Testing Shape Restrictions in LASSO Regression

Matúš Maciak
Joint work with Ivan Mizera

ROBUST 2016
Jeseníky

Motivation: Shape Constraints Workshop in Leiden, 2015

Motivation: ROBUST 2014, Jetřichovice

- **Change-points**
  - relaxing some shape/smoothness/continuity assumptions;
  - additional modeling flexibility in assumed models;
- **Shape-constraints**
  - limiting the overall flexibility in assumed models;
  - posing additional shape constrains/restrictions;

Model: A General Setup

- random sample $\{(Y_i, X_i);\ i = 1, \dots, N\}$ from population $(\mathscr{X}, \mathscr{Y})$;
- we consider a standard regression problem formulated as

$$Y_i = m(X_i) + \varepsilon_i, \quad \text{for } i = 1, \dots, N,$$

  for independent random error terms $\varepsilon_i \sim N(0, \sigma^2)$ and some $\sigma^2 > 0$;

- common problem considered under various sets of conditions by many authors from different statistical perspectives;
- primary focus on estimation & statistical inference in the model;

- only very mild assumptions for the unknown functional dependence
    - no a-priori parametric shape restrictions;
    - no strict continuity or smoothness properties required;
- the unknown functional dependence structure decomposition

$$m(x) = m_0(x) + \sum_{j=0}^{p-1} s_j(x), \qquad x \in \mathscr{D} \subset \mathbb{R};$$

- for a reasonably smooth function m_0 (of the order $p \in \mathbb{N}$) and some background shock processes $s_0, \dots, s_{p-1}$;

- random sample $\{(X_i, Y_i);\ i = 1, \dots, N\}$, where $X_i < X_{i+1}$;
- the underlying model structure

$$Y_i = a_i + b_i X_i + \varepsilon_i, \quad a_i, b_i \in \mathbb{R};$$

- under the continuity condition $a_i + b_i X_i = a_{i+1} + b_{i+1} X_i$;
- sparsity in b_i's as $b_i \neq b_{i+1}$ only for some few indexes;
- optionally under some shape constraints (e.g. monotonisity);
- motivated by the paper of Harchaoui and Lévy-Leduc (2010);
- the same problem considered by (e.g.) Bosetti et al. (2008); Kim et al. (2009); Qui et al. (2009); Maciak and Mizera (2016), etc.;

Utilization: Step by Step

- estimation of the unknown regression function and its components;
  (change-points estimation - the locations and magnitudes)

- statistical inference about the unknown dependence structure;
  (statistical inference with respect to the estimates of parameters)

- statistical inference about the estimated change-points;
  (significance of the change-points and structural breaks occurrences)

- statistical inference about the assumed shape constraints;
  (specifically monotone and isotonic properties are interesting)

Estimation: LASSO Regularized Shape Constrained Fit

- the estimate for the model can be obtained by solving

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^N}{Argmin} \quad \|\boldsymbol{Y} - \mathbb{X}_N \boldsymbol{\beta}\|_2^2 + \lambda_N \left\|\boldsymbol{\beta}_{(-2)}\right\|_1,$$

- wrt. to additional constraints, e.g. $A\boldsymbol{\beta}_{(-1)} \succeq 0$ (non-decreasing);
- for the unknown parameters $\boldsymbol{\beta} = (\beta_0, \underbrace{\beta_1, \boldsymbol{\beta}_{(-2)}^\top}_{\boldsymbol{\beta}_{(-1)}})^\top \in \mathbb{R}^n$ and

$$\mathbb{X}_N = \begin{pmatrix} 1 & X_1 & 0 & \dots & 0 \\ 1 & X_2 & 0 & \dots & 0 \\ 1 & X_3 & (X_3 - X_2) & 0 & 0 \\ \vdots & \dots & \dots & \vdots & 0 \\ 1 & X_N & (X_3 - X_2) & \dots & (X_N - X_{N-1}) \end{pmatrix} \qquad \boldsymbol{A} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots \\ \vdots & \dots & \dots & \vdots \\ 1 & \dots & 1 & 0 \\ 1 & \dots & \dots & 1 \end{pmatrix}$$

