Valid confidence intervals for post-model-selection predictors

François Bachoc, Hannes Leeb et Benedikt M. Pötscher

University of Vienna

Full linear model

(Full) linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}$$

- $\boldsymbol{Y}$ of size $n \times 1$: observation vector
- $\boldsymbol{X}$ of size $n \times p$: design matrix
- $\boldsymbol{\beta}$ of size $p \times 1$: regression coefficients
- $\boldsymbol{U} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$
- $p < n$

$\Longrightarrow$ Working distribution $P_{n,\boldsymbol{\beta},\sigma}$

Least square estimator :

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

Standard variance estimator :

$$\hat{\sigma}^2 = \frac{1}{n-p}||\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}||^2$$

Linear submodels

Subsets $M \subset \{1, ..., p\}$ of the columns of $\boldsymbol{X}$. Give

$$\boldsymbol{Y} = \boldsymbol{X}[M]\boldsymbol{v} + \boldsymbol{U}$$

- M of cardinality m
- $\boldsymbol{X}[M]$ of size $n \times m$: only the columns of $\boldsymbol{X}$ that are in M
- $\boldsymbol{v}$ of size $m \times 1$: needs to be defined/estimated to give the best representation of the full linear model

Non-standard regression coefficient vector

$$\beta_M^{(n)} = \underset{\boldsymbol{v}}{\operatorname{argmin}} \, ||\boldsymbol{X}\beta - \boldsymbol{X}[M]\boldsymbol{v}||$$

$$\beta_M^{(n)} = \beta[M] + \left(\boldsymbol{X}'[M]\boldsymbol{X}[M]\right)^{-1} \boldsymbol{X}'[M]\boldsymbol{X}[M^c]\beta[M^c],$$

- $\beta[M]$ of size $m \times 1$: components of β in M

Restricted least square estimator

$$\hat{\beta}_M = \left(\boldsymbol{X}'[M]\boldsymbol{X}[M]\right)^{-1} \boldsymbol{X}'[M]Y$$

Model selection procedure

Data-driven selection of the model with $\hat{M}(\boldsymbol{Y}) = \hat{M}$
Ex. : sequential testing, AIC, BIC, LASSO

Berk et al., 2013, Annals of Statistics consider the non-standard target

$$\boldsymbol{\beta}_{\hat{M}}^{(n)}$$

as their target for confidence intervals

Comments :

- Model selector $\hat{M}$ is "imposed"
- Objective : best coefficients in this imposed model
- Random target

Predictors

Let $\boldsymbol{x}_0$ be a fixed $p \times 1$ vector and consider

$$y_0 = \boldsymbol{x}_0' \boldsymbol{\beta} + u_0$$

- $u_0 \sim \mathcal{N}(0, \sigma^2)$

We consider the design-dependent non-standard target

$$\boldsymbol{x}_0'[\hat{M}]\boldsymbol{\beta}_{\hat{M}}^{(n)}$$

Optimality property : when $\boldsymbol{x}_0$ is random and follows the empirical distribution given by the lines of $\boldsymbol{X}$:

$$\mathbb{E}_{n,\boldsymbol{\beta},\sigma}\left(\left[y_0 - \mathbf{x}_0'[\hat{M}]\boldsymbol{\beta}_{\hat{M}}^{(n)}\right]^2\right) \leq \mathbb{E}_{n,\boldsymbol{\beta},\sigma}\left(\left[y_0 - \mathbf{x}_0'[\hat{M}]v(Y)\right]^2\right),$$

for any function $v(Y) \in \mathbb{R}^{|\hat{M}|}$.

