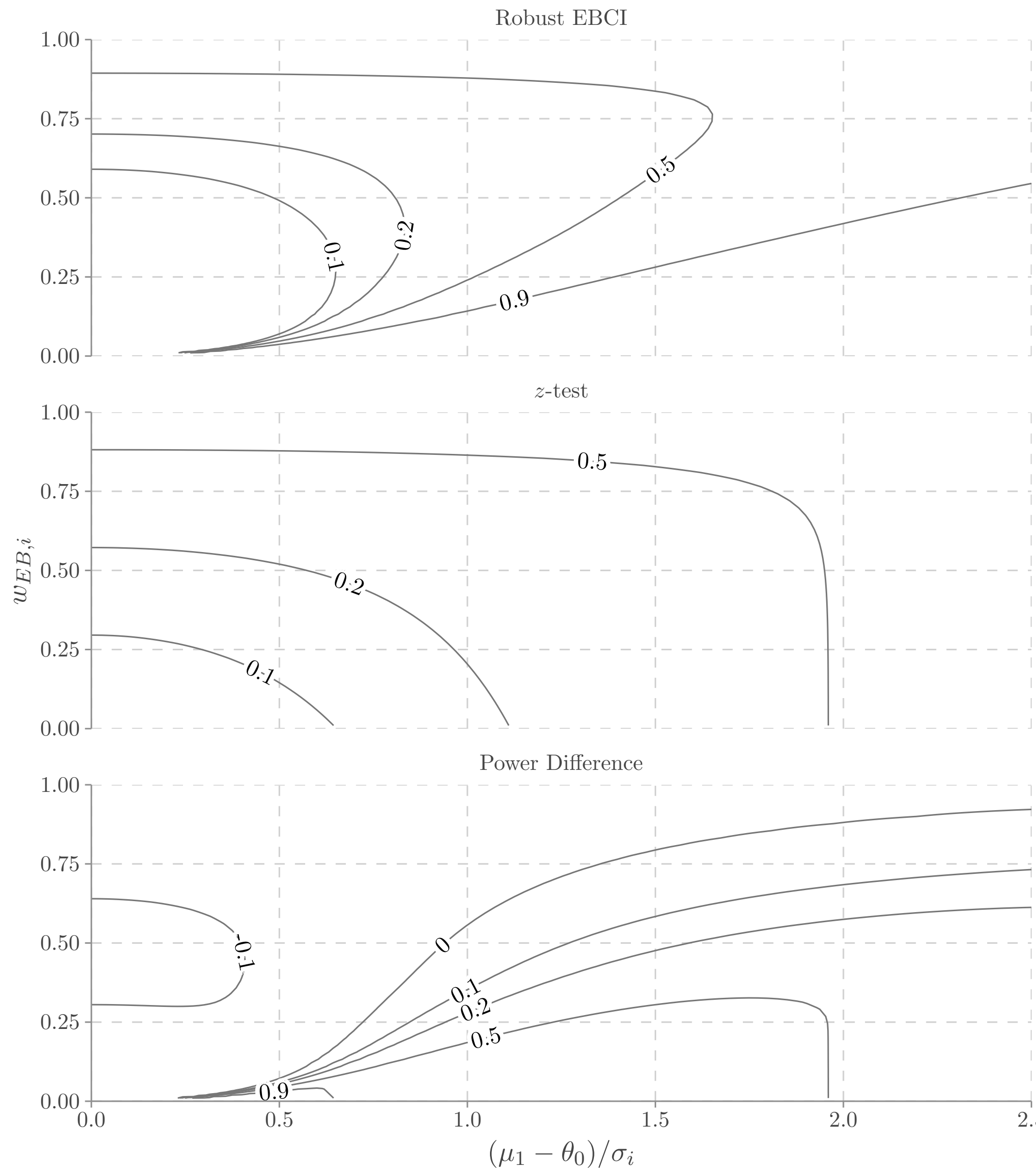

Figure S1: Average power of the robust EBCI and the z-test based on the unshrunk estimate as a function of the normalized average distance to the null and of the shrinkage $w_{EB,i}$.

