Robust Empirical Bayes Confidence Intervals

Timothy B. Armstrong
University of Southern California

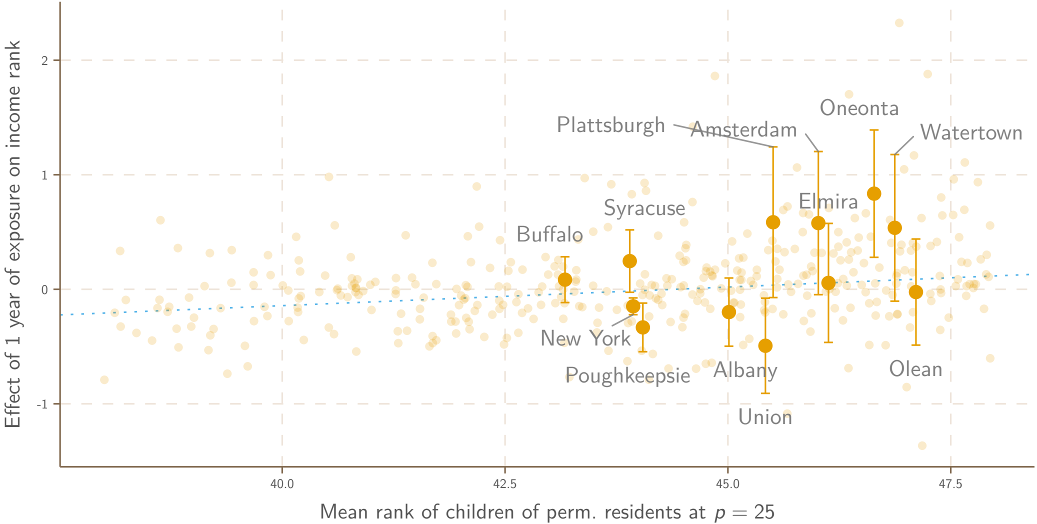
Michal Kolesár
Princeton University

Mikkel Plagborg-Møller
Princeton University

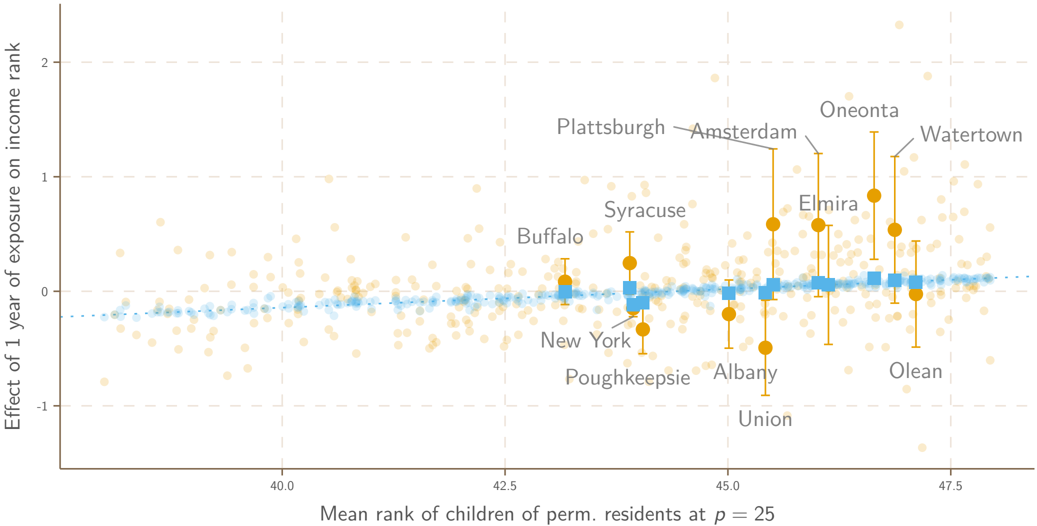
November 2, 2021

Linear shrinkage estimators

- Often interested in estimating effects θ_i for many individuals/units i.
  - Value-added of teacher/school/hospital/neighborhood/politician/patient.
    Jacob & Lefgren (2008); Kane & Staiger (2008); Chetty, Friedman & Rockoff (2014); Angrist, Hull, Pathak & Walters (2017); Finkelstein, Gentzkow, Hull & Williams (2017); Chetty & Hendren (2018); Hull (2020); Easterly & Pennings (2021)
  - Subgroup analysis: split results by countries, sectors, occupations, etc.
- Common to (linearly) shrink noisy unbiased estimates toward baseline values.
  - Empirical Bayes (EB) motivation: Bayesian/random-effects model with $\theta_i \sim N$.
  - MSE gain over unshrunk estimate robust to failure of Bayesian model. James & Stein (1961)



Neighborhood effects: unshrunk estimates and 90% CIs
(replicates Chetty & Hendren, 2018, Fig. 1, for NY CZ)



Neighborhood effects: shrinkage estimates
(replicates Chetty & Hendren, 2018, Fig. 2, for NY CZ)

This paper: How to construct CIs for linear shrinkage estimates?

- Parametric empirical Bayes confidence interval (EBCI) (Morris 1983a,b): Bayesian credible set, treating estimated normal distribution of θ_i's as prior.
- Existing theoretical justification requires correct distribution for θ_i's.
- In contrast, for point estimation, get MSE improvement even if θ_i's non-random.

