UCDAVIS
DEPARTMENT OF STATISTICS

Uniformly Valid Inference Based on the Lasso in Linear Mixed Models

joint work with U. Schneider and T. Krivobokova

Peter Kramlinger

Lima, SAE 2024

Model Selection and Inference

General Linear Model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}_n\{\mathbf{0}_n, \mathbf{V}(\boldsymbol{\theta}_0)\}$$

for $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^r$ unknown, $n, p, r \in \mathbb{N}$, $n > p$

Goal: Inference on $\boldsymbol{\beta}_0$ after model selection.

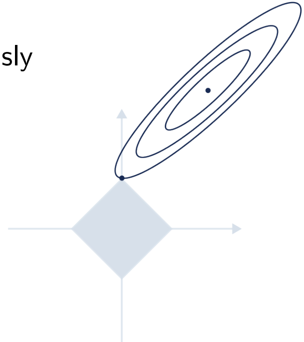
- Model selection is data driven
- Selection and estimation are two sources of uncertainty
- Naive inference after model selection is not valid

Least Absolute Shrinkage and Selection Operator (Lasso)

for $\lambda_j > 0$, $j = 1, \ldots, p$ and REML estimator $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$,

$$\widehat{\boldsymbol{\beta}}_L = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \ell(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}) + 2 \sum_{j=1}^{p} \lambda_j \, |\beta_j| \right\}$$

- Selection and estimation simultaneously
- Lasso not uniformly consistent
  Pötscher and Leeb [2009]



Related Work

In mixed models:

- Penalization of $\boldsymbol{\theta}_0$: Bondell et al. [2010], Ibrahim et al. [2011]
- Penalization of the random effects: Peng and Lu [2012]
- Many instances of penalized $\boldsymbol{\beta}_0$ in mixed models.

In linear models:

- Uniformly valid inference based on the Lasso if $n > p$
  Ewald and Schneider [2018]

$$\inf_{\boldsymbol{\beta}_0} \mathrm{P}_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0 \in S) = \min_{\mathbf{d} \in \{-1,1\}^p} \mathrm{P}(\boldsymbol{\beta}_0 \in S_{\mathbf{d}}) = 1 - \alpha$$

Lasso

for $\lambda_j > 0$, $j = 1, \ldots, p$ and REML estimator $\widehat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$,

$$\widehat{\boldsymbol{\beta}}_L = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \ell(\boldsymbol{\beta}, \widehat{\boldsymbol{\theta}}) + 2 \sum_{j=1}^{p} \lambda_j \, |\beta_j| \right\}$$

Goal: Uniformly valid inference on $\boldsymbol{\beta}_0$ based on the Lasso.

- Find $S \subset \mathbb{R}^p$ such that

$$\inf_{\boldsymbol{\beta}_0, \boldsymbol{\theta}_0} \mathrm{P}_{\boldsymbol{\beta}_0, \boldsymbol{\theta}_0}(\widehat{\boldsymbol{\beta}}_L \in S) \approx 1 - \alpha$$

- Idea: Separate estimation for $\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}_0$
- Requires that $\widehat{\boldsymbol{\theta}}$ is uniformly bounded

Restricted Maximum Likelihood (REML) Methodology

Estimate $\boldsymbol{\theta}_0$ from $\mathbf{A}^t\mathbf{y}$ instead of $\mathbf{y}$, for $\mathbf{A} \in \mathbb{R}^{n\times(n-p)}$ such that $\mathbf{A}^t\mathbf{X} = \mathbf{0}_{n-p}$. Then,

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}}\, \ell_R(\boldsymbol{\theta}),$$

where $\mathbf{P}(\boldsymbol{\theta}) = \mathbf{A}\{\mathbf{A}^t\mathbf{V}(\boldsymbol{\theta})\mathbf{A}\}^{-1}\mathbf{A}^t$ and

$$\ell_R(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} -\frac{1}{2}\ln|\mathbf{V}(\boldsymbol{\theta})| - \frac{1}{2}\ln|\mathbf{X}^t\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}^t\mathbf{P}(\boldsymbol{\theta})\mathbf{y}$$

- Requires low-dimensional setting $n > p$
- $\widehat{\boldsymbol{\theta}}$ is independent of $\boldsymbol{\beta}_0$
- *Uniform* consistency has to be proved
- Maximum not unique, hence consider *Cramér* consistency

Theorem 1: Cramér Consistency of REML Estimator

Under some regularity conditions, set

$$\nu_k(\boldsymbol{\theta}_0) = \frac{\mathrm{tr}\left\{\mathbf{V}(\boldsymbol{\theta}_0)^{-1}\frac{\partial \mathbf{V}}{\partial \theta_k}\right\}}{\sqrt{\mathrm{rk}\left(\frac{\partial \mathbf{V}}{\partial \theta_k}\right)}}, \qquad k = 1, \ldots, r.$$

