

Valid confidence intervals post-model-selection

François Bachoc

Institut de Mathématiques de Toulouse
Université Paul Sabatier

Joint work with **Hannes Leeb** (Vienna), **Benedikt Pötscher** (Vienna), **David Preinerstorfer** (Brussels) and **Lukas Steinberger** (Vienna)

Séminaire LJK-Données et Aléatoire Théorie & Applications
Grenoble
January 2020

- 1 Post-model-selection inference with Gaussian linear models
- 2 Confidence intervals
- 3 Extension to linear predictors
- 4 Extension to non-linear non-Gaussian settings

Location model

$$Y = \mu + U.$$

- Y of size $n \times 1$: **observation vector**.
- μ of size $n \times 1$: **unknown mean vector**.
- $U \sim \mathcal{N}(0, \sigma^2 I_n)$.
- σ^2 **known** in the first three sections for simplicity of exposition.

Linear models

- Design matrix X of size $n \times p$.
 - $p < n$ or $p \geq n$.
- Universe \mathcal{M} of models/submodels.

$$\mathcal{M} \subseteq \{M \subseteq \{1, \dots, p\}\}.$$

- Each M is a set of selected columns of X .
- Write $|M|$ for the cardinality of M .
- Write $X[M]$ of size $n \times |M|$: only the columns of X that are in M .
- Restricted least square estimator

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

- For $M \in \mathcal{M}$.
- Assuming $X[M]$ has full column rank for $M \in \mathcal{M}$.
- Implies $|M| \leq n$.

⇒ We consider subsets of selected variables and construct linear models from them.

Examples of universes of models \mathcal{M}

■ All non-empty models:

$$\mathcal{M} = \{M \subseteq \{1, \dots, p\}; M \neq \emptyset\},$$

- only when $p \leq n$.

■ All models containing the first variable:

$$\mathcal{M} = \{M \subseteq \{1, \dots, p\}; M \neq \emptyset, 1 \in M\},$$

- only when $p \leq n$,
- e.g. first variable is an intercept (first column of X composed of 1s).

■ s -sparse models:

$$\mathcal{M} = \{M \subseteq \{1, \dots, p\}; |M| \leq s\},$$

- allows for $n < p$,
- $1 \leq s \leq n$ is the sparsity parameter.

Regression coefficients of interest

- The projection-based target: Let for $M \in \mathcal{M}$,

$$\begin{aligned}\beta_M^{(n)} &= \underset{|M| \times 1 \text{ vector } v}{\operatorname{argmin}} \quad \|\mu - X[M]v\| \\ &= (X'[M]X[M])^{-1} X'[M]\mu.\end{aligned}$$

- ⇒ $\beta_M^{(n)}$ is a target of inference in this talk.
- ⇒ Motivated in [Berk et al 2013](#).
- ⇒ Subsequently considered in [Lee et al 2016](#), [Tibshirani et al 2018](#), [Kuchibhotla et al 2019](#),...
- ⇒ When $p < n$ and $\mu \notin \operatorname{span}(X)$: links to extensive literature on misspecified parametric models (e.g. [Huber 1967](#), [White 1982](#)).

Illustration (1/2)

- $n = 50, p = 2$



$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

- **well-specified case:**

$$\mu_i = 1/2 + x_i$$

for $i = 1, \dots, n$.

- $\mu \in \text{span}(X)$.

- **misspecified case:**

$$\mu_i = -1/2 + x_i + 4x_i^2$$

for $i = 1, \dots, n$.

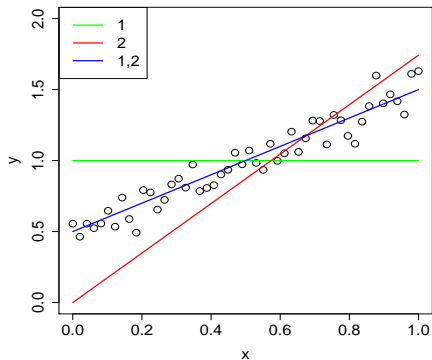
- $\mu \notin \text{span}(X)$.

