Supplemental Materials for
“Robust Empirical Bayes Confidence Intervals”

Timothy B. Armstrong* Michal Kolesár†

Yale University Princeton University

Mikkel Plagborg-Møller‡

Princeton University

June 15, 2020

This supplement is organized as follows. Supplemental Appendix D gives proofs for the formal results in the main text, technical details on t-statistic shrinkage and on Assumption C.5. Supplemental Appendix E gives details for simulations and empirical applications.

Appendix D Theoretical details and proofs

Supplemental Appendices D.1 and D.2 give technical details on t-statistic shrinkage and on Assumption C.5, respectively. The remainder of this section provides the proofs of all results in the main paper and in this supplement.

D.1 t-statistic shrinkage

We here provide details on the t-statistic shrinkage approach discussed in Remark 3.8. Let $W_i = Y_i/\hat{\sigma}_i$ and let $\tau_i = \theta_i/\sigma_i$. Let $\hat{X}_i'\hat{\delta}$ be a regression estimate where $\hat{\delta}$ is an estimate of a regression parameter δ (typically the limit or probability limit of $(\sum_{i=1}^n X_i X_i')^{-1} \sum_{i=1}^n X_i \tau_i$, although we do not impose this). We apply the approach in Appendix C.3 with W_i in place of Y_i and τ_i in place of θ_i, which leads to the estimate

$$\hat{\tau}_i = \hat{X}_i'\hat{\delta} + \hat{w} \cdot (W_i - \hat{X}_i'\hat{\delta})$$

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

for τ_i, where $\hat{w}$ is a shrinkage coefficient, which is an estimate of some unknown constant $w > 0$ (for example, the choice $\hat{w} = \hat{\mu}_2/(\hat{\mu}_2 + 1)$, with $\hat{\mu}_2$ an estimate of the second moment of $\theta_i/\sigma_i - X_i'\delta$, optimizes mean squared error for estimating θ_i/τ_i). The standard error of this estimate is $\hat{w}$, so that an interval for τ_i with critical value χ is given by $\{\hat{\tau}_i \pm \hat{w} \cdot \chi\}$. This leads to the interval $\{\hat{\theta}_i \pm \mathrm{se}_i \cdot \chi\}$ where $\hat{\theta}_i = \hat{\tau}_i\hat{\sigma}_i$ and $\mathrm{se}_i = \hat{w}\hat{\sigma}_i$. Then

$$Z_i = \frac{\hat{\theta}_i - \theta_i}{\mathrm{se}_i} = \frac{\hat{\sigma}_i \cdot \hat{X}_i'\hat{\delta} + \hat{\sigma}_i \cdot \hat{w} \cdot (W_i - \hat{X}_i'\hat{\delta}) - \theta_i}{\hat{w}\hat{\sigma}_i} = \frac{Y_i - \theta_i}{\hat{\sigma}_i} + \frac{\hat{w} - 1}{\hat{w}}\left(\frac{\theta_i}{\hat{\sigma}_i} - \hat{X}_i'\hat{\delta}\right).$$

Let $\tilde{b}_i = \frac{\hat{w}-1}{\hat{w}}\left(\frac{\theta_i}{\hat{\sigma}_i} - \hat{X}_i'\hat{\delta}\right)$ and let $b_{i,n} = \frac{w-1}{w}\left(\frac{\theta_i}{\sigma_i} - X_i'\delta\right)$. Let $g(b) = (b^{\ell_1}, \dots, b^{\ell_p})'$ where $\ell_1, \dots, \ell_p$ are as in Appendix C.3. Let $\hat{\mu}$ be an estimate of the (unconditional) moments of $\frac{\theta_i}{\sigma_i} - X_i'\delta$. This leads to the estimate $\hat{m}_j = [(\hat{w} - 1)/\hat{w}]^{\ell_j}\hat{\mu}_{\ell_j}$ of the ℓ_jth moment of the $b_{i,n}$'s. Let $\hat{m} = (\hat{m}_1, \dots, \hat{m}_p)'$ and let $\hat{\chi} = \mathrm{cva}_{\alpha,g}(\hat{m})$. We consider unconditional average coverage, and we verify the conditions of Theorem C.1 with $\mathcal{A}$ containing only one set, which contains all observations.

We use conditions similar to those in Appendix C.3, but we replace Assumption C.7 with the following assumption, which does not impose any independence between the conditional moments of the $b_{i,n}$ and σ_i.

