Numerical methods for estimating the tuning parameter in penalized least squares problems

Paraskevi Roupa

Department of Mathematics, National and Kapodistrian University of Athens

parask_roupa@math.uoa.gr

Joint work with: Christos Koukouvinos, Angeliki Lappa

Department of Mathematics, National Technical University of Athens

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

NASCA' 18, Kalamata, Greece, 2-6 July 2018

Outline of the talk

Introduction

We consider the **linear regression model**

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

- X is the **design matrix** of order $p \times d$,
- $\mathbf{y}$ is the **data vector** of length p,
- $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$ is the **error vector** of length p.

The vector $\boldsymbol{\beta}$ can be considered as the minimizer of

$$\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \|^2 .$$

Introduction (cont.)

However, there are many situations where the solution of the above minimization problem is not a very good estimator, such as:

- X is high-dimensional, i.e.
  d (number of characteristics) $> p$ (number of samples)
- columns of X are highly correlated, i.e. X is rank deficient and the matrix $X^T X$ is singular.

Therefore, we introduce the corresponding **penalized least squares problem** of the form

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^d} \{ \| \mathbf{y} - X\boldsymbol{\beta} \|^2 + \lambda \| \boldsymbol{\beta} \|^2 \},$$

which depends on an appropriately chosen value of the parameter λ.

Task: Estimation of the tuning parameter λ.

The GCV function

The **generalized cross-validation (GCV) function** $V(\lambda)$ is given by

$$V(\lambda) = \frac{\| (I_p - A_\lambda)\,\mathbf{y} \|^2}{(Tr(I_p - A_\lambda))^2} = \frac{\| \mathbf{y} - X\hat{\beta}(\lambda) \|^2}{(Tr(I_p - A_\lambda))^2},$$

where

- $A_\lambda = X(X^T X + \lambda I_d)^{-1} X^T$ is the influence matrix
- $\hat{\beta}(\lambda) = (X^T X + \lambda I_d)^{-1} X^T \mathbf{y}$ is the solution of the corresponding penalized least squares problem
- λ is the **tuning parameter**.

P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numerische Mathematik 31, 377-403 (1979).

The GCV function (cont.)

By setting $B = XX^T + \lambda I_p$, the GCV function is expressed as

$$V(\lambda) = \frac{\mathbf{y}^T B^{-2} \mathbf{y}}{(Tr(B^{-1}))^2}.$$

The **minimization of the GCV function** over λ leads to an **approximation** of the tuning parameter λ.

L. Reichel, G. Rodriguez, S. Seatzu, Error estimates for large-scale ill-posed problems, Numer. Algorithms 51(3), 341-361 (2009).

Estimation of the GCV function

The **estimation of the GCV function** $V(\lambda)$ is based on:

- an **extrapolation procedure** for approximating the quadratic forms of the type $\mathbf{x}^T B^{-r}\mathbf{x}$, $r = 1, 2$,
- the **Hutchinson' trace estimator**.

P. Fika , M. Mitrouli, Estimation of the bilinear form $\mathbf{y}^* f(A)\mathbf{x}$ for Hermitian matrices, Linear Algebra Appl. 502, 140-158 (2016).

M. F. Hutchinson, A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines, Commun. Statist., Simula. 19, 433-450 (1990).

The moments

For any integer n and a vector $\mathbf{x} \in \mathbb{R}^p$, we can define the moments of the matrix $B = XX^T + \lambda I_p \in \mathbb{R}^{p \times p}$ as

$$c_n = c_n(B, \mathbf{x}) = (\mathbf{x}, B^n \mathbf{x}).$$

We also consider the moments of the matrix XX^T as

$$s_n = s_n(XX^T, \mathbf{x}) = (\mathbf{x}, (XX^T)^n \mathbf{x}), \quad n \in \mathbb{Z}.$$

Estimation of the GCV function

The mathematical tools (cont.)

