SFB 649 Discussion Paper 2017-003

FRM: a Financial Risk Meter based on penalizing tail events occurrence

Lining Yu*
Wolfgang Karl Härdle*
Lukas Borke*
Thijs Benschop*

*Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

FRM: a Financial Risk Meter based on penalizing tail events occurrence *

Lining Yu† Wolfgang Karl Härdle‡ Lukas Borke §
Thijs Benschop ¶

Abstract

In this paper we propose a new measure for systemic risk: the Financial Risk Meter (FRM). This measure is based on the penalization parameter (λ) of a linear quantile lasso regression. The FRM is calculated by taking the average of the penalization parameters over the 100 largest US publicly traded financial institutions. We demonstrate the suitability of this risk measure by comparing the proposed FRM to other measures for systemic risk, such as VIX, SRISK and Google Trends. We find that mutual Granger causality exists between the FRM and these measures, which indicates the validity of the FRM as a systemic risk measure. The implementation of this project is carried out using parallel computing, the codes are published on www.quantlet.de with keyword FRM. The R package **RiskAnalytics** is another tool with the purpose of integrating and facilitating the research, calculation and analysis methods around the FRM project. The visualization and the up-to-date FRM can be found on `http://frm.wiwi.hu-berlin.de`.

Keywords: Systemic Risk, Quantile Regression, Value at Risk, Lasso, Parallel Computing

JEL: C21, C51, G01, G18, G32, G38.

---

*Financial support from the Deutsche Forschungsgemeinschaft (DFG) via SFB 649 "Ökonomisches Risiko", IRTG 1792 "High-Dimensional Non-Stationary Times Series" and Sim Kee Boon Institute for Financial Economics, Singapore Management University are gratefully acknowledged.

†Research associate at Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Econometrics, IRTG 1792, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Email: yulining@wiwi.hu-berlin.de.

‡Professor at Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Econometrics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Email: haerdle@wiwi.hu-berlin.de.

§Research associate at Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Econometrics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Email: lukas@borke.net.

¶Research associate at Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for Applied Statistics and Econometrics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Email: thijs.benschop@hu-berlin.de.

1 Introduction

Systemic risk is dangerous for the stability of financial markets, since the bankruptcy of one firm may have an impact on the stability of other firms too. There are various definitions of systemic risk. One of the most popular definitions is introduced in Schwarcz (2008). He defined systemic risk as a trigger event, such as an economic shock or institutional failure, causing a chain of bad economic consequences, sometimes referred to as domino effect. This definition indicates that interlinkages and interdependencies in a system or market are very crucial for controlling systemic risk. The financial crisis in 2008 is an example. After the bankruptcy of Lehman Brothers, several more financial cooperations bankrupted as a result of their interlinkages with Lehman Brothers. Consequently, there has been a surge in the interest in measuring and controlling systemic risk since the 2008 crisis, which has led to an increase in the research on this topic.

Several methodologies for measuring systemic risk have been proposed. Adrian and Brunnermeier (2016) proposed CoVaR, the value at risk of financial institutions conditional on the other institutions being under distress, which uses two linear quantile regressions. Hautsch et al. (2015) refined this algorithm by introducing linear quantile lasso regression with a fixed penalization parameter λ for each company to select the relevant risk drivers. Fan et al. (2016) and Härdle et al. (2016) use a nonlinear Single Index Model (SIM) combined with a variable selection technique to select the risk factors. We are inspired by the early version of the work of Fan et al. (2016)[1]. In their application, they use data on 200 financial companies and 7 macro variables to estimate the CoVaR. During the estimation procedure, they generate the time-varying penalization parameter λ. This series has a striking pattern: the higher values correspond to the financial crises and the lower values correspond to financial stable periods. This observation has led to the idea to use the penalization parameter λ itself as a measure for systemic risk. The time-varying feature of λ is specific to Fan et al. (2016) and different from Hautsch et al. (2015), who applied a fixed λ for each firm, but not time varying.

Fan et al. (2016) provide the λ series for single companies. In contrast, we would like to see the behavior of λ for all firms. Härdle et al. (2016) compare the linear quantile lasso model and SIM, and conclude that SIM is better than the linear model, but that the linear quantile lasso model is also valid in terms of backtesting. The problem is that the SIM algorithm is computationally intensive and time-consuming. Härdle et al. (2016) generated λ series for 100 firms with less than 300 observations each. The application of SIM is not realistic for large datasets with more than thousand observations. Since linear quantile lasso is easier to apply and time saving, we decided to apply linear quantile lasso regression to compute our risk measure. In the application, we estimate the λ's for all firms individually and take the average over all firms.

We use log return data from the 100 largest US publicly traded financial institutions as well as 6 macro variables. Our model is based on daily log returns of these financial institutions. The time period under consideration runs from April 5, 2007 until September 23, 2016 and covers several documented financial crises (2008, 2011). We observe that

[1] Their slides are available from `https://www.wiwi.hu-berlin.de/de/professuren/quantitativ/statistik/members/personalpages/wh/talks/20130314FanHaeWanZhuYuQRandSIM.pdf`

the pattern of this risk measure is more precise and robust to measure financial risk than the λ series of a single firm. The shape and volatility of the series correspond to the market volatility and financial events with a large impact on systemic risk are clearly visible. Therefore, we propose this series as a new measure for systemic risk and call it Financial Risk Meter (FRM). The webpage of the FRM was released in the end of 2014 and updated weekly since. Currently, Zbonakova et al. (2016) apply linear quantile lasso regression to analyze the behavior of the λ series. They find that λ is sensitive to the changes of volatility, which provide the theoretical evidence for the FRM to be a systemic risk measure, as high volatility indicates high risk.

In this paper we introduce the methodology of the FRM, describe the risk levels, the computational implementation as well as the visualization of the webpage. To show the suitability of the FRM we compare it with other systemic risk measures, such as VIX (see Hallett, 2009), SRISK (see Brownlees and Engle, 2016) as well as the Google trends of key words related to financial crises (see Preis et al., 2013). We find that the FRM and these risk measures mutually Granger cause, which indicates the validity of the FRM as a systemic risk measure.

The remainder of this paper is organized as follows. In Section 2 the methodology used to construct our FRM, which is quantile lasso modeling, is presented. Section 3 presents the data, computational challenge and the visualization of the results. Section 4 shows the validity of our FRM as a measure for financial risk by comparing with other financial risk measures. Section 5 describes the R package **RiskAnalytics** (Borke, 2017) facilitating real-time processing of Nasdaq and Yahoo finance data and parallelized quantile lasso regression methods. Section 6 concludes, the financial institutions applied in this paper is listed in Section 6 Appendix. All the R programs for this paper can be found on www.quantlet.de (Borke and Härdle, 2017a).