Illustration (2/2)

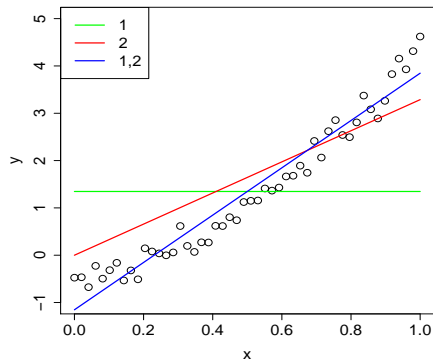
Plot of

- Observations Y_1, \dots, Y_n ,
- $(X[M]\beta_M^{(n)})_i, i = 1, \dots, n$, for

$M = \{1\}$, $M = \{2\}$ and $M = \{1, 2\}$.



(well-specified)



(misspecified)

- **Model selection procedure:** data-driven selection of the model with

$$\hat{M}(Y) = \hat{M} \in \mathcal{M}.$$

- Sequential testing, AIC, BIC, LASSO, SCAD (Fan and Li 2001), MCP (Zhang 2010),...
- In Berk et al 2013, target for inference is $\beta_{\hat{M}}^{(n)}$ and \hat{M} can be any model selection procedure.
 - Model selector \hat{M} is **imposed**.
 - Objective: best coefficients in this imposed model.

This is what we call the **post-model-selection inference** setting.

Discussion (1/2)

- Motivated by the following common practice in applications:
 - 1 Select model \hat{M} from data Y .
 - 2 Apply usual confidence intervals/tests with design matrix $X[\hat{M}]$.

⇒ Invalid because \hat{M} is data-dependent.

⇒ Aim at changing tests/confidence intervals so that they become valid.
- Motivation for considering target $\beta_{\hat{M}}^{(n)}$:
 - Best we can do once \hat{M} is fixed.
 - Relevant in misspecified case when $\mu \notin \text{span}(X)$ (when $p < n$).
 - Relevant when $p > n$ and no sparse representation of μ in X .

Discussion (2/2)

- Aim for procedures that work for any function $Y \rightarrow \hat{M}(Y)$.
- In practice \hat{M} can be
 - not formally defined,
 - imposed.
- Robustness to
 - hunting for significance,
 - data snooping.
- This talk is not about how to select a “good” model \hat{M} .

- We consider the setting of [Berk et al 2013](#),
 - confidence intervals for $\beta_{\hat{M}}^{(n)}$ for any \hat{M} ,
 - current related work [Zhang 2017](#), [Kuchibhotla et al 2019](#),...
- In [Lee et al 2016](#), [Tibshirani et al 2018](#),...,
 - confidence intervals for $\beta_{\hat{M}}^{(n)}$,
 - \hat{M} is **specific**: LASSO, sequential testing,...
 - valid (coverage probability) **conditionally** to \hat{M} .
- In [Van der Geer et al 2014](#), the LASSO model selector is used for confidence intervals in **sparse well-specified** models in high-dimension.
- Some **intrinsic difficulties** in post-model-selection inference were discussed earlier in [Leeb and Pötscher 2005, 2006](#),...

1 Post-model-selection inference with Gaussian linear models

2 Confidence intervals

3 Extension to linear predictors

4 Extension to non-linear non-Gaussian settings

Confidence intervals

We consider confidence intervals for

$$\left(\beta_{\hat{M}}^{(n)}\right)_j,$$

for $j = 1, \dots, |\hat{M}|$, of the form

$$CI_{\hat{M},j} = \left(\hat{\beta}_{\hat{M}}\right)_j \pm K \|s_{\hat{M},j}\| \sigma,$$

with

$$s'_{\hat{M},j} = \text{row } j \text{ of } \left(X'[\hat{M}]X[\hat{M}]\right)^{-1} X'[\hat{M}].$$

Interpretation:

- For **fixed** M and j ,

$$\left(\hat{\beta}_M\right)_j - \left(\beta_M^{(n)}\right)_j \sim \mathcal{N}(0, \|s_{M,j}\|^2 \sigma^2).$$

- Thus, selecting K as a Gaussian quantile is valid when M is deterministic.
- When \hat{M} is random, K needs to be larger to **account for model selection**.

⇒ **Question:** choosing K .