Questions

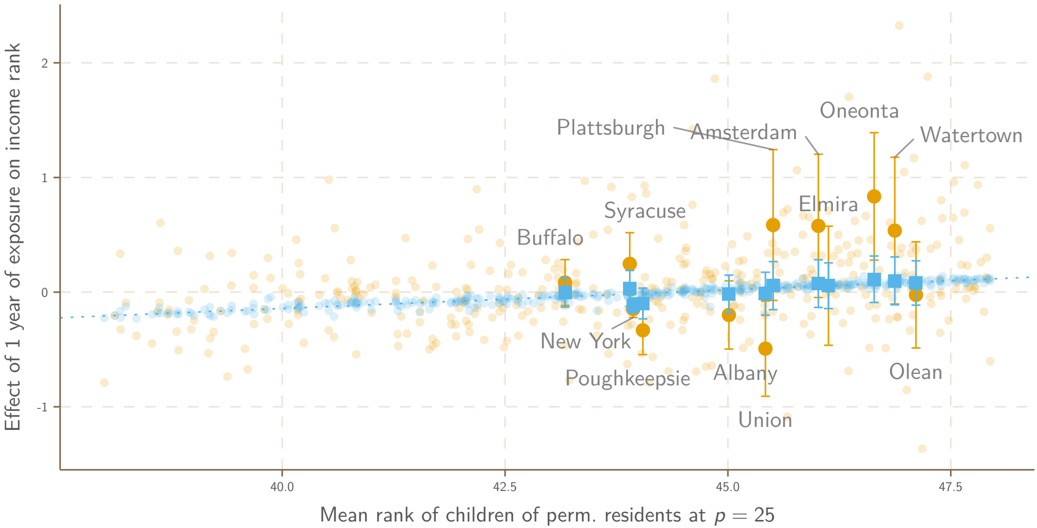
1. Is parametric EBCI robust to failure of assumption on distribution of θ_i's?
2. If not, can we "robustify" it (and keep it short)?
3. Does robust EBCI have frequentist coverage properties (nonrandom θ_i's)?

Question 1: Is parametric EBCI robust to failure of $\theta_i \sim N$ assumption?

- In general, no: Coverage of 95% parametric EBCI can be as low as 74% under repeated sampling of $(\theta_i, \text{data}_i)$.

Question 2: Can we robustify parametric EBCI while keeping it short?

- Yes, we provide critical values. Only input is moment estimates already used to compute shrinkage estimator.
  - Idea: Use estimated moments of bias to bound non-coverage.
- Guaranteed coverage under repeated sampling of $(\theta_i, \text{data}_i)$.
- If in fact $\theta_i \sim N$, then robust EBCI is not much wider than parametric EBCI.



Neighborhood effects: shrunk estimates and 90% CIs

Question 3: Does robust EBCI have frequentist coverage properties (nonrandom θ_i's)?

- Yes, controls average coverage as $n \to \infty$:

$$\frac{1}{n}\sum_{i=1}^{n} P(\theta_i \in CI_i \mid \theta) \geq 1 - \alpha.$$

- Usual CI centered at unshrunk estimate also has this property, but is wider.
- Improvement in CI length possible because average coverage is weaker requirement than usual frequentist coverage for each i separately.
    - Intuition: Easier to estimate the average effect of shrinkage bias on coverage (using moments) than to estimate bias for each i separately.

Related literature

- Our paper: robust uncertainty quantification for linear shrinkage estimator; near-efficient when $\theta_i \sim N$. (Also give extensions to non-linear shrinkage.)
- Do not attempt to recover full distribution of θ_i to improve estimator/CI.
  - "Flexible parametric" and nonparametric EB literature focuses on point estimation. Robbins (1951); Jiang & Zhang (2009); Koenker & Mizera (2014); Efron (2016)
- EB in econometrics: Hansen (2016); Abadie & Kasy (2019); Cheng, Liao & Shi (2019); Fessler & Kasy (2019); Bonhomme & Weidner (2021); Ignatiadis & Wager (2021); Liu, Moon & Schorfheide (2021)
- Average coverage in nonparam. regression: Wahba (1983); Nychka (1988); Wasserman (2006); Cai, Low & Ma (2014)
- Shrinkage confidence balls: Casella & Hwang (2012)

Outline

1. Overview of results
2. Practical implementation
3. Simulation study
4. Application
5. Extension: non-linear shrinkage
6. Summary

Empirical Bayes set-up and notation

- Observe initial estimates $Y_1, \dots, Y_n$ of unknown scalar parameters $\theta_1, \dots, \theta_n$.
  - Treat θ as random throughout. $P(\cdot)$: probability under joint distribution of $\{(\theta_i, Y_i)\}_{i=1}^n$.
  - Statements involving $P(\cdot \mid \theta)$ don't actually require $\theta = (\theta_1, \dots, \theta_n)$ to be random, but we maintain conditioning for notational clarity.
- Linear shrinkage estimator: $\hat{\theta}_i = (1 - w)a + wY_i$.
  - w: tuning parameter, chosen based on data.
  - a: baseline value or pooled estimate.
- We will later allow: (i) heteroskedastic Y_i's; (ii) a and w that depend on covariates and $\hat{\sigma}_i$; (iii) asymptotic (rather than exact) normality. For now, consider simple setting...