There exists a sequence $\widehat{\boldsymbol{\theta}}$ *of local maximizers of* ℓ_R*, such that*

$$\nu_k(\boldsymbol{\theta}_0)\left|\widehat{\theta}_k - \theta_{0,k}\right| = O_P(1)$$

uniformly over $\boldsymbol{\theta}_0 \in \Theta$*.*

- If $\boldsymbol{\theta}_0$ were fixed, $\nu_k(\boldsymbol{\theta}_0) \equiv \sqrt{n}$
- Note that $-\mathrm{E}_{\boldsymbol{\theta}_0}\{\partial^2 \ell_R(\boldsymbol{\theta})/\partial \theta_k^2|_{\boldsymbol{\theta}_0}\} = O\{\nu_k(\boldsymbol{\theta}_0)^2\}$

Theorem 2: Uniformly Valid Confidence Sets

There exists a sequence $\widehat{\boldsymbol{\theta}}$ of local maximizers of ℓ_R such that for $E(\widehat{\mathbf{C}},\widehat{\tau}) = \{\mathbf{z} \in \mathbb{R}^p \mid \mathbf{z}^t\widehat{\mathbf{C}}\mathbf{z} \leq \widehat{\tau}\}$, with $\widehat{\mathbf{C}} = n^{-1}\mathbf{X}^t\mathbf{V}(\widehat{\theta})^{-1}\mathbf{X}$ and

$$\widehat{\tau} = \max_{\mathbf{d}\in\{-1,1\}^p} \chi^2_{p,1-\alpha}\left\{n^{-1}\left\|\widehat{\mathbf{C}}^{-1/2} diag(\lambda_1,\ldots,\lambda_p)\mathbf{d}\right\|^2\right\},$$

a quantile of a non-central χ^2_p-distribution, it holds that

$$\inf_{\substack{\boldsymbol{\beta}_0\in\mathbb{R}^p\\ \boldsymbol{\theta}_0\in\Theta}} \mathrm{P}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}\left\{\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_0\right) \in E\left(\widehat{\mathbf{C}},\widehat{\tau}\right)\right\} = 1-\alpha+O\left(\frac{1}{\sqrt{n}}\right).$$

- Idea:

$$\inf_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0} \mathrm{P}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\beta}_0 \in S) = \inf_{\boldsymbol{\theta}_0}\min_{\mathbf{d}\in\{-1,1\}^p} \mathrm{P}_{\boldsymbol{\beta}_0,\boldsymbol{\theta}_0}(\boldsymbol{\beta}_0 \in S_{\mathbf{d}})$$

- Confidence set: $M_L = \left\{\boldsymbol{\beta} \in \mathbb{R}^p\colon \quad n\left\|\widehat{\mathbf{C}}^{1/2}\left(\widehat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}\right)\right\|^2 \leq \widehat{\tau}\right\}$

Simulation Design

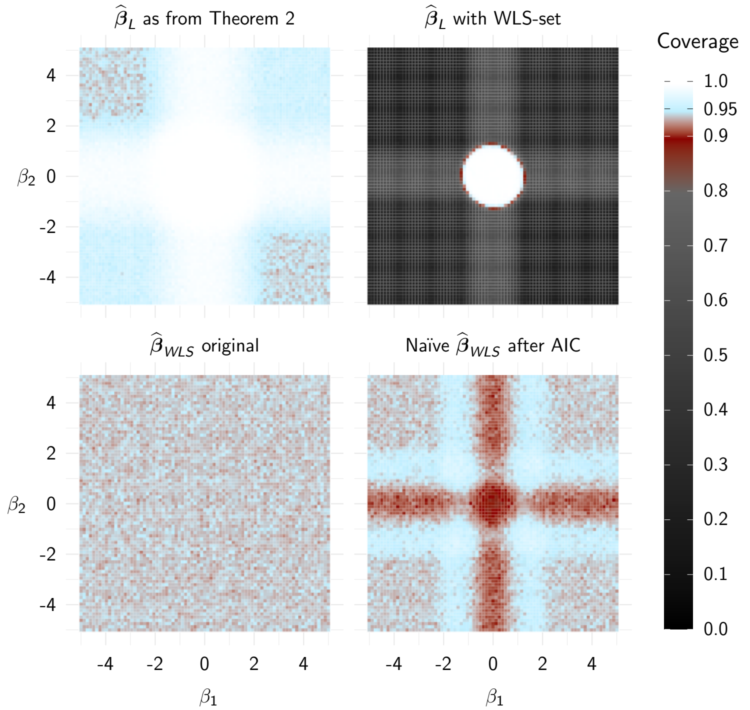
Random Intercept Model

For $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$, let

$$y_{ij} = \mathbf{x}_{ij}^t \boldsymbol{\beta}_0 + v_i + u_{ij}, \qquad u_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_u^2), \quad v_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_v^2)$$

and $m = 20$, $n = 400$, $n_i = 20$, $\sigma_u = \sigma_v = 4$, $\lambda_i = n^{1/2}/2$, $p = 2$.