Reduction to a simultaneous coverage problem

Berk et al 2013:

The coverage

$$\text{for } j = 1, \dots, |\hat{M}|, \quad \left(\beta_{\hat{M}}^{(n)} \right)_j \in CI_{\hat{M},j}$$

holds if the **simultaneous coverage**

$$\text{for } M \in \mathcal{M} \text{ and } j = 1, \dots, |M|, \quad \left(\beta_M^{(n)} \right)_j \in CI_{M,j}$$

holds. This is equivalent to

$$\text{for } M \in \mathcal{M} \text{ and } j = 1, \dots, |M|, \quad \frac{\left| \left(\hat{\beta}_M \right)_j - \left(\beta_M^{(n)} \right)_j \right|}{\|s_{M,j}\| \sigma} \leq K$$

(convention $0/0 = 0$).

POSI (post-selection inference) constant

The last event can be rewritten as

$$\max_{M \in \mathcal{M}, j=1, \dots, |M|} \left| \frac{s'_{M,j} (Y - \mu)}{\|s_{M,j}\| \sigma} \right| \leq K.$$

- Distribution of the maximum does not depend on μ, σ .

⇒ Taking

$$K = K_{1-\alpha}(X)$$

(POSI constant) as the $1 - \alpha$ quantile of this maximum yields

$$\mathbb{P} \left(\text{for } j = 1, \dots, |\hat{M}|, \left(\beta_{\hat{M}}^{(n)} \right)_j \in CI_{\hat{M},j} \right) \geq 1 - \alpha,$$

for all $n, p, \mu \in \mathbb{R}^n, \sigma > 0$.

⇒ Uniformly valid confidence intervals (Berk et al 2013).

⇒ $K_{1-\alpha}(X)$ is optimal to guarantee this property.

$K_{1-\alpha}(X)$ quantile $1 - \alpha$ of

$$\max_{M \in \mathcal{M}, j=1, \dots, |M|} \left| \frac{s'_{M,j}}{\|s_{M,j}\|} (U/\sigma) \right|$$

with $U/\sigma \sim \mathcal{N}(0, I_n)$.

- Supremum norm of a **large centered Gaussian vector**,
 - dimension $\sum_{M \in \mathcal{M}} |M|$,
 - up to $p2^{p-1}$ when $p < n$.
- With unit variances.
- Alternatively: **many one-dimensional projections** of a standard Gaussian vector.

Computation of the POSI constant

- When p not too large, $K_{1-\alpha}(X)$ can be estimated by [Monte Carlo](#),
 - say $p < 30$ when \mathcal{M} is unrestricted,
 - larger p for sparse models,
 - R package PoSI.
- But cost usually **exponential in p** .
- Upper bound

$$B_{1-\alpha} \geq K_{1-\alpha}(X)$$

suggested in [Berk et al 2013](#),

- computation complexity \approx **constant** w.r.t. n, p ,
- can be used in practice for large n, p .

How large are the POSI constant and its upper bound?

(Berk et al 2013:)

- Fixed model, $\mathcal{M} = \{M_0\}$:

$$\sup_{X \text{ } n \times p \text{ matrix}} K_{1-\alpha}(X) = O(1).$$

- All models, $p < n$, $\mathcal{M} = \{M \subseteq \{1, \dots, p\}\}$:

$$\inf_{X \text{ } n \times p \text{ matrix}} K_{1-\alpha}(X) = \sqrt{2 \log(p)}(1 + o(1)),$$

$$0.6363\sqrt{p}(1 + o(1)) \leq \sup_{X \text{ } n \times p \text{ matrix}} K_{1-\alpha}(X) \leq 0.866\sqrt{p}(1 + o(1)).$$

$\implies K_{1-\alpha}(X)$ depends on X in a complex way.

Upper bound:

- Sparse models, $n < p$, $\mathcal{M} = \{M \subseteq \{1, \dots, p\}; |M| \leq s\}$:

$$B_{1-\alpha} = O\left(\sqrt{s \log\left(\frac{p}{s}\right)}\right).$$

1 Post-model-selection inference with Gaussian linear models

2 Confidence intervals

3 Extension to linear predictors

4 Extension to non-linear non-Gaussian settings

This section is based on the paper:



Bachoc, F., Leeb, H., & Pötscher, B.M., Valid confidence intervals for post-model-selection predictors, *Annals of Statistics*, 47(3), 1475-1504, 2019.