Simple empirical Bayes model

- Homoskedastic normal location model with known σ^2:

$$(Y_i \mid \theta) \sim N(\theta_i, \sigma^2), \quad i = 1, \dots, n.$$

- Shrinkage estimator $\hat{\theta}_i = wY_i$ (shrink toward 0).
- How to choose shrinkage constant w?
  - Working model: $\theta_i \sim N(0, \mu_2)$.
  - MSE-optimal estimate: posterior mean $\hat{\theta}_i = w_{EB} Y_i$, where $w_{EB} = \mu_2/(\sigma^2 + \mu_2)$.
  - Feasible version: replace μ_2 with consistent estimator $\hat{\mu}_2$, e.g., $\hat{\mu}_2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i^2 - \sigma^2)$.

MSE of empirical Bayes point estimator

- MSE gain of EB estimator robust to failure of working assumption $\theta_i \sim N(0, \mu_2)$.
  - EB estimate $\hat{\theta}_i = \frac{\hat{\mu}_2}{\hat{\mu}_2+\sigma^2} Y_i$ has lower "frequentist" compound MSE

$$\sum_{i=1}^{n} E[(\hat{\theta}_i - \theta_i)^2 \mid \theta]$$

    than unshrunk estimate Y_i whenever $n \geq 3$. James & Stein (1961)
  - Thus, MSE improvement holds even if θ is nonrandom...
  - ... or if θ is random, but working assumption is wrong.
- Intuition: MSE improvement only depends on $E[\theta_i^2] = \mu_2$, not distribution of θ_i's.

Empirical Bayes confidence intervals

- Following Morris (1983a) and Carlin & Louis (2000, Ch. 3.5), we say that CI_i is $1-\alpha$ empirical Bayes confidence interval (EBCI) if

$$P(\theta_i \in CI_i) \geq 1-\alpha,$$

  where $P(\cdot)$ denotes joint distribution of (θ_i, CI_i).

Ex post

- Parametric EBCI: Assume working model $\theta_i \sim N(0, \mu_2)$ and use Bayesian credible interval $\hat{\theta}_i \pm z_{1-\alpha/2}\sqrt{w_{EB}}\sigma$.
  - Feasible version: plug in $\hat{\mu}_2$ for μ_2. Morris (1983b)
- We consider EB coverage as $n \to \infty$, so assume $\mu_2 = E[\theta_i^2]$ known for now.
- Parametric EBCI valid if working model correct. Can we robustify it?

Robust EBCI construction

- Consider CI centered at $\hat{\theta}_i = w_{EB} Y_i$, by inverting t-statistic

$$\left(\frac{w_{EB} Y_i - \theta_i}{w_{EB}\sigma} \,\middle|\, \theta \right) \sim N(b_i, 1), \quad \text{where} \quad b_i = \frac{w_{EB} - 1}{w_{EB}\sigma}\theta_i \quad \text{is conditional (scaled) bias.}$$

- With critical value χ, non-coverage given θ is

$$P\left(\theta_i \notin \{w_{EB} Y_i \pm w_{EB}\sigma\chi\} \mid \theta\right) = P_{Z \sim N(0,1)}\left(|Z + b_i| > \chi\right) \equiv r(b_i, \chi).$$

- Averaging over θ:

$$P\left(\theta_i \notin \{w_{EB} Y_i \pm w_{EB}\sigma\chi\}\right) = E[r(b_i, \chi)].$$

- How to choose χ so that this is $\leq \alpha$?

Robust EBCI construction: critical value

- Want to choose χ to bound non-coverage probability

$$E[r(b_i, \chi)], \quad \text{where} \quad b_i = \frac{w_{EB} - 1}{w_{EB}\sigma}\theta_i.$$

- Since $E[\theta_i^2] = \mu_2$, we have

$$E[b_i^2] = \frac{(w_{EB} - 1)^2}{w_{EB}^2\sigma^2}\mu_2 = \frac{\sigma^2}{\mu_2}.$$

- Therefore non-coverage is bounded above by

$$\rho(\sigma^2/\mu_2, \chi) \equiv \sup_F E_{b\sim F}[r(b, \chi)] \quad \text{s.t.} \quad E_{b\sim F}[b^2] = \sigma^2/\mu_2.$$

- Robust EB critical value: Choose χ so that $\rho(\sigma^2/\mu_2, \chi) = \alpha$.

Robust EBCI

- Leads to robust EBCI:
$$\hat{\theta}_i \pm \text{cva}_\alpha(\sigma^2/\mu_2) w_{EB}\sigma,$$
where $\text{cva}_\alpha(t) = \rho^{-1}(t, \alpha)$ (inverse is in second argument), and
$$\rho(t, \chi) \equiv \sup_F E_{b\sim F}[r(b, \chi)] \quad \text{s.t.} \quad E_{b\sim F}[b^2] = t.$$

- Easy to compute $\rho(t, \chi)$: linear program in F.

- F that achieves the maximum ("least favorable distribution") concentrates on three points. Get closed-form formula for $\rho(t, \chi)$.

- Can tighten EBCI using higher moments of bias b_i (more later).

Average coverage

- Robust EBCI has frequentist (conditional on θ) average coverage property:

$$\frac{1}{n}\sum_{i=1}^{n} P(\theta_i \notin \{\hat{\theta}_i \pm w_{EB}\sigma\chi\} \mid \theta) = \frac{1}{n}\sum_{i=1}^{n} r(b_i, \chi) \leq \alpha + o(1)$$

  if we use the critical value $\chi = \text{cva}_\alpha(\sigma^2/\mu_2)$.