Assumption D.1. *For some μ_0 such that $(\mu_{0,\ell_1}, \dots, \mu_{0,\ell_p})$ is in the interior of the set of values of $\int g(b)\,dF(b)$ where F ranges over probability measures on $\mathbb{R}$ where $g_j(b) = b^{\ell_j}$ and some constant K, we have, for each $j = 1, \dots, p$*

$$\frac{1}{n}\sum_{i=1}^{n}(\theta_i/\sigma_i - \delta_0'X_i)^{\ell_j} \to \mu_{0,\ell_j}, \quad \limsup_n \frac{1}{n}\sum_{i=1}^{n}|\theta_i|^{\ell_j} \le K, \quad \limsup_n \frac{1}{n}\sum_{i=1}^{n}\|X_i\|^{\ell_j} \le K.$$

and $\hat{\mu}_{\ell_j}$ converges in probability to μ_{0,ℓ_j} under $\tilde{P}$.

Theorem D.1. *Let $\hat{\theta}_i$, se_i and $\hat{\chi}_i$ be defined above, and suppose that Assumptions D.1, C.5 and C.6 hold. Suppose $\hat{w}$ converges in probability to $w > 0$ under $\tilde{P}$. Then $\frac{1}{n}\sum_{i=1}^{n}\tilde{P}(\theta_i \notin \{\hat{\theta}_i \pm se_i \cdot \hat{\chi}\}) \le \alpha + o(1)$. If, in addition, $(Y_i, \hat{\sigma}_i)$ is independent over i under $\tilde{P}$, then $\frac{1}{n}\sum_{i=1}^{n}\mathrm{I}\{\theta_i \notin \{\hat{\theta}_i \pm se_i \cdot \hat{\chi}\}\} \le \alpha + o_{\tilde{P}}(1)$.*

To prove Theorem D.1, we verify the conditions of Theorem C.1. The first part of Assumption C.1 is immediate from Assumption C.5. For the second part, we have

$$\tilde{b}_i - b_{i,n} = \frac{\hat{w} - 1}{\hat{w}}\left(\frac{\theta_i}{\hat{\sigma}_i} - \hat{X}_i'\hat{\delta}\right) - \frac{w - 1}{w}\left(\frac{\theta_i}{\sigma_i} - X_i'\delta\right) = f(\hat{w}, \hat{\sigma}_i, \hat{X}_i, \hat{\delta}, \theta_i) - f(w, \sigma_i, X_i, \delta, \theta_i)$$

where $f(w, \sigma_i, X_i, \delta, \theta_i) = \frac{w-1}{w}\left(\frac{\theta_i}{\sigma_i} - X_i'\delta\right)$ is uniformly continuous on any compact set on which σ_i is bounded away from zero. Let $C > 0$ be given. It follows that, for any $\varepsilon > 0$, there exists η such that $\|(\hat{\sigma}_i - \sigma_i, \hat{w} - w, \hat{X}_i' - X_i', \hat{\delta}' - \delta')'\| \leq \eta$ and $\|\theta_i\| + \|X_i\| \leq C$ implies $|\tilde{b}_i - b_{i,n}| < \varepsilon$ (where we use the fact that $\hat{\sigma}_i$ and σ_i are bounded away from zero once η is small enough by Assumption C.6). Thus,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \tilde{P}(|\tilde{b}_i - b_{i,n}| \geq \varepsilon) \\
&\qquad \leq \frac{1}{n}\sum_{i=1}^{n} \tilde{P}(\|(\hat{\sigma}_i - \sigma_i, \hat{w} - w, \hat{X}_i' - X_i', \hat{\delta}' - \delta')'\| > \eta)\, \mathrm{I}\{\|\theta_i\| + \|X_i\| \leq C\|\} \\
&\qquad\qquad + \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\{\|\theta_i\| + \|X_i\| > C\|\}.
\end{aligned}
$$

The first term converges to zero by Assumption C.6 and the assumption that $\hat{w}$ converges in probability to w. The last term can be made arbitrarily small by Assumption D.1. This completes the verification of Assumption C.1.

For Assumption C.2, letting $m_j = [(w-1)/w]^{\ell_j} \mu_{0,\ell_j}$, it is immediate from Assumption D.1 and the assumption that $\hat{w}$ converges in probability to $\hat{w}$ that $\hat{m}$ converges to m under $\tilde{P}$, which gives the second part of Assumption C.2. Furthermore, the first part of Assumption C.2 holds since

$$
\frac{1}{n}\sum_{i=1}^{n} b_{n,i}^{\ell_j} - m_j = \left(\frac{w-1}{w}\right)^{\ell_j} \frac{1}{n}\sum_{i=1}^{n} \left[(\theta_i/\sigma_i - X_i'\delta)^{\ell_j} - \mu_{0,\ell_j}\right] \to 0
$$

by Assumption D.1.

Assumption C.3 is vacuous since there is no covariate $\tilde{X}_i$ and $m = m(\tilde{X}_i)$ takes on only one value. Assumption C.4 holds by Lemma D.8 in Supplemental Appendix D.7 below. This completes the proof of Theorem D.1.

