City Research Online

City, University of London Institutional Repository

This is the published version of the paper.

This version of the publication may differ from the final published version.

Department of Economics

Time Varying Quantile Lasso

Lenka Zboňáková[1]
Humboldt-Universität zu Berlin

Wolfgang Karl Härdle
Humboldt-Universität zu Berlin

Weining Wang
City, University of London

[1] Corresponding author: Lenka Zboňáková, C.A.S.E – Center for Applied Statistics & Economics, Humboldt-Universität zu Berlin, 10178 Berlin, Germany.
Email: zbonakle@hu-berlin.de

Time Varying Quantile Lasso[*]

Lenka Zboňáková[†], Wolfgang Karl Härdle [‡] and Weining Wang [§]

November 03, 2016

Abstract

In the present paper we study the dynamics of penalization parameter λ of the least absolute shrinkage and selection operator (Lasso) method proposed by Tibshirani (1996) and extended into quantile regression context by Li and Zhu (2008). The dynamic behaviour of the parameter λ can be observed when the model is assumed to vary over time and therefore the fitting is performed with the use of moving windows. The proposal of investigating time series of λ and its dependency on model characteristics was brought into focus by Härdle et al. (2016), which was a foundation of FinancialRiskMeter (`http://frm.wiwi.hu-berlin.de`). Following the ideas behind the two aforementioned projects, we use the derivation of the formula for the penalization parameter λ as a result of the optimization problem. This reveals three possible effects driving λ; variance of the error term, correlation structure of the covariates and number of nonzero coefficients of the model. Our aim is to disentangle these three effect and investigate their relationship with the tuning parameter λ, which is conducted by a simulation study. After dealing with the theoretical impact of the three model characteristics on λ, empirical application is performed and the idea of implementing the parameter λ into a systemic risk measure is presented. The codes used to obtain the results included in this work are available on `http://quantlet.de/d3/ia/`.

1   Introduction

The least absolute shrinkage and selection operator (Lasso) method as proposed by Tibshirani (1996) has been widely used and extended during recent years. The literature presents a method which simultaneously completes the task of model selection and parameter estimation, while studying its consistency. A key factor for the estimation precision is choosing a tuning parameter which controls the degree of penalization. Although there is much literature on Lasso, including a time series context, the time variation of the tuning parameter remains unexplored.

Here we explain dynamics of the penalization parameter λ and how it can be used in financial practice, particularly when dealing with systemic risk. Let us assume for the moment a linear model with a vector of responses $Y = (Y_1, Y_2, \ldots, Y_n)^\top$, a vector of parameters $\beta = (\beta_1, \ldots, \beta_p)^\top$, an $(n \times p)$ design matrix X, which might be either fixed or random, and a vector of independent identically distributed errors ε with zero mean and variance σ^2. Then the objective function of Lasso is

$$\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{n} \left(Y_i - X_i^\top \beta \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{1}$$

with tuning parameter $\lambda \geq 0$ and X_i, $0 \leq i \leq n$, denoting row vectors of X. In (1) one assumes that the columns of the matrix $X = (x_{ij})_{i=1,\ldots,n,j=1,\ldots,p}$ have been standardized, i.e. $n^{-1} \sum_{i=1}^{n} x_{ij} = 0$ and $n^{-1} \sum_{i=1}^{n} x_{ij}^2 = 1$. Solving this type of penalized least squares problem with L_1-penalization allows some of the coefficients of the model to shrink to 0. This is a highly advantageous property when dealing with high-dimensional data, since variable selection and shrinkage of coefficients are performed simultaneously. Shrinking some of the coefficients to exactly 0 also improves the interpretability of the fitted model.

Modification of Lasso in quantile regression (Koenker and Basset (1978)) studied by Li and Zhu (2008) and Belloni and Chernozhukov (2011) solves the optimization problem with

$$\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{n} \rho_\tau \left(Y_i - X_i^\top \beta \right) + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{2}$$

where $\tau \in (0, 1)$ and $\rho_\tau(\cdot)$ is the check function

$$\rho_\tau(x) = \begin{cases} \tau \cdot x & \text{if } x > 0; \\ -(1 - \tau) \cdot x & \text{otherwise.} \end{cases} \tag{3}$$

The Lasso models described above account for independent observations. However, there is much literature on the Lasso in time series context as well. For the univariate case we refer to Wang et al. (2007), Nardi and Rinaldo (2011) and Chen and Chan (2011). The case of multivariate time series, particularly vector autoregression, was covered by e.g. Hsu et al. (2008).

Lasso in quantile regression has been used by Härdle et al. (2016) to model tail event dependencies among U.S. financial companies. Based on the penalization parameters the FinancialRiskMeter (FRM), `http://frm.wiwi.hu-berlin.de`, was developed, see Figure 1.

The value of the averaged penalization parameter λ was elevated during the financial crises. This fact led us to the question we indicated above; what drives the penalization parameter λ and what are the dynamics of λ? We investigate this by simulation study and empirical application.
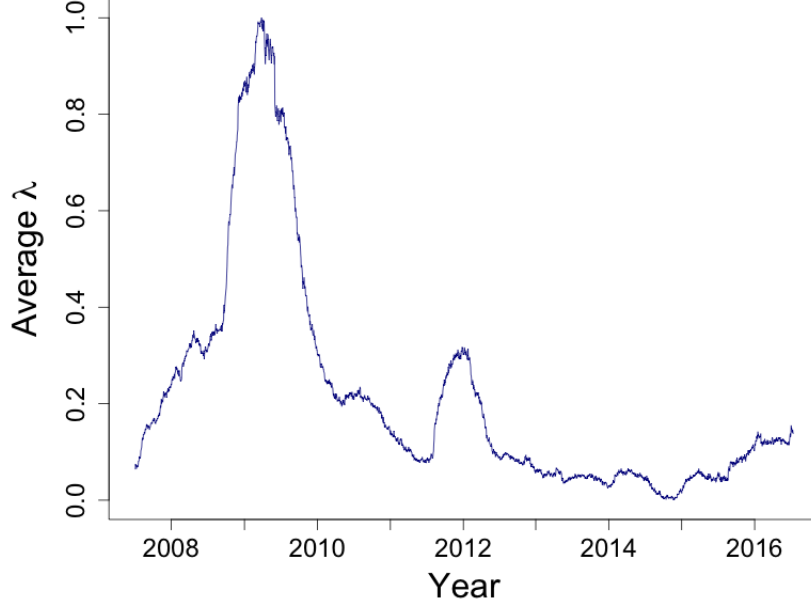
Figure 1: Time series of λ taken from FinancialRiskMeter (`http://frm.wiwi.hu-berlin.de`), normalized to interval (0,1).

The computations included in this paper were performed in the environment of R software developed by R Core Team (2014) and the codes are available on `http://quantlet.de/d3/ia/`.

2 Lasso method

2.1 Lasso as an optimization problem

In this section we firstly follow Osborne et al. (2000) to derive formula for the penalization parameter λ of the Lasso method when applied in linear regression problems. Then we aim our focus on the representation of λ in penalized quantile regression.

If we treat λ as a fixed value in the objective function of the penalized regression

$$f(\beta, \lambda) = \left\{ \frac{1}{2} \sum_{i=1}^{n} \left(Y_i - X_i^{\top} \beta \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{4}$$

then the function $f(\beta, \lambda)$ is convex in parameter β. Moreover, with diverging β we observe that $f(\beta, \lambda) \to \infty$. Hence there exists at least one minimum of the function $f(\cdot, \lambda)$. According to Osborne (1985) this minimum is attained in $\widehat{\beta}(\lambda)$ if and only if the null-vector $0 \in \mathbb{R}^p$ is an element of the subdifferential

$$\frac{\partial f(\beta, \lambda)}{\partial \beta} = -X^{\top}(Y - X\beta) + \lambda u(\beta), \tag{5}$$

where $u(\beta) = (u_1(\beta), \ldots, u_p(\beta))^\top$ is defined as $u_j(\beta) = 1$ if $\beta_j > 0$, $u_j(\beta) = -1$ if $\beta_j < 0$ and $u_j(\beta) \in [-1, 1]$ if $\beta_j = 0$. Then, for $\widehat{\beta}(\lambda)$ as a minimizer of $f(\beta, \lambda)$ the following has to be satisfied

$$0 = -X^\top \{Y - X\widehat{\beta}(\lambda)\} + \lambda u(\widehat{\beta}(\lambda)), \tag{6}$$

Here we denote the estimator of a parameter vector β as a function of the penalization parameter λ. This dependency follows from the formulation of the penalized regression method and its objective function (4), where we first select λ and then search for $\widehat{\beta}(\lambda)$ which minimizes (4). Using the fact that $u(\beta)^\top \beta = \sum_{j=1}^p |\beta_j| = ||\beta||_1$, where $||\cdot||_1$ denotes L_1-norm of a p-dimensional vector, (6) can be further rewritten in the formula

$$\lambda = \frac{\{Y - X\widehat{\beta}(\lambda)\}^\top X\widehat{\beta}(\lambda)}{||\widehat{\beta}(\lambda)||_1}. \tag{7}$$

The identity (7) leads us to consider possible constituents which influence the value of parameter λ and therein its dynamics when treated in a time-dependent framework. Here we propose to study three effects which are related to the size of λ:

1. size of residuals of the model;

2. absolute size of the coefficients of the model, $||\beta||_1$;

3. singularity of a matrix $X^\top X$.