  - Holds because

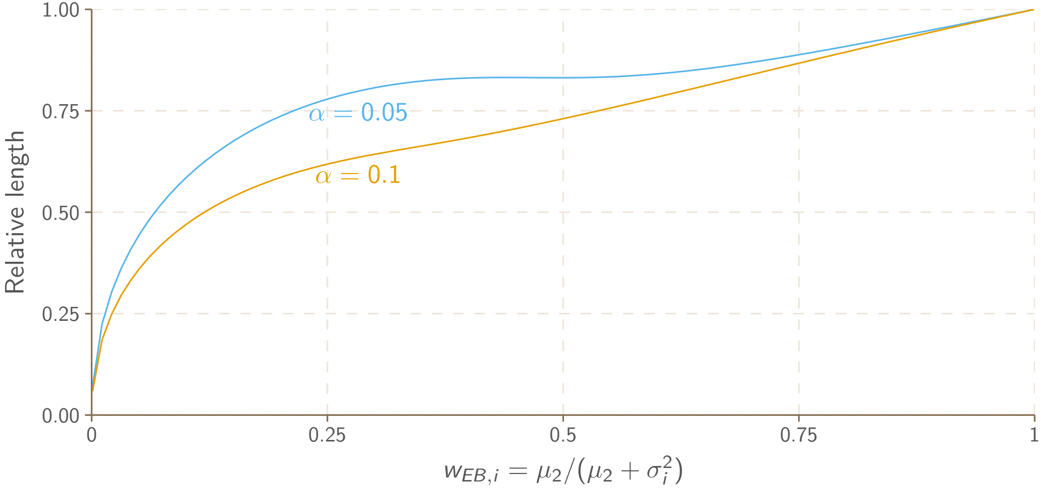
$$\tfrac{1}{n}\textstyle\sum_{i=1}^{n} b_i^2 = E_{b\sim F_n}[b^2] = \sigma^2/\mu_2 + o_P(1),$$

    where F_n is the empirical distribution of the b_i's. Holds in finite samples if $\frac{1}{n}\sum_{i=1}^{n}\theta_i^2 = \mu_2$.

- In fact, can show $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\theta_i \notin \{\hat{\theta}_i \pm w_{EB}\sigma\chi\}) \leq \alpha + o_{P(\cdot|\theta)}(1)$.

- Unshrunk CI $Y_i \pm \sigma z_{1-\alpha/2}$ also satisfies avg. coverage property, but is wider (next slide).

Efficiency of robust EBCI relative to unshrunk CI



Average coverage versus usual coverage notion

- Usual frequentist coverage stronger, cannot use shrinkage to tighten CI.
  Pratt (1961); Armstrong & Kolesár (2018)

$$\underbrace{\text{simultaneous coverage}}_{P(\forall i:\ \theta_i \in CI_i \mid \theta) \geq 1-\alpha} \implies \underbrace{\text{usual coverage}}_{\forall i:\ P(\theta_i \in CI_i \mid \theta) \geq 1-\alpha} \implies \underbrace{\text{average coverage}}_{\frac{1}{n}\sum_{i=1}^{n} P(\theta_i \in CI_i \mid \theta) \geq 1-\alpha}$$

- Avg. coverage allows us to borrow strength from other i: Can't get accurate data-driven bound on each b_i, but can bound "average effect" of b_i on coverage, using moments.

- Is average coverage a sensible criterion?

  1. We already agreed on compound loss for estimation (want small MSE on average). Worries about undercoverage for particular i analogous to worries about bad MSE for particular i.

  2. Easy interpretation, even to a layperson: $100 \times (1-\alpha)\%$ of the n EBCIs contain true θ_i.

Undercoverage of parametric EBCI

- Parametric EBCI (Bayesian credible interval with $\theta_i \sim N(0, \mu_2)$ prior)

$$\hat{\theta}_i \pm \sqrt{w_{EB}}\sigma z_{1-\alpha/2}$$

  has no robust coverage guarantee. How bad can EB coverage get?

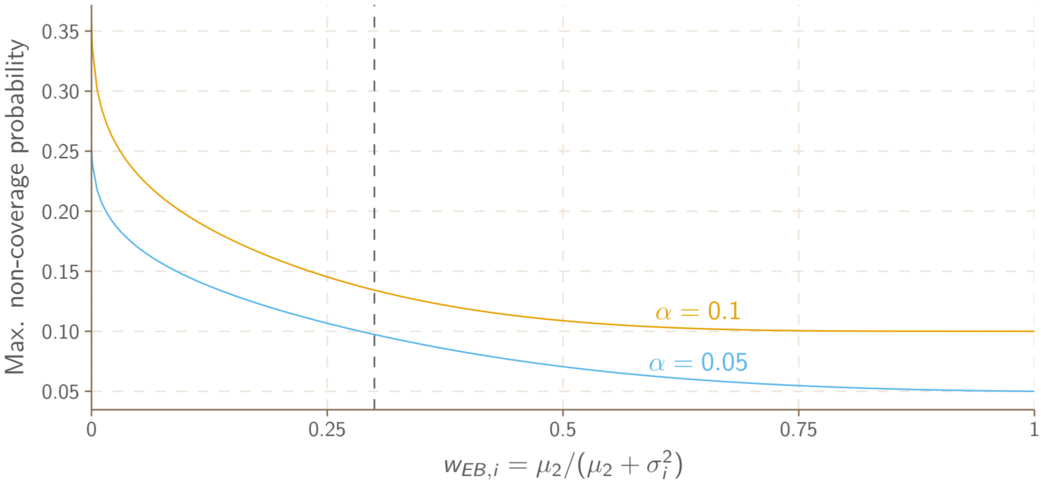
- Corresponds to EBCI with critical value $\chi = z_{1-\alpha/2}/\sqrt{w_{EB}}$. Hence, the worst-case EB coverage consistent with $E[\theta_i^2] = \mu_2$ is given by

$$\rho(\sigma^2/\mu_2, z_{1-\alpha/2}/\sqrt{w_{EB}}).$$

- Rule of thumb: Coverage at least 90% for nominal 95% CI when $w_{EB} \geq 0.3$ (next slide).