- We consider a $p \times 1$ vector x_0 ,
 - new explanatory variables.
- Define $x_0[M]$: subvector of x_0 with indices in M ,
 - for $M \in \mathcal{M}$.
- We want to cover the post-model-selection predictor

$$x_0[\hat{M}]' \beta_{\hat{M}}^{(n)}.$$

Adaptation of [Berk et al 2013](#):

- Confidence interval:

$$CI_{\hat{M},x_0} = x_0[\hat{M}]'\hat{\beta}_{\hat{M}} \pm K_{1-\alpha}(X, x_0) \|s_{\hat{M},x_0}\| \sigma,$$

- with

$$s'_{\hat{M},x_0} = x_0[\hat{M}]' \left(X'[\hat{M}]X[\hat{M}] \right)^{-1} X'[\hat{M}],$$

- with $K_{1-\alpha}(X, x_0)$ the $1 - \alpha$ quantile of

$$\max_{M \in \mathcal{M}} \left| \frac{s'_{M,x_0}}{\|s_{M,x_0}\|} \frac{(Y - \mu)}{\sigma} \right|.$$

- We still have an upper bound

$$B'_{1-\alpha} \geq K_{1-\alpha}(X, x_0).$$

Case of partially observed x_0

- Can frequently happen that
 - x_0 not observed entirely,
 - only $x_0[\hat{M}]$ is observed,
 - variable selection for cost reasons.
- In this case $K_{1-\alpha}(X, x_0)$ is **unavailable**.
- We still have the upper bound $B'_{1-\alpha}$.
- We also suggest

$$K_{2,1-\alpha}(X, x_0[\hat{M}], \hat{M}) = \sup_{x_0[\hat{M}^c]} K_{1-\alpha}(X, x_0),$$

- very hard to compute,
- but theoretically interesting.



$$K_{1-\alpha}(X, x_0) \leq K_{2,1-\alpha}(X, x_0[\hat{M}], \hat{M}) \leq B'_{1-\alpha}.$$

Large p analysis for orthogonal design matrices (1/2)

- When X has orthogonal columns, $K_{1-\alpha}(X)$ has rate $\sqrt{\log(p)}$ (Berk et al. 2013).
- From that we deduce that $K_{1-\alpha}(X, x_0)$ has rate $\sqrt{\log(p)}$ when x_0 is a sequence of basis vectors.

Large p analysis for orthogonal design matrices (2/2)

We show:

Proposition

Let \mathcal{M} be the power set of $\{1, \dots, p\}$.

(a) Let X have orthogonal columns. There exists a sequence of vectors x_0 such that $K_{1-\alpha}(X, x_0)$ satisfies

$$\liminf_{p \rightarrow \infty} K_{1-\alpha}(X, x_0) / \sqrt{p} \geq 0.6363.$$

(b) Let $\gamma \in [0, 1)$ be given. Then $K_{2,1-\alpha}(X, x_0[M], M)$ satisfies

$$\liminf_{p \rightarrow \infty} \inf_{x_0 \in \mathbb{R}^p} \inf_{X \in \mathcal{X}(p)} \inf_{M \in \mathcal{M}, |M| \leq \gamma p} K_{2,1-\alpha}(X, x_0[M], M) / \sqrt{p} \geq 0.6363 \sqrt{1 - \gamma},$$

where $\mathcal{X}(p) = \bigcup_{n \geq p} \{X : X \text{ is } n \times p \text{ with non-zero orthogonal columns}\}$.

\Rightarrow Strong impact of x_0 on $K_{1-\alpha}(X, x_0)$.

\Rightarrow Price to pay when only $x_0[\hat{M}]$ is observed.

Summary of another contribution of the paper

- We consider the **random regressors** setting.
- The rows of X and x_0 are realizations from a **distribution \mathcal{L}** .
- We define the post-model-selection predictor

$$x_0[\hat{M}]' \beta_{\hat{M}}^{(*)}$$

defined based on \mathcal{L} rather than on X .

- We show that the same confidence intervals as before work **asymptotically**.
 - p fixed, $n \rightarrow \infty$ here.

⇒ Recent work on the distinction between fixed and random regressors, **Buja et al 2019, Kuchibothla et al 2018**.

1 Post-model-selection inference with Gaussian linear models

2 Confidence intervals

3 Extension to linear predictors

4 Extension to non-linear non-Gaussian settings

This section is based on the paper:



Bachoc, F., Preinerstorfer, D. & Steinberger, L., Uniformly valid confidence intervals post-model-selection, *Annals of Statistics*, forthcoming, arxiv.org/abs/1611.01043, 2019.

