Robust Empirical Bayes Confidence Intervals

Michal Kolesár

August 26, 2021

The package `ebci` implements robust empirical Bayes confidence intervals (EBCIs) proposed by Armstrong et al. [2020] for inference in a normal means model $Y_i \sim N(\theta_i, \sigma_i^2)$, $i = 1, \dots, n$.

Setup

Suppose we use an empirical Bayes estimator of θ_i that shrinks toward the predictor based on the regression of θ_i onto X_i (equivalently, regression of Y_i onto X_i),

$$\hat{\theta}_i = X_i'\delta + w_{EB,i}(Y_i - X_i'\delta), \tag{1}$$

where $\delta = E[X_i X_i']^{-1} E[X_i \theta_i]$, $w_{EB,i} = \frac{\mu_2}{\mu_2 + \sigma_i^2}$ is the degree of shrinakge, and

$$\mu_2 = E[(\theta_i - X_i'\delta)^2 \mid X_i, \sigma_i]. \tag{2}$$

is the second moment of the regression residual. We assume that μ_2 doesn't depend on σ_i. Under this setup, Morris [1983] proposes to use the *parametric EBCI*

$$\hat{\theta}_i \pm \frac{z_{1-\alpha/2}}{\sqrt{w_{EB,i}}} w_{EB,i} \sigma_i.$$

The critical value $z_{1-\alpha/2}/\sqrt{w_{EB,i}}$ is larger than the usual critical value $z_{1-\alpha/2} =$`qnorm(1-alpha/2)` if the estimator was unbiased conditional on θ_i. This CI is justified if we strengthen the assumption (2) by making the normality assumption $\theta_i \mid X_i, \sigma_i \sim N(X_i'\delta, \mu_2)$. However, if the residual $\theta_i - X_i'\delta$ is not normally distributed, this CI will undercover. Armstrong et al. [2020] derive a *robust EBCI* that is only uses (2) and not the normality assumption. The EBCI takes the form

$$\hat{\theta}_i \pm cva_\alpha(m_{2i}, \infty) w_{EB,i} \sigma_i, \;\; m_{2i} = (1 - 1/w_{EB,i})^2 \mu_2/\sigma_i^2 = \sigma_i^2/\mu_2, \tag{3}$$

where the critical value cva_α is derived in Armstrong et al. [2020]. Here m_{2i} is the second moment of the conditioanl bias of $\hat{\theta}_i$, scaled by the standard error (normalized bias, henceforth). This critical value imposes a constraint (2) on the second moment of θ_i, but no constraints on higher moments. We can make the critical value smaller by also imposing a constraint on the kurtosis of θ_i (or equivalently, the kurtosis of the normalized bias)

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

In analogy to (2), we assume here that the conditional fourth moment of $\theta_i - X_i'\delta$ doesn't depend on (X_i, σ_i). In this case, the robust EBCI takes the form

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

These critical values are implemented in the package by the `cva` function:

```
library("ebci")
## If m_2=0, then we get the usual critical value
cva(m2=0, kappa=Inf, alpha=0.05)$cv
#> [1] 1.959964
## Otherwise the critical value is larger:
cva(m2=4, kappa=Inf, alpha=0.05)$cv
#> [1] 7.216351
## Imposing a constraint on kurtosis tightens it
cva(m2=4, kappa=3, alpha=0.05)$cv
#> [1] 4.619513
```

In practice, the parameters δ, μ_2, and κ are unknown. To implement the EBCI, the package replaces them with consistent estimates, following the baseline implementation in Armstrong et al. [2020]. We illustrate this in the next section.

Example

Here we illustrate the use of the package using a dataset from Chetty and Hendren [2018] (CH hereafter). The dataset is included in the package as the list `cz`. Run `?cz` for a full description of the dataset. As in Chetty and Hendren [2018], we use precision weights proportional to the inverse of the squared standard error to compute (δ, μ_2, κ).

```
## As Y_i, use fixed effect estimate theta25 of causal effect of neighborhood
## for children with parents at the 25th percentile of income distribution. The
## standard error for this estimate is se25. As predictors use average outcome
## for permanent residents (stayers), stayer25. Let us use 90% CIs, as in
## Armstrong et al
r <- ebci(formula=theta25~stayer25, data=cz, se=se25, weights=1/se25^2,
          alpha=0.1)
```

For shrinkage toward the grand mean, or toward zero, use the specification `theta25 ~ 1`, and `theta25 ~ 0`, respectively, in the `formula` argument of `ebci`.