D.2 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 $\beta_{i,n}$ with rate of convergence κ_n and asymptotic variance estimate $\hat{\sigma}_i^2$. Suppose that

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

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

As an example, consider the case where the estimate is a sample mean: $\tilde{Y}_i = \bar{X}_i = \frac{1}{T_{i,n}} \sum_{i=1}^{T_{i,n}} X_{i,t}$, where $X_{i,t}$ is a sequence of random variables that is independent across both i and t and identically distributed across i with the same t, with mean $\beta_{i,n} = EX_{i,t}$. Letting s_i^2 denote the variance of $X_{i,t}$ and $\hat{s}_i^2$ the sample variance, we can then define $\kappa_n^2 = \bar{T}_n = \frac{1}{n}\sum_{i=1}^n T_{i,n}$ and $\hat{\sigma}_i^2 = \hat{s}_i^2 \bar{T}_n / T_{i,n}$ so that

$$\frac{\kappa_n(\tilde{Y}_i - \beta_{i,n})}{\hat{\sigma}_i} = \frac{\sqrt{T_{i,n}}(\tilde{Y}_i - \beta_{i,n})}{\hat{s}_i}.$$

If $\min_{1 \le i \le n} T_{i,n} \to \infty$ and the family of distributions of $X_{i,t} - \beta_{i,n}$ satisfy the uniform integrability condition (11.77) in Lehmann and Romano (2005), then (S1) holds by applying Theorem 11.4.4 in Lehmann and Romano (2005) along any sequence i_n.

D.3 Proof of Lemma 4.1

Part (i). Let $\Gamma(m)$ denote the space of probability measures on $\mathbb{R}$ with second moment bounded above by $m > 0$. By definition of the maximal non-coverage probability,

$$\tilde{\rho}(w) = \sup_{F \in \Gamma(1/w-1)} E_{b\sim F}\left[P(|b - Z| > z/\sqrt{w} \mid b)\right] = \sup_{F \in \Gamma(1/w-1)} P_{b \sim F}\big(\sqrt{w}|b - Z| > z\big), \quad \text{(S2)}$$

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 (S2)—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 $\frac{w_1}{w_0} \times \frac{1-w_1}{w_1} + \frac{w_1 - w_0}{w_0} = \frac{1-w_0}{w_0}$, 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

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

Part (ii). It follows from Proposition B.1 that, if we define $r(b,\chi) = \Phi(-\chi-b)+\Phi(-\chi+b)$, then

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

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

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

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 $\lambda = (\tilde{z}(1-w)^{-1/2}+\varepsilon(w))^{-2} \in [0,1]$,

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

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

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

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 above equals

$$\max\left\{\sup_{0\le\lambda\le(z-\varepsilon(w))^{-2}}\lambda\Phi\left(\lambda^{-1/2}(1/w-1)^{1/2}-zw^{-1/2}\right),\right.$$
$$\left.\sup_{(z-\varepsilon(w))^{-2}<\lambda\le 1}\lambda\Phi\left(\lambda^{-1/2}(1/w-1)^{1/2}-zw^{-1/2}\right)\right\}. \quad \text{(S3)}$$

The first argument to the maximum above is bounded above by

$$\sup_{0\le\lambda\le(z-\varepsilon(w))^{-2}}\lambda=(z-\varepsilon(w))^{-2}\to\frac{1}{z^2}.$$

The second argument to the maximum in (S3) tends to 0 as $w\to 0$, since

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

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

D.4 Proof of Proposition B.1

Let $r_0(t,\chi)=r(\sqrt{t},\chi)$. Since $r(b,\chi)$ is symmetric in b, Eq. (21) 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 $\rho(m_2,\chi)\le\bar\rho(m_2,\chi)$, 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 $\rho(m_2,\chi)\le\bar r(m_2,\chi)$.

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

$$\bar r(t,\chi)=\sup\{\lambda r_0(x_1,\chi)+(1-\lambda)r_0(x_2,\chi)\colon \lambda x_1+(1-\lambda)x_2=t,\ 0\le x_1\le x_2,\lambda\in[0,1]\}, \quad \text{(S4)}$$

which corresponds to the problem in Eq. (21), 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 $\rho(m_2,\chi)\ge\bar r(m_2,\chi)$. Thus, $\rho(m_2,\chi)=\bar r(m_2,\chi)$. The proposition then follows by Lemma D.2 below.