2 FRM methodology and estimation

In this section we describe the methodology and algorithm used to compute the proposed FRM, which is the average over the series of the selected penalization term λ for the companies under consideration. Since the penalization parameters are computed based on an L_1-norm (LASSO) quantile linear regression, this regression framework is introduced first. Within this framework, the penalization parameter λ is exogenous. Since the FRM consists of the selected penalization parameter, we subsequently discuss the method used to select λ. We use the generalized approximate cross-validation criterion (GACV) proposed by Yuan and Lin (2006) to determine the optimal λ. The determination of the penalization parameter is pivotal to the methodology of the FRM.

2.1 Linear Quantile Lasso Regression Model

Following Härdle et al. (2016), we introduce the quantile lasso regression model. Let m be the number of macro variables describing the state of the economy, k the number of firms under consideration, $j \in \{1, \dots, k\}$. Then $p = k + m - 1$ represents the number of covariates. $t \in \{1, \dots, T\}$ is the time point with T the total number of observations (days). s is the index of moving window, $s \in \{1, \dots, (T - (n-1))\}$, where n is the length of window size. Then the quantile lasso regression is defined as:

$$X_{j,t}^s = \alpha_j^s + A_{j,t}^{s,\top} \beta_j^s + \varepsilon_{j,t}^s, \tag{1}$$

where $A_{j,t}^s \stackrel{def}{=} \begin{bmatrix} M_{t-1}^s \\ X_{-j,t}^s \end{bmatrix}$, M_{t-1}^s the m dimensional vector of macro variables, $X_{-j,t}^s$ is the $p - m$ dimensional vector of log returns of all other firms except firm j at time t and in moving window s, α_j^s is a constant term and β_j^s is a $p \times 1$ vector defined for moving window s.

The regression is performed using L_1-norm quantile regression proposed by Li and Zhu (2008), which is defined as:

$$\min_{\alpha_j^s, \beta_j^s} \left\{ n^{-1} \sum_{t=s}^{s+(n-1)} \rho_\tau \left(X_{j,t}^s - \alpha_j^s - A_{j,t}^{s,\top} \beta_j^s \right) + \lambda_j^s \parallel \beta_j^s \parallel_1 \right\}, \tag{2}$$

where λ_j^s is the penalization parameter, and the check function $\rho_\tau(u)$ is defined as:

$$\rho_\tau(u) = |u|^c |\mathbf{I}(u \leq 0) - \tau|,$$

where $c = 1$ corresponds to quantile regression. The L_1-norm quantile linear regression can be used to select relevant covariates (other firms and macro state variables) for each firm.

2.2 Penalization Parameter λ

Since Equation (2) has a L_1 loss function and an L_1-norm penalty term, the optimization problem is an L_1-norm quantile regression estimation problem. The choice of the penalization parameter λ_j^s is crucial. There are several options to select λ_j^s, e.g. with the Bayesian Information Criterion (BIC) or using the Generalized Approximate Cross-Validation criterion (GACV). Yuan (2006) conducted simulations and concluded that GACV outperforms BIC in terms of statistical efficiency. Therefore, we determine λ_j^s with the GACV criterion in the FRM model and set λ_j^s as the solution of the following minimization problem:

$$\min GACV(\lambda_j^s) = \min \frac{\sum_{t=s}^{s+n} \rho_\tau \left(X_{j,t}^s - \alpha_j^s - A_{j,t}^{s\top} \beta_j^s \right)}{n - df},$$

where df is a measure of the effective dimensionality of the fitted model. df is the trace of the hat matrix with the t, o entry $\partial(\alpha_j^s - A_{j,t}^{s\top} \beta_j^s)/\partial X_{j,o}^s$, and $o \in \{1, \dots, T\}$. The advantage of GACV is that it also works for $p > n$, which can be important for the FRM if the moving window size is small.

To compute the FRM, we perform the regression analysis as described above and select the optimal penalization parameter $\lambda_j^{s,*}$ for each firm j using GACV. This yields a lambda series for each firm. The average of these λ_j^*'s constitutes our proposed risk measure. The Financial Risk Meter is defined as the average lambdas over the set of k firms for all windows:

$$FRM \stackrel{def}{=} \frac{1}{k}\sum_{j=1}^{k}\lambda_j^*$$

3 Computational challenges and visualization

3.1 Data

To compute the FRM, we use data from 100 US publicly traded financial institutions as well as six macro variables. The selection of financial companies is based on the NASDAQ company list[2] and based on the market capitalization. The selected companies are the 100 US publicly traded financial institutions with the largest market capitalization, see Table 15 in Appendix.

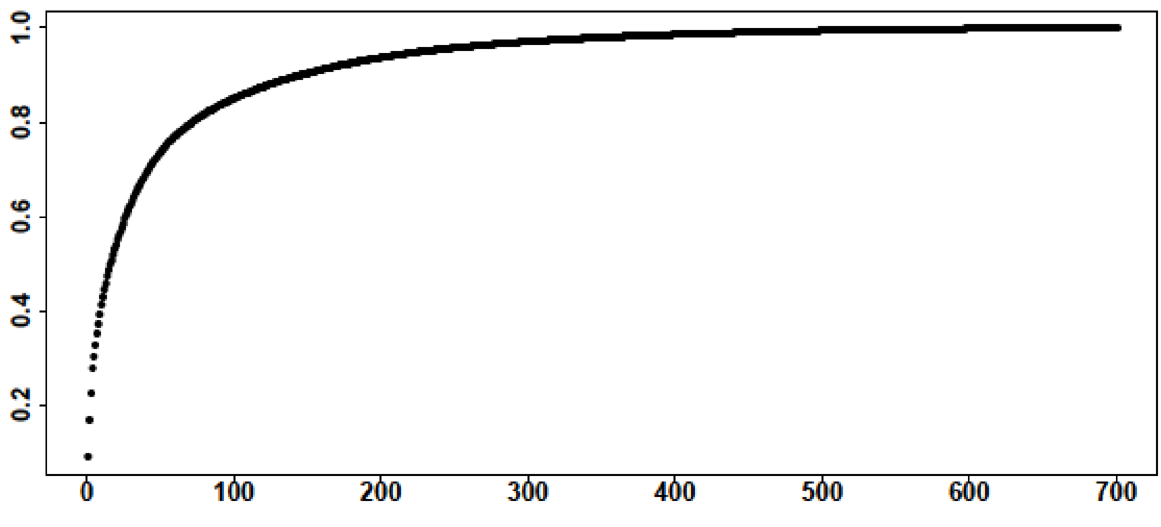

Figure 1: The x-axis represents the number of firms ordered by market capitalization and the y-axis the percentage of total market capitalization.