The second effect can also be translated into the effect of a number of nonzero parameters the so-called active set of the model, $q = ||\beta||_0 = \sum_{j=1}^p \mathbf{I}(\beta_j \neq 0)$, where $||\cdot||_0$ stands for L_0-norm on $\mathbb{R}^p$ and $\mathbf{I}(\cdot)$ is an indicator function. As a measure of the third structure, the condition number $\kappa(X^\top X)$ defined as the ratio $\phi_{\max}(X^\top X)/\phi_{\min}(X^\top X)$, the maximum and the minimum eigenvalue of the matrix $X^\top X$, can be used.

Similarly, one can derive formulae for the penalization parameter λ in a quantile regression problem (2) and (3). Following Li and Zhu (2008)

$$\lambda = \frac{\theta^\top X\widehat{\beta}(\lambda)}{||\widehat{\beta}(\lambda)||_1}, \tag{8}$$

where $\theta = (\theta_1, \ldots, \theta_n)^\top$ satisfies the following

$$\theta_i = \begin{cases} \tau & \text{if } Y_i - X_i^\top \widehat{\beta}(\lambda) > 0; \\ -(1-\tau) & \text{if } Y_i - X_i^\top \widehat{\beta}(\lambda) < 0; \\ \in (-(1-\tau), \tau) & \text{if } Y_i - X_i^\top \widehat{\beta}(\lambda) = 0. \end{cases} \tag{9}$$

Hence, we observe that λ depends on cardinality of the active set q, which is again influenced by the correlation structure of the design matrix. Direct impact of the variance of residuals disappears and only the sign of the residuals stays in effect. However, when looking at Figure 2 one can see similarities between the time series of λ and historic values of the implied volatility index (VIX) reported by the Chicago Board Options Exchange. This fact leads us to believe that the dynamics of λ is also influenced by the changes in the variance of model residuals.
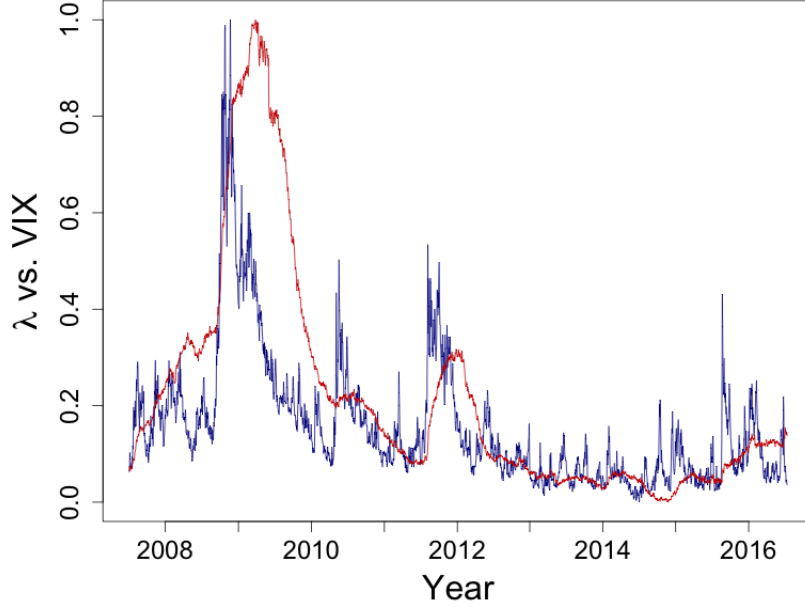
Figure 2: Normalized implied volatility index (blue) and λ from FinancialRiskMeter (red).

XFGTVP_LambdaVIX

2.2 Choosing the penalization parameter

In theory the equalities (7) and (8) hold for every solution of the Lasso optimization problems
(1) and (2) respectively, since first λ is chosen and afterwards the model is fitted according to
the given value of the penalization parameter. One of the commonly used methods of choosing
estimator of λ is cross-validation in its three forms; k-fold, leave-one-out and generalized
cross-validation method, see e.g. Tibshirani (1996). As pointed out in Hastie et al. (2009),
cross-validation is a widely used method for estimation of prediction error. This feature is used
when estimating λ in Lasso method, where, on a grid of penalization parameters λ, the one
which minimizes estimated prediction error is chosen. However, as Leng et al. (2006) argued
in their work, methods of choosing penalization parameter based on prediction accuracy are
in general not consistent when variable selection is considered. The same argument was used
by Wang et al. (2009) where they compared the asymptotic behaviour of the generalized
cross-validation to the one of Akaike's information criterion (AIC); it is efficient if one is
interested in the model error, but inconsistent in selecting the true model.

The second widely used method of estimating λ is the Bayesian information criterion (BIC).
By $\beta_0 = (\beta_{01}, \ldots, \beta_{0p})^\top$ we denote the true vector of coefficients of the regression model and
q_0 defines the number of its nonzero elements, i.e. $\beta_{0j} \neq 0$ for $1 \leq j \leq q_0$ and $\beta_{0j} = 0$ for
$j > q_0$. The permutation of the elements of β_0 is performed without loss of generality, so the
previous notation holds. Secondly, by $\mathcal{S} = \{j_1, \ldots, j_q\}$ we denote an arbitrary model with
$X_{\mathcal{S}} = (X_{j_1}, \ldots, X_{j_q}) \in \mathbb{R}^{n \times q}$ as a design matrix associated with it. Vector of coefficients of
a model $\mathcal{S}$ is $\beta_{\mathcal{S}} = (\beta_{j_1}, \ldots, \beta_{j_q})^\top$ and the model size is $|\mathcal{S}| = q$. The true model is referred
to by $\mathcal{S}_0$.

5

Using the notation from above, the BIC is written in the following form

$$\text{BIC}_{\mathcal{S}} = \log(\widehat{\sigma}_{\mathcal{S}}^2) + |\mathcal{S}|\frac{\log(n)}{n}C_n, \tag{10}$$

with $\widehat{\sigma}_{\mathcal{S}}^2 = n^{-1}\,\text{SSE}_{\mathcal{S}} = \inf_{\beta_{\mathcal{S}}}(n^{-1}||Y - X_{\mathcal{S}}\beta_{\mathcal{S}}||_2^2)$ where $||\cdot||_2$ denotes L_2-norm of a vector and C_n is some positive constant. Wang and Leng (2007) prove the consistency of (10) in selecting a true model also for a diverging parameter vector dimension p and a true number of nonzero coefficients q_0. This is shown in unpenalized as well as in penalized regression models.

Modification of (10) in terms of a tuning parameter leads to

$$\text{BIC}_{\lambda} = \log(\widehat{\sigma}_{\lambda}^2) + |\mathcal{S}_{\lambda}|\frac{\log(n)}{n}C_n, \tag{11}$$

where $\widehat{\sigma}_{\lambda}^2 = n^{-1}\,\text{SSE}_{\lambda} = n^{-1}||Y - X\widehat{\beta}(\lambda)||_2^2$ and $\mathcal{S}_{\lambda} = \{j : \widehat{\beta}(\lambda)_j \neq 0\}$. The estimation of the tuning parameter $\widehat{\lambda}$ is then chosen by minimizing (11) with $C_n = \log\{\log(p)\}$ or $C_n = \sqrt{n}/p$, see Chand (2012).

Consistency of the BIC_{λ} selector holds for the penalized regression methods such as smoothly clipped absolute deviation (SCAD) method defined by Fan and Li (2001) and adaptive Lasso introduced by Zou (2006). For the regular Lasso method by Tibshirani (1996) the additional assumption on a design matrix X called irrepresentable condition has to be fulfilled.

The aforementioned condition was presented by Zhao and Yu (2006). Firstly they assumed that $n^{-1}X^{\top}X \xrightarrow{p} C$, with C a positive definite matrix

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}. \tag{12}$$

Here C_{11} is a $(q_0 \times q_0)$ matrix that corresponds to the q_0 active predictors and is assumed to be invertible. Then the formulation of the irrepresentable condition is

$$\left|\left[C_{21}C_{11}^{-1}\text{sgn}(\beta_{\mathcal{S}_0})\right]_k\right| \leq 1, \quad k = 1, \ldots, p - q_0. \tag{13}$$

Adopting the notation from above, q_0 is a number of nonzero parameters in the true model $\mathcal{S}_0$ and $\text{sgn}(\beta_{\mathcal{S}_0}) = (\text{sgn}(\beta_{01}), \ldots, \text{sgn}(\beta_{0q_0}))^{\top}$ with sign function $\text{sgn}(\beta_j) = 1$ if $\beta_j > 0$, $\text{sgn}(\beta_j) = -1$ if $\beta_j < 0$ and $\text{sgn}(\beta_j) = 0$ if $\beta_j = 0$.