Proposition

The moments c_n satisfy the relation

$$c_n = \sum_{k=0}^{n} \binom{n}{k} \lambda^k s_{n-k}, \;\; n \in \mathbb{N}.$$

For negative integers, the moments of the matrix B satisfy the relation

$$c_{-n} = \sum_{k=0}^{\infty} (-1)^k \binom{n+k-1}{k} \lambda^{-n-k} s_k, \;\; n \in \mathbb{N},$$

where $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$ is the binomial coefficient.

The **spectral factorization** of the matrix $B \in \mathbb{R}^{p \times p}$ is

$$B = U \Lambda U^T,$$

$U = \begin{bmatrix} u_1 & u_2 & \ldots & u_p \end{bmatrix}$ is an **orthonormal** matrix,
$\Lambda = diag(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^{p \times p}$.

The **matrix** B^{-r} can be written as

$$B^{-r} = U \Lambda^{-r} U^T = \sum_{i=1}^{p} \lambda_i^{-r} u_i u_i^T.$$

Therefore, the **moments** c_n can be expressed as

$$c_n = c_n(B, x) = \sum_{i=1}^{p} \lambda_i^n |\alpha_i|^2, \;\; n \in \mathbb{Z},$$

where $\alpha_i = (x, u_i)$.

Estimation of the GCV function

The mathematical tools (cont.)

Generation of **one-** and **two-term estimates** for the **GCV function**:

- **Approximation of quadratic forms** of the type $\mathbf{x}^T B^{-r} \mathbf{x}$, $r = 1, 2$.
  - We keep an appropriate number of terms of the summation of the moments c_n.
  - We solve the system of suitable imposing interpolation conditions.
- **Hutchinson' trace estimator**, i.e.

$$\mathcal{T} = \frac{\sum_{i=1}^{N} \mathbf{v_i}^T B^{-1} \mathbf{v_i}}{N},$$

where $\{\mathbf{v_i},\ i = 1, 2, \ldots, N\}$ is a sample of N random p-vectors with elements ± 1 with probability of appearing $1/2$.

M. Mitrouli, P. R., Estimates for the generalized cross-validation function via an extrapolation and statistical approach, Calcolo, (2018) 55: 24. https://doi.org/10.1007/s10092-018-0266-3.

Tuning parameter selection via the GCV estimates

One-term GCV estimates

We consider

- $$\rho(\mathbf{x}) = \frac{s_0(\mathbf{x})(s_2(\mathbf{x}) + 2\lambda s_1(\mathbf{x}) + \lambda^2 s_0(\mathbf{x}))}{(s_1(\mathbf{x}) + \lambda s_0(\mathbf{x}))^2}, \qquad (1)$$

- a family of **one-term estimates** for the **quadratic form** $\boldsymbol{v_i}^T B^{-1} \boldsymbol{v_i}$

$$e_{\nu_2}(\boldsymbol{v_i}) = \rho(\boldsymbol{v_i})^{-\nu_2} \frac{s_0(\boldsymbol{v_i})^2}{s_1(\boldsymbol{v_i}) + \lambda s_0(\boldsymbol{v_i})}, \quad \nu_2 \in \mathbb{C},$$

- an **one-term estimate** for the **trace** of B^{-1} given by

$$\mathcal{T}_1 = \frac{\sum_{i=1}^{N} e_{\nu_2}(\boldsymbol{v_i})}{N}, \qquad (2)$$

where N is the sample size of the random vectors.