FRM_per_cap

Initially, we used data on the 200 US publicly traded financial institutions with the largest market capitalization to compute the FRM. However, the smaller companies in

[2]See the NASDAQ webpage:
http://www.nasdaq.com/screening/companies-by-industry.aspx?industry=Finance

this set change regularly over the time period under consideration (2007-2016) due to, for instance, bankruptcies. This leads to issues with automatic downloading of the data and therefore we use only 100 firms. Figure 1 shows the cumulative market capitalization of US financial firms. The x-axis represents the firms ordered by market capitalization and the y-axis the cumulative market capitalization. We observe that the largest 100 firms cover more than 85% of the total market capitalization of all companies in the US financial market and are therefore can restrict our analysis to 100 firms. Furthermore, the results of estimating the FRM based on 100 or 200 firms are very similar if the moving window size is the same. Figure 2 plots both FRM series with the window size $n = 126$, the shape and the trends of them are similar.

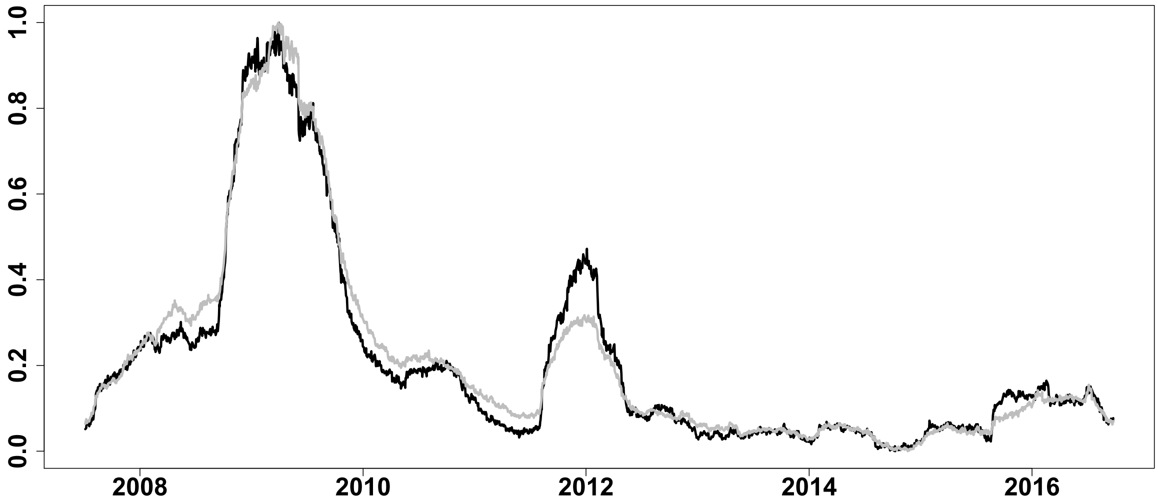

Figure 2: FRM with 100 firms (black) and FRM with 200 firms (grey), moving window size $n = 126$.
FRM_compare_nf

We select six macro state variables to represent the general state of the economy: 1) the implied volatility index, VIX from Yahoo Finance; (2) the changes in the three-month Treasury bill rate from the Federal Reserve Bank of St. Louis; (3) the 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 from the Federal Reserve Bank of St. Louis; (4) the changes in the credit spread between BAA-rated bonds and the Treasury rate from the Federal Reserve Bank of St. Louis; (5) the daily S&P500 index returns from Yahoo Finance, and (6) the daily Dow Jones US Real Estate index returns from Yahoo Finance.

To compute the FRM we use the algorithm as described in Section 2 and with parameter $\tau = 0.05$, i.e. at the tail level. To find the optimal window size, n, we have to make a trade-off. We find that the lasso selection technique performs worse if the window size is too small. Since we use daily data, the moving window size should be larger than 50, so that the estimation for each window is more precise. The results of using different window sizes (we have considered window sizes $n = 63$ (one quarter) and $n = 126$ (half a year)) are shown in Figure 3. The larger the window size, the more lagged, but also the smoother the plot is. Cross correlation can be used to determine the time delay of a

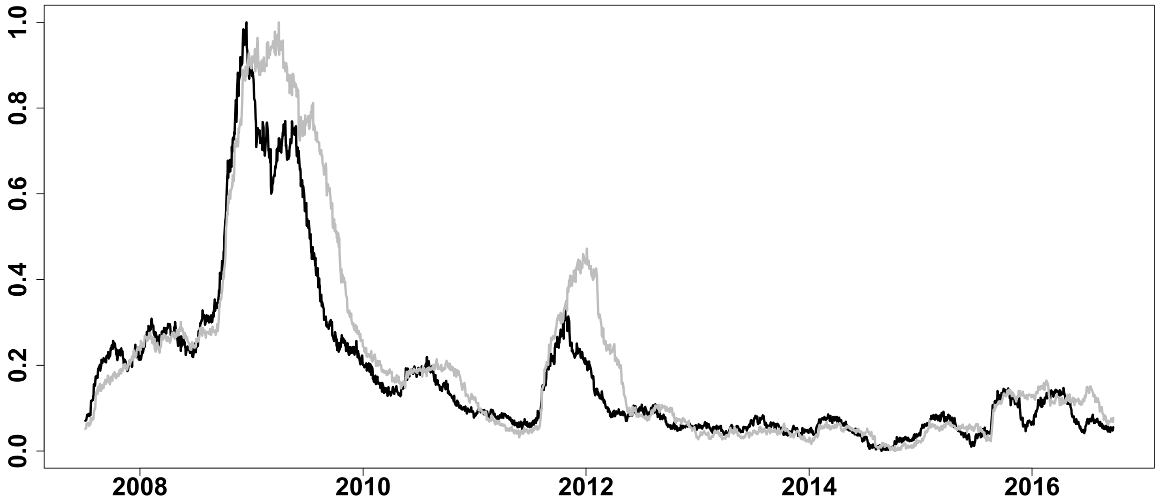

Figure 3: FRM with different moving window size, $n = 63$ (black) and $n = 126$ (grey), both series are scaled into the interval [0,1], from July 6, 2007 until September 23, 2016.

FRM_compare_ws

time series, which we apply here for the estimate of the FRM with $n = 63$ and the FRM with $n = 126$. In Figure 4 and Table 1, the largest autocorrelation between FRM with $n = 63$ and the lagged FRM with $n = 126$ is 0.967 from lag -29 to lag -22. We conclude that the FRM with $n = 63$ leads the FRM $n = 126$ by at least 22 periods. From all the preceding we set the moving window size to $n = 63$.