Let a nominal level $1 - \alpha \in (0, 1)$ be fixed

We consider confidence intervals for $\mathbf{x}_0'[\hat{M}]\beta_{\hat{M}}^{(n)}$ of the form

$$CI = \mathbf{x}_0'[\hat{M}]\hat{\beta}_{\hat{M}} \pm K||\mathbf{s}_{\hat{M}}||\hat{\sigma},$$

with

$$\mathbf{s}_M' = \mathbf{x}_0'[M]\left(\mathbf{X}'[M]\mathbf{X}[M]\right)^{-1}\mathbf{X}'[M]$$

Interpretation

- "Constant" K does not depend on Y (but on $\boldsymbol{X}, \mathbf{x}_0, \hat{M}$)
- For fixed M,
$$\mathbf{x}_0'[M]\hat{\beta}_M - \mathbf{x}_0'[M]\beta_M^{(n)} \sim \mathcal{N}(0, ||\mathbf{s}_M||\sigma^2)$$
- Thus, $K_{naive} = q_{S,n-p,1-\alpha/2}$ (Student quantile) is valid when M is deterministic
- When $\hat{M}$ is random, K needs to be larger (e.g. Leeb et al. 2015, Statistical Science)

$\Longrightarrow$ Main issue : choosing K

The construction of Berk et al.

Observe that

$$\mathbf{x}_0'[\hat{M}]\hat{\beta}_{\hat{M}} - \mathbf{x}_0'[\hat{M}]\beta_{\hat{M}}^{(n)} = \mathbf{s}_{\hat{M}}' \left(\mathbf{Y} - \mathbf{X}\beta\right)$$

Then, we have

$$\left| \frac{\mathbf{s}_{\hat{M}}'}{||\mathbf{s}_{\hat{M}}||\hat{\sigma}} \left(Y - \boldsymbol{X}\beta\right) \right| \leq \max_{M \subseteq \{1,\ldots,p\}} \left| \frac{\mathbf{s}_M'}{||\mathbf{s}_M||\hat{\sigma}} \left(Y - \boldsymbol{X}\beta\right) \right|$$

Distribution of the upper-bound does not depend on $\beta, \sigma \implies$ let K_1 be its $(1-\alpha)$ quantile

The CI given by K_1 satisfies

$$\inf_{\boldsymbol{\beta} \in \mathbb{R}^p, \sigma > 0} P_{n,\boldsymbol{\beta},\sigma} \left(\mathbf{x}_0'[\hat{M}]\beta_{\hat{M}}^{(n)} \in CI \right) \geq 1 - \alpha$$

$\implies$ **Uniformly valid** confidence interval

The constant K_1 depends on all the components of $\mathbf{x}_0$

It can happen that only $\mathbf{x}_0[\hat{M}]$ is observed

- model selection for cost reason

We construct other constants (see the paper for details)

$$K_1 \leq K_2 \leq K_3 \leq K_4$$

(The CIs given by K_2, K_3, K_4 are hence universally valid)

Remark : The case where only $\mathbf{x}_0[\hat{M}]$ is observed motivates all the more the study of $\mathbf{x}_0'[\hat{M}]\beta_{\hat{M}}^{(n)}$ as opposed to $\mathbf{x}_0'\beta$

Issue : The target $\mathbf{x}_0'[\hat{M}]\beta_{\hat{M}}^{(n)}$ depends on $\mathbf{X}$

Issue is solved when lines of $\boldsymbol{X}$ and $\mathbf{x}_0'$ are realizations from the same distribution $\mathcal{L}$

Let, for $\mathbf{x}' \sim \mathcal{L}$, $\mathbf{\Sigma} = \mathbb{E}(\mathbf{x}\mathbf{x}')$. Then, define the design-independent non-standard target by

$$\mathbf{x}_0[\hat{M}]'\beta_{\hat{M}}^{(\star)} = \mathbf{x}_0[\hat{M}]'\beta[\hat{M}] + \mathbf{x}_0[\hat{M}]' \left(\mathbf{\Sigma}[\hat{M},\hat{M}]\right)^{-1} \mathbf{\Sigma}[\hat{M},\hat{M}^c]\beta[\hat{M}^c],$$