Proposition

An ***one-term GCV estimate*** $\hat{\lambda}_1$ *for the* ***tuning parameter*** λ *is given by*

$$\hat{\lambda}_1 = \underset{\lambda}{\text{argmin}} \left(g_{\nu_1,\nu_2}(\lambda) \right),$$

where $g_{\nu_1,\nu_2}(\lambda) = \dfrac{\rho(\mathbf{y})^{-2\nu_1} \dfrac{s_0(\mathbf{y})^3}{(s_1(\mathbf{y}) + \lambda s_0(\mathbf{y}))^2}}{\mathcal{T}_1^2}$, $\nu_1, \nu_2 \in \mathbb{C}$ *and* $\rho(\mathbf{y})$, $\mathcal{T}_1$ *are given by formulae (1), (2) respectively.*

We consider

- $|\alpha_1(\boldsymbol{x})|^2 = \dfrac{s_0(\boldsymbol{x})l_2(\boldsymbol{x}) - (s_1(\boldsymbol{x}) + \lambda s_0(\boldsymbol{x}))}{l_2(\boldsymbol{x}) - l_1(\boldsymbol{x})}$,
- $|\alpha_2(\boldsymbol{x})|^2 = \dfrac{s_1(\boldsymbol{x}) + \lambda s_0(\boldsymbol{x}) - s_0(\boldsymbol{x})l_1(\boldsymbol{x})}{l_2(\boldsymbol{x}) - l_1(\boldsymbol{x})}$,
- $l_{1,2}(\boldsymbol{x}) = \dfrac{r(\boldsymbol{x}) \pm \sqrt{r(\boldsymbol{x})^2 - 4t(\boldsymbol{x})}}{2}, \; l_1(\boldsymbol{x}) \neq l_2(\boldsymbol{x})$,
- $r(\boldsymbol{x}) = \dfrac{s_0(\boldsymbol{x})s_3(\boldsymbol{x}) - s_1(\boldsymbol{x})s_2(\boldsymbol{x})}{s_0(\boldsymbol{x})s_2(\boldsymbol{x}) - s_1(\boldsymbol{x})^2} + 2\lambda$,
- $t(\boldsymbol{x}) = \dfrac{s_1(\boldsymbol{x})s_3(\boldsymbol{x}) - s_2(\boldsymbol{x})^2}{s_0(\boldsymbol{x})s_2(\boldsymbol{x}) - s_1(\boldsymbol{x})^2} + \lambda r(\boldsymbol{x}) - \lambda^2$.

- A **two-term estimator** $\mathcal{T}_2$ for $Tr(B^{-1})$ is

$$\mathcal{T}_2 = \frac{\sum_{i=1}^{N} \left(l_1(\mathbf{v_i})^{-1} |\alpha_1(\mathbf{v_i})|^2 + l_2(\mathbf{v_i})^{-1} |\alpha_2(\mathbf{v_i})|^2 \right)}{N}.$$

Proposition

A ***two-term GCV estimate*** $\hat{\lambda}_2$ *for the* ***tuning parameter*** λ *is given by*

$$\hat{\lambda}_2 = \underset{\lambda}{\operatorname{argmin}} \left(\tilde{g}(\lambda) \right),$$

where $\tilde{g}(\lambda) = \dfrac{l_1(\mathbf{y})^{-2} |\alpha_1(\mathbf{y})|^2 + l_2(\mathbf{y})^{-2} |\alpha_2(\mathbf{y})|^2}{\mathcal{T}_2^2}$ *and* $|\alpha_{1,2}(\mathbf{y})|^2$, $l_{1,2}(\mathbf{y})$, $\mathcal{T}_2$ *are given by the above formulae.*

Penalty functions

We consider the **penalized least squares problem** of the form

$$\frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})^T(\mathbf{y} - X\boldsymbol{\beta}) + n\sum_{j=1}^{d} p_\lambda(|\beta_j|),$$

where

- $p_\lambda(.)$ is a penalty function
- λ is an unknown tuning parameter.

A **good penalty function** should result in an estimator which satisfies three properties:

- **Unbiasedness:** produce nearly unbiased estimates for large coefficients
- **Sparsity:** many estimated coefficients are zero
- **Continuity:** for stability of model selection

Penalty functions (cont.)