LASSO: Various Regularization Approaches

- **classical LASSO** minimization problem for $N, p \in \mathbb{N}$ and $\lambda > 0$ (Tibshirani, 1996)

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^p}{Minimize} \quad \|Y - \mathbb{X}_N \boldsymbol{\beta}\|_2^2 + \lambda_N \|\boldsymbol{\beta}\|_1$$

- **generalized LASSO** for $N, p \in \mathbb{N}$ and $\lambda > 0$ (Tibshirani and Taylor, 2011)

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^p}{Minimize} \quad \|Y - \mathbb{X}_N \boldsymbol{\beta}\|_2^2 + \lambda_N \|\boldsymbol{A}\boldsymbol{\beta}\|_1$$

- **constrained LASSO** for $N, p \in \mathbb{N}$ and $\lambda > 0$ (James, Paulson and Rusmevichientong, 2012)

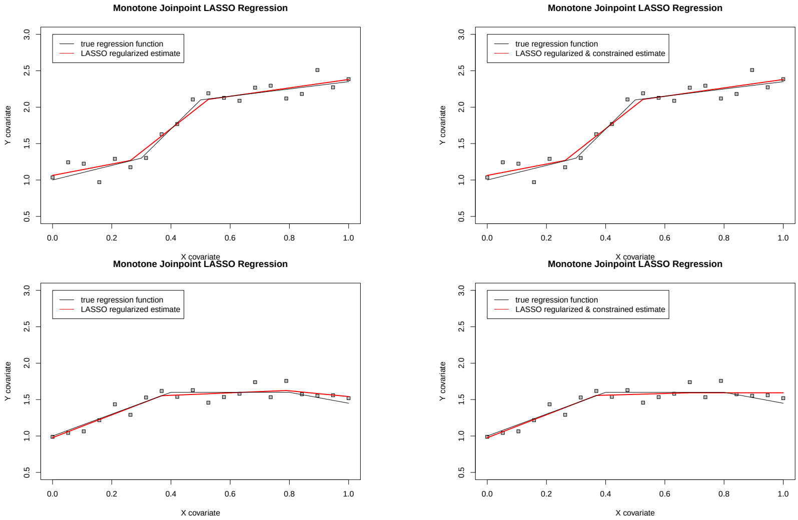
$$\underset{\boldsymbol{\beta} \in \mathbb{R}^p}{Minimize} \quad \|Y - \mathbb{X}_N \boldsymbol{\beta}\|_2^2 + \lambda_N \|\boldsymbol{\beta}\|_1 \quad \text{subject to } \boldsymbol{A}\boldsymbol{\beta} \succeq \boldsymbol{\xi}$$

Problems: Standard Estimation Approaches Not Applicable

- for classical LASSO problem $\Rightarrow$ LARS-LASSO algorithm;
  (this is however, not applicable under any additional constraints)
- no straightforward generalization of the LARS algorithm either;
  (only degenerate solutions at the boundaries with no interpretation)
- same reasoning applies for the coordinate descent algorithm;
  (Friedman, Hastie, and Tibshirani, 2010)
- the full solution paths can not be easily obtained as well;
  (iterative procedures for different values of the $\lambda > 0$ parameter)
- however, still a **CONVEX PROBLEM** which can be solved;
  (Mosek optimization toolbox and the R package `Rmosek`)

Example: Constrained vs. Unconstrained Fit

Example: Constrained vs. Unconstrained Fit



- once the vector of parameter estimates $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\boldsymbol{\beta}}_{(-2)}^\top)^\top \in \mathbb{R}^N$ is obtained we can try to construct a test for the set of hypothesis

$$H_0: \; A\boldsymbol{\beta}_{(-1)} \succeq 0$$
$$H_1: \; \neg H_0$$

BUT ...