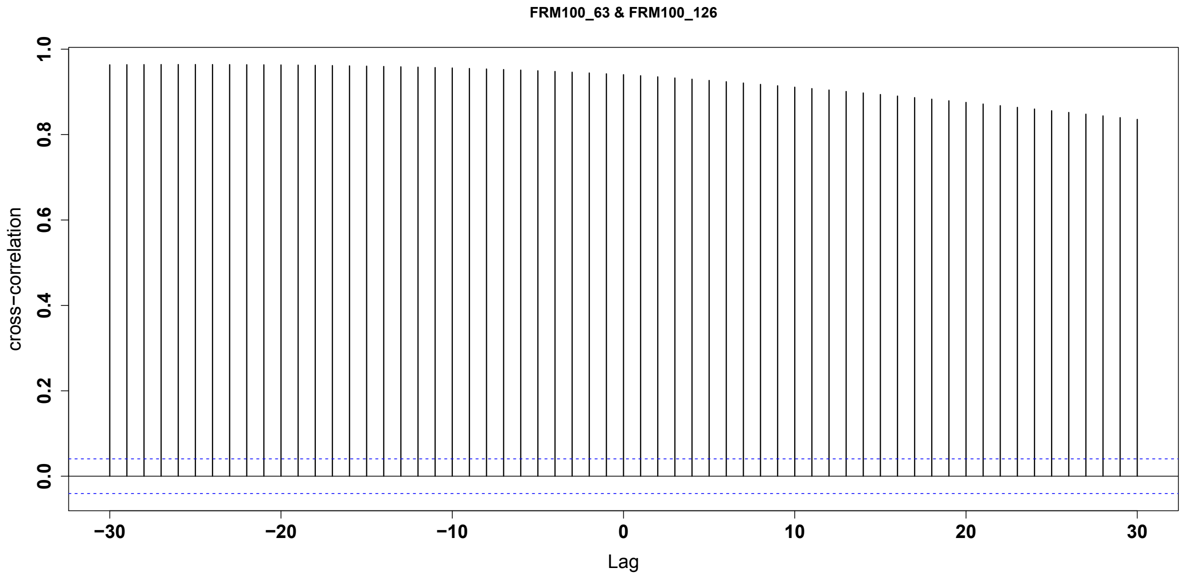

Figure 4: Cross correlation between FRM with $n = 63$ and FRM with $n = 126$, where the number of firms is 100.

FRM_compare_ws

For each firm we have $2,386$ daily observations and 105 covariates (99 firms and 6 macro-state variables). The FRM is the average of the λ's computed for the 100 individual firms.

The λ's for the individual firms are more volatile and less smooth than the average over 100 firms and therefore robuster to reflect the impact from financial events on systemic risk. Figure 5 illustrates this by plotting the λ of firm Wells Fargo (the largest firm by market capitalization) and the FRM.

Lag	-30	-29	-28	-27	-26	-25	-24	-23	-22	-21
Cross correlation	0.963	0.964	0.964	0.964	0.964	0.964	0.964	0.964	0.964	0.963

Table 1: Cross correlation between the estimates of the FRM with $n = 63$ and FRM with $n = 126$.

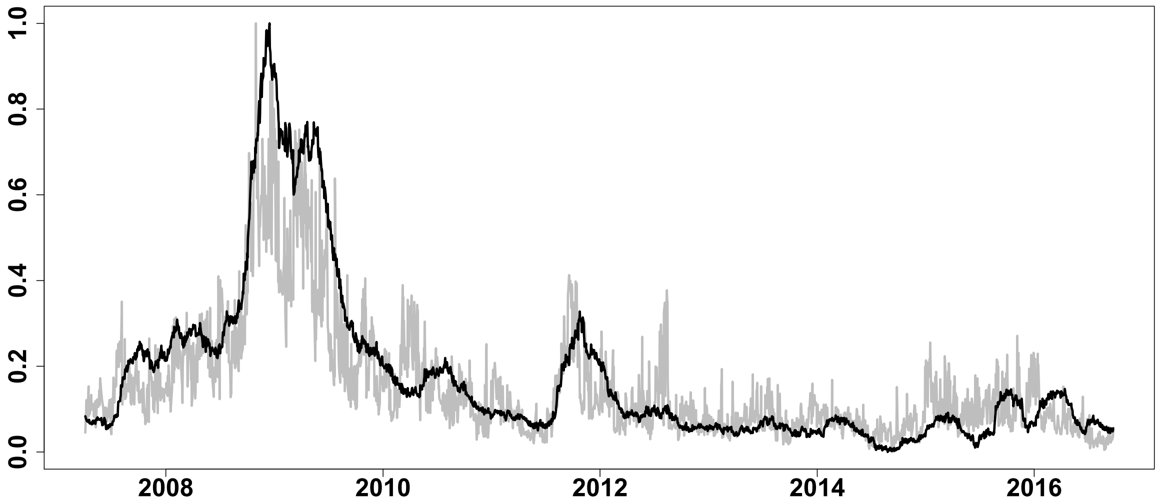

Figure 5: FRM (black) and λ of Wells Fargo (grey), both series are scaled into interval [0,1], from April 5, 2007 until September 23, 2016.

FRM_compare_of

3.2 Computational challenges

We wrote a script to automatically download the data from Yahoo Finance and Federal Reserve Bank of St. Louis. The R package *quantmod* is used. More details and the script are available from Quantnet (FRM_download_data).

The L_1-norm quantile regression used to generate the λ series is computationally intensive and therefore time-consuming, if applied sequentially for a large number of firms, see for instance the code from Quantnet (FRM_lambda_series). Therefore, we consider parallel computing in R to reduce the computation time. R offers several algorithms for performance computing, such as *lapply*, *mclapply*, *parLapply*, *for* and *foreach* [3]. For our purposes the *foreach* loops is the fastest solution, which we use for implementation.

We use the *doParallel* and *foreach* packages in R as developed and proposed by Calaway, Weston, Tenenbaum and Analytics (2015) and Calaway, Weston and Analytics (2015),

[3] The webpage http://www.parallelr.com/r-with-parallel-computing/ provides an overview.

see also Kane et al. (2013). Since we have 100 financial firms and for each firm we have to do the moving window estimation, we use the *foreach* loops twice: the first loop is for the 100 financial firms with the second loop nested in the first loop to perform the moving window estimation. The speed of computation is increased considerably, the script is available from Quantnet: FRM_parallel_compute.

Without the use of parallel computing, i.e. using a processor with four cores for each moving window, it requires around two minutes to generate the FRM estimate for one day. The Research Data Center (RDC) of Humboldt-Universität zu Berlin has provided access to there multi-core servers. Their servers have respectively 24, 32, and 40 cores. By using these servers combined with parallel computing, the average computation time is reduced approximately 12 seconds to obtain a daily value for the FRM. The FRM webpage is updated weekly, which takes only 1 minute to generate the FRM series for five working days.