Data:

- We consider a triangular array of independent $1 \times l$ random vectors $y_{1,n}, \dots, y_{n,n}$.
- We let $\mathbb{P}_n = \bigotimes_{i=1}^n \mathbb{P}_{i,n}$ be the distribution of $y_n = (y'_{1,n}, \dots, y'_{n,n})'$, where $\mathbb{P}_{i,n}$ is the distribution of $y_{i,n}$.

Models:

- We now consider a set $\mathbb{M}_n = \{\mathbb{M}_{1,n}, \dots, \mathbb{M}_{d,n}\}$ composed of d models.
- $\mathbb{M}_{i,n}$ is a set of distributions on $\mathbb{R}^{n \times \ell}$.
- d does not depend on n (fixed-dimensional asymptotics).

\implies We do not assume that the observation distribution \mathbb{P}_n belongs to one of the $\{\mathbb{M}_{1,n}, \dots, \mathbb{M}_{d,n}\}$. The set of models can be **misspecified**.

Parameters and estimators

Parameters:

- We define for each model $\mathbb{M} \in \mathbb{M}_n$ an **optimal parameter** $\theta_{\mathbb{M},n}^* = \theta_{\mathbb{M},n}^*(\mathbb{P}_n)$, that we assume to be non-random and of fixed dimension $m(\mathbb{M})$.
- In the case of linear models:
 - each $\mathbb{M} \in \mathbb{M}_n$ corresponds to a $M \subseteq \{1, \dots, p\}$,
 - $\theta_{\mathbb{M},n}^* = \beta_M^{(n)}$.
- The optimal parameter $\theta_{\mathbb{M},n}^*$ is specific to the model \mathbb{M} .

Estimators:

- We consider, for each $\mathbb{M} \in \mathbb{M}_n$, an estimator $\hat{\theta}_{\mathbb{M},n}$ of the optimal parameter $\theta_{\mathbb{M},n}^*$.

Model selection:

- We consider a **model selection procedure**: a function $\hat{M}_n : \mathbb{R}^{n \times \ell} \rightarrow M_n$.
- We are hence interested in constructing **confidence intervals** for the random quantity of interest $\theta_{\hat{M}_n, n}^*$.

Main idea and notation

Main idea:

- We aim at showing a **joint asymptotic normality** of $\{\hat{\theta}_{\mathbb{M},n} - \theta_{\mathbb{M},n}^*\}_{\mathbb{M} \in \mathbb{M}_n}$.
- We then use the same construction as in the **linear Gaussian case** for the confidence intervals.
- Additional difficulty: we do not know the **asymptotic covariance matrix**.

Notation:

- $\hat{\theta}_n = (\hat{\theta}'_{\mathbb{M}_1,n}, \dots, \hat{\theta}'_{\mathbb{M}_d,n})'$.
- $\theta_n^* = (\theta_{\mathbb{M}_1,n}^*, \dots, \theta_{\mathbb{M}_d,n}^*)'$.
- Let $k = \sum_{j=1}^d m(\mathbb{M}_j,n)$, be the dimension of $\hat{\theta}_n$.

Joint Asymptotic normality

- Let $r_n = \hat{\theta}_n - \theta_n^*$.
- Let $S_n = \text{VC}_n(r_n)$ (covariance matrix).
- Let d_w be a distance generating the topology of weak convergence for distributions on an Euclidean space.
- Let $\text{corr}(\Sigma)$ be the correlation matrix obtained from a covariance matrix Σ .
- Let $\text{diag}(\Sigma)$ be obtained by setting the off-diagonal elements of Σ to 0.

Lemma

Under some conditions, we have

$$d_w \left(\text{law of } \text{diag}(S_n)^{-1/2} \left(\hat{\theta}_n - \theta_n^* \right), N(0, \text{corr}(S_n)) \right) \rightarrow 0.$$

Some notation

- For $\alpha \in (0, 1)$ and for a covariance matrix Γ , let $K_{1-\alpha}(\Gamma)$ be the $1 - \alpha$ -quantile of $\|Z\|_\infty$ for $Z \sim N(0, \Gamma)$.
 \implies Very similar to above **POSI constant**.
- For $\mathbb{M} = \mathbb{M}_{i,n} \in \mathbb{M}_n$ and $j \in \{1, \dots, m(\mathbb{M})\}$ let

$$j \star \mathbb{M} := \sum_{l=1}^{i-1} m(\mathbb{M}_{l,n}) + j$$

($j \star \mathbb{M}$ is the index of $(\theta_{\mathbb{M},n}^*)_j$ in $(\theta_{\mathbb{M}_1,n}^*, \dots, \theta_{\mathbb{M}_d,n}^*)'$).