- Proposition: Worst-case coverage over all w_{EB} is $1 - 1/\max\{z_{1-\alpha/2}^2, 1\}$. Equals 74% for nominal 95% EBCI. Obtains as $w_{EB} \to 0$ (i.e., $\mu_2/\sigma^2 \to 0$).

Undercoverage of parametric EBCI



Maximal non-coverage probability of parametric EBCI. Vertical line: rule of thumb $w_{EB} = 0.3$.

Outline

Baseline model

- Allow for covariates and heteroskedasticity:

$$(Y_i \mid \theta_i, X_i, \sigma_i) \sim N(\theta_i, \sigma_i^2).$$

- Working assumption (not actually imposed later):

$$(\theta_i \mid X_i, \sigma_i) \sim N(\mu_{1,i}, \mu_2), \quad \text{where} \quad \mu_{1,i} = X_i'\delta.$$

- Suggests posterior mean shrinkage estimator

$$\hat{\theta}_i = X_i'\delta + w_{EB,i}(Y_i - X_i'\delta), \quad \text{where} \quad w_{EB,i} = \frac{\mu_2}{\mu_2 + \sigma_i^2}.$$

- Assume moment independence (also needed for MSE gain): Xie, Kou & Brown (2012)

$$E[(\theta_i - X_i'\delta)^2 \mid X_i, \sigma_i] = \mu_2, \quad E[(\theta_i - X_i'\delta)^4 \mid X_i, \sigma_i] = \kappa\mu_2^2.$$

In paper: relax using nonparametrics.

Practical implementation of robust EBCI

- Tighter coverage bound by also imposing kurtosis of cond'l bias $b_i = \frac{(1-w_{EB,i})(\theta_i - X_i'\delta)}{w_{EB,i}\sigma_i}$:

$$\rho(m_2, \kappa, \chi) = \sup_F E_{b\sim F}[r(b,\chi)] \quad \text{s.t.} \quad E_{b\sim F}[b^2] = m_2,\ E_{b\sim F}[b^4] = \kappa m_2^2.$$

  - Linear program in F. Optimum has 5 support points. Recast as 2 nested univariate optimiz's.

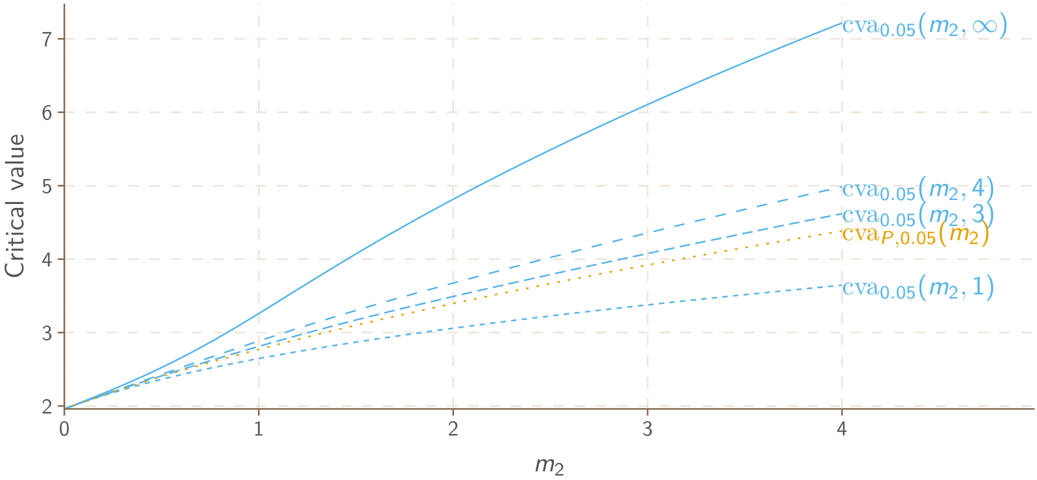
- Critical value $\text{cva}_\alpha(m_{2,i}, \kappa) = \rho^{-1}(m_{2,i}, \kappa, \alpha)$ (inverse is in last argument), with $m_{2,i} = E[b_i^2 \mid X_i, \sigma_i] = \sigma_i^2/\mu_2$.

- Robust EBCI with $1-\alpha$ EB coverage, conditional on (X_i, σ_i):

$$\hat{\theta}_i \pm w_{EB,i}\sigma_i \,\text{cva}_\alpha(m_{2,i}, \kappa).$$

- Feasible version: Replace δ with OLS, (μ_2, κ) with (trimmed) moment estimates.