The most common **penalties** are

- the L_1 penalty

$$p_\lambda(|\beta|) = \lambda|\beta|,$$

  which results in the Least Absolute Shrinkage and Selection Operator method (**LASSO**)

- the hard thresholding penalty (**Hard**)

$$p_\lambda(|\beta|) = \lambda^2 - (|\beta| - \lambda)^2 I(|\beta| < \lambda),$$

  where $I(\cdot)$ is an indicator function

- the Smoothly Clipped Absolute Deviation penalty (**SCAD**)

$$p'_\lambda(\beta) = \lambda\Big\{I(\beta \leq \lambda) + \frac{(\alpha\lambda - \beta)_+}{(\alpha - 1)\lambda} I(\beta > \lambda)\Big\},$$

  for some $\alpha > 2$ and $\beta > 0$, with $p_\lambda(0) = 0$.

Stepwise variable selection

Stepwise variable selection is a method that allows dropping or adding variables.

- **Forward stepwise selection:**
  - We start with no variables in the model.
  - We proceed introducing other variables or omitting existing ones according to the F-statistic for testing the significance of the corresponding regression coefficients.

    (calculated value > F-to-enter)

- **Backward stepwise selection:**
  - We start with all the independent variables in the model.
  - We proceed by omitting variables according to the F-statistic for testing the significance of the corresponding regression coefficients.

    (calculated value < F-to-remove)

The numerical method

- Define a **grid of values** for λ.
- Set the **level of significance**, f-to-enter (f_{in}) and f-to-remove (f_{out}), in the stepwise variable selection procedure for choosing an initial value for the penalized least squares method with SCAD, LASSO and Hard penalties.
- Over the defined grid of λ, **minimize**:
  - the one-term estimate for the GCV function, i.e. $g_{\nu_1,\nu_2}(\lambda)$,
  - the two-term estimate for the GCV function, i.e. $\tilde{g}(\lambda)$.

  The minimizers of these estimates are the one- and two-term GCV estimates for λ, respectively.

The numerical method (cont.)

- For each used penalty function and for the derived value of λ, compute the solution of the penalized least squares problem, which is

$$\boldsymbol{\beta}^{(1)} = [X^T X + n \sum_{\lambda} (\boldsymbol{\beta}^{(0)})]^{-1} X^T \mathbf{y},$$

  where $\boldsymbol{\beta}^{(0)}$ is an initial value with $\sum_{\lambda}(\boldsymbol{\beta}^{(0)}) = \mathrm{diag}\{p'_{\lambda}(|\beta_1^{(0)}|)/|\beta_1^{(0)}|, \ldots, p'_{\lambda}(|\beta_{d^*}^{(0)}|)/|\beta_{d^*}^{(0)}|\}$ and d^* is the number of statistical significant factors, for the other factors the corresponding penalty is 0.

Simulation study

We **compare the behaviour** of:

- one-term GCV estimates for $\nu_1 = -3/2$ and $\nu_2 = -1$
  $\hookrightarrow$ gcv-L1
- two-term GCV estimates
  $\hookrightarrow$ gcv-L2
- evaluation and minimization of the exact GCV function

J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assos., 96, 1348-1360 (2001).

- error estimates for linear systems
  $\hookrightarrow \eta_\nu(\lambda)$

E. Androulakis, C. Koukouvinos, K. Mylona, Tuning Parameter Estimation in Penalized Least Squares Methodology, Communications in Statistics - Simulation and Computation, 40:9, 1444-1457 (2011).

Simulation study (cont.)

- **1000 linear models** $\mathbf{y} = X\beta + \varepsilon$, where
  - random **error** $\varepsilon_i \sim N(0,1)$, $i = 0, 1, \ldots, n$ considered for each observation y_i,
  - the **coefficients** are randomly selected from -5 to 5.
    If a generated coefficient was "almost zero", it was replaced by a 50% of the maximum coefficient.
- **Grid of λ values:** 1000 linearly spaced points from 10^{-12} to 10
- Required **parameter** for **SCAD** penalty function: $\alpha = 3.7$

Error rates:

- **Type I**: the cost of declaring an inert effect to be active
- **Type II**: the cost of declaring an active effect to be inactive

Simulation results

Test designs:

1. OA(9,4,3,2) $\hookrightarrow 9 \times 8$
   three level orthogonal array of strength two
2. Dataset 1 $\hookrightarrow 100 \times 10$
   the first six columns of which are standard normal and the last four are independently identically distributed as a Bernoulli distribution with probability of success 0.5.