Lemma D.1. *Let $r_0(t,\chi)=r(\sqrt{t},\chi)$. If $\chi\le\sqrt{3}$, 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

$$
\chi u_0 + u_0^2 + 1 = e^{2\chi u_0}(u_0^2 - \chi u_0 + 1) \tag{S5}
$$

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. (S5) and simplifying the expression yields

$$
\begin{aligned}
f'(u_0) &= \frac{(u_0^2 - \chi u_0 + 1)(2u_0 + \chi) - (\chi u_0 + u_0^2 + 1)(2\chi(u_0^2 - \chi u_0 + 1) + 2u_0 - \chi)}{u_0^2 - \chi u_0 + 1} \\
&= \frac{2u_0^2 \chi(\chi^2 - 3 - u_0^2)}{u_0^2 - \chi u_0 + 1}.
\end{aligned} \tag{S6}
$$

Suppose $\chi^2 < 3$. Then $f'(u_0) < 0$ at all positive roots u_0 by Eq. (S6). 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. (S6), this implies $u_1 \geq \sqrt{\chi^2 - 3}$. As a result, again by Eq. (S6), $f(u_i) \leq 0$ 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.2. *The problem in Eq. (S4) can be written as*

$$\bar{r}(t, \chi) = \sup_{u \geq t} \{(1 - t/u) r_0(0, \chi) + \frac{t}{u} r_0(u, \chi)\}. \tag{S7}$$

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. (S7) satisfies $u = t$ for $t > t_0$ and $u = t_0$ otherwise.

Proof. If in the optimization problem in Eq. (S4), 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. (S7) 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

$$r_0(x_2) - (x_2 - x_1) r_0'(x_1) - r_0(x_1) = \mu \frac{(x_2 - x_1)^2}{(x_2 - t)}, \tag{S8}$$

$$r_0(x_2) + (x_1 - x_2) r_0'(x_2) - r_0(x_1) = 0, \tag{S9}$$

where $\mu \geq 0$ is the Lagrange multiplier on the constraint that $x_1 \geq 0$. Subtracting Eq. (S9) from Eq. (S8) and applying the fundamental theorem of calculus then yields

$$\mu \frac{x_2 - x_1}{(x_2 - t)} = r_0'(x_2) - r_0'(x_1) = \int_{x_1}^{x_2} r_0''(t)\, dt > 0, \tag{S10}$$

which implies that $\mu > 0$. Here the last inequality follows because by Taylor's theorem, Eq. (S9) 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.1, it follows that $x_1 \leq t_1 \leq x_2$, and hence that

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

Finally Eq. (S10) implies that $\mu > 0$, so that $x_1 = 0$ at the optimum. Consequently, the problem in Eq. (S4) takes the form in Eq. (S7) as claimed.

To show the second part of Lemma D.2, note that by Lemma D.1, if $\chi \leq \sqrt{3}$, r_0 is concave, so that we can put $u = t$ in Eq. (S7). Otherwise, let $\mu \geq 0$ denote the Lagrange multiplier associated with the constraint $u \geq t$ in the optimization problem in Eq. (S7). The first-order condition is then given by

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

Let $f(u) = r_0(0) - r_0(u) + u r_0'(u)$. Since $f'(u) = u r_0''(u)$, it follows from Lemma D.1 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.5 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

$$
\rho(m_2, \kappa, \chi) = \sup_F E_F[r_0(t,\chi)] \quad \text{s.t.} \quad E_F[t] = m_2,\ E_F[t^2] = \kappa m_2^2, \tag{S11}
$$

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.2. 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 $E_F[t^2] = \kappa m_2^2$ doesn't 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. (S11) 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

$$\begin{aligned} \min_{x_0 > 0, \lambda_2} \; & r_0(x_0) + r_0'(x_0)(m_2 - x_0) + \lambda_2((x_0 - m_2)^2 + (\kappa - 1)m_2^2) \\ \text{s.t.} \quad & r_0(x_0) + r_0'(x_0)(x - x_0) + \lambda_2(x - x_0)^2 \geq r_0(x). \end{aligned} \tag{S12}$$

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.1, $r_0'(t)$ 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.1, it follows that the maximum of $h(x)$ occurs at $x = x_0$, and equals 0. However, Eq. (S12) 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. (S12) 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. (S12) binds at any x that maximizes δ. However, by the claim above, the constraint cannot bind for $x > t_0$.

D.6 Proof of Theorem C.1

To prove this theorem, we begin with some lemmas.

Lemma D.3. *Under Assumption C.1, we have, for any deterministic $\chi_1, \dots, \chi_n$, 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.4. *$\rho_g(\chi; m)$ 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.5 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.5. *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. $\square$

Lemma D.6. *Let M be any 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 that $\lim_{b\to\infty} g_j(b) = \lim_{b\to-\infty} g_j(b) = \infty$ and $\inf_b g_j(b) \geq 0$ for some j. Then $\lim_{\chi\to\infty} \sup_{m\in M} \rho_g(\chi; m) = 0$ 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.4, $\rho_g(\chi; m)$ 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. $\square$

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.7. *Let M be any 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 and 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.6, 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

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

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.3, 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.7 and the fact that $\underline{\rho}_g(\underline{\chi}_j; \mu_j, 2\varepsilon) \leq \alpha$ by construction.

