

# Package ‘ebci’

August 26, 2021

**Title** Robust Empirical Bayes Confidence Intervals

**Version** 0.0.0.9002

**Description** Computes empirical Bayes confidence estimators and confidence intervals (EBCIs) in a normal means model. The EBCIs are robust in the sense that they achieve correct coverage regardless of the distribution of the means. If the means are treated as fixed, the EBCIs have an average coverage guarantee.

**Depends** R ( $\geq 4.1.0$ )

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Suggests** spelling,  
testthat ( $\geq 2.1.0$ ),  
lpSolve,  
knitr,  
rmarkdown

**Language** en-US

**URL** <https://github.com/kolesarm/ebci>

**BugReports** <https://github.com/kolesarm/ebci/issues>

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.1.1

**VignetteBuilder** knitr

## R topics documented:

ct	2
cva	3
cva_tbl	4
cz	4
ebci	5

Index	8
-------	---

ct

*Neighborhood effects data from Chetty and Hendren (2018)***Description**

This dataset contains a subset of the publicly available data from Chetty and Hendren (2018). It contains raw estimates and standard errors of neighborhood effects at the county level

**Usage**

ct

**Format**

A data frame with 741 rows corresponding to counties and 12 columns corresponding to the variables:

**ct** County ID

**cz** ID of commuting zone to which the county belongs.

**ctname** County name

**state** 2-digit state code

**pop** Population of the county according to the year 2000 Census

**pop\_cz** Population of the CZ to which the county belongs according to the year 2000 Census

**theta25** Fixed-effect estimate of the causal effect of living in the county for one year on children's percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for children growing up with parents at the 25th percentile of national income distribution

**theta75** Fixed-effect estimate of the causal effect of living in the county for one year on children's percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for children growing up with parents at the 75th percentile of national income distribution

**se25** Standard error of theta25

**se75** Standard error of theta75

**stayer25** Average percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for stayers (children who grew up in the county and did not move) with parents at the 25th percentile of national income distribution.

**stayer75** Average percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for stayers (children who grew up in the county and did not move) with parents at the 75th percentile of national income distribution.

**Source**

[https://opportunityinsights.org/data/?paper\\_id=599](https://opportunityinsights.org/data/?paper_id=599)

## References

Chetty, R., & Hendren, N. (2018). *The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*. *The Quarterly Journal of Economics*, 133(3), 1163–1228. <https://doi.org/10.1093/qje/qjy006>

---

cva	Compute average coverage critical value under moment constraints.
-----	---

---

## Description

Computes the critical value  $cva_{\alpha}(m_2, \kappa)$  from Armstrong, Kolesár, and Plagborg-Møller (2020).

## Usage

```
cva(m2, kappa = Inf, alpha = 0.05, check = TRUE)
```

## Arguments

m2	Bound on second moment of the normalized bias, $m_2$
kappa	Bound on the kurtosis of the normalized bias, $\kappa$
alpha	Determines confidence level, $1 - \alpha$ .
check	If TRUE, verify accuracy of the solution by checking that the implied least favorable distribution satisfies the m2 and kappa constraints and yields the same non-coverage rate. If this fails (perhaps due to numerical accuracy issues), solve a finite-grid approximation (by discretizing the support of the normalized bias) to the primal linear programming problem, and check that it agrees with the dual solution.

## Value

Returns a list with 4 components:

cv Critical value for constructing two-sided confidence intervals.  
 alpha The argument alpha.  
 x Support points for the least favorable distribution for the squared normalized bias,  $b^2$ .  
 p Probabilities associated with the support points.

## References

Armstrong, Timothy B., Kolesár, Michal, and Plagborg-Møller, Mikkel (2020): *Robust Empirical Bayes Confidence Intervals*, <https://arxiv.org/abs/2004.03448>

## Examples

```
## Critical value without imposing a constraint on kurtosis
cva(1, kappa=Inf)
## With a constraint
cva(1, kappa=3)
```

---

cva_tbl	<i>Table of pre-computed critical values</i>
---------	--

---

**Description**

Table of pre-computed critical values

**Usage**

cva\_tbl

**Format**

A data frame with 4 variables:

m2 Bound on the second moment of the normalized bias

kappa Bound on the kurtosis of the normalized bias

cv Critical value

alpha Determines confidence level,  $1 - \alpha$ .

**Source**

Computed using the cva function in this package

---

cz	<i>Neighborhood effects data from Chetty and Hendren (2018)</i>
----	---

---

**Description**

This dataset contains a subset of the publicly available data from Chetty and Hendren (2018). It contains raw estimates and standard errors of neighborhood effects at the commuting zone level

**Usage**

cz

**Format**

A data frame with 741 rows corresponding to commuting zones (CZ) and 10 columns corresponding to the variables:

**cz** Commuting zone ID

**czname** Name of CZ

**state** 2-digit state code

**pop** Population according to the year 2000 Census

**theta25** Fixed-effect estimate of the causal effect of living in the CZ for one year on children's percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for children growing up with parents at the 25th percentile of national income distribution

**theta75** Fixed-effect estimate of the causal effect of living in the CZ for one year on children's percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for children growing up with parents at the 75th percentile of national income distribution

**se25** Standard error of theta25

**se75** Standard error of theta75

**stayer25** Average percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for stayers (children who grew up in the CZ and did not move) with parents at the 25th percentile of national income distribution.