Modified selection criteria for penalized quantile regression which were used by Li and Zhu (2008) are BIC for quantile regression presented by Koenker et al. (1994) and generalized approximate cross-validation criterion (GACV) introduced by Yuan (2006)

$$\text{BIC}(\lambda) = \log\left[n^{-1}\sum_{i=1}^{n}\rho_{\tau}\{Y_i - X_i^{\top}\widehat{\beta}(\lambda)\}\right] + \frac{\log(n)}{2n}\,\widehat{\text{df}}(\lambda), \tag{14}$$

$$\text{GACV}(\lambda) = \frac{\sum_{i=1}^{n}\rho_{\tau}\{Y_i - X_i^{\top}\widehat{\beta}(\lambda)\}}{n - \widehat{\text{df}}(\lambda)}, \tag{15}$$

where $\widehat{\text{df}}(\lambda)$ stands for the estimated effective dimension of the fitted model. Li and Zhu (2008) argued that number of interpolated observations Y_i denoted by $\mathcal{E}$ is a plausible measure for this quantity, i.e. $\widehat{\text{df}}(\lambda) = |\mathcal{E}|$.

2.3 Algorithms to solve Lasso

Finding a feasible solution of the optimization problems (1) and (2) can be computationally demanding, since one has to check all of the combinations of values of the tuning parameter λ and its respective model parameter estimates $\widehat{\beta}(\lambda)$. Only after all of the possible combinations are found, the particular method of choosing $\widehat{\lambda}$ can be applied.

The first algorithm for finding solution of Lasso was presented by Tibshirani (1996) in his work introducing the Lasso method itself. Then Osborne et al. (2000) developed an algorithm which works not only for the case where $p < n$ but also $n > p$. In order to make the computation more efficient, Efron et al. (2004) proposed the use of the least angle regression algorithm (LARS). The latter procedure is as efficient as a single least squares fit and can also be used in cases where number of parameters of the investigated model is much larger than the number of observations. As a selection criterion of $\widehat{\lambda}$ for LARS, Efron et al. (2004) suggested to use C_p-type selection criterion. Zou et al. (2007) then defined model selection criteria such as C_p, Akaike information criterion (AIC) and BIC suitable for the Lasso framework.

Another approaches to find a path of Lasso solutions, particularly for the quantile regression, were proposed by Belloni and Chernozhukov (2011) and Li and Zhu (2008). The second one comes into focus in this paper, since one is interested in modeling tail event dependencies when dealing with systemic risk evaluations.

3 Simulation study

As derived in the previous section, the penalization parameter λ of the Lasso regression depends on three effects. The factors driving its dynamics are variance of the error term of the model, conditionality of the matrix $X^\top X$ and absolute size of the coefficients of the model, $||\beta||_1$. In this section we conduct simulations which describe the relationships between these three effects and the parameter λ focusing mainly on a quantile regression case. Our aim is to disentangle these effects and find the way to explain behaviour of λ in dependency of the three aforementioned elements.

3.1 Penalty λ dependent on variance σ^2

Firstly we investigate the effect of the size of variance σ^2 of the error term ε on the penalty parameter λ. According to the identity (7) λ is supposed to rise with higher σ^2 and vice versa. This holds for the linear regression problem, and as discussed previously for the quantile regression as well. The evidence is visible from Figure 2, whereas when considering the formula (8) this dependency is not straightforward to follow.

In our simulation study we use quantile regression model $Y = X\beta + \varepsilon$ with a vector of responses $Y = (Y_1, \ldots, Y_n)^\top$, a vector of parameters $\beta = (\beta_1, \ldots, \beta_p)^\top$, an $(n \times p)$ design matrix X and *iid* error term $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ such that $P(\varepsilon_i \leq 0 | X_i = x) = \tau$ for almost every $x \in \mathbb{R}^p$ with $\tau \in (0, 1)$ denoting conditional quantile of Y.

The design matrix X is simulated from the p-dimensional normal distribution

$$\{X_i\}_{i=1}^n \sim \mathrm{N}_p(0, \Sigma), \tag{16}$$

where the elements of $(p \times p)$ covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1}^p$ are defined as follows

$$\sigma_{ij} = \rho^{|i-j|} \text{ for } i, j = 1, \dots, p, \tag{17}$$

with $\rho = 0.5$ as in Tibshirani (1996). Here we select $n = 600$ and $p = 100$. In order to study the effect of increased dispersion (in the error term ε) on λ, the vector of parameters is set to

$$\beta_{0(100 \times 1)} = (1, 1, 1, 1, 1, 0, \dots, 0)^\top. \tag{18}$$

The error term is simulated such that its variance changes after the observation $i_0 = 300$. We assume ε_i for $i = 1, \dots, n$ to be independently distributed with asymmetric Laplace distribution

$$\varepsilon_i \sim \begin{cases} \mathrm{ALD}(0, 1, 0.05), & \text{if } i \le i_0 \\ \mathrm{ALD}(0, 2, 0.05), & i > i_0 \end{cases}. \tag{19}$$

The density of asymmetric Laplace distribution is

$$f(x|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\frac{\rho_\tau(x - \mu)}{\sigma} \right\}, \tag{20}$$

with location parameter μ, scale parameter $\sigma > 0$, skewness parameter $\tau \in (0, 1)$ and the check function $\rho_\tau(\cdot)$ as defined in (3). The idea to use this type of distribution comes from Lee et al. (2014).

We simulate 50 scenarios using the algorithm designed by Li and Zhu (2008) and select $\widehat{\lambda}$ according to BIC (14). For model fitting we apply moving windows technique to capture the dynamics of the tuning parameter λ. The size of the moving window is set to be $w = 80$. Resulting values of $\widehat{\lambda}$ obtained by simulation settings above are, together with other model characteristics of interest, captured in Figure 3.

As can be seen from Figure 3, the values of the estimated tuning parameter $\widehat{\lambda}$ are indeed increasing with higher variation σ^2 of the error term. Number of nonzero parameters $q_0 = ||\beta_0||_0$ was set to be constant over all $n = 600$ observations and also the level around which the condition number $\kappa(X^\top X)$ fluctuates stays constant. However, the L_1-norm of estimated model coefficients $||\widehat{\beta}(\widehat{\lambda})||_1$ changes with higher values of $\widehat{\lambda}$. Since that is an idea of the Lasso method itself, this can be seen as a natural effect.

In order to study the size of impact of σ^2 on λ we conducted a set of simulations, where different values of scale parameter σ were used after the change point i_0. The starting value was defined as in the previous case, $\sigma = 1$, and the relative and absolute change of average $\widehat{\lambda}$ were examined. Observed changes are noted in Table 1.

From Table 1 one can see that the penalization parameter $\widehat{\lambda}$ increases in dependency of the change in the scale parameter σ of the distribution of the error term in the assumed model. This conclusion of course corresponds to what we see from Figure 2. Again we use BIC as a selection criterion. However, as discussed before, theoretically other methods yield the same dependency structure.

(a) Penalty parameter $\widehat{\lambda}$

(b) Cardinality of active set $\widehat{q}$

(c) L_2-norm of residuals

(d) L_1-norm of coefficients

Figure 3: Time series of $\widehat{\lambda}$ (blue), other model characteristics and their respective averages (red) drawn from the 50 simulations with change of σ_i after $i_0 = 300$, moving windows of length 80.

XFGTVP_LambdaSim

σ_i $i > i_0$	$\dfrac{\bar{\bar{\lambda}}_{end}}{\bar{\bar{\lambda}}_{start}}$	$\bar{\bar{\lambda}}_{end} - \bar{\bar{\lambda}}_{start}$
1.1	1.061	0.027
1.2	1.084	0.037
1.3	1.112	0.050
1.4	1.135	0.060
1.5	1.144	0.064
1.6	1.162	0.072
1.7	1.169	0.075
1.8	1.177	0.079
1.9	1.187	0.083
2.0	1.199	0.089

Table 1: Relative and absolute change in averaged values of $\widehat{\lambda}$ before and after the change point $i_0 = 300$ with starting value of the scale parameter $\sigma_i = 1$ for $i \leq i_0$.

3.2 Penalty λ dependent on model size q

The second effect driving the size of the penalization parameter λ is the number of nonzero parameters q. In order to study this case, the design matrix X was again set as in (16) and (17) with $\rho = 0.5$. The error term ε_i was simulated to have scale $\sigma = 1$ for all $1 \le i \le n$ and the change in vector of model parameters β came into focus. The number of nonzero parameters of the model was defined by setting β_0 to have the form

$$\beta_{0i} = \begin{cases} (1, 1, 1, 1, 1, 0, \ldots, 0)^\top, & i \le i_0 \\ \underbrace{(1, 1, \ldots, 1}_{10\times}, 0, \ldots, 0)^\top, & i > i_0. \end{cases} \tag{21}$$

Thus, the first i_0 simulated observations have five active parameters and the rest has ten of them.