D.7 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

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

For $\|\theta_i\| + \|X_i\| \leq C$, 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 $\|(\hat{\sigma}_i - \sigma_i, \hat{\gamma} - \gamma, \hat{\delta}' - \delta', \hat{X}_i' - X_i')'\| \leq \eta$ implies that the absolute value of the above display is less than ε. Thus, for any $\mathcal{X} \in \mathcal{A}$,

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

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,

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

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,

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

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.8

below. This completes the proof of Theorem C.2.

Lemma D.8. *Suppose that, as F ranges over all probability measures with respect to the Borel sigma algebra, $(\mu_{\ell_1}, \dots, \mu_{\ell_p})'$ is interior to the set of values of $\int (b^{\ell_1}, \dots, b^{\ell_p})' \, dF(b)$. Let $c \in \mathbb{R}$. Then, as F ranges over all probability measures with respect to the Borel sigma algebra, $(c^{\ell_1}\mu_{\ell_1}, \dots, c^{\ell_p}\mu_{\ell_p})'$ is also in the interior of the set of values of $\int (b^{\ell_1}, \dots, b^{\ell_p})' \, dF(b)$.*

Proof. We need to show that, for any vector r with $\|r\|$ small enough, there exists a probability measure F such that $\int (b^{\ell_1}, \dots, b^{\ell_p})' \, dF(b) = (c^{\ell_1}\mu_{\ell_1} + r_1, \dots, c^{\ell_p}\mu_{\ell_p} + r_p)'$. 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. ☐

Appendix E Details for simulations and applications

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 6.1. Supplemental Appendices E.3 and E.4 gives additional details on the neighborhood effects and factor model applications in Sections 6.1 and 6.2, respectively.

E.1 Details for homoskedastic simulation designs

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

1. Normal (kurtosis $\kappa = 3$): $\theta_i \sim N(0, \mu_2)$.
2. Scaled chi-squared ($\kappa = 15$): $\theta_i \sim \sqrt{\mu_2/2} \times \chi^2(1)$.
3. 2-point ($\kappa = 1/(0.9 \times 0.1) - 3 \approx 8.11$):

$$\theta_i \sim \begin{cases} 0 & \text{w.p. } 0.9, \\ \sqrt{\mu_2/(0.9 \times 0.1)} & \text{w.p. } 0.1. \end{cases}$$

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) 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 $\chi = \text{cva}_\alpha(m_2)$. 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 6.1. 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 $E_n[(\hat{\theta}_i - \bar{\theta})^2] \times E_n[1/\hat{\sigma}_i^2] = 0.6698$, 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 3.2. 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 $(\hat{\theta}_j)$, $j = 1, \dots, n$.

2. Sample σ_i, $i = 1, \dots, n$, with replacement from the empirical distribution $(\hat{\sigma}_j)$, $j = 1, \dots, n$.

3. Compute $\theta_i = \bar{\theta} + \sqrt{m/c} \times (\tilde{\theta}_i - \bar{\theta})$, $i = 1, \dots, n$. 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 6.1.

Table S1: Monte Carlo simulation results: heteroskedastic design.

n	Robust, μ_2 only: Oracle	Robust, μ_2 only: Baseline	Robust, μ_2 & κ: Oracle	Robust, μ_2 & κ: Baseline	Parametric: Oracle	Parametric: 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 $i = 1, \ldots, n$ and across 5,000 Monte Carlo repetitions. Average CI length is measured relative to the oracle robust EBCI that exploits μ_2 and κ.

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 under-covers by 9.3 percentage points.

E.3 Neighborhood effects

Figure S1 gives a plot analogous to that in Figure 6, but for children with parents at the 75th percentile of the income distribution.

Figure S2 gives a plot of the density of the t-statistics $Y_i/\hat{\sigma}_i$, with a normal density overlaid. It is clear from the figure that the left tail has more mass than the normal density, which explains the high kurtosis estimates in Table 2.

E.4 Structural change in the Eurozone

E.4.1 Data

To enhance cross-country comparability, we use only data sets maintained by Eurostat, the BIS, and the OECD, rather than supplementing with data from national statistical agencies. We discard any time series that are available for fewer than 15 years. Tables S2 and S3 list the variables and countries in the data set, respectively. If not already seasonally adjusted, we apply the X13-ARIMA-SEATS seasonal adjustment procedure to all variables other than asset prices. In the pre-2008 sample, 4.2% of the observations are missing (24% of series have