Comparison of critical values ($\alpha = 0.05$)

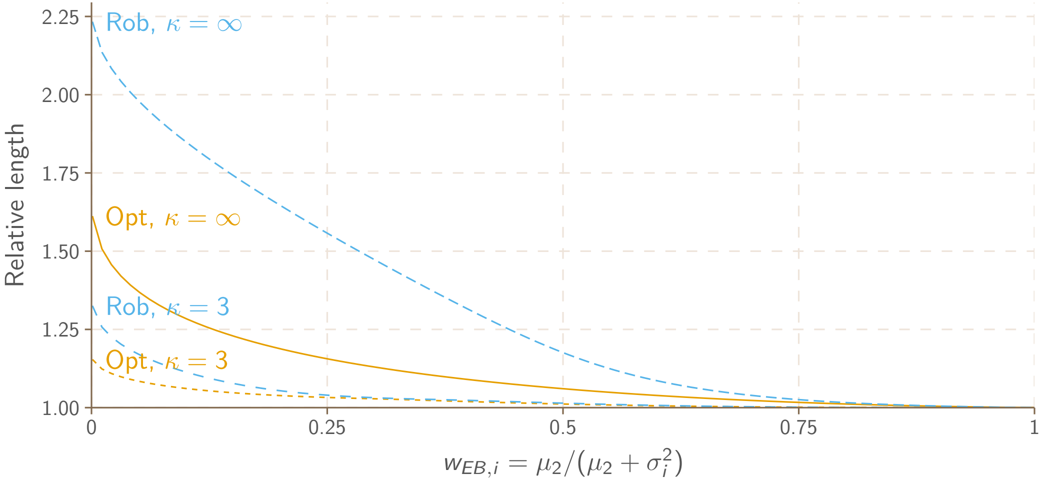


Critical value when b_i has 2nd moment m_2 and kurtosis κ. $\text{cva}_{P,0.05}$ assumes $\theta_i \sim N$.

Efficiency of robust EBCI

- Efficiency relative to unshrunk CI:
  - Already showed efficiency gain for $\kappa = \infty$.
  - Even greater gain when $\kappa < \infty$.
- Efficiency relative to parametric EBCI:
  - Robust EBCI not much wider than parametric EBCI when indeed $\theta_i \sim N$.
  - To verify claim, compare lengths when $\kappa = 3$ (kurtosis of normal distribution) – next slide.
  - Extension: Gain additional efficiency by optimizing shrinkage coefficient w for EBCI length rather than MSE.

Efficiency relative to parametric EBCI ($\alpha = 0.05$)



Length of robust EBCI and length-optimal robust EBCI relative to parametric EBCI.

Outline

Simulation study

- Panel data model: $W_{it} = \theta_i + U_{it}$, U_{it} i.i.d. mean zero, $i = 1, \dots, n$, $t = 1, \dots, T$.
- Unshrunk estimator of θ_i: $Y_i = T^{-1} \sum_{t=1}^{T} W_{it}$, with usual unbiased squared s.e. $\hat{\sigma}_i^2$.
- Effect distributions $\theta_i \overset{i.i.d.}{\sim} \Pi$:

(i) normal ($\kappa = 3$)	(ii) scaled χ_1^2 ($\kappa = 15$)
(iii) two-point ($\kappa \approx 8.11$)	(iv) three-point ($\kappa = 2$)
(v) LFD for robust EBCI (μ_2 only)	(vi) LFD for parametric EBCI

- For all distributions, consider $\mu_2 / \operatorname{Var}(Y_i \mid \theta_i) \in \{0.1, 0.5, 1, 2\}$.
- Covariates: $X_i = 1$ (shrinkage towards grand mean).
- Compare "oracle" EBCI (uses true values for σ_i, μ_2, κ) to our baseline procedure.

Monte Carlo results (nominal $\alpha = 0.05$)

	Robust, μ_2 only				Robust, μ_2 & κ				Parametric			
T	10	20	∞	ora	10	20	∞	ora	10	20	∞	ora
Panel A: Average coverage (%), minimum across 24 DGPs												
$n = 100$	92.1	93.7	94.0	95.0	91.8	93.2	93.2	94.6	79.2	79.7	79.3	86.9
$n = 200$	91.9	93.4	92.9	95.0	91.8	93.3	92.9	94.8	80.7	80.3	81.0	86.3
$n = 500$	91.9	93.6	94.8	95.0	91.9	93.5	94.3	94.9	84.2	85.1	85.1	85.6
Panel B: Relative average length, average across 24 DGPs												
$n = 100$	1.09	1.10	1.11	1.16	1.03	1.02	1.02	1.00	0.81	0.82	0.83	0.86
$n = 200$	1.09	1.10	1.12	1.16	1.02	1.02	1.01	1.00	0.81	0.82	0.84	0.86
$n = 500$	1.10	1.11	1.13	1.16	1.04	1.03	1.01	1.00	0.82	0.83	0.84	0.86