Level of significance:

- $f_{in} = f_{out} = 0.1$
- $f_{in} = f_{out} = 0.05$
- $f_{in} = 0.05$ and $f_{out} = 0.1$

Simulation results (cont.)

1. OA(9,4,3,2)

- **Level of significance:** $f_{in} = f_{out} = 0.1$

q	SCAD(GCV)		SCAD(gcv-L1)		SCAD(gcv-L2)		SCAD($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0788	0.0210	0.0717	0.0220	0.0765	0.0300	0.0454	0.0240
2	0.0564	0.0215	0.0521	0.0215	0.0561	0.0340	0.0269	0.0630
3	0.0310	0.0217	0.0288	0.0220	0.0310	0.0217	0.0168	0.0597

q	LASSO(GCV)		LASSO(gcv-L1)		LASSO(gcv-L2)		LASSO($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.1346	0.0100	0.1153	0.0130	0.1293	0.0270	0.0800	0.0140
2	0.0946	0.0105	0.0807	0.0120	0.0924	0.0205	0.0610	0.0395
3	0.0565	0.0147	0.0510	0.0157	0.0565	0.0153	0.0450	0.0287

q	Hard(GCV)		Hard(gcv-L1)		Hard(gcv-L2)		Hard($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0788	0.0210	0.0779	0.0210	0.0783	0.0250	0.0698	0.1860
2	0.0564	0.0215	0.0561	0.0215	0.0564	0.0245	0.0501	0.1080
3	0.0310	0.0217	0.0310	0.0217	0.0310	0.0217	0.0252	0.0907

Simulation results (cont.)

- **Level of significance:** $f_{in} = f_{out} = 0.05$

q	SCAD(GCV)		SCAD(gcv-L1)		SCAD(gcv-L2)		SCAD($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0432	0.0290	0.0423	0.0290	0.0396	0.0370	0.0236	0.0350
2	0.0413	0.0265	0.0397	0.0265	0.0407	0.0540	0.0231	0.0620
3	0.0322	0.0407	0.0312	0.0410	0.0320	0.0440	0.0182	0.0763

q	LASSO(GCV)		LASSO(gcv-L1)		LASSO(gcv-L2)		LASSO($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0659	0.0170	0.0643	0.0170	0.0569	0.0320	0.0410	0.0230
2	0.0644	0.0165	0.0614	0.0170	0.0636	0.0490	0.0464	0.0400
3	0.0502	0.0293	0.0480	0.0297	0.0500	0.0317	0.0425	0.0400

q	Hard(GCV)		Hard(gcv-L1)		Hard(gcv-L2)		Hard($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0432	0.0290	0.0432	0.0290	0.0419	0.0340	0.0366	0.2640
2	0.0413	0.0265	0.0410	0.0265	0.0413	0.0280	0.0351	0.1520
3	0.0322	0.0407	0.0322	0.0407	0.0322	0.0417	0.0247	0.1213

Simulation results (cont.)