Let $\alpha \in (0, 1)$. Let \hat{S}_n be such that, with $\|A\|$ the largest singular value of A ,

$$\|\text{corr}(\hat{S}_n) - \text{corr}(\text{VC}_n(r_n))\| + \|\text{diag}(\text{VC}_n(r_n))^{-1} \text{diag}(\hat{S}_n) - I_k\| \rightarrow_p 0.$$

Consider, for $\mathbb{M} \in \mathbb{M}_n$ and $j = 1, \dots, m(\mathbb{M})$, the **confidence interval**

$$\text{CI}_{1-\alpha, \mathbb{M}}^{(j), \text{est}} = \left[\hat{\theta}_{\mathbb{M}, n} \right]_j \pm \sqrt{[\hat{S}_n]_{j^* \mathbb{M}, j^* \mathbb{M}}} K_{1-\alpha} \left(\text{corr}(\hat{S}_n) \right).$$

We show:

Theorem

Then, $\mathbb{P}_n \left(\left[\theta_{\mathbb{M}, n}^* \right]_j \in \text{CI}_{1-\alpha, \mathbb{M}}^{(j), \text{est}} \text{ for all } \mathbb{M} \in \mathbb{M}_n \text{ and } j = 1, \dots, m(\mathbb{M}) \right)$ goes to $1 - \alpha$ as $n \rightarrow \infty$. In particular, for any model selection procedure $\hat{\mathbb{M}}_n$, we have

$$\liminf_{n \rightarrow \infty} \mathbb{P}_n \left(\left[\theta_{\hat{\mathbb{M}}_n, n}^* \right]_j \in \text{CI}_{1-\alpha, \hat{\mathbb{M}}_n}^{(j), \text{est}} \text{ for all } j = 1, \dots, m(\hat{\mathbb{M}}_n) \right) \geq 1 - \alpha.$$

- When the models are misspecified it may not be possible to estimate $\mathbb{V}\mathbb{C}_n(r_n)$ consistently.
- We show how to **overestimate** the diagonal components of $\mathbb{V}\mathbb{C}_n(r_n)$.
- This is based on overestimating $\mathbb{V}(y_{i,n})$ based on

$$\mathbb{V}(y_{i,n}) \leq \mathbb{E}((y_{i,n} - \hat{y}_{i,n})^2)$$

where $\hat{y}_{i,n}$ is obtained from a misspecified model \mathbb{M} .

- Also there exists an **upper-bound** of $K_{1-\alpha}(\text{corr}(S_n))$ (similar to $B_{1-\alpha}$).

⇒ We obtain the same asymptotic guarantee as before with more conservative confidence intervals.

- We have seen a general method that can be applied to specific situations on a case by case basis.
- Need **uniform central limit theorems** for fixed models in misspecified cases (sandwich rule).
- Need to consistently overestimate variances.
- In the paper, we provide applications to
 - homoscedastic linear models with homoscedastic data,
 - heteroscedastic linear models with heteroscedastic data,
 - binary regression models with binary data.

Some simulation results

In a Monte Carlo simulation (1000 repetitions) for logistic regression ($p = 10, n = 30, 100$), we compare

- CI coverage for a nominal level at 0.9 (cov. 0.9),
- CI median length (med.),
- CI 90% quantile length (qua.)

for

- our post-selection inference CI (P),
- the CI by [Taylor and Tibshirani, 2017](#), specific to the lasso (L),
- the naive CI that ignores the presence of model selection (N).

model selector	cov. 0.9			med.			qua.		
	P	L	N	P	L	N	P	L	N
lasso (1)	0.99	0.89	0.84	4.26	7.44	2.09	6.97	43.33	3.42
lasso (2)	1.00	0.85	0.68	1.63	2.31	0.74	1.90	13.52	0.84
lasso (3)	1.00	0.25	0.98	2.22	1.23	1.01	2.83	3.50	1.24
sig. hun.	0.95		0.39	4.40		2.63	6.22		3.63

Some simulation results in high dimension (1/2)

Monte Carlo simulation (1000 repetitions) for homoscedastic linear models ($p = 1000, n = 50$).

- The model selector is [forward stepwise](#).