3.3 Visualization

To implement the visualization of the FRM, we use the JavaScript framework D3.js (or just D3 for Data-Driven Documents), which is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. The QuantNetXploRer is a good example of D3 in power. More information about the D3 architecture, its various designs and the D3-based QuantNetXploRer can be found in Bostock et al. (2011) and Borke and Härdle (2017b). The repository `https://github.com/Quantlet/D3Genesis` contains the development of the main D3 components for the QuantNet visualization together with live examples on GitHub pages.

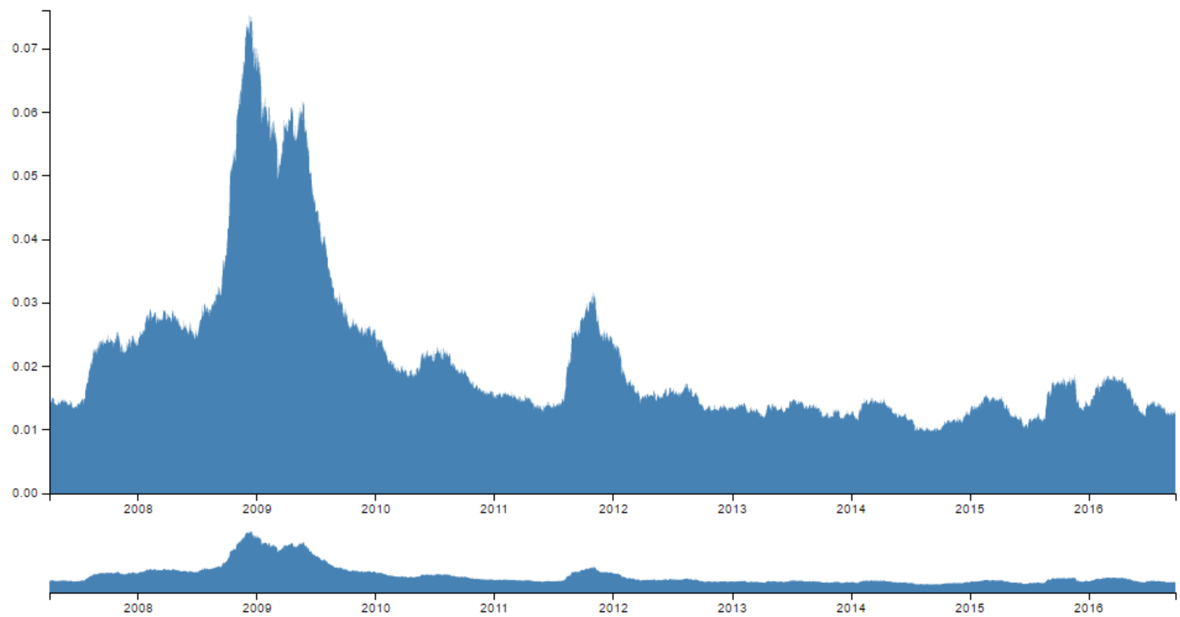

Figure 6: The graph of Financial Risk Meter (FRM).
FRM_parallel_compute

Figure 6 illustrates the D3-based FRM visualization, and more examples are available on the FRM webpage: http://frm.wiwi.hu-berlin.de/. There are two time varying graphs on the FRM webpage. The upper one is the overview of the full FRM series. While the y-axis represents the value of the FRM, the x-axis represents time. The lower graph serves as an interface tool for the upper one. By selecting a time horizon in the lower graph, the upper graph zooms in on the FRM series in this time frame.

3.3.1 Descriptive statistics

Figure 6 shows the FRM series from April 5, 2007 through September 23, 2016. The FRM has no theoretical upper bound. In the time frame under consideration, the maximum value is 0.075, which occurred on December 15, 2008 and the mean value is 0.021. We observe several peaks in the FRM series, which correspond to crises and other events in these periods. Two peaks correspond to the financial crises in 2008 and 2010. The peak in the first quarter of 2009 is at the height of the Great Recession: 800 thousand jobs were lost and the unemployment rate rose to 7.8% in the US, which was the highest since June 1992. Another peak around the fourth quarter of 2011 coincides with the decline in stock markets in August 2011, which was due to fears of contagion of the European sovereign debt crisis to Spain and Italy.

Therefore, the peaks of FRM series identify financial events and their impact on financial and systemic risk. The minimum of the FRM series in the time period under consideration is observed in August 26, 2014, with a value of 0.009. This was a relatively stable period. In this sense, we conclude that the higher value of FRM indicates of higher systemic risk for the US financial market.

3.3.2 Risk levels

We divide risk into five levels with different classifications and colors. The levels of risk are defined as different intervals of quantiles of the FRM. These quantiles are computed based on the past values of the FRM. The color codes are similar to those used by the US Homeland Security Advisory System for the terrorism threat advisory scale. As shown in Figure 7, we have five levels of risk with five color codes. The current risk level is determined by the quantile based on all past FRM observations into which the current λ falls. Table 2 presents the risk levels as well as the colors, descriptions and quantiles of the risk levels.

As an example, on September 23, 2016 the value of FRM was 0.013. Since the maximum of FRM series up to that date was 0.075, the quantile level of the risk measure on September 23, 2016 was 17.3%. Since this is less than the 20%-quantile, we classify the risk on that day as low risk of crisis in the financial market with color green. On the website the current risk level is marked with a cross as shown in Figure 7 for this example.

Figure 7: Risk levels of FRM

Color	Risk level description	FRM quantile
Green	Low risk of crisis in the financial market.	<20
	The incidence of a crisis is less likely than usual.	
Blue	General risk of crisis in the financial market.	20-40
	There is no specific risk of a crisis.	
Yellow	Elevated risk of crisis in the financial market.	40-60
	The incidence of a crisis is somewhat higher than usual.	
Orange	High risk of crisis in the financial market.	60-80
	A crisis might occur very soon.	
Red	Severe risk of a crisis in the financial market.	>80
	A financial crisis is imminent or happening right now.	

Table 2: Risk levels, color codes and quantiles for FRM

4 Causality of FRM and other systemic risk measures

Zbonakova et al. (2016) analyze the factors affecting the value of lambda and summarize that lambda depends on three major factors: the variance of the error term, the correlation structure of the covariates and the number of non-zero coefficients of the model. Since high volatility indicates high risk in finance and the number of non-zero coefficients is related to the connectedness of the financial firms, they provide more theoretical evidence for the FRM as a risk measure. In their application, they find the co-integration relationship between $\hat{\lambda}$ and other systemic risk measures. We extend their idea and use Granger causality analysis to validate our FRM as a systemic risk measure. We select three measures: VIX (see Hallett, 2009), SRISK (see Brownlees and Engle, 2016) as well as the Google trends of the key word "financial crisis" (see Preis et al., 2013).