- **Level of significance:** $f_{in} = 0.05$ and $f_{out} = 0.1$

q	SCAD(GCV)		SCAD(gcv-L1)		SCAD(gcv-L2)		SCAD($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0525	0.0210	0.0488	0.0210	0.0500	0.0320	0.0280	0.0240
2	0.0557	0.0240	0.0514	0.0240	0.0554	0.0370	0.0263	0.0650
3	0.0333	0.0360	0.0323	0.0363	0.0332	0.0367	0.0190	0.0737

q	LASSO(GCV)		LASSO(gcv-L1)		LASSO(gcv-L2)		LASSO($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0759	0.0140	0.0683	0.0150	0.0703	0.0190	0.0490	0.0170
2	0.0921	0.0175	0.0784	0.0180	0.0903	0.0270	0.0601	0.0410
3	0.0560	0.0233	0.0535	0.0253	0.0557	0.0240	0.0442	0.0373

q	Hard(GCV)		Hard(gcv-L1)		Hard(gcv-L2)		Hard($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0525	0.0210	0.0521	0.0210	0.0511	0.0240	0.0464	0.2310
2	0.0557	0.0240	0.0554	0.0240	0.0557	0.0270	0.0494	0.1115
3	0.0333	0.0360	0.0333	0.0360	0.0333	0.0360	0.0257	0.1003

Simulation results (cont.)

2. Dataset 1

- **Level of significance:** $f_{in} = f_{out} = 0.1$

q	SCAD(GCV)		SCAD(gcv-L1)		SCAD(gcv-L2)		SCAD($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0041	0.0000	0.0041	0.0000	0.0041	0.0020	0.0017	0.0000
2	0.0047	0.0000	0.0038	0.0000	0.0047	0.0000	0.0002	0.0005
3	0.0043	0.0000	0.0036	0.0000	0.0043	0.0000	0.0005	0.0007
4	0.0046	0.0000	0.0034	0.0000	0.0044	0.0000	0.0000	0.0025
5	0.0038	0.0000	0.0015	0.0000	0.0038	0.0000	0.0000	0.0024
6	0.0022	0.0000	0.0010	0.0000	0.0022	0.0000	0.0000	0.0033
7	0.0053	0.0000	0.0027	0.0000	0.0053	0.0000	0.0000	0.0034
8	0.0047	0.0000	0.0010	0.0000	0.0047	0.0000	0.0000	0.0045
9	0.0060	0.0000	0.0010	0.0000	0.0050	0.0000	0.0000	0.0049

Simulation results (cont.)

q	LASSO(GCV)		LASSO(gcv-L1)		LASSO(gcv-L2)		LASSO($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0175	0.0000	0.0135	0.0000	0.0168	0.0010	0.0000	0.0000
2	0.0196	0.0000	0.0142	0.0000	0.0189	0.0000	0.0000	0.0000
3	0.0174	0.0000	0.0094	0.0000	0.0160	0.0000	0.0000	0.0000
4	0.0184	0.0000	0.0100	0.0000	0.0170	0.0000	0.0000	0.0003
5	0.0168	0.0000	0.0057	0.0000	0.0137	0.0000	0.0000	0.0004
6	0.0142	0.0000	0.0034	0.0000	0.0118	0.0000	0.0000	0.0000
7	0.0160	0.0000	0.0047	0.0000	0.0123	0.0000	0.0000	0.0004
8	0.0150	0.0000	0.0047	0.0000	0.0110	0.0000	0.0007	0.0001
9	0.0155	0.0000	0.0050	0.0003	0.0115	0.0000	0.0000	0.0002

q	Hard(GCV)		Hard(gcv-L1)		Hard(gcv-L2)		Hard($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0041	0.0000	0.0041	0.0000	0.0041	0.0010	0.0041	0.0080
2	0.0047	0.0000	0.0047	0.0000	0.0047	0.0000	0.0047	0.0000
3	0.0043	0.0000	0.0043	0.0000	0.0043	0.0000	0.0043	0.0000
4	0.0046	0.0000	0.0046	0.0000	0.0046	0.0000	0.0046	0.0003
5	0.0038	0.0000	0.0037	0.0000	0.0038	0.0000	0.0038	0.0004
6	0.0022	0.0000	0.0022	0.0000	0.0022	0.0000	0.0022	0.0000
7	0.0053	0.0000	0.0047	0.0000	0.0053	0.0000	0.0053	0.0000
8	0.0047	0.0000	0.0037	0.0000	0.0047	0.0000	0.0047	0.0005
9	0.0060	0.0000	0.0060	0.0000	0.0060	0.0000	0.0060	0.0006

Simulation results (cont.)