The paths of the values of $\widehat{\lambda}$ obtained from the aforementioned simulation settings are plotted in Figure 4. Visible are also other characteristics of the model which we are interested in to examine.

(a) Penalty parameter $\widehat{\lambda}$

(b) Cardinality of active set $\widehat{q}$

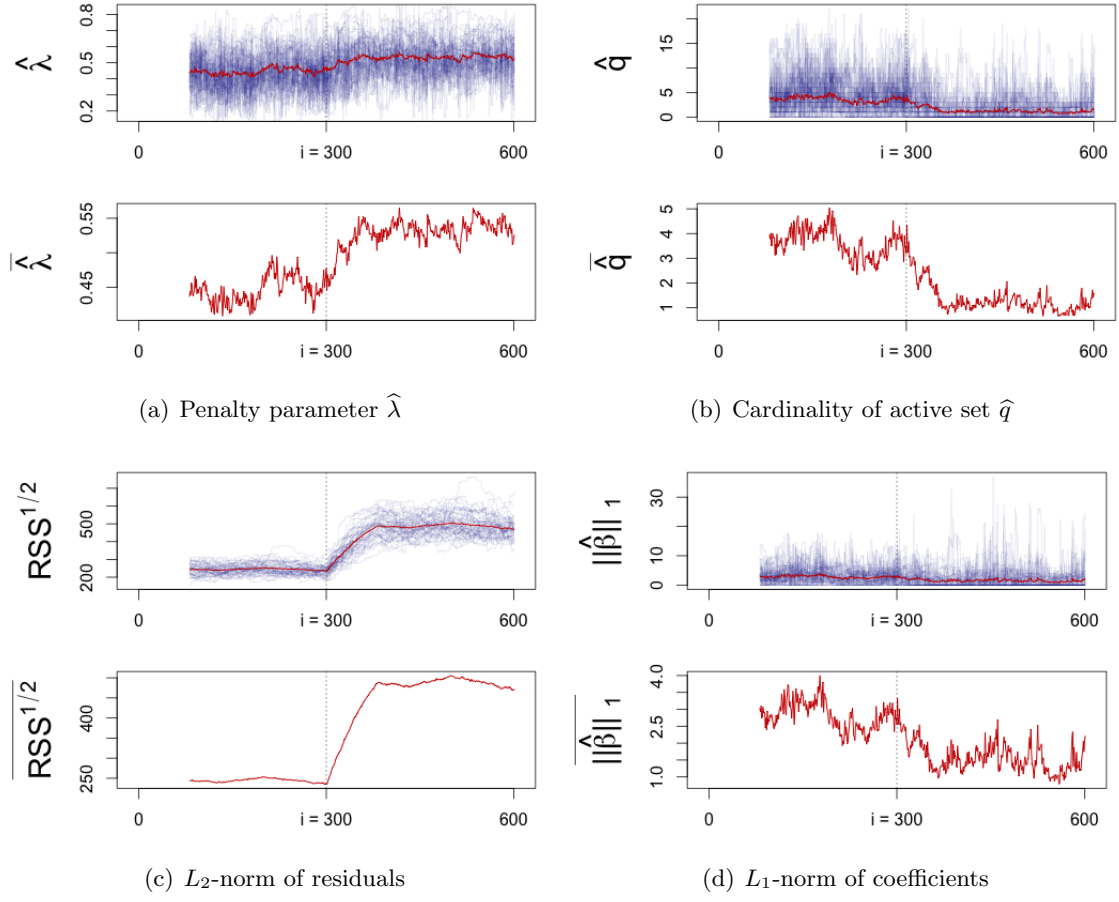(c) L_2-norm of residuals

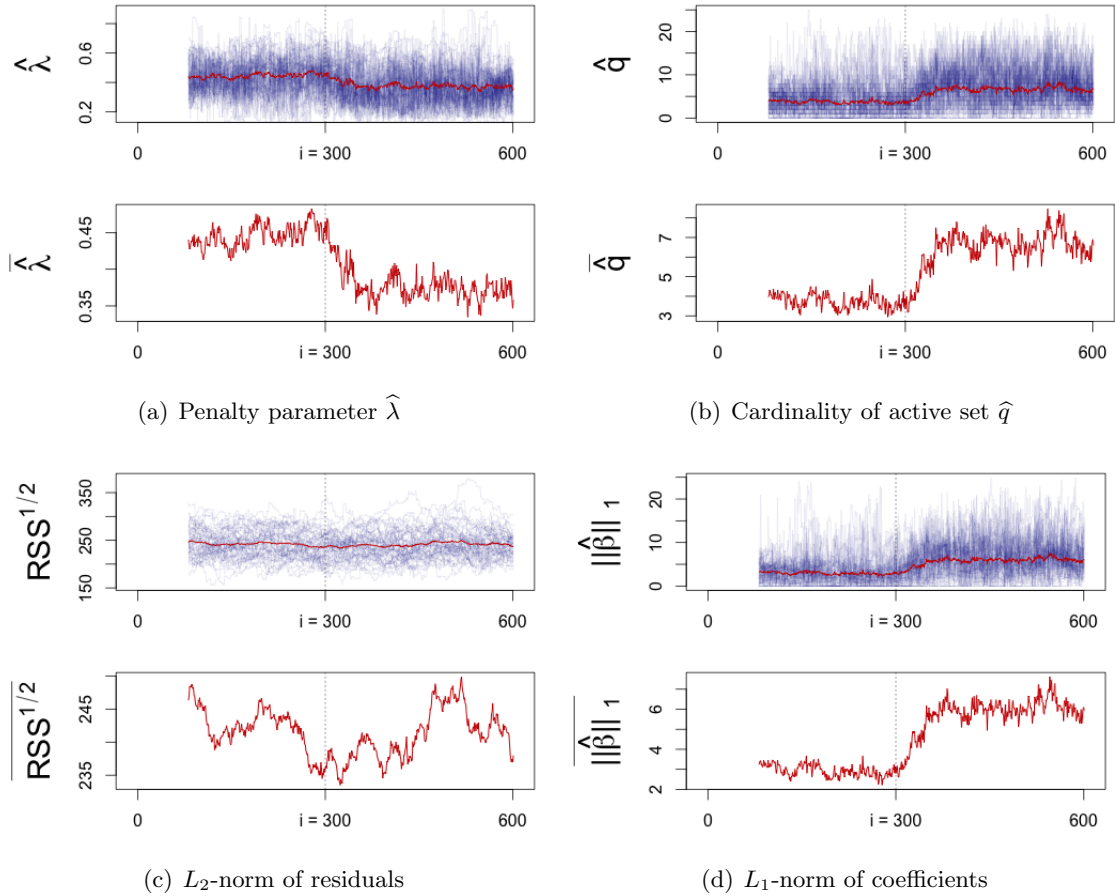(d) L_1-norm of coefficients

Figure 4: Time series of $\widehat{\lambda}$ (blue), other model characteristics and their respective averages (red) drawn from the 50 simulations with change of q_0 after $i_0 = 300$, moving windows of length 80.

Q XFGTVP_LambdaSim

As expected from (8) defining λ, an increasing value of $||\widehat{\beta}(\lambda)||_1$ or q results in a decreasing

value of the tuning parameter λ. In this specific case $||\beta_0||_1 = q_0$. From Figure 4 one can see that the value of $\widehat{\lambda}$ decreased with higher $\widehat{q}$.

To study the reaction of λ on the cardinality of the active set q, we performed simulations with different changes of q after the observation i_0, the starting value was always $q_0 = 5$. The results are summarized in Table 2. From equation (8) the relationship between λ and $||\widehat{\beta}(\lambda)||_0$ as well as q is inversely proportional and values in Table 2 correspond to this statement.

q_{0i} $i > i_0$	$\dfrac{\bar{\bar{\lambda}}_{end}}{\bar{\bar{\lambda}}_{start}}$	$\bar{\bar{\lambda}}_{end} - \bar{\bar{\lambda}}_{start}$
6	0.952	-0.021
7	0.922	-0.035
8	0.905	-0.043
9	0.862	-0.062
10	0.837	-0.073
15	0.736	-0.118

Table 2: Relative and absolute change in averaged values of $\widehat{\lambda}$ before and after the change point $i_0 = 300$ with starting number of nonzero parameters $q_{0i} = 5$ for $i \leq i_0$.