We compare

- CI coverage for a nominal level at 0.9 (cov.),
- CI median length (med.),
- CI 90% quantile length (qua.)

for

- our post-selection inference CI (P),
- the CI by [Tibshirani et al. 2016](#), specific to forward-stepwise (FS),
- the naive CI that ignores the presence of model selection (N).

Some simulation results in high dimension (2/2)

	cov.	Step 1 med.	qua.	cov.	Step 2 med.	qua.	cov.	Step 3 med.	qua.	Simult. cov.
P	0.99	8.33	9.38	1.00	10.39	12.73	1.00	11.49	14.35	0.99
FS	0.94	11.66	55.76	0.88	786.92	Inf	0.90	1754.00	Inf	0.77
N	0.58	3.54	3.98	0.49	3.33	4.08	0.45	3.22	4.03	0.08
P	0.91	7.24	8.07	1.00	9.34	12.15	1.00	10.36	13.68	0.91
FS	0.93	15.15	72.67	0.88	752.74	Inf	0.90	1582.32	Inf	0.76
N	0.00	3.07	3.43	0.12	3.00	3.90	0.19	2.91	3.84	0.00







Remarks:

- Top 3 rows: design matrix X has independent columns.
- Bottom 3 rows: design matrix X has correlated columns.
- The CI's P and FS use the knowledge that k variables are selected at step k .






- Inference for targets that depend on selected models.
- Simultaneous coverage of many correlated and normalized errors.
- Exact for Gaussian case \implies asymptotic for more general cases.
- **Open questions:** high-dimensional asymptotics, computational aspects.

Thank you for your attention!






Bibliography

-  Berk, R., Brown, L., Buja, A., Zhang, K. & Zhao, L., Valid post-selection inference, *The Annals of Statistics*, 41(2), 802-837, 2013.
-  Lee, J. D., Sun, D. L., Sun, Y. & Taylor, J. E., Exact post-selection inference, with application to the lasso, *The Annals of Statistics*, 44(3), 907-927, 2016.
-  Huber, P. J., The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221-233, 1967.
-  White, H., Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 125, 1982.
-  Fan, J. & Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, 96, 1348-1360, 2001.
-  Zhang, C., Nearly unbiased variable selection under minimax concave penalty, *Annals of Statistics*, 38, 894-942, 2010.

Bibliography

-  Zhang, K., Spherical cap packing asymptotics and rank-extreme detection, *IEEE Transactions on Information Theory*, 63, 2017.
-  Kuchibhotla, A. K., Brown, L. D., Buja, A., Cai, J., George E.I. & Zhao, L., Valid post-selection inference in model-free linear regression, *Annals of Statistics*, forthcoming, 2019.
-  Tibshirani, R. J., Rinaldo, A., Tibshirani, R., & Wasserman, L., Uniform asymptotic inference and the bootstrap after model selection, *The Annals of Statistics*, 46(3), 1255-1287, 2018.
-  van de Geer, S., Buhlmann, P., Ritov, Y. & Dezeure, R., On asymptotically optimal confidence regions and tests for high-dimensional models, *Annals of Statistics*, 42, 1166-1202, 2014.
-  Leeb, H. & Pötscher, B. M., Model selection and inference: facts and fiction, *Econometric Theory*, 21, 2159, 2005.

Bibliography

-  Leeb, H. & Pötscher, B. M., Performance limits for estimators of the risk or distribution of shrinkage-type estimators, and some general lower risk-bound results, *Econometric Theory*, 22, 6997, 2006.
-  Buja, A., Brown, L. D., Kuchibhotla, A. K., Berk, R., George, E.I. & Zhao, L., Models as approximations—part II: a general theory of model-robust regression, *Statistical Science*, to appear (with discussion), 2019.
-  Kuchibhotla, A.K., Brown, L.D., Buja, A., George, E.I. & Zhao, L., A model free perspective for linear regression: uniform-in-model bounds for post selection inference, *arXiv:1802.05801*, 2018.
-  Taylor, J., & Tibshirani, R., Post-selection inference for penalized likelihood models, *Canadian Journal of Statistics*, 46(1), 41-61, 2018.
-  Tibshirani, R. J., Taylor, J., Lockhart, R. & Tibshirani, R., Exact post-selection inference for sequential regression procedures, *Journal of the American Statistical Association*, 111 (514), 600-620, 2016.