For the causality analysis we first need to introduce the Vector Autoregression (VAR) model briefly. Lütkepohl (2005) proposes the VAR(P) model as follows:

$$y_t = \alpha + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_P y_{t-P} + \mathbf{u}_t, \tag{3}$$

where $y_t \stackrel{def}{=} (y_{1t}, \ldots, y_{Kt})^\top$, A_i are fixed $(K \times K)$ coefficient matrices, $\mathbf{u}_t$ is a K dimensional process. The coefficients could be estimated by applying multivariate least squares estimation. In order to perform the Granger causality test, the vector of endogenous variables y_t is split into two subvectors y_{1t} and y_{2t} with dimensions $(K_1 \times 1)$ and $(K_2 \times 1)$ and $K = K_1 + K_2$. Then the VAR(P) model can be rewritten as follows:

$$\begin{aligned} y_t &= \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} A_{11,1} & A_{12,1} \\ A_{21,1} & A_{22,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots \\ &\quad + \begin{pmatrix} A_{11,P} & A_{12,P} \\ A_{21,P} & A_{22,P} \end{pmatrix} \begin{pmatrix} y_{1,t-P} \\ y_{2,t-P} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \end{aligned} \tag{4}$$

The null hypothesis of the Granger causality test is that the subvector y_{1t} does not Granger-cause y_{2t}, which is defined as $A_{21,i} = 0$ for $i = 1, 2, \ldots, P$. The alternative hypothesis states that the subvector y_{1t} Granger-causes y_{2t} and is defined as: $\exists\, A_{21,i} \neq 0$ for $i = 1, 2, \ldots, P$. The test statistic follows an F distributions with PK_1K_2 and $KJ - n^*$ degrees of freedom, where J is the sample size and n^* equals the total number of parameters in the above VAR(P) model.

4.1 FRM versus VIX

The VIX series represents the market volatility which can be interpreted as a measure for systemic risk (Hallett, 2009). For reasons of comparability, we standardize these two series by setting the lowest value in the sample to zero and the highest to one. Figure 8 plots the standardized FRM series (thick black line) and the VIX series (thin red line).

The plot clearly shows that both indicators move in the same direction, where the VIX series is more volatile. We also get some evidence of some financial events by observing the corresponding volatility levels of the FRM and VIX. For example, in the end of 2008 there is a sharp upward trend of FRM, whereas the upward trends dominates VIX as well, which corresponds to the bankruptcy of Lehman Brothers on September 15, 2008. Both FRM and VIX have higher values between 2008 and 2010, which corresponds to the time period of the financial crises. After 2013 the values of FRM are relative stable at a low level, while there is similar pattern of VIX, which shows signs of the slow recovery of the global economy from the recession.

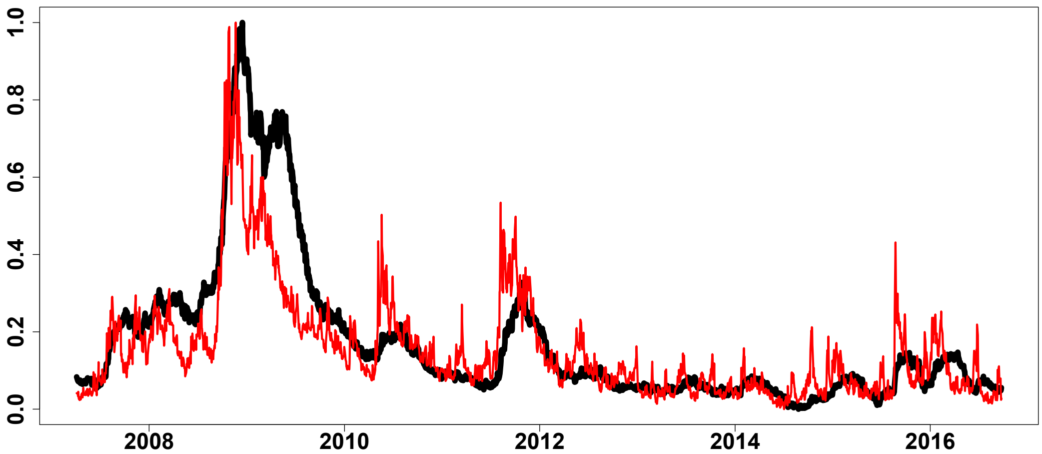

Figure 8: Scaled FRM (thick black line) and VIX (thin red line)

FRM_VIX

Before we perform the Granger causality test, we should test for stationarity of both time series with the Augmented Dickey-Fuller (ADF) test. The null hypothesis of the ADF test is the presence of a unit root in the time series. The results of the test are shown in Table 3. For the FRM series, the p value is larger than 0.05, so we cannot reject the null hypothesis, i.e. the FRM series has a unit root and is non-stationary. We reject the null hypothesis for the VIX series with a p value smaller than 0.05 and conclude that the VIX series is stationary. We do not need to consider the co-integration problem, since only if both series are non-stationary, we should take into account the co-integration. There is a trade-off between using the original data and the transformed (differenced) data to find the causality relationship. Sims (1980) prefers to use the original data. He argues that VAR with non-stationary variables may provide important insights, if one is interested in the nature of relationships between variables. Brooks (2014) also states that differencing will destroy information on any long-run relationships between the series. However, other people argue that the original non-stationary data might lead to untrusted estimation, see Yule (1926) and Granger and Newbold (1974). In our case, we consider both the original data and transformed data.

Firstly, we consider the original data. We choose the VAR order according to four criteria: the Akaike information criterion (AIC), the Hannan-Quinn information criterion (HQ),

Series	p values
FRM	0.28
VIX	0.01
DFRM	0.01

Table 3: p values of ADF test for stationarity

FRM_VIX

Model	AIC	HQ	SC	FPE
FRM and VIX	20	3	3	20
DFRM and VIX	19	8	5	19

Table 4: Suggested order for VAR process by different criteria

Model	Order VAR	PT (asymptotic)	PT (adjusted)	BG	ES
FRM and VIX	3	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	1.1×10^{-07}	1.0×10^{-07}
	11	2.5×10^{-07}	2.0×10^{-07}	1.6×10^{-01}	1.7×10^{-01}
	20	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	3.1×10^{-08}	4.1×10^{-08}
DFRM and VIX	5	2.2×10^{-16}	2.2×10^{-16}	3.2×10^{-08}	3.1×10^{-08}
	8	6.7×10^{-12}	4.9×10^{-12}	1.4×10^{-06}	1.5×10^{-06}
	11	2.3×10^{-09}	1.8×10^{-09}	1.5×10^{-03}	1.7×10^{-03}
	19	1.7×10^{-03}	1.6×10^{-03}	5.5×10^{-08}	7.2×10^{-08}