- For each $\boldsymbol{\beta}_0 \in [-4, 4]^2$, perform $3,000$ replications
- For each replications, check if $\boldsymbol{\beta}_0 \in M_L$



$\widehat{\beta}_L$ as from Theorem 2
$\widehat{\beta}_L$ with WLS-set
$\widehat{\beta}_{WLS}$ original
Naïve $\widehat{\beta}_{WLS}$ after AIC
Coverage
1.0
0.95
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
β_2
β_1
4
2
0
-2
-4

Real Data Example

Acidity in US Lakes Opsomer et al. [2009]

$$\mathbf{y} = \mathbf{1}_n\beta_1 + \mathbf{x}\beta_2 + \mathbf{Z}\mathbf{v} + \mathbf{D}\mathbf{u} + \boldsymbol{\epsilon},$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}_m, \sigma_v^2\mathbf{I}_m), \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}_K, \sigma_u^2\mathbf{I}_K) \quad \text{and} \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_n, \sigma_e^2\mathbf{I}_n)$$

- WLS: $\widehat{\boldsymbol{\beta}}_{WLS} = (483, -1.20)^t$; $sd(\widehat{\boldsymbol{\beta}}_{WLS}) = (1718, 0.14)^t$
- Lasso: $\widehat{\boldsymbol{\beta}}_L = (0, -1.12)^t$
- Lasso-based confidence set $\sim$ 200 times larger, but

$$M_L(\beta_2) = [-1.49, -0.75] \qquad M_{WLS}(\beta_2) = [-1.57, -0.83]$$

H. Bondell, A. Krishna, and S. Ghosh. Joint Variable Selection for Fixed and Random Effects in Linear Mixed-Effects Models. *Biometrics*, 66:1069–1077, 2010.

K. Ewald and U. Schneider. Uniformly Valid Confidence Sets Based on the Lasso. *Electronic Journal of Statistics*, 12:1358–1387, 2018.

J. Ibrahim, H. Zhu, R. Garcia, and R. Guo. Fixed and Random Effects Selection in Mixed Effects Models. *Biometrics*, 67(2):1358–1387, 2011.

P. Kramlinger, U. Schneider, and T. Krivobokova. Uniformly valid inference based on the lasso in linear mixed models. *Journal of Multivariate Analysis*, 198:105230, 2023.

J. D. Opsomer, Claeskens, G., M. G. Ranalli, G. Kauermann, and F. J. Breidt. Non-parametric small area estimation using penalized spline regression. *JRSS B*, 70:265286, 2009.

H. Peng and Y. Lu. Model Selection in Linear Mixed Models. *Journal of Multivariate Analysis*, 109:109–129, 2012.

B. Pötscher and H. Leeb. On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding. *Journal of Multivariate Analysis*, 100:2065–2082, 2009.

L. Weiss. Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases. *Journal of the American Statistical Association*, 66: 345–350, 1971.

Conclusions

- REML estimator for $\boldsymbol{\theta}_0$ is uniformly Cramér consistent
- Inference on $\boldsymbol{\beta}_0$ based on $\widehat{\boldsymbol{\beta}}_L$ uniformly valid over $\mathbb{R}^p \times \Theta$
- Resulting coverage probability depends on $\boldsymbol{\beta}_0$

- If interest in inference only, do no selection
- Lasso-based confidence sets are large
- Valid inference (for $n > p$) if nuisance and target parameters can be separated

Regularity Conditions

(A) $\boldsymbol{\theta}_0 \in \Theta = \left\{\boldsymbol{\theta} \in \mathbb{R}^r_{>0} \mid \max(\boldsymbol{\theta})/\min(\boldsymbol{\theta}) \le c\right\}$; $\infty > c$ const.

(B) $\mathbf{V}(\boldsymbol{\theta}_0) = \sum_{k=1}^r \theta_{0,k}\mathbf{H}_k$; $\mathbf{H}_k \ge 0$, $k = 1\ldots, r-1$ and $\mathbf{H}_r > 0$

(C) $\mathrm{rk}(\mathbf{X}) = p < n$

(D) For constants $0 < \underline{\omega} \le \overline{\omega} < \infty$ it holds

$$\underline{\omega} \le \frac{\mathrm{tr}\left\{\mathbf{V}(\mathbf{1}_r)^{-1}\mathbf{H}_k\right\}}{\mathrm{rk}(\mathbf{H}_k)} \le \overline{\omega},$$

where $\mathbf{1}_r = (1, \ldots, 1)' \in \mathbb{R}^r$

(E) $m = \min\{\mathrm{rk}(\mathbf{H}_1), \ldots, \mathrm{rk}(\mathbf{H}_r)\} \to \infty$