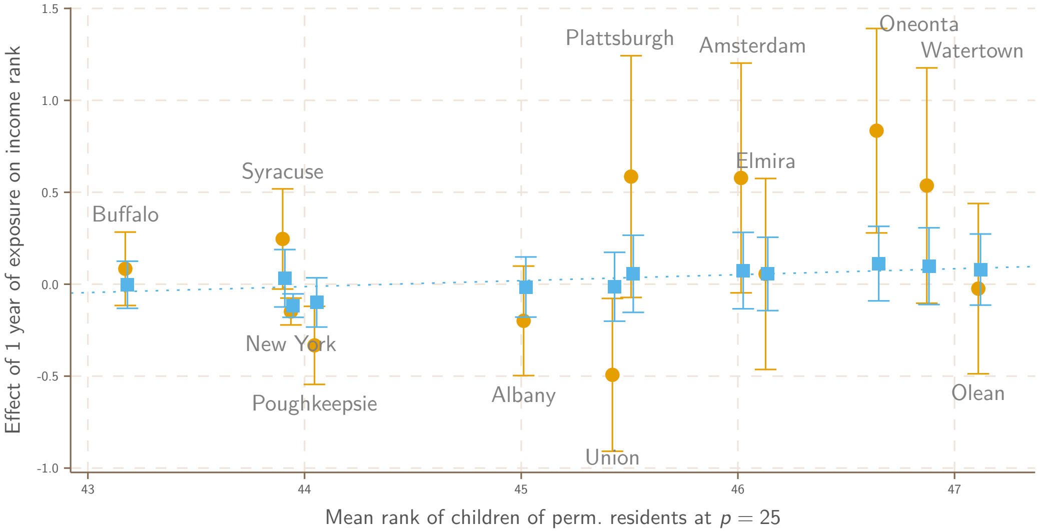
Normally distributed errors U_{it}. In paper: $U_{it} \sim \chi^2$.

Outline

Neighborhood effects

- Chetty & Hendren (2018): EB estimates of effects of neighborhoods on intergenerational mobility.
- θ_i: effect on adult income of living in commuting zone (CZ) i for one year as child (relative to average CZ).
- Y_i: fixed effect estimate of θ_i, unbiased under as'n that timing of a move is exogenous.
  - Essentially only uses data on families that move between CZs ("movers"), so it is noisy.
- To lower MSE, Chetty & Hendren regress Y_i on income X_i for permanent residents, and shrink Y_i toward this regression estimate.
- We construct robust EBCIs centered at these estimates for children in 25th percentile of household income.

Neighborhood effects for NY CZs with 90% robust EBCIs



Neighborhood effects: efficiency gain

	$E_n[\text{half-length}_i]$
Robust EBCI	0.195
Optimal robust EBCI	0.149
Parametric EBCI	0.123
Unshrunk CI	0.786

- Robustification widens the parametric EBCI, but still much shorter than unshrunk CI.
- Effect of one childhood year spent in given location, using \$818 income per percentile: Chetty & Hendren (2018, p. 1183)
  - Robust EBCI: $\pm\$818 \times 0.195 = \pm\$160$.
  - Unshrunk CI: $\pm\$818 \times 0.786 = \pm\$643$.

Neighborhood effects: fragility of parametric EBCI

Summary statistics	
κ	778.5
$E_n[\mu_2/\sigma_i^2]$	0.142
$E_n[w_{EB,i}]$	0.093
$E_n[w_{opt,i}]$	0.191
$E_n[\text{non-cov of parametric EBCI}_i]$	0.227

- Large κ and small w $\Rightarrow$ large potential undercoverage of parametric EBCI.
  - Average of 77.3% worst-case EB coverage for nominal 90% CI.
- Consistent with "rule of thumb" ($w_{EB} < 0.3$).

Outline

Local vs. global efficiency

- Our EBCI is globally valid and locally nearly efficient (when $\theta_i \sim N$).
  - Analogous to robust standard errors for OLS: only efficient under normal homoskedastic errors.
- In our model, all moments of θ_i are identified. Can in principle use to further tighten CI and center at more efficient estimator.
  - Analogous to OLS: WLS more efficient under heteroskedasticity.
  - Several nonparametric EB point estimators available. Kiefer and Wolfowitz (1956); Brown and Greenshtein (2009); Jiang and Zhang (2009); Koenker and Mizera (2014); Efron (2019)
- Challenging to achieve global optimality while allowing for (i) covariates, (ii) heteroskedasticity, and (iii) potential dependence across i, and (iv) maintaining good finite-sample performance.

Non-linear shrinkage

- Instead of going fully nonparametric, our approach can be adapted to non-linear shrinkage settings that are motivated by a specific (non-normal) effects distribution.

- Example: soft thresholding in the normal model $(Y_i \mid \theta) \sim N(\theta_i, \sigma^2)$.

  - $\hat{\theta}_i = \text{sign}(Y_i) \max\{|Y_i| - \sqrt{2\sigma^2/\mu_2}, 0\}$ is the MAP estimator under Laplace prior.

  - Obtain corresponding EBCI by calibrating HPD set

$$\mathcal{S}(Y_i; \chi) = \{\theta_i \colon \log \underbrace{\pi(\theta_i \mid Y_i)}_{\text{posterior under Laplace prior}} + \log \chi \geq 0\}.$$

  - For robust EB coverage, choose χ such that $\rho(\mu_2, \chi) = \alpha$, where

$$\rho(\mu_2, \chi) = \sup_F E_F\left[P(\theta_i \notin \mathcal{S}(Y_i; \chi) \mid \theta_i)\right] \quad \text{s.t.} \quad E_F[\theta_i^2] = \mu_2.$$

    Approximate with finely discretized linear program.

General shrinkage

- In paper: When $\theta_i \sim$ Laplace, robust soft thresholding EBCI has shorter average length than (i) unshrunk CI and (ii) robust linear EBCI.
- General idea on previous slide applicable even to non-normal sampling models $(Y_i \mid \theta)$.
  - Given some choice of family of EBCIs $\mathcal{S}(Y_i; \chi)$, just need a way to evaluate conditional non-coverage probability
$$P(\theta_i \notin \mathcal{S}(Y_i; \chi) \mid \theta_i),$$
potentially by numerical integration or simulation.
  - Example in paper: EBCI for rate parameter θ_i in Poisson sampling model.