We may conclude that the cardinality of the active set q has a real impact on change in value of λ. Since in (8) the effect of q is captured by the effect of $||\widehat{\beta}(\lambda)||_1$, this is also of our interest. Another simulation was conducted to investigate the impact of the L_1-norm of the model coefficients. Previously the coefficients were hard thresholded, i.e. cut off abruptly and set to be zero. Now the parameters are allowed to decrease to zero more smoothly

$$\beta_{0i} = \begin{cases} \underbrace{(1, 1, \ldots, 1}_{10\times}, 0, \ldots, 0)^\top, & i \leq i_0 \\ (1, 0.9, 0.8, \ldots, 0.2, 0.1, 0, \ldots, 0)^\top, & i > i_0, \end{cases} \tag{22}$$

i.e. $||\beta_{0i}||_1 = 10$ for $i \leq i_0$ and $||\beta_{0i}||_1 = 5.5$ for $i > i_0$.

We put this simulation setting forward, because it seems more natural that the effect of particular covariates fades away rather than disappears. Time series of model characteristics of this case are to be found in Figure 5. The relative and absolute change of average $\widehat{\lambda}$ after the point $i_0 = 300$ is 1.245 and 0.091 respectively.

3.3 Penalty λ dependent on design

We examine the dependency of the parameter λ on the design matrix X of the given model through the characteristics called condition number of a matrix:

$$\kappa(X^\top X) = \frac{\phi_{\max}(X^\top X)}{\phi_{\min}(X^\top X)},$$

where $\phi_{\max}(\cdot)$ and $\phi_{\min}(\cdot)$ are the largest and the smallest eigenvalues of a matrix. If the condition number κ is low the problem is called well-conditioned, matrices with higher κ values are referred to as ill-conditioned. The condition number can help to diagnose a multicollinearity issue. With the presence of multicollinearity, one can expect more coefficients to be incorrectly defined as significant and therefore values of q and $||\beta||_1$ to rise. This is in
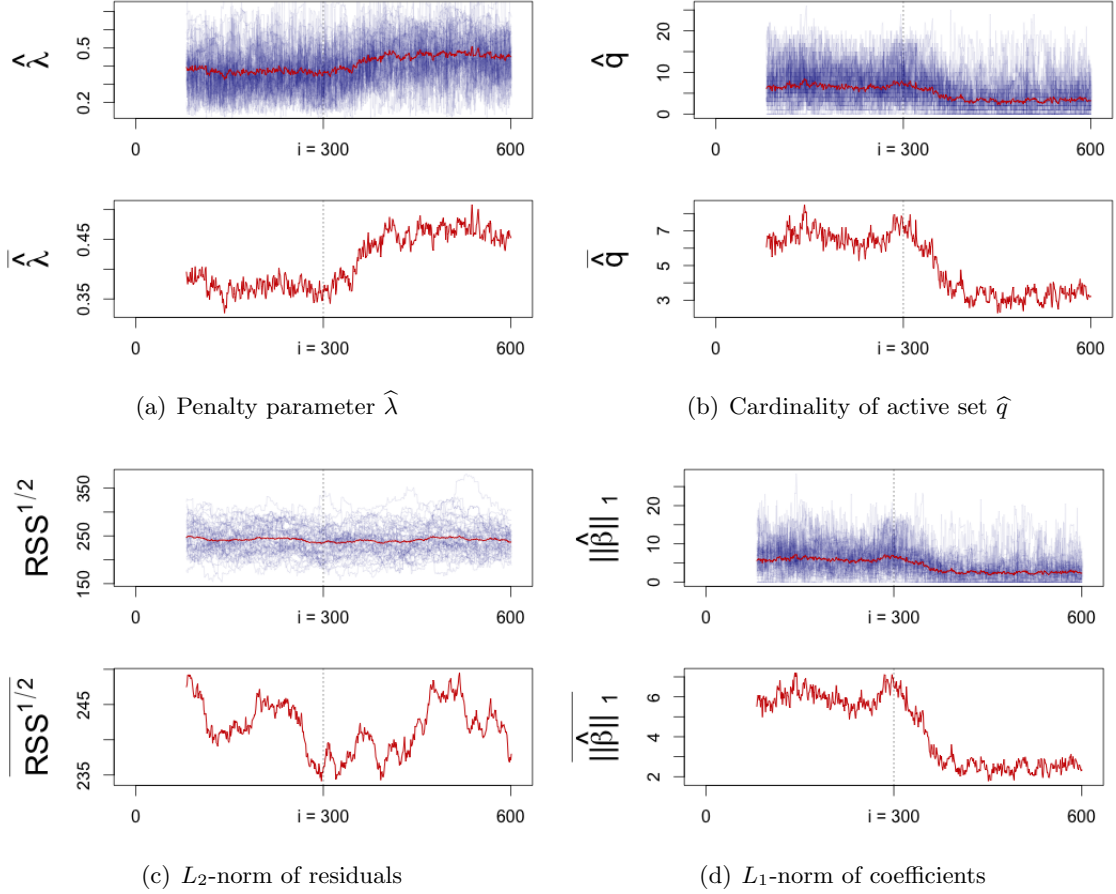
(a) Penalty parameter $\widehat{\lambda}$ (b) Cardinality of active set $\widehat{q}$

(c) L_2-norm of residuals   (d) L_1-norm of coefficients

Figure 5: Time series of $\widehat{\lambda}$ (blue), other model characteristics and their respective averages (red) drawn from the 50 simulations with change of $||\beta_{0i}||_1$ after $i_0 = 300$, moving windows of length 80.

Q XFGTVP_BetaChange

analogy to the situation described in the previous subsection and regarding the formula (8) we expect the tuning parameter λ to decrease with higher condition number of the matrix $X^\top X$.

The simulation settings are as follows; parameter β_0 as in (18) and the error term is *iid* with $\varepsilon_i \sim \mathrm{ALD}(0, 1, 0.05)$ for $0 \le i \le n$. The design matrix X is simulated from (16) and (17), but here the parameter ρ is allowed to change after the point $i_0 = 300$. The case where $\rho_i = 0$ for $i \le i_0$ and $\rho_i = 0.5$ for $i > i_0$ is illustrated in Figure 6.

Indeed, our expectations presented above hold true. Increased correlation between the covariates and with that increased condition number $\kappa(X^\top X)$ result in decreasing values of the estimated tuning parameter $\widehat{\lambda}$. This case together with other simulated changes in correlation structure between covariates are summarized in Table 3. Starting value of ρ from (17) is always 0.

(a) Penalty parameter $\widehat{\lambda}$

(b) Cardinality of active set $\widehat{q}$

(c) L_2-norm of residuals

(d) L_1-norm of coefficients

Figure 6: Time series of $\widehat{\lambda}$ (blue), other model characteristics and their respective averages (red) drawn from the 50 simulations with change of ρ_i after $i_0 = 300$, moving windows of length 80.

Q XFGTVP_LambdaSim

ρ_i $i > i_0$	$\dfrac{\bar{\bar{\lambda}}_{end}}{\bar{\bar{\lambda}}_{start}}$	$\bar{\bar{\lambda}}_{end} - \bar{\bar{\lambda}}_{start}$
0.1	1.023	0.012
0.3	0.943	-0.028
0.5	0.890	-0.055
0.7	0.692	-0.155
0.9	0.750	-0.126

Table 3: Relative and absolute change in averaged values of $\widehat{\lambda}$ before and after the change point $i_0 = 300$ with starting number of nonzero parameters $\rho_i = 0$ for $i \leq i_0$.

3.4 All factors affecting the value of λ

So far we investigated the effect of the change in the variance of error term σ^2, in structure of the vector of parameters β and in the correlation structure of the covariates ceteris paribus. In this subsection we focus on all of the factors driving dynamics of λ at once and examine the strength of their impact when combined together.

For each of the elements driving the dynamics of the penalization parameter λ we simulated three cases. The values of interest either stayed constant, increased or decreased after the point $i_0 = 300$. If constant, the scale parameter σ of the distribution of the error term was set to be 1. Otherwise it increased from the value of 1 to 2 or decreased from 2 to the value of 1. Number of nonzero parameters was either $q_0 = 5$ for all $n = 600$ observations or it increased to the value $q_0 = 10$ or decreased from $q_0 = 10$ to $q_0 = 5$ after the point i_0. The change of the design matrix was again defined by the change of the correlation structure between corresponding covariates, i.e. change of ρ from (17). For the constant case it was set to be $\rho = 0.5$, when increased it had value 0.9 after the i_0-th observation and for the decreasing case it was $\rho = 0.9$ for $i \leq i_0$ and $\rho = 0.5$ for $i > i_0$.

Results of all combinations of the changes in the factors having impact on λ are summarized in Table 4. There we can see that the effects can overpower each other when combined. This holds particularly for the cases, when the condition number κ is increased and number of nonzero parameters q_0 decreased and vice versa. This fact can be explained by the issue of multicollinearity as discussed before.