- where $\boldsymbol{\beta}_{(-1)} = (\beta_1, \boldsymbol{\beta}_{(-2)}^\top)^\top \in \mathbb{R}^{N-1}$;
- $\hat{\beta}_1$ is a kind of LS estimate;
- $\hat{\boldsymbol{\beta}}_{(-2)}$ is a LASSO shrunk estimate;

- many different proposals in the area of post selection inference;
- some are quite intuitive some are not;

HOWEVER

- the main principle is behind the fact that LASSO regularized parameters enter the model at random;
- standard regression inference methods (e.g. comparing two nested models) do not take this fact into account;
- standard approaches are too much liberal causing the I type error much larger than the required nominal level;

- consider some test statistic T_k based on the LASSO regularized parameter selection at some step $k \in \mathbb{N}$.
- for the given null hypothesis the conditioning on the selection would take the form $P_{H_0}\left(T_k \leq x | \text{well defined LASSO history}\right)$ which corrects for too much liberal performance of classical approaches;

- **Polyhedral Lemma**

  (Lee et al., 2016; Tibshirani et. al, 2016)

  Conditioning on the LASSO selection history can be equivalently expressed using some additional linear constraints in a form of a polyhedra $\{\boldsymbol{y} \in \mathbb{R}^n;\ \boldsymbol{B}\boldsymbol{y} \geq 0\}$;

Test: Verifying Shape Restrictions in the Model

- using the LASSO selection history and conditioning with respect to the polyheral set defined before...
- considering a general null hypothesis $H_0 : \boldsymbol{v}^\top \boldsymbol{\beta}_{(-2)} \geq 0$ against the alternative $H_1 : \boldsymbol{v}^\top \boldsymbol{\beta}_{(-2)} < 0$ for some $\boldsymbol{v} \in \mathbb{R}^{n-2}$...
- we can define a test statistic T_k as a normally distributed $N(\mu, \sigma^2)$ random variable truncated to some interval $[a, b] \subset \mathbb{R}$...
- where the interval $[a, b]$ is fully defined by the history conditioning, respectively using the polyheral set $\{\boldsymbol{y} \in \mathbb{R}^n;\ \boldsymbol{B}\boldsymbol{y} \geq 0\}$;
- finally, it can be shown that the distribution of the test statistic T_k under the null hypothesis H_0 is (exactly) uniform, which means

$$P_{H_0}\left(T_k \leq \alpha | \boldsymbol{B}\boldsymbol{y} \geq 0\right) = \alpha \in (0, 1)$$

Correction: Holm-Bonferroni Testing Approach

- a simple test uniformly more powerful than the Bonferroni correction;
- suitable for tests which are not independent or positively dependent;

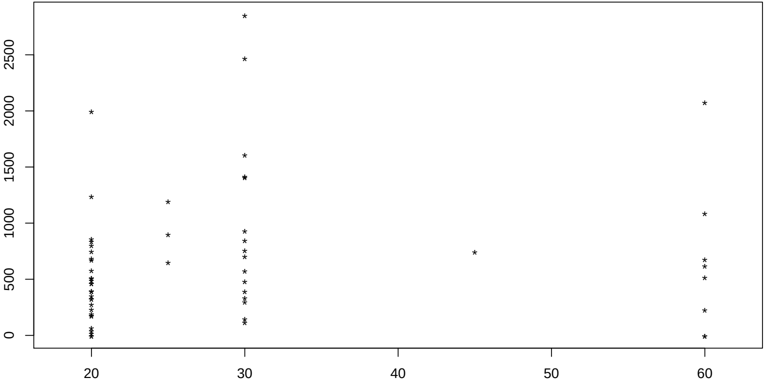
- for a set of null hypothesis $H_1, \dots, H_k$ and the corresponding p-values $p_1, \dots, p_k$ we order the hypothesis with respect to increasing p-values $p_{(1)} \leq \dots \leq p_{(k)}$;
- for a given $\alpha \in (0,1)$ find the smallest $j \in \{1, \dots, k\}$ such that