Table 5: p values of model selection tests

Cause	Effect	p values
FRM	VIX	4.0×10^{-08}
VIX	FRM	6.1×10^{-11}
DFRM	VIX	6.6×10^{-11}
VIX	DFRM	8.7×10^{-13}

Table 6: p values of Granger causality test

FRM_VIX

the Schwarz criterion (SC) and the Prediction Error Criterion (FPE), see Table 4. While

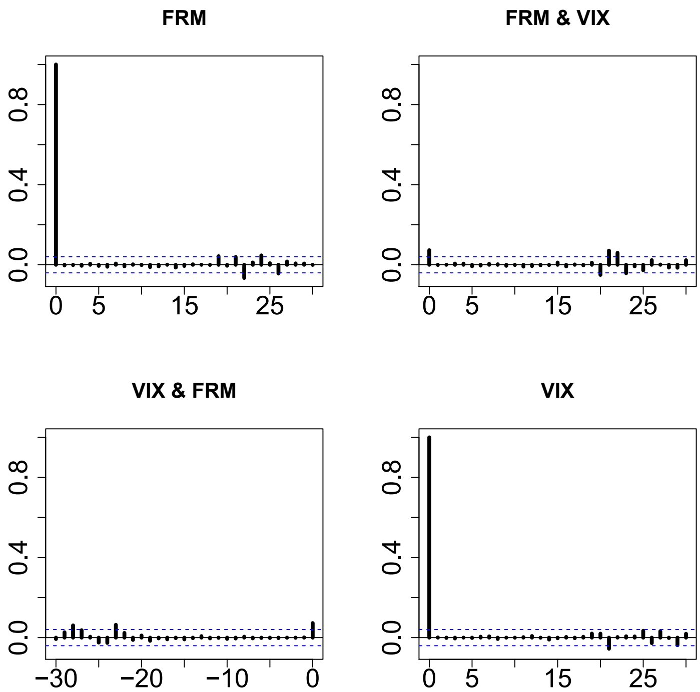

Figure 9: Autoregression functions of FRM and VIX

FRM_VIX

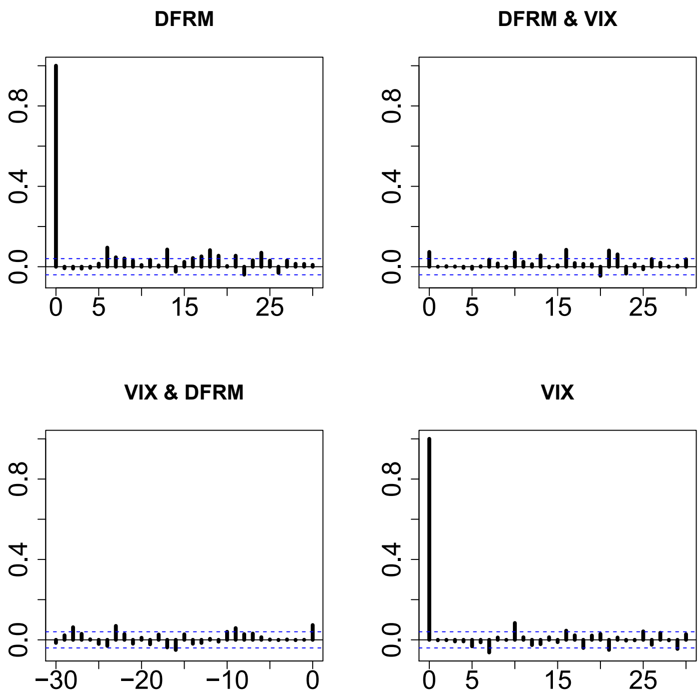

Figure 10: Autoregression functions of DFRM and VIX

FRM_VIX

HQ and SC suggest an order 3 VAR process, AIC and FPE suggest an order 20 process. We fit both VAR models with order 3 and order 20. Next, we check the autocorrelation of the residuals to decide the optimal order. Four tests are carried out: the asymptotic Portmanteau Test, the adjusted Portmanteau Test, the Breusch-Godfrey LM test and the Edgerton-Shukur F test. The null hypothesis of these tests is that there is no first order autocorrelation among residuals. Choosing order 3 and 20 leads to the rejection of all these tests (cf. Table 5). Subsequently we try the other orders and find that with order 11 both the Breusch-Godfrey LM test and the Edgerton-Shukur F tests are passed. Therefore, we select order 11. The autocorrelation function of the residuals is plotted in Figure 9. Table 6 shows the results of the Granger causality test. All p values are smaller than 0.05 which indicates that the null hypothesis is rejected. Therefore, FRM Granger causes VIX, and also VIX Granger causes FRM.

Next, we consider the transformed series. Since FRM is non-stationary, we take the first difference. The transformed series is called as DFRM. In Table 3 we see that DFRM is stationary. Then the same procedure as before is performed. While HQ suggests an order 8 process, SC suggest an order 5, and AIC and FPE both suggest an order 19 (cf. Table 4). After checking the four tests for autocorrelation of the residuals, we conclude that the optimal order is 19. Although it does not pass the autocorrelation test, the p value is close to the critical value 0.05, and the autocorrelation function confirms this result (cf. Table 5 and Figure 10). The result of the Granger causality test is summarized in Table 6. We find that all p values are significantly smaller than 0.05, which indicates that the null hypothesis is rejected. Therefore we conclude that DFRM Granger causes VIX, and also VIX Granger causes DFRM.

4.2 FRM versus SRISK

SRISK is a macro-finance measure of systemic risk (Acharya et al., 2012; Brownlees and Engle, 2016). Our data on SRISK for the US are obtained from V-Lab [4]. We also standardize SRISK, so that both series are comparable on the same scale. Figure 11 plots the standardized FRM series (thick black line) and the SRISK series (thin blue line). We see that there is a peak in the first quarter of 2008 for SRISK, but afterwards FRM and SRISK have similar patterns. Especially during the beginning of 2010 and the beginning of 2012, the two series have a similar shape.