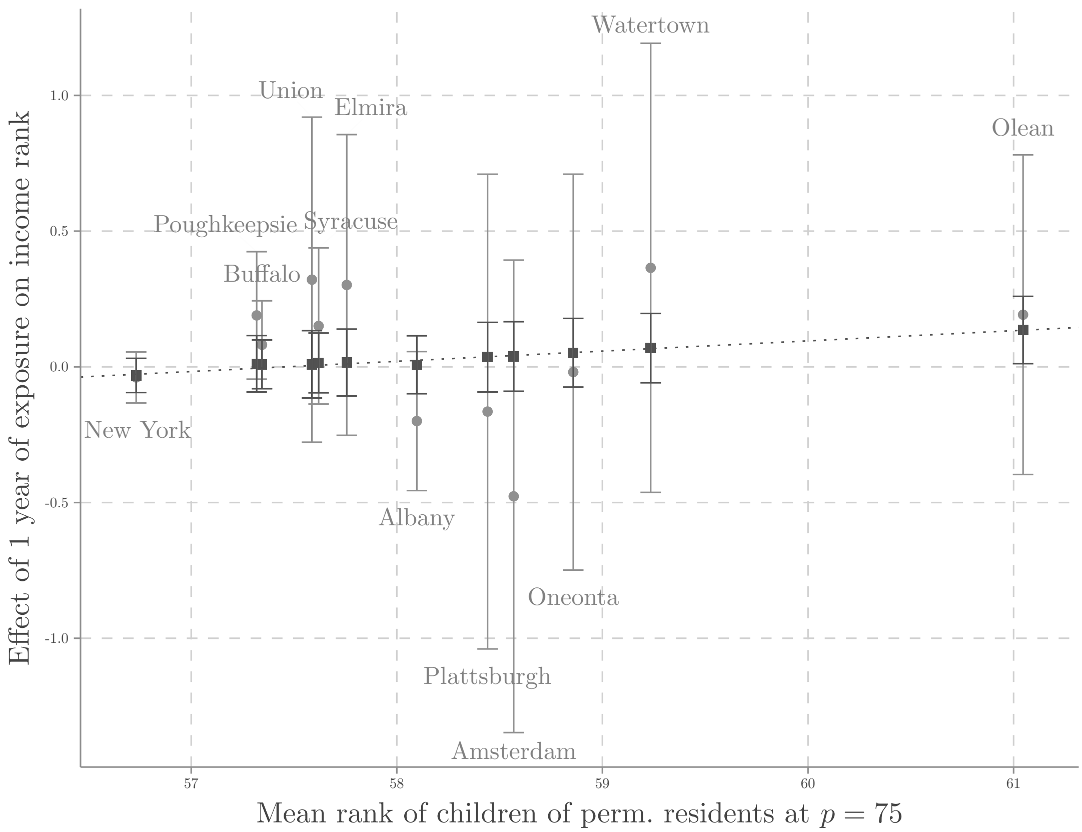

Figure S1: Neighborhood effects for New York and 90% robust EBCIs for children with parents at the $p = 75$ percentile of national income distribution, plotted against mean outcomes of permanent residents. Gray lines correspond to CIs based on unshrunk estimates represented by circles, and black lines correspond to robust EBCIs based on EB estimates represented by squares that shrink towards a dotted regression line based on permanent residents' outcomes. Baseline implementation as in Section 3.2.

at least one missing observation). We impute the missing observations in the way suggested by Stock and Watson (2016, section 2.3.4.1): First, we estimate a DFM on the non-missing data, then we regress the series with missing observations on the newly obtained principal components and impute the missing observations using this regression, then we go back and re-estimate the DFM on the now-balanced panel, then we re-impute the originally missing observations through regressions on the new factors, and so on until numerical convergence. For simplicity, we ignore any error induced by the imputation.