	$\sigma^2 \nearrow$			$\sigma^2 \rightarrow$			$\sigma^2 \searrow$		
	$\kappa \nearrow$	$\kappa \rightarrow$	$\kappa \searrow$	$\kappa \nearrow$	$\kappa \rightarrow$	$\kappa \searrow$	$\kappa \nearrow$	$\kappa \rightarrow$	$\kappa \searrow$
$q_0 \nearrow$	0.884	1.101	1.311	0.783	0.843	1.003	0.659	0.710	0.841
$q_0 \rightarrow$	0.992	1.198	1.425	0.854	1.001	1.191	0.719	0.843	0.998
$q_0 \searrow$	1.162	1.403	1.555	1.000	1.172	1.300	0.759	0.889	1.125

Table 4: Relative changes $\widehat{\lambda}_{end}/\widehat{\lambda}_{start}$ as a result of combinations of changes in a model. Blue and red colours indicate increased and decreased values of $\widehat{\lambda}$ after the change point respectively, and white colour denotes cases when there is no change in $\widehat{\lambda}$ visible.

Empirically, when considering the situation on financial markets (particularly modeling of stock prices), increased volatility indicates elevated risk. Parameter λ is sensitive to the changes in degree of variation and therefore can be bound to the risk evaluation problem. Another aspect indicating time series of λ as a measure of systemic risk is its dependency on interconnectedness of financial institutions, which can be measured by the number of nonzero parameters in estimated model and their magnitude.

4   Empirical analysis

4.1   Data description

In order to be able to apply our insight to the FinancialRiskMeter (`http://frm.wiwi.hu-berlin.de`), we closely follow the choice of data of Härdle et al. (2016). Due to the computational efficiency, our dataset consists of daily stock returns of the first 100 largest U.S. financial companies ordered by market capitalization according to NASDAQ company list. In the FRM case it is 200. The stock returns are downloaded from Yahoo Finance and the list of the corresponding companies is to be found in Table 6.

As a characterization of the general state of the economy, six macroprudential variables are used as covariates in our model settings. These are implied volatility index reported by

the Chicago Board Options Exchange, daily S&P500 index returns, daily Dow Jones U.S. Real Estate index returns, changes in the three-month Treasury bill rate, changes in the slope of the yield curve corresponding to the yield spread between the ten-year Treasury rate and the three-month bill rate and, finally, changes in the credit spread between BAA-rated bonds and the Treasury rate. The former three are obtained from Yahoo Finance and the latter three from the Federal Reserve Board. The macro state variables are summarized in Table 7. The data are downloaded with help of Q FRM_download_data.

All of the variables are recorded in the time interval from 03 January 2007 to 17 August 2016. For the macroprudential variables we use 1 day lagged values.

4.2   Construction of time series of $\widehat{\lambda}$

In order to capture interdependencies among the companies and to reduce the dimensionality of the data set into single time series of the penalization parameter λ of the Lasso regression, we proceed as follows.

We take each of the 100 companies as a dependent variable and use the remaining 99 together with the macro variables as predictors, i.e. $p = 105$. This way we get hundred regression models, which are then fitted with help of the quantile Lasso method by Li and Zhu (2008). To record the dynamics of $\widehat{\lambda}$, we use moving windows of size 63 observations ($n = 63$) which in this case represents 3 months.

Within each window algorithm designed by Li and Zhu (2008) is used to fit the Lasso model. Then the best fit and with it also the tuning parameter $\widehat{\lambda}$ are chosen with help of the BIC criterion (14). We obtain time series of tuning parameters $\widehat{\lambda}_k$ for each of the hundred regressed companies. These are plotted in Figure 7(a) together with the average over all estimated parameters $\widehat{\lambda}_k$, $k = 1, \ldots, 100$, which we are interested in.

Indeed as suggested in our previous simulation study, $\widehat{\lambda}$ is driven by characteristics of an investigated model. From Figure 7 we can see that its values are higher when the residuals of the model are higher, too. There are several peaks in time series of $\widehat{\lambda}$, which correspond to time periods of financial crises. This fact drives us to the conclusion that the dynamics of $\widehat{\lambda}$ can serve as an indicator of a systemic risk.

4.3 $\widehat{\lambda}$ and systemic risk measures

In the past decade, much attention has been paid to measuring of systemic risk, particularly after the financial crisis between 2007 and 2009. It has uncovered the cross-sectional dependencies among financial institutions to be important when determining the risk on the market. Adrian and Brunnermeier (2016), Hautsch et al. (2015) and Härdle et al. (2016), just to mention a few, dealt with evaluating systemic risk according to the relevance of each financial institution itself. This inspired us to connect the Lasso parameter λ with the systemic risk, since it depends not only on the volatility but also on the size of model parameters and the correlation structure of the design matrix. The latter two effects can be translated into the connectedness of financial institutions throughout the market.
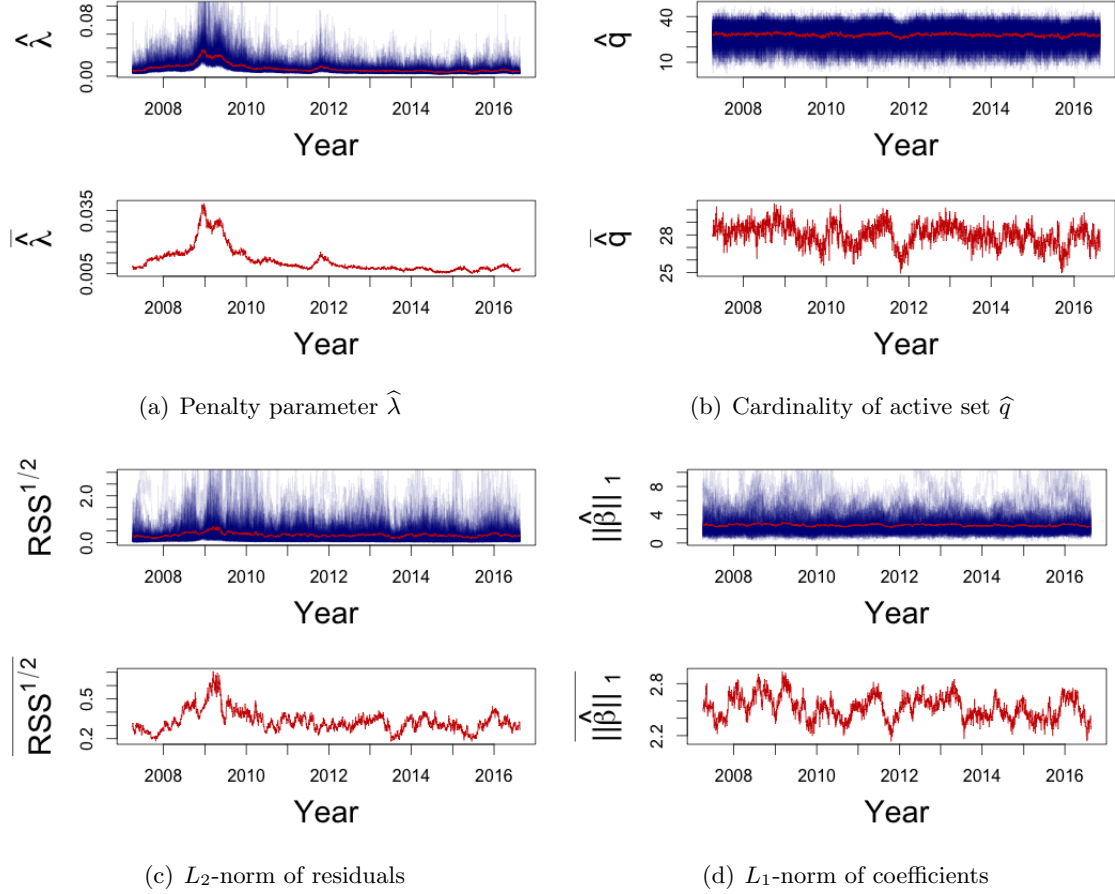
(a) Penalty parameter $\widehat{\lambda}$ (b) Cardinality of active set $\widehat{q}$

(c) L_2-norm of residuals   (d) L_1-norm of coefficients

Figure 7: Time series of $\widehat{\lambda}_k$ (blue) and other model characteristics and their respective averages (red) when fitted to given dataset, moving windows of length 63.

🔷 XFGTVP_FRM

To illustrate the connection between $\widehat{\lambda}$ computed according to the method mentioned previously and other systemic risk measures, we plotted their common time development starting from 3 April 2007 to 17 August 2016, see Figure 8.