The return value contains (see `?ebci` for full description)

1. The least squares estimate of δ:

   ```
   r$delta
   #> (Intercept)    stayer25
   #> -1.44075193  0.03244676
   ```

2. Estimates of μ_2 and κ. The estimate used for EBCI calculations (`estimate`) is obtained by applying a finite-sample correction to an initial method of moments estimate (`uncorrected_estimate`). This correction ensures that we don't shrink all the way to zero (or

past zero) if the method-of-moments estimate of μ_2 is negative (see Armstrong et al. [2020] for details):

```
c(r$mu2, r$kappa)
#>             estimate uncorrected_estimate             estimate
#>         6.243867e-03         6.243867e-03         7.785337e+02
#> uncorrected_estimate
#>         3.453191e+02
```

3. The parameter α determining the confidence level, `r$alpha`, and the matrix of regressors, `r$X`.
4. A data frame with columns:

```
names(r$df)
#>  [1] "w_eb"      "w_opt"     "ncov_pa"   "len_eb"    "len_op"    "len_pa"
#>  [7] "len_us"    "th_us"     "th_eb"     "th_op"     "se"        "weights"
#> [13] "residuals"
```

The columns of the data frame refer to:

- `w_eb` Empirical Bayes shrinkage factor $w_{EB,i} = \mu_2/(\mu_2 + \sigma_i^2)$.
- `th_eb` Empirical Bayes estimator $\hat{\theta}_i$ given in (1)
- `len_eb` Half-length $cva_\alpha(m_2, \kappa)w_i\sigma_i$ of the robust EBCI, so that the lower endpoint of the EBCIs are given by `th_eb-len_eb`, and the upper endpoint by `th_eb+len_eb`. Let us verify this for the first observation:

```
cva(r$df$se[1]^2/r$mu2[1], r$kappa[1], alpha=0.1)$cv*r$df$w_eb[1]*r$df$se[1]
#> [1] 0.1916245
r$df$len_eb[1]
#> [1] 0.1916245
```

- `len_pa` Half-length $z_{1-\alpha/2}\sqrt{w_i}\sigma_i$ of the parametric EBCI.
- `w_opt` Shrinkage factor that optimizes the length of the resulting confidence interval. In other words, instead of using $w_{EB,i}$ in (3), we use shrinkage w_i that minimizes $cva_\alpha((1 - 1/w_{EB,i})^2\mu_2/\sigma_i^2, \kappa)w_i\sigma_i$. See Armstrong et al. [2020] for details. The vector is missing here since the default option, `wopt=FALSE`, is to skip computation of the length-optimal CIs to speed up the calculations.
- `th_op` Estimator based on the length-optimal shrinkage factor `w_opt` (missing here since the default is `wopt=FALSE`)
- `len_op` Half-length $cva_\alpha((1 - 1/w_{EB,i})^2\mu_2/\sigma_i^2, \kappa)w_i\sigma_i$ of the length-optimal EBCI (missing here since we specified `wopt=FALSE`).
- `th_us` The unshrunk estimate Y_i, as specified in the `formula` argument of the function `ebci`.
- `len_us` Half-length $z_{1-\alpha/2}\sigma_i$ of the CI based on the unshrunk estimate
- `se` The standard error σ_i, as specified by the argument `se` of the `ebci` function.
- `ncov_pa` maximal non-coverage of the parametric EBCI.

Using the data frame, we can give a table summarizing the results. Let us show the results for the CZ in New York:

```
df <- (cbind(cz[!is.na(cz$se25), ], r$df))
df <- df[df$state=="NY", ]

knitr::kable(data.frame(cz=df$czname, unshrunk_estimate=df$theta25,
             estimate=df$th_eb,
             lower_ci=df$th_eb-df$len_eb, upper_ci=df$th_eb+df$len_eb),
             digits=3)
```

cz	unshrunk_estimate	estimate	lower_ci	upper_ci
Syracuse	0.246	0.032	-0.124	0.189
Oneonta	0.835	0.112	-0.091	0.315
Union	-0.493	-0.014	-0.201	0.173
Buffalo	0.084	-0.003	-0.131	0.125
Elmira	0.056	0.056	-0.143	0.255
Olean	-0.024	0.080	-0.114	0.273
Watertown	0.537	0.098	-0.111	0.307
Plattsburgh	0.585	0.056	-0.153	0.266
Amsterdam	0.578	0.074	-0.134	0.282
Albany	-0.199	-0.015	-0.179	0.148
Poughkeepsie	-0.333	-0.099	-0.233	0.035
New York	-0.148	-0.116	-0.180	-0.053

Figure 6 in Armstrong et al. [2020] presents the same information in graphical form.

Finally, let us compute some summary statistics as in Table 3 in Armstrong et al. [2020]. Average length of the robust, parametric, and unshrunk CI:

```
mean(r$df$len_eb)
#> [1] 0.1952409
mean(r$df$len_pa)
#> [1] 0.1234121
mean(r$df$len_us)
#> [1] 0.7858001
```

Therefore, the efficiency of the parametric and unshrunk CI relative to the robust EBCI is given by

```
mean(r$df$len_eb)/mean(r$df$len_pa)
#> [1] 1.582023
mean(r$df$len_eb)/mean(r$df$len_us)
#> [1] 0.2484613
```

While the parametric EBCI is shorter on average, it yields CIs that may violate the 90% coverage requirement. In particular, the average maximal non-coverage probability at the estimated value of (μ_2, κ) is given by

```r
mean(r$df$ncov_pa)
#> [1] 0.2274703
```

References

Tim Armstrong, Michal Kolesár, and Mikkel Plagborg-Møller. Robust empirical Bayes confidence intervals. ArXiv: 2004.03448, April 2020. URL https://arxiv.org/abs/1606.01200.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228, August 2018. doi: 10.1093/qje/qjy006.

Carl N. Morris. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381):47–55, March 1983. doi: 10.1080/01621459.1983.10477920.