$$p_{(j)} > \frac{\alpha}{k + 1 - j}$$

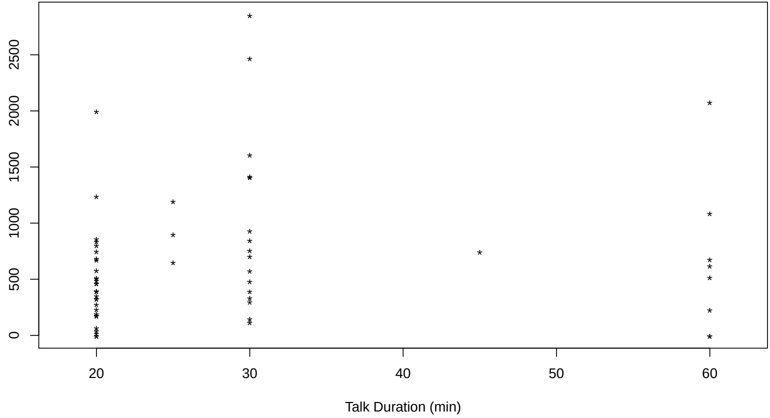
- reject the null hypothesis $H_1, \dots, H_{j-1}$ and do not reject $H_j, \dots, H_k$;
- if $k = 1 \Longrightarrow$ do not reject any of the null hypothesis $H_1, \dots, H_k$;

- an analogous approach can be even more straightforwardly applied to testing significance of change-points occurring in the model.
- more easier scenario as one just directly assumes some null hypothesis $H_0 : \beta_{j_l} = 0$ against some general alternative $H_1 : \beta_{j_l} \neq= 0$;
- having the LASSO estimate $\hat{\beta}_{j_l}(\lambda(n))$ one can directly apply the approach based on the polyhedral lemma to make conclusions;
- of a p-value smaller than the given critical value we reject the null hypothesis $\Rightarrow$ significant occurrence of the change-point;

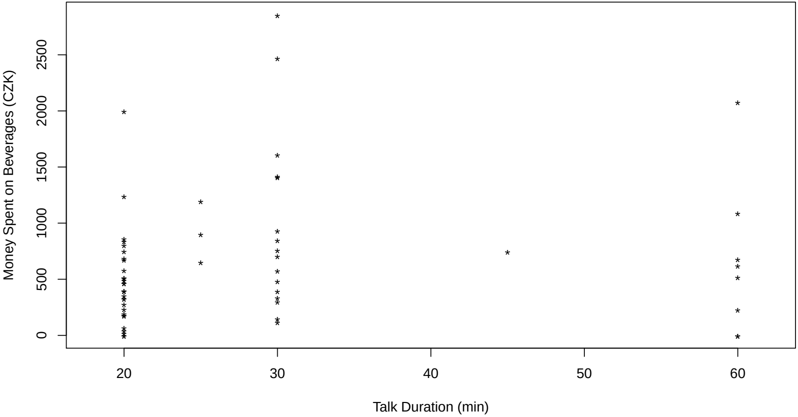
Example: One from the past...

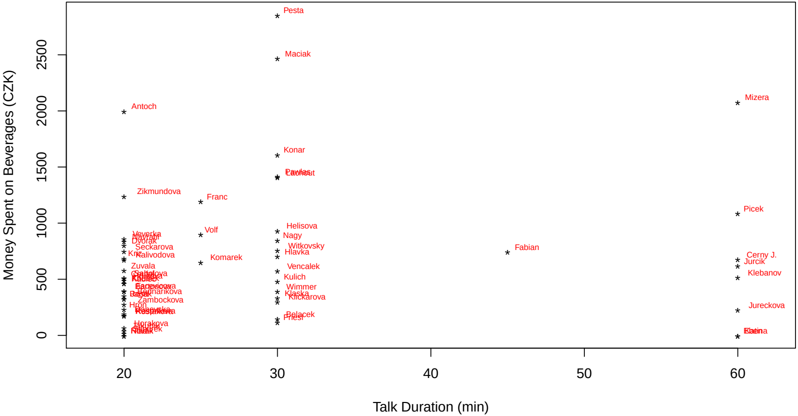


Example: One from the past...