**stayer75** Average percentile rank in the national distribution of household earnings at age 26 relative to others in the same birth cohort for stayers (children who grew up in the CZ and did not move) with parents at the 75th percentile of national income distribution.

## Source

[https://opportunityinsights.org/data/?paper\\_id=599](https://opportunityinsights.org/data/?paper_id=599)

## References

Chetty, R., & Hendren, N. (2018). *The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*. *The Quarterly Journal of Economics*, 133(3), 1163–1228. <https://doi.org/10.1093/qje/qjy006>

---

ebci	<i>Compute empirical Bayes confidence intervals by shrinking toward regression</i>
------	--

---

## Description

Computes empirical Bayes estimators based on shrinking towards a regression, and associated robust empirical Bayes confidence intervals (EBCIs), as well as length-optimal robust EBCIs.

## Usage

```
ebci(
  formula,
  data,
  se,
  weights = NULL,
  alpha = 0.1,
```

```

kappa = NULL,
wopt = FALSE,
fs_correction = "PMT"
)

```

### Arguments

formula	object of class "formula" (or one that can be coerced to that class) of the form $Y \sim \text{predictors}$ , where $Y$ is a preliminary unbiased estimator, and predictors are predictors $X$ that guide the direction of shrinkage. For shrinking toward the grand mean, use $Y \sim 1$ , and for shrinking toward $\emptyset$ use $Y \sim \emptyset$
data	optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the preliminary estimator $Y$ and the predictors. If not found in data, these variables are taken from <code>environment(formula)</code> , typically the environment from which the function is called.
se	Standard errors $\sigma$ associated with the preliminary estimates $Y$
weights	An optional vector of weights to be used in the fitting process in computing $\delta$ , $\mu_2$ and $\kappa$ . Should be <code>NULL</code> or a numeric vector.
alpha	Determines confidence level, $1 - \alpha$ .
kappa	If non- <code>NULL</code> , use pre-specified value for the kurtosis $\kappa$ of $\theta - X'\delta$ (such as <code>Inf</code> ), instead of computing it.
wopt	If <code>TRUE</code> , also compute length-optimal robust EBCIs. These are robust EBCIs centered at estimates with the shrinkage factor $w_i$ chosen to minimize the length of the resulting EBCI.
fs_correction	Finite-sample correction method used to compute $\mu_2$ and $\kappa$ . These corrections ensure that we do not shrink the preliminary estimates $Y$ all the way to zero. If "PMT", use posterior mean truncation, if "FPLIB" use limited information Bayesian approach with a flat prior, and if "none", truncate the estimates at $\emptyset$ for $\mu_2$ and 1 for $\kappa$ .

### Value

Returns a list with the following components:

**mu2** Estimated second moment of  $\theta - X'\delta$ ,  $\mu_2$ . Vector of length 2, the first element corresponds to the estimate after the finite-sample correction as specified by `fs_correction`, the second element is the uncorrected estimate.

**kappa** Estimated kurtosis  $\kappa$  of  $\theta - X'\delta$ . Vector of length 2 with the same structure as **mu2**.

**delta** Estimated regression coefficients  $\delta$

**df** Data frame with components described below.

**df** has the following components:

**w\_eb** EB shrinkage factors,  $\mu_2/(\mu_2 + \sigma_i^2)$

**w\_opt** Length-optimal shrinkage factors

**ncov\_pa** Maximal non-coverage of parametric EBCIs

`len_eb` Half-length of robust EBCIs based on EB shrinkage, so that the intervals take the form `cbind(th_eb-len_eb, th_eb+len_eb)`  
`len_op` Half-length of robust EBCIs based on length-optimal shrinkage, so that the intervals take the form `cbind(th_op-len_op, th_op+len_op)`  
`len_pa` Half-length of parametric EBCIs, which take the form `cbind(th_eb-len_pa, th_eb+len_a)`  
`len_us` Half-length of unshrunk CIs, which take the form `cbind(th_us-len_us, th_us+len_us)`  
`th_us` Unshrunk estimate  $Y$   
`th_eb` EB estimate.  
`th_op` Estimate based on length-optimal shrinkage.  
`se` Standard error  $\sigma$ , as supplied by the argument `se`  
`weights` Weights used  
`residuals` The residuals  $Y_i - X_i\delta$

## References

Armstrong, Timothy B., Kolesár, Michal, and Plagborg-Møller, Mikkel (2020): *Robust Empirical Bayes Confidence Intervals*, <https://arxiv.org/abs/2004.03448>

## Examples

```
## Using only commuting zones in New York
ebci(theta25 ~ stayer25, cz[cz$state=="NY", ],
      se25, 1/se25^2)
```

# Index

## \* datasets

ct, [2](#)

cva\_tbl, [4](#)

cz, [4](#)

ct, [2](#)

cva, [3](#)

cva\_tbl, [4](#)

cz, [4](#)

ebci, [5](#)