- **Level of significance:** $f_{in} = f_{out} = 0.05$

q	SCAD(GCV)		SCAD(gcv-L1)		SCAD(gcv-L2)		SCAD($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0052	0.0000	0.0048	0.0000	0.0052	0.0000	0.0023	0.0000
2	0.0037	0.0000	0.0036	0.0000	0.0037	0.0010	0.0004	0.0005
3	0.0045	0.0000	0.0043	0.0000	0.0045	0.0000	0.0003	0.0007
4	0.0031	0.0000	0.0020	0.0000	0.0031	0.0000	0.0000	0.0025
5	0.0033	0.0000	0.0015	0.0000	0.0033	0.0000	0.0000	0.0024
6	0.0038	0.0000	0.0020	0.0000	0.0038	0.0000	0.0000	0.0028
7	0.0045	0.0000	0.0018	0.0000	0.0040	0.0000	0.0000	0.0046
8	0.0033	0.0000	0.0017	0.0000	0.0030	0.0000	0.0000	0.0034
9	0.0060	0.0000	0.0010	0.0000	0.0050	0.0000	0.0000	0.0049

Simulation results (cont.)

q	LASSO(GCV)		LASSO(gcv-L1)		LASSO(gcv-L2)		LASSO($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0181	0.0000	0.0164	0.0000	0.0180	0.0010	0.0000	0.0000
2	0.0161	0.0000	0.0129	0.0000	0.0158	0.0010	0.0000	0.0000
3	0.0170	0.0000	0.0121	0.0000	0.0163	0.0000	0.0000	0.0003
4	0.0146	0.0000	0.0076	0.0000	0.0131	0.0000	0.0000	0.0000
5	0.0158	0.0000	0.0060	0.0000	0.0137	0.0000	0.0000	0.0004
6	0.0154	0.0000	0.0052	0.0000	0.0130	0.0000	0.0000	0.0000
7	0.0190	0.0000	0.0055	0.0000	0.0150	0.0000	0.0000	0.0001
8	0.0157	0.0000	0.0030	0.0001	0.0113	0.0000	0.0003	0.0000
9	0.0155	0.0000	0.0050	0.0004	0.0115	0.0000	0.0000	0.0002

q	Hard(GCV)		Hard(gcv-L1)		Hard(gcv-L2)		Hard($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0052	0.0000	0.0052	0.0000	0.0052	0.0000	0.0052	0.0080
2	0.0037	0.0000	0.0037	0.0000	0.0037	0.0000	0.0037	0.0000
3	0.0045	0.0000	0.0045	0.0000	0.0045	0.0000	0.0045	0.0000
4	0.0031	0.0000	0.0031	0.0000	0.0031	0.0000	0.0031	0.0003
5	0.0033	0.0000	0.0032	0.0000	0.0033	0.0000	0.0033	0.0004
6	0.0038	0.0000	0.0038	0.0000	0.0038	0.0000	0.0038	0.0000
7	0.0045	0.0000	0.0040	0.0000	0.0045	0.0000	0.0045	0.0000
8	0.0033	0.0000	0.0027	0.0000	0.0033	0.0000	0.0033	0.0003
9	0.0060	0.0000	0.0060	0.0000	0.0060	0.0000	0.0060	0.0006

Simulation results (cont.)