We chose VIX to show the dependency between $\widehat{\lambda}$ and volatility observed on the financial market. The Standard & Poor's 500 stock market index (S&P500) moves in opposite direction of $\widehat{\lambda}$, which can also provide some information about behaviour of $\widehat{\lambda}$ in connection to the situation on financial markets. Another systemic risk measure is CoVaR presented by Adrian and Brunnermeier (2016) and extended by Härdle et al. (2016), where a single index model for generalized quantile regression instead of linear quantile regression was employed. The data for CoVaR$_S$ were downloaded from 🔷 TENET_VaR_CoVaR where only weekly data between 7 December 2007 and 4 January 2013 were available. Financial turbulence as a risk measure was proposed by Kritzman and Li (2010). Its comovement with the time series of $\widehat{\lambda}$ is visible from the Figure 8(d). A composite indicator of systemic risk (CISS) is an indicator of contemporaneous stress in the financial system developed by Holló et al. (2012) and computed for the area of Europe on weekly basis. Even when considering another financial market, particularly collecting data from another countries, periods where CISS was elevated correspond to the periods of higher $\widehat{\lambda}$ values. And, finally, credit spread, i.e. changes in

16

(a) $\widehat{\lambda}$ and VIX

(b) $\widehat{\lambda}$ and S&P500

(c) $\widehat{\lambda}$ and CoVaR$_S$

(d) $\widehat{\lambda}$ and Turbulence

(e) $\widehat{\lambda}$ and CISS
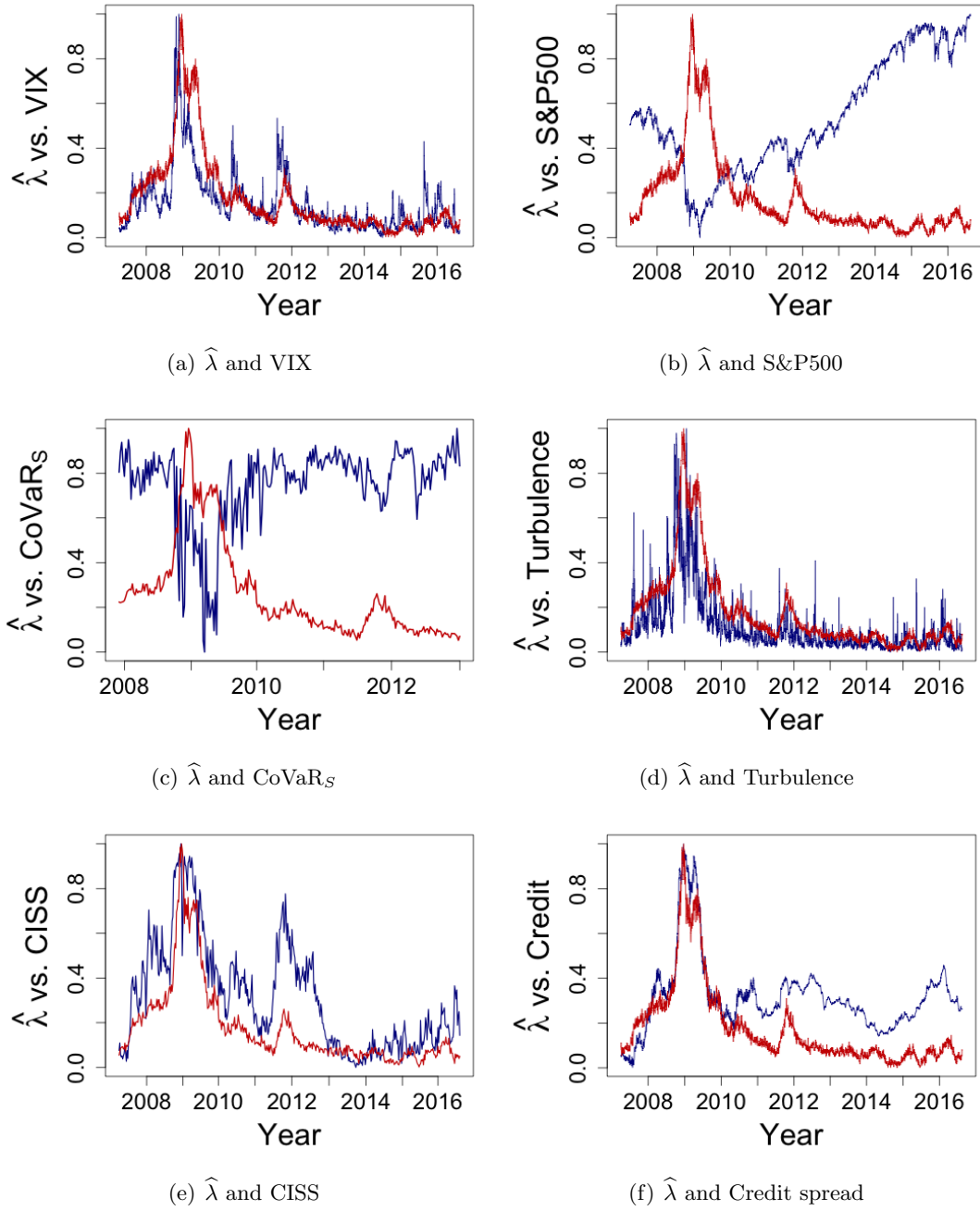
(f) $\widehat{\lambda}$ and Credit spread

Figure 8: Time series of $\widehat{\lambda}$ (red) and various systemic risk measures (blue).

XFGTVP_LambdaSysRisk

the credit spread between BAA-rated bonds and the Treasury rate, suggested by Giglio et al. (2016), was used to relate $\widehat{\lambda}$ to systemic risk level.

From Figure 8 it is visible, that $\widehat{\lambda}$ has a common trend with some of the aforementioned systemic risk measures. For CoVaR$_S$ and S&P500 index it holds, that their time development goes in opposite direction compared to $\widehat{\lambda}$.

In order to show there is a comovement between $\widehat{\lambda}$ and other systemic risk measures also from the statistical point of view, we conducted several cointegration tests. When looking

17

at Figure 8 one can see, that the time series of observed measures are nonstationary, however, there may exist cointegration relations between them which would make it a stationary stochastic process.

As a testing procedure we chose the Johansen (1991) test, where we used its eigenvalue type. In Table 5 there are stated resulting values of test statistics and their corresponding critical values on significance levels 10 % and 5%. Variable r corresponds to a number of cointegration relations found between the two investigated nonstationary time series, i.e. for the valid inference we require that $r = 1$.

	H_0	Test statistic	10 %	5 %
VIX	$r \leq 1$	4.80	7.52	9.24
	$r = 0$	87.43	13.75	15.67
S&P500	$r \leq 1$	7.59	10.49	12.25
	$r = 0$	9.20	16.85	18.96
CoVaR$_S$	$r \leq 1$	4.52	10.49	12.25
	$r = 0$	50.58	16.85	18.96
CoVaR$_L$	$r \leq 1$	4.59	10.49	12.25
	$r = 0$	57.15	16.85	18.96
Turbulence	$r \leq 1$	8.94	10.49	12.25
	$r = 0$	212.24	16.85	18.96
CISS	$r \leq 1$	6.90	10.49	12.25
	$r = 0$	31.12	16.85	18.96
Volatility Connectedness	$r \leq 1$	9.48	10.49	12.25
	$r = 0$	10.51	16.85	18.96
Yield Slope	$r \leq 1$	7.20	10.49	12.25
	$r = 0$	13.63	16.85	18.96
Credit Spread	$r \leq 1$	5.45	10.49	12.25
	$r = 0$	42.29	16.85	18.96

Table 5: Cointegration of $\widehat{\lambda}$ with systemic risk measures, r is number of cointegration relations in Johansen procedure, measures cointegrated with $\widehat{\lambda}$ are written in bold.

In Table 5 we included 3 more systemic risk measures. We chose also CoVaR computed with variable selection based on linear quantile regression (CoVaR$_L$). Another systemic risk measure is the volatility connectedness index designed by Diebold and Yilmaz (2014) and accessed from `http://financialconnectedness.org`. Yield slope denotes changes in the slope of the yield curve corresponding to the yield spread between the 10-year Treasury rate and the 3-month bill rate.

As we can see, many of the measures are cointegrated with the estimated Lasso parameter $\widehat{\lambda}$.