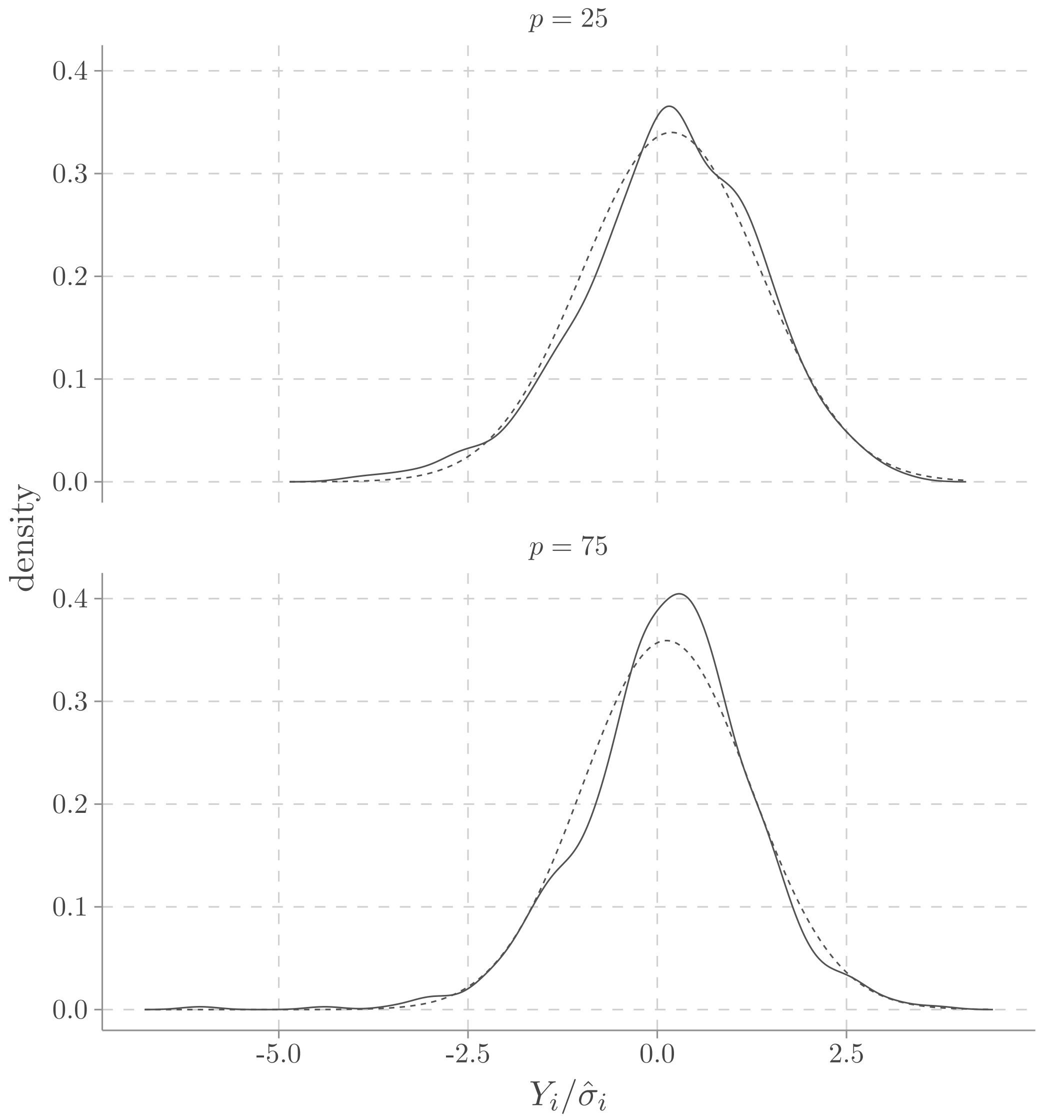

Figure S2: Estimated density of t-statistics $Y_i/\hat{\sigma}_i$ for neighborhood effects for children with parents at the $p = 25$ and $p = 75$ percentile of national income distribution. Dashed line overlays a normal density. Bandwidth selected using the Silverman rule of thumb.

Table S2: Variables in the data set used in the Eurozone DFM application.

Code	Series	Source	Stat. transf.	#countries
CA	current account divided by GDP	Eurostat	Δ	16
CAPUTIL	capacity utilization	Eurostat	none	18
CONS	real consumption	Eurostat	Δlog	18
CONSCONF	consumer confidence index	Eurostat	none	19
CPI	consumer price index	Eurostat	Δ^2log	19
CREDITHH	credit to households divided by GDP	BIS	Δlog	12
CREDITNFB	credit to non-fin. bus. divided by GDP	BIS	Δlog	12
EUR	euro exchange rate	Eurostat	Δlog	5 currencies
GDP	real GDP	Eurostat	Δlog	18
GOVBOND	10-yr gov't rate vs. 3-month EZ rate	Eurostat	Δ	18
HOUSEP	nominal house price index	BIS	Δlog	11
INT3M	3-month EZ interest rate	Eurostat	Δ	1 (EZ)
INTDD	overnight EZ interest rate	Eurostat	Δ	1 (EZ)
OILBRENT	Brent crude oil price	U.S. EIA	Δlog	1 (EZ)
STOCKP	stock price index	OECD	Δlog	15
UNEMPRATE	unemployment rate	Eurostat	none	19
WAGE	nominal wage index	Eurostat	Δlog	18

Notes: Code: series codes. *Stat. transf.*: stationarity transformation method. The 5 currencies for the euro exchange rates (EUR) are against the Swiss Franc (CHF), Chinese Yuan (CNY), British Pound (GBP), Japanese Yen (JPY), and United States Dollar (USD).

Table S3: Countries in the dataset used in the Eurozone DFM application.

Code	Country	#series	Code	Country	#series
AT	Austria	13	IT	Italy	13
BE	Belgium	13	LT	Lithuania	10
CY	Cyprus	8	LU	Luxembourg	12
DE	Germany	12	LV	Latvia	9
EE	Estonia	9	MT	Malta	8
ES	Spain	13	NL	Netherlands	13
FI	Finland	13	PT	Portugal	12
FR	France	13	SI	Slovenia	10
GR	Greece	12	SK	Slovakia	7
IE	Ireland	13	EZ	Eurozone	8

Notes: The "EZ" row lists the number of Eurozone-wide series. The "#series" columns list the number of country-specific series.