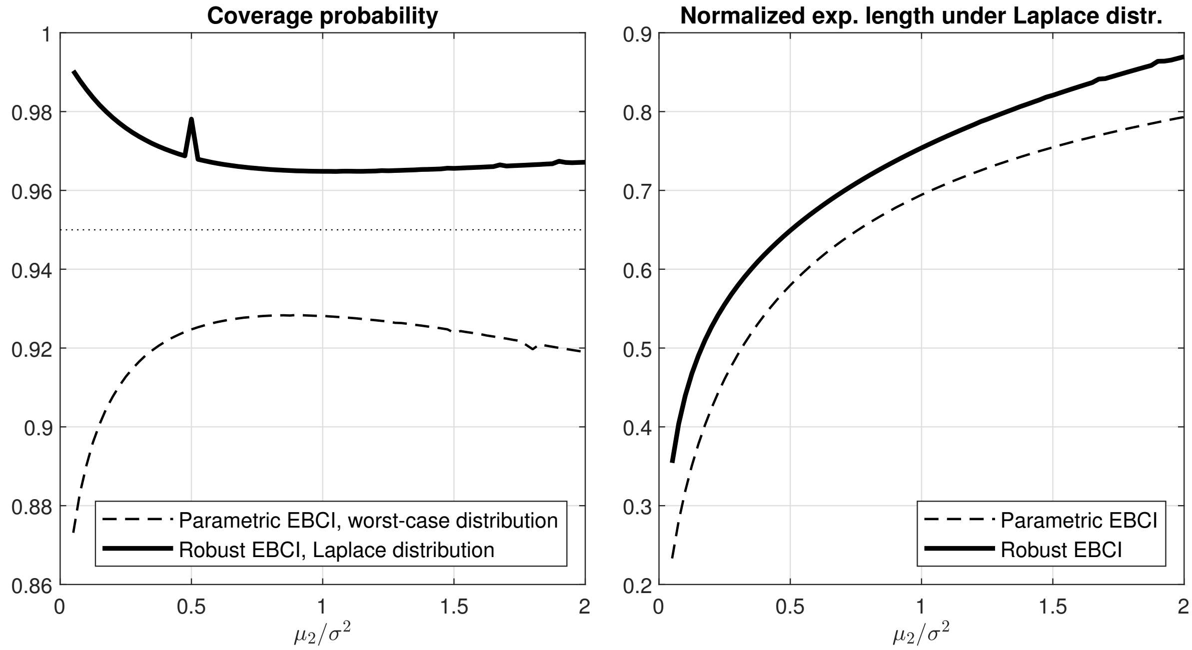

Figure S2: Soft thresholding EBCIs in the normal means model, $\alpha = 0.05$. The expected length is normalized by the length of the unshrunk CI. The grid for θ_i for the linear program in Eq. (19) is given by 500 points equally spaced on $[-10, 10]$. Integrals over the Y_i distribution are truncated at the endpoints -10 and 10.

can similarly compute a parametric soft thresholding EBCI, with χ solving $E_F[r(\theta_i, \chi)] = \alpha$; here F is the Laplace distribution with second moment μ_2.

We now compute the coverage and expected length of the soft thresholding EBCIs. We consider an asymptotic setting where $\mu_2 = E[\theta^2]$ is known, and this is the only constraint imposed when we compute the robust EBCI. Figure S2 shows the coverage and expected length of the parametric and robust EBCIs with $\alpha = 0.05$. The worst-case coverage (over all θ_i-distributions with second moment μ_2) of the nominal 95% parametric EBCI is below 88% for small signal-to-noise ratios μ_2/σ^2. When θ_i is in fact Laplace-distributed, both the parametric and robust soft thresholding EBCIs deliver substantial expected length improvements relative to the unshrunk EBCI $Y_i \pm z_{1-\alpha/2}\sigma$. For small values of μ_2/σ^2, the length improvement exceeds that of the linear EBCIs shown in Figure 3.

G.2 Poisson data

Suppose now that Y_i has a Poisson distribution with rate parameter θ_i, conditional on θ_i. As a baseline prior for θ_i, we use the conjugate gamma distribution with shape parameter k and scale parameter λ. Let $\Gamma^{-1}(\alpha; k, \lambda)$ denote the α-quantile of this distribution. As candidate sets $\mathcal{S}(y; \chi)$, we use a modification of the equal-tailed posterior credible set for θ_i under the