Then, we have for $\mathbf{x}_0 \sim \mathcal{L}$,

$$\mathbb{E}\left(\left[y_0 - \mathbf{x}_0'[\hat{M}]\beta_{\hat{M}}^{(\star)}\right]^2\right) \leq \mathbb{E}\left(\left[y_0 - \mathbf{x}_0'[\hat{M}]\mathbf{v}(\mathbf{Y})\right]^2\right),$$

for any function $\mathbf{v}(\mathbf{Y}) \in \mathbb{R}^{|\hat{M}|}$

Asymptotic coverage when p is fixed and $n \to \infty$

Observe that

$$\left(\mathbf{x}_0[\hat{M}]' \beta^{(\star)}_{\hat{M}} - \mathbf{x}_0[\hat{M}]' \beta^{(n)}_{\hat{M}} \right) =$$

$$\mathbf{x}_0'[\hat{M}] \left(\left(\mathbf{X}'[\hat{M}] \mathbf{X}[\hat{M}] \right)^{-1} \mathbf{X}'[\hat{M}] \mathbf{X}[\hat{M}^c] - \left(\mathbf{\Sigma}[\hat{M}, \hat{M}] \right)^{-1} \mathbf{\Sigma}[\hat{M}, \hat{M}^c] \right) \beta[\hat{M}^c]$$

Theorem

Assume that

$$\sqrt{n} \left[(\mathbf{X}'\mathbf{X}/n) - \mathbf{\Sigma} \right] = O_p(1)$$

and that for any M with $|M| < p$ and for any $\delta > 0$,

$$\sup \left\{ P_{n,\boldsymbol{\beta},\sigma}(\hat{M} = M | \boldsymbol{X}) : \boldsymbol{\beta} \in \mathbb{R}^p, \sigma > 0, \left\| \beta[M^c] \right\| / \sigma \geq \delta \right\} = o_p(1)$$

Then, for CI obtained by K_1, K_2, K_3, K_4,

$$\inf_{\boldsymbol{\beta} \in \mathbb{R}^p, \sigma > 0} P_{n,\boldsymbol{\beta},\sigma} \left(\mathbf{x}_0'[\hat{M}] \beta^{(\star)}_{\hat{M}} \in CI \middle| \boldsymbol{X} \right) \geq (1 - \alpha) + o_p(1)$$

For $\alpha = 0.05$ and $p = 10$ we evaluate

$$\inf_{\boldsymbol{\beta} \in \mathbb{R}^p, \sigma > 0} P_{n,\boldsymbol{\beta},\sigma} \left(\mathbf{x}_0'[\hat{M}]\beta_{\hat{M}}^{(n,\star)} \in CI \middle| \boldsymbol{X} \right),$$

for one realization of **X**

Results :

n	model selector	target							
		design-dependent $\mathbf{x}_0[\hat{M}]'\beta_{\hat{M}}^{(n)}$				design-independent $\mathbf{x}_0[\hat{M}]'\beta_{\hat{M}}^{(\star)}$			
		K_{naive}	K_1	K_3	K_4	K_{naive}	K_1	K_3	K_4
20	AIC	0.84	0.99	1.00	1.00	0.79	0.97	0.99	0.99
20	BIC	0.84	0.99	1.00	1.00	0.74	0.96	0.98	0.98
20	LASSO	0.90	1.00	1.00	1.00	0.18	0.48	0.61	0.61
100	AIC	0.87	0.99	1.00	1.00	0.88	0.99	1.00	1.00
100	BIC	0.88	0.99	1.00	1.00	0.87	0.99	1.00	1.00
100	LASSO	0.88	0.99	1.00	1.00	0.87	0.99	1.00	1.00

- It is known that in the classical case (estimation of β), it is difficult to construct valid post-model-selection confidence intervals
- Recently, alternative coefficients have been studied Berk et al. 2013
  - this removes some obstacles
  - but naive procedures still fail
- We extend the confidence intervals to prediction
  - exact coverage of the design-dependent target
  - asymptotic coverage of the design-independent target

The paper :

F. Bachoc, H. Leeb, B.M. Pötscher. Valid confidence intervals for post-model-selection predictors,
`http://arxiv.org/abs/1412.4605`

Thank you for your attention !