E.4.2 Model, estimation and inference

We assume that the n observed times series $z_{i,t}$ are driven by a small number r of common factors $f_t = (f_{1,t}, \dots, f_{r,t})'$, where $i = 1, \dots, n$ and $t = 1, \dots, T$. The data $z_{i,t}$ are given by the $n = 221$ country-specific or Eurozone-wide series described above. Specifically, we consider a Dynamic Factor model with a potential break in the factor loadings at a known date $t = T_0 + 1$ (= 2009q1):

$$z_{i,t} = \begin{cases} \lambda_i^{(0)\prime} f_t + \epsilon_{i,t} & \text{if } t = 1, \dots, T_0, \\ \lambda_i^{(1)\prime} f_t + \epsilon_{i,t} & \text{if } t = T_0 + 1, \dots, T. \end{cases}$$

Here $\lambda_i^{(0)}, \lambda_i^{(1)} \in \mathbb{R}^r$ are the pre- and post-break factor loadings, respectively. In our data, $T_0 = T - T_0 = 40$ quarters. We assume that the idiosyncratic errors $\epsilon_{i,t}$ are weakly correlated across t and across i, as well as weakly correlated with the common factors f_t (see Bai and Ng, 2008, for standard assumptions). We do not require $\epsilon_{i,t}$ to be stationary across the break date, and its variance, for example, is allowed to change after time T_0.

We normalize the first factor as being the latent factor driving Eurozone-wide GDP growth (the "named factor" normalization, cf. Stock and Watson, 2016). Let the scalar time series s_t denote real aggregate GDP growth in the 19 current Eurozone countries. Then

$$s_t = f_{1,t} + u_t, \quad t = 1, \dots, T, \tag{S13}$$

where u_t is weakly stationary within the two subsamples and uncorrelated with all factors f_t and idiosyncratic errors $\epsilon_{i,t}$. That is, we identify Eurozone-wide GDP growth s_t as being driven solely (and one-for-one) by the first latent factor $f_{1,t}$, which we thus interpret as an Eurozone-wide real activity factor. Because we are only interested in the loadings on this factor, we do not need further normalizations, except that we impose the conventional assumption that the r factors are mutually uncorrelated.

Our parameters of interest are the structural breaks in the loadings of each series on the Eurozone-wide real activity factor $f_{1,t}$,

$$\theta_i = \lambda_{i,1}^{(1)} - \lambda_{i,1}^{(0)}, \quad i = 1, \dots, n.$$

Following conventional practice (Stock and Watson, 2016), before analysis, all series $\{s_t\}$ and $\{z_{i,t}\}_t$, $i = 1, \dots, n$, have been standardized to each have sample mean 0 and sample variance 1. Hence, the magnitudes of θ_i can be meaningfully compared across different series i. We estimate θ_i as follows:

1. Estimate the DFM separately on the two subsamples (before and after T_0) by applying principal components to the data $\{z_{i,t}\}_{i,t}$, cf. Stock and Watson (2016). We choose the number r of factors using the "IC_{p2}" information criterion of Bai and Ng (2002). This criterion selects 5 and 4 factors on the early and late subsample, respectively, although the scree plot is flat around the optimum. Thus, we conservatively set $r = 6$ on both subsamples. Let $\hat{f}_t^{(j)}$ denote the principal component factor estimates from subsamples $j = 0, 1$.

2. For each series $i = 1, \dots, n$ and each subsample $j = 0, 1$, estimate $\lambda_i^{(j)}$ by running a two-stage least squares (2SLS) regression of $z_{i,t}$ on s_t, with the r instruments given by $\hat{f}_t$. Call the coefficient estimate $\hat{\lambda}_i^{(j)}$.

3. Compute the preliminary estimator $Y_i = \hat{\lambda}_{i,1}^{(1)} - \hat{\lambda}_{i,1}^{(0)}$, $i = 1, \dots, n$.

This estimator is consistent as $n, T_0, (T - T_0) \to \infty$ under conditions similar to those stated in Bai and Ng (2008), since (i) the principal components $\hat{f}_t$ consistently estimate the linear space spanned by the true factors f_t, and (ii) the fitted value from the first stage of the 2SLS regression is a consistent estimate of $f_{1,t}$ by the normalization (S13). We compute standard errors for $\hat{\lambda}_{i,1}^{(j)}$ using the usual 2SLS formula, with a Newey-West correction for serial correlation of u_t (bandwidth = 8 lags). The standard errors $\hat{\sigma}_i$ for Y_i are obtained by assuming independence of the two subsamples (weak dependence would suffice in practice).

References

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2008). Large Dimensional Factor Analysis. *Foundations and Trends in Econometrics*, 3(2):89–163.

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

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer, New York, NY, 3rd edition.

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. (2016). Factor models and structural vector autoregressions in macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 415–525. Elsevier.

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