- **Level of significance:** $f_{in} = 0.1$ and $f_{out} = 0.05$

q	SCAD(GCV)		SCAD(gcv-L1)		SCAD(gcv-L2)		SCAD($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0045	0.0000	0.0043	0.0000	0.0045	0.0020	0.0017	0.0000
2	0.0046	0.0000	0.0039	0.0000	0.0046	0.0000	0.0007	0.0000
3	0.0043	0.0000	0.0036	0.0000	0.0043	0.0000	0.0005	0.0007
4	0.0046	0.0000	0.0023	0.0000	0.0041	0.0000	0.0000	0.0018
5	0.0038	0.0000	0.0015	0.0000	0.0038	0.0000	0.0000	0.0024
6	0.0036	0.0000	0.0020	0.0000	0.0034	0.0000	0.0000	0.0028
7	0.0045	0.0000	0.0018	0.0000	0.0045	0.0000	0.0000	0.0037
8	0.0047	0.0000	0.0010	0.0000	0.0047	0.0000	0.0000	0.0045
9	0.0060	0.0000	0.0010	0.0000	0.0050	0.0000	0.0000	0.0049

Simulation results (cont.)

q	LASSO(GCV)		LASSO(gcv-L1)		LASSO(gcv-L2)		LASSO($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0192	0.0000	0.0149	0.0000	0.0186	0.0040	0.0000	0.0000
2	0.0200	0.0000	0.0156	0.0000	0.0192	0.0000	0.0000	0.0000
3	0.0174	0.0000	0.0094	0.0000	0.0160	0.0000	0.0000	0.0000
4	0.0181	0.0000	0.0091	0.0000	0.0167	0.0000	0.0000	0.0000
5	0.0168	0.0000	0.0057	0.0000	0.0137	0.0000	0.0000	0.0004
6	0.0164	0.0000	0.0054	0.0000	0.0134	0.0000	0.0000	0.0000
7	0.0170	0.0000	0.0065	0.0000	0.0135	0.0000	0.0005	0.0000
8	0.0150	0.0000	0.0047	0.0000	0.0110	0.0000	0.0007	0.0001
9	0.0155	0.0000	0.0050	0.0003	0.0115	0.0000	0.0000	0.0002
q	Hard(GCV)		Hard(gcv-L1)		Hard(gcv-L2)		Hard($\eta_\nu(\lambda)$)	
Factors	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
1	0.0045	0.0000	0.0045	0.0000	0.0045	0.0010	0.0045	0.0080
2	0.0046	0.0000	0.0046	0.0000	0.0046	0.0000	0.0046	0.0015
3	0.0043	0.0000	0.0043	0.0000	0.0043	0.0000	0.0043	0.0000
4	0.0046	0.0000	0.0046	0.0000	0.0046	0.0000	0.0046	0.0000
5	0.0038	0.0000	0.0037	0.0000	0.0038	0.0000	0.0038	0.0004
6	0.0036	0.0000	0.0036	0.0000	0.0036	0.0000	0.0036	0.0000
7	0.0045	0.0000	0.0045	0.0000	0.0045	0.0000	0.0045	0.0001
8	0.0047	0.0000	0.0037	0.0000	0.0047	0.0000	0.0047	0.0005
9	0.0060	0.0000	0.0060	0.0000	0.0060	0.0000	0.0060	0.0006

Conclusions and future work

- Derivation of GCV estimates for the tuning parameter based on an extrapolation procedure.
- Comparison between the behaviour of the proposed method with other methods concerning the Type I and Type II error rates.

Future work:

- Employment of Gaussian quadrature techniques for linear regression modelling.

References

- A. Antoniadis, Wavelets in statistics: a review (with discussion), J. Italian Statist. Assos. 6, 97-144 (1997).
- G.H. Golub, M. Heath, G. Wahba, Generalized Cross-Validation as a method for choosing a good ridge parameter, Technometrics 21, 215-223 (1979).
- G.H. Golub, G. Meurant, Matrices, Moments and Quadrature with Applications. Princeton University Press, Princeton (2010).
- C. Koukouvinos, A. Lappa, P. R., Numerical methods for estimating the tuning parameter in penalized least squares problems, submitted.
- R.J. Tibshirani, Regression shrinkage and selection via the LASSO, J. Roy. Statist. Soc. Ser. 58, 267-288 (1996).

Thank you!