Outline

1. Overview of results
2. Practical implementation
3. Simulation study
4. Application
5. Extension: non-linear shrinkage
6. Summary

Summary

- Construct robust empirical Bayes CIs: centered at usual EB estimator, critical value easy to compute (Matlab/R/Stata code on GitHub).
- Coverage guarantees without strong assumptions on distribution of θ_i's:
  1. Empirical Bayes coverage (repeated sampling of θ_i and data).
  2. Frequentist average coverage (fixed θ).
- Narrower than usual unshrunk CI due to weaker – but sensible – coverage requirement.
- Robust EBCI not much wider than parametric EBCI (Morris, 1983b) when parametric assumption holds.

Summary

- Construct robust empirical Bayes CIs: centered at usual EB estimator, critical value easy to compute (Matlab/R/Stata code on GitHub).
- Coverage guarantees without strong assumptions on distribution of θ_i's:
  1. Empirical Bayes coverage (repeated sampling of θ_i and data).
  2. Frequentist average coverage (fixed θ).
- Narrower than usual unshrunk CI due to weaker – but sensible – coverage requirement.
- Robust EBCI not much wider than parametric EBCI (Morris, 1983b) when parametric assumption holds.

Thank you!

Appendix

Comparison to ex-post robust Bayes

- Robust EBCI has coverage across repeated samples of (θ_i, Y_i), regardless of "prior" on θ_i.
- Instance of (asymptotically) ex-ante Γ-minimax:

$$P_{\theta\sim\pi}(\theta_i \in CI_i) \geq 1 - \alpha \quad \text{for all } \pi \in \Gamma,$$

where Γ denotes all distributions with second moment μ_2.

- Stronger requirement: ex-post Γ-minimax. Giacomini, Kitagawa & Uhlig (2019)

$$P_{\theta\sim\pi}(\theta_i \in CI_i \mid \text{data}) \geq 1 - \alpha \quad \text{for all } \pi \in \Gamma \text{ and data realizations.}$$

In our setting, this leads to reporting entire parameter space (up to moment bound).

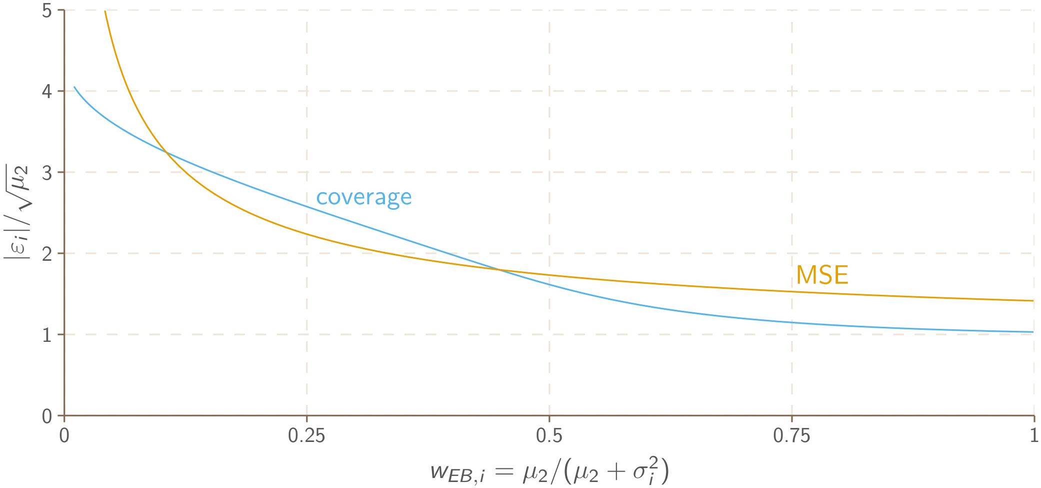
Moment estimates

- Trim moment estimates $\hat{\mu}_2$ and $\hat{\kappa}$ from below to avoid coverage problems when $\hat{w}_{EB,i} \approx 0$.
- Defining $\hat{\varepsilon}_i = Y_i - X_i'\hat{\delta}$, we use

$$\hat{\mu}_2 = \max\left\{ E_n[\hat{\varepsilon}_i^2 - \hat{\sigma}_i^2], \frac{2}{n}\frac{E_n[\hat{\sigma}_i^4]}{E_n[\hat{\sigma}_i^2]} \right\}, \quad \hat{\kappa} = \max\left\{ \frac{E_n[\hat{\varepsilon}_i^4 - 6\hat{\sigma}_i^2\hat{\varepsilon}_i^2 + 3\hat{\sigma}_i^4]}{\hat{\mu}_2^2}, 1 + \frac{32}{n\hat{\mu}_2^2}\frac{E_n[\hat{\sigma}_i^8]}{E_n[\hat{\sigma}_i^4]} \right\}.$$

- Trimming interpretation: lower bound on posterior mean under flat prior on $\mu_2 \in [0, \infty)$ or $\mu_4 - \mu_2^2 \in [0, \infty)$, in large samples when μ_2 or κ small. Morris (1983a,b)
- Actual posterior mean estimates more complicated, perform similarly in simulations.

Coverage and MSE conditional on θ_i



Value of $\varepsilon_i = \theta_i - X_i'\delta$ such that conditional coverage of EBCI equals 0.95 or such that conditional MSE of shrinkage estimator $\hat{\theta}_i$ equals that of MLE Y_i.

Soft thresholding EBCI