WFC	Wells Fargo & Company	ALL	Allstate Corporation (The)
JPM	JP Morgan Chase & Co.	BEN	Franklin Resources, Inc.
BAC	Bank of America Corporation	STI	SunTrust Banks, Inc.
C	Citigroup Inc.	MCO	Moody's Corporation
AIG	American International Group, Inc.	PGR	Progressive Corporation (The)
GS	Goldman Sachs Group, Inc. (The)	AMP	AMERIPRISE FINANCIAL SERVICES, INC.
USB	U.S. Bancorp	AMTD	TD Ameritrade Holding Corporation
AXP	American Express Company	HIG	Hartford Financial Services Group, Inc. (The)
MS	Morgan Stanley	TROW	T. Rowe Price Group, Inc.
BLK	BlackRock, Inc.	NTRS	Northern Trust Corporation
MET	MetLife, Inc.	MTB	M&T Bank Corporation
PNC	PNC Financial Services Group, Inc. (The)	FITB	Fifth Third Bancorp
BK	Bank Of New York Mellon Corporation (The)	IVZ	Invesco Plc
SCHW	The Charles Schwab Corporation	L	Loews Corporation
COF	Capital One Financial Corporation	EFX	Equifax, Inc.
PRU	Prudential Financial, Inc.	PFG	Principal Financial Group Inc
TRV	The Travelers Companies, Inc.	RF	Regions Financial Corporation
CME	CME Group Inc.	MKL	Markel Corporation
CB	Chubb Corporation (The)	LNC	Lincoln National Corporation
MMC	Marsh & McLennan Companies, Inc.	CBG	CBRE Group, Inc.
BBT	BB&T Corporation	KEY	KeyCorp
ICE	Intercontinental Exchange Inc.	NDAQ	The NASDAQ OMX Group, Inc.
STT	State Street Corporation	CINF	Cincinnati Financial Corporation
AFL	Aflac Incorporated	CNA	CNA Financial Corporation
AON	Aon plc	HBAN	Huntington Bancshares Incorporated
SEIC	SEI Investments Company	ERIE	Erie Indemnity Company
ETFC	E*TRADE Financial Corporation	OZRK	Bank of the Ozarks
AMG	Affiliated Managers Group, Inc.	WTM	White Mountains Insurance Group, Ltd.
RJF	Raymond James Financial, Inc.	SNV	Synovus Financial Corp.
UNM	Unum Group	ISBC	Investors Bancorp, Inc.
NYCB	New York Community Bancorp, Inc.	MKTX	MarketAxess Holdings, Inc.
Y	Alleghany Corporation	LM	Legg Mason, Inc.
SBNY	Signature Bank	CBSH	Commerce Bancshares, Inc.
CMA	Comerica Incorporated	BOKF	BOK Financial Corporation
AJG	Arthur J. Gallagher & Co.	EEFT	Euronet Worldwide, Inc.
JLL	Jones Lang LaSalle Incorporated	DNB	Dun & Bradstreet Corporation (The)
TMK	Torchmark Corporation	WAL	Western Alliance Bancorporation
WRB	W.R. Berkley Corporation	EV	Eaton Vance Corporation
AFG	American Financial Group, Inc.	CFR	Cullen/Frost Bankers, Inc.
SIVB	SVB Financial Group	MORN	Morningstar, Inc.
EWBC	East West Bancorp, Inc.	THG	The Hanover Insurance Group, Inc.
ROL	Rollins, Inc.	UMPQ	Umpqua Holdings Corporation
ZION	Zions Bancorporation	CNO	CNO Financial Group, Inc.
AIZ	Assurant, Inc.	FHN	First Horizon National Corporation
PACW	PacWest Bancorp	WBS	Webster Financial Corporation
AFSI	AmTrust Financial Services, Inc.	PB	Prosperity Bancshares, Inc.
ORI	Old Republic International Corporation	PVTB	PrivateBancorp, Inc.
PBCT	People's United Financial, Inc.	SEB	Seaboard Corporation
CACC	Credit Acceptance Corporation	FCNCA	First Citizens BancShares, Inc.
BRO	Brown & Brown, Inc.	MTG	MGIC Investment Corporation

Table 6: List of 100 U.S. largest financial companies

1.	VIX
2.	Daily change in the 3-month Treasury maturities
3.	Change in the slope of the yield curve
4.	Change in the credit spread
5.	Daily Dow Jones U.S. Real Estate index returns
6.	Daily S&P500 index returns

Table 7: List of macro state variables.

5 Summary

In the present paper we proposed to study dynamics of the penalization parameter of the Lasso regression by Tibshirani (1996) and its quantile regression extension by Li and Zhu (2008). We focused on three effects driving the time-dependent behaviour of the penalization parameter λ, particularly variation of the model residuals, size of active set of the model and the covariance structure of its respective design matrix.

In the simulation study we justified our expectations about relationships between the aforementioned effects and λ. These results led us to relate the dynamics of λ to the evaluation of systemic risk, since both interdependency among covariates and variability of the error term are neatly connected to the problem of risk controlling.

For the empirical study we chose to follow the setup of the risk control time series Financial-RiskMeter (http://frm.wiwi.hu-berlin.de) and focused on 100 largest U.S. financial companies and 6 macroprudential variables to obtain time series of the estimated Lasso parameter λ. Consequently, we found cointegration relations between estimated λ and several systemic risk measures.

Thus, we may conclude there is a substantiated reason to study time series of λ and build theoretical model, which would provide further insight into prediction of the Lasso parameter λ and systemic risk simultaneously. Furthermore, implementing our work into time series context or the network framework might be of interest.

References

Adrian, T. and Brunnermeier, M. K. (2016). CoVaR, American Economic Review **106**: 1705–1741.

Belloni, A. and Chernozhukov, V. (2011). l_1-Penalized Quantile Regression in High-Dimensional Sparse Models, The Annals of Statistics **39**: 82–130.

Chand, S. (2012). On Tuning Parameter Selection of Lasso-Type Methods - A Monte Carlo Study, Proceedings of 9th International Bhurban Conference on Applied Sciences & Technology, 120–129.

Chen, K. and Chan, K. S. (2011). Subset ARMA Selection via the Adaptive Lasso, Statistics and Its Interface **4**: 197–205.

Diebold, F. X. and Yilmaz, K. (2014). On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms, Journal of Econometrics **182**: 119–134.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least Angle Regression, The Annals of Statistics **32**: 407–499.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties, Journal of the American Statistical Association **96**: 1348–1360.

Giglio, S., Kelly, B. and Pruitt, S. (2016). Systemic Risk and the Macroeconomy: An Empirical Evaluation, Journal of Financial Economics **119**: 457–471.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning: Data mining, Inference and Prediction, 2nd ed, Springer Verlag, New York.

Hautsch, N., Schaumburg, J. and Schienle, M. (2015). Financial Network Systemic Risk Contributions, Review of Finance **19**: 685–738.

Härdle, W. K., Wang, W. and Yu, L. (2016). TENET: Tail-Event Driven NETwork Risk, Journal of Econometrics **192**: 499–513.

Hsu, N. J., Hung, H. L. and Chang, Y. M. (2008). Subset Selection for Vector Autoregressive Processes Using Lasso, Computational Statistics and Data Analysis **52**: 3645–3657.

Holló, D., Kremer, M. and Lo Duca, M. (2012). CISS - A Composite Indicator of Systemic Stress in the Financial System, ECB Working Paper Series No. 1426.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, Econometrica **59**: 1551–1580.

Koenker, R. and Basset, G. (1978). Regression Quantiles, Econometrica **46**: 33–50.

Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile Smoothing Splines, Biometrica **81**: 673–680.

Kritzman, M. and Li, Y. (2010). Skulls, Financial Turbulence, ans Risk Management, Financial Analysts Journal **66**: 30–41.

Lee, E. R., Noh, H. and Park, B. U. (2014). Model Selection via Bayesian Information Criterion for Quantile Regression Models, Journal of the American Statistical Association **109**: 216–229.

Leng, C., Lin, Y. and Wahba, G. (2006). A Note on the Lasso and Related Procedures in Model Selection, Statistica Sinica **16**: 1273–1284.

Li, Y. and Zhu, J. (2008). L_1-Norm Quantile Regression, Journal of Computational and Graphical Statistics **17**: 1–23.

Nardi, Y. and Rinaldo, A. (2011). Autoregressive Process Modeling via the Lasso Procedure, Journal of Multivariate Analysis **102**: 528–549.

Osborne, M. R. (1985). Finite Algorithms in Optimization and Data Analysis, Wiley Series in Probability and Mathematical Statistics, Chichester, Wiley.

Osborne, M. R., Presnell, B. and Turlach, B. A. (2000). On the LASSO and its Dual, Journal of Computational and Graphical Statistics **9**: 319–337.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. `http://www.R-project.org/` (Accessed: 15th April 2015).

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society: Series B **58**: 267–288.

Wang, H. and Leng, C. (2007). Unified LASSO Estimation by Least Squares Approximation, Journal of the American Statistical Association **102**: 1039–1048.

Wang, H., Li, G. and Tsai, C. L. (2007). Regression Coefficient and Autoregressive Order Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society: Series B **69**: 63–78.

Wang, H., Li, B. and Leng, C. (2009). Shrinkage Tuning Parameter Selection with a Diverging Number of Parameters, Journal of the Royal Statistical Society: Series B **71**: 671–683.

Yuan, M. (2006). GACV for Quantile Smoothing Splines, Computational Statistics & Data Analysis **50**: 813–829.

Zhao, P. and Yu, B. (2006). On Model Selection Consistency of Lasso, Journal of Machine Learning Research **7**: 2541–2563.

Zou, H. (2006). The Adaptive Lasso and its Oracle Properties, Journal of the American Statistical Association **101**: 1418–1429.

Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "Degrees of Freedom" of the Lasso, The Annals of Statistics **35**: 2173–2192.