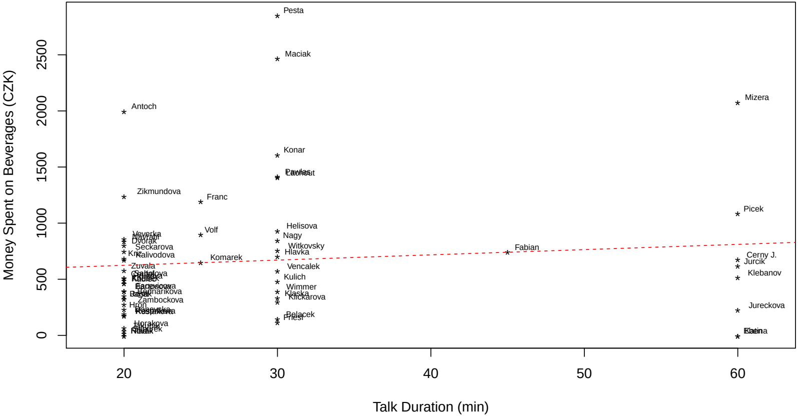


Example: One from the past...

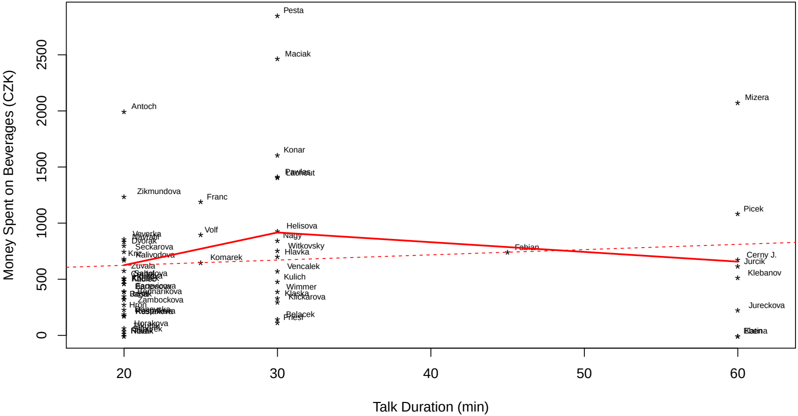



Money Spent on Beverages (CZK)
Talk Duration (min)
0
500
1000
1500
2000
2500
20
30
40
50
60
Pesta
Maciak
Mizera
Antoch
Konar
Zikmundova
Franc
Picek
Helisova
Volf
Nagy
Veverka
Dvorak
Seckarova
Witkovsky
Hlavka
Kalivodova
Fabian
Komarek
Cerny J.
Jurcik
Zuvala
Vencalek
Klebanov
Kulich
Wimmer
Klickarova
Zambockova
Hron
Jureckova
Belacek
Horakova

Example: Thank you all for contributing!



Example: Any conclusion after all?



Thank you

Any Questions?

Thank you

Any Questions?

✍ Lockhart, M., Taylor, J., Tibshirani, R., and Tibshirani. R. (2014). A Significance Test on LASSO. *Annals of Statistics*, Vol.42, 413–468.

✍ Tibhisrani, R., Taylor, J., Lockhart, R., and Tibshirani, R. (2015). Exact Post-Selection Inference for Sequential Regression Problems. JASA (to appear).

✍ Lee, J, Sun, D, Su, Y., and Taylor, J. (2016). Exaxt Post-selection Inference with Application to the LASSO. *Annals of Statistics*,Vol. 44, No. 3, 907–927.

✍ Gareth, J., Paulson, C., and Rusmevichientong, P. (2012). The Constrained LASSO.

✍ Maciak, M. and MIzera, I. (2016). Regularization Techniques in Change-point Estimation, *Statistical Papers*, (accepted).

✍ Maciak, M. (2016). Testing Shape Constrains in Lasso Regularized Joinpoint Regression, *AMISTAT 2016 Springer Proceedings*, (accepted).