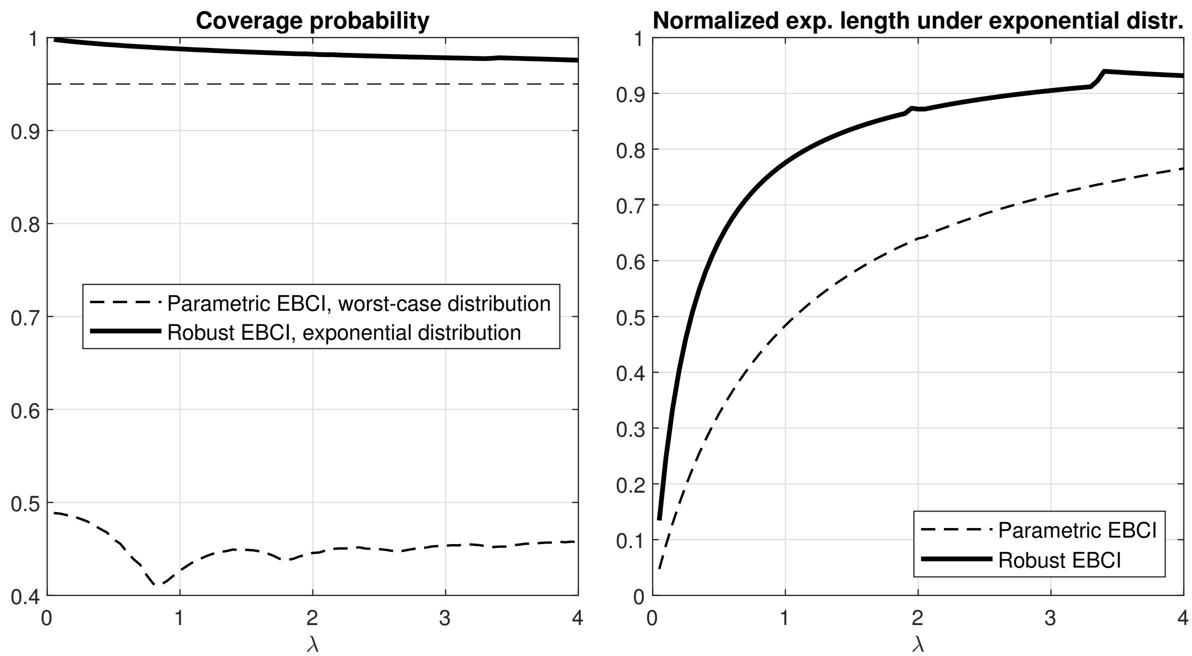

Figure S3: Poisson EBCIs, $\alpha = 0.05$. The expected length is normalized by that of the unshrunk Garwood (1936) CI. The grid for θ_i for the linear program in Eq. (19) is given by 500 points equally spaced on $[10^{-6}, \Gamma^{-1}(0.999; 1, \lambda)]$. The support for Y_i is truncated above at 30.

baseline prior,

$$\mathcal{S}(y;\chi) = \left[\Gamma^{-1}\left(\frac{\alpha}{2}; e^{-\chi}k + y, \frac{\lambda}{e^{-\chi}+\lambda}\right), \Gamma^{-1}\left(1-\frac{\alpha}{2}; 1 + e^{-\chi}(k-1) + y, \frac{\lambda}{e^{-\chi}+\lambda}\right)\right],$$

where $1-\alpha$ is the nominal confidence level. For $\chi = 0$, this corresponds to the equal-tailed posterior credible interval under the baseline prior; we call this the parametric EBCI. As $\chi \to \infty$, the interval converges to the "unshrunk" Garwood (1936) confidence interval for the Poisson parameter θ_i, which has coverage at least $1-\alpha$ conditional on θ_i. We compute the value $\hat{\chi} \in (0, \infty)$ that leads to a robust EBCI numerically as in Supplemental Appendix G.1, except that we replace integrals over the distribution of Y_i with (truncated) sums.

Figure S3 displays the coverage and expected length for $k = 1$, i.e., when the baseline θ_i-distribution is exponential with mean λ. We consider the asymptotic limit where the first two moments of θ_i are known.[2] We set $\alpha = 0.05$. The worst-case coverage (over all θ_i-distributions with the same first and second moments as the exponential distribution) of the nominal 95% parametric EBCI is disastrously low for all values of λ considered here. At the same time, the robust EBCI is over 50% shorter on average than the unshrunk Garwood

[2] These moments are easily obtained from the first and second marginal moments of the data: $E[\theta] = E[Y]$ and $E[\theta^2] = E[Y^2] - E[Y]$. They equal $E[\theta] = k\lambda$ and $E[\theta^2] = k(k+1)\lambda^2$ under the baseline distribution.

(1936) CI when $\lambda \leq 0.3$, and more than 25% shorter when $\lambda \leq 0.85$.

References

Boyd, S. P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Garwood, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika*, 28(3/4):437–442.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

Smith, J. E. (1995). Generalized Chebychev inequalities: Theory and applications in decision analysis. *Operations Research*, 43(5):807–825.

Stock, J. H. and Watson, M. W. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1):155–174.

van der Vaart, A. (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.