Variables	p-values
FRM	0.48
SRISK	0.10

Table 7: p values of ADF test for stationarity for FRM and SRISK

We perform the same procedure as in section 4.1. The results of the ADF test for the SRISK series in Table 7 show that the series is non-stationary. Since the FRM series

[4] See the Systemic Risk Analysis Welcome Page: https://vlab.stern.nyu.edu/welcome/risk/

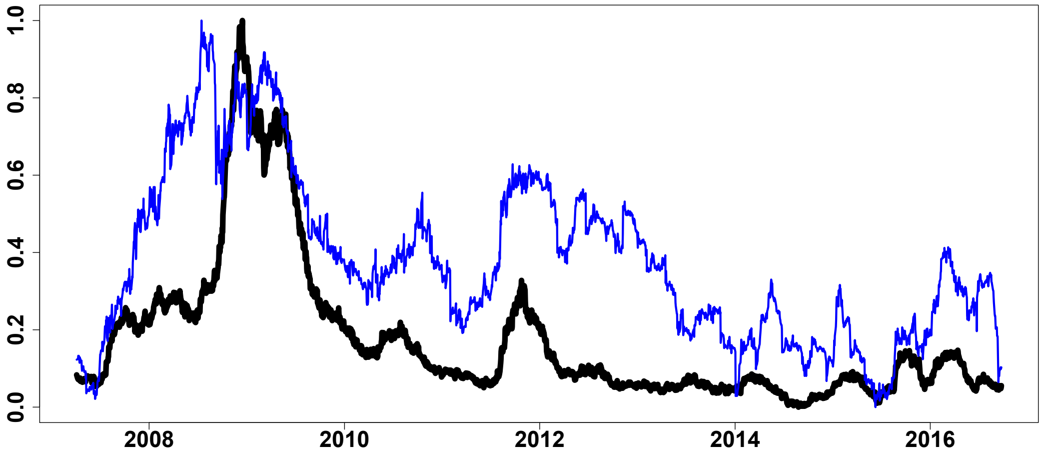

Figure 11: Scaled FRM (thick black line) and SRISK (thin blue line)

FRM_SRISK

Explanatory (Cause)	Response (Effect)	Value of test-statistic	Critical value at 5%
FRM	SRISK	-3.1	-1.95
SRISK	FRM	-2.7	-1.95

Table 8: Results of Engle Granger 2-step co-integration test

FRM_SRISK

is neither stationary, we consider the co-integration of them. From Granger (1988) we know that if both series are co-integrated, then there must be Granger causality between them in at least one way. We perform the Engle Granger 2-step test for co-integration, which is suitable for bivariate time series. In the first step, the linear regression of FRM on SRISK is carried out, i.e. FRM is the explanatory variable and SRISK the response variable. In the second step, we test the residuals of the aforementioned linear regression. If these residuals are stationary, then there is co-integration of FRM and SRISK. The null hypothesis of this test is that the residuals are non-stationary. The result of this test are summarized in Table 8. We conclude that FRM and SRISK are co-integrated, in other words, FRM Granger causes SRISK. If we regress SRISK on FRM, i.e. SRISK is the explanatory variable and FRM the response variable, we also conclude that SRISK and FRM are co-integrated, which indicates that SRISK Granger causes FRM. We thus conclude that there is mutual causality between FRM and SRISK.

4.3 FRM versus Google Trends

Finally, we analyze the relationship between FRM and Google Trends (GT) for the keyword "financial crisis". Google Trends provides data on the search volume of particular words and phrases relative to the total search volume. This can be disaggregated by countries. If a keyword is more frequently searched for, this might indicate a particular interest. Preis et al. (2013) analyzed the data related to finance from Google Trends, and find that Google Trends data did not only reflect the current state of the stock markets, but may have also been able to forecast certain future trends. We use Google Trends for the keyword "financial crisis", assuming that more people will search for this term if they feel the risk for a financial crisis is high. The Google Trends data are weekly data. To allow for comparison with the FRM we apply cubic interpolation to estimate daily data from the weekly Google Trends series. This series is compared with the daily FRM series. Figure 12 plots both the daily FRM series as well as the cubic interpolated Google Trends daily series. Both series are standardized to the interval zero-one for comparison. We observe some co-movement between both series.

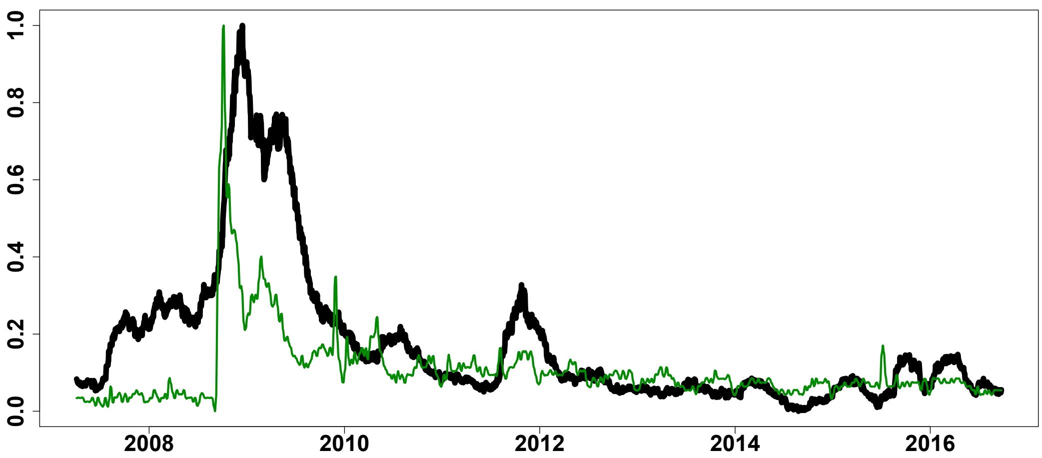

Figure 12: Scaled FRM (thick black line) and Google Trends (thin green line)
FRM_GT

The ADF test shows that the GT series is stationary (cf. Table 9). We perform two tests for the relationship between the two series. Firstly, we consider the original data of FRM, then we consider the transformed data. We perform four criteria to find the optimal order of VAR model. As the results in Table 10 show, all the criteria suggest an order 20 VAR process. Therefore, we apply an order 20 VAR model. Next, the autocorrelation of the residuals is tested. Although none of the tests can be passed (cf. Table 11), we have no better choice for the order than 20. The autocorrelation function of residuals are plotted in Figure 13. Table 12 shows the results of the Granger causality test. All p values are significantly smaller than 0.05, which indicates that the null hypothesis is rejected. Therefore, FRM Granger causes GT, and GT Granger causes FRM.

For the first differenced FRM, i.e. DFRM, the same procedure is used. In Table 10 all the criteria suggest an order 20 VAR process. The result of the autocorrelation tests are presented in Table 10. Although none of the tests is passed, we still use order 20. The autocorrelation function of the residuals is shown in Figure 14. Table 12 shows the results of the Granger causality test. All p values are significantly smaller than 0.05, which indicates that the null hypothesis is rejected. Therefore, DFRM Granger causes GT, and GT Granger causes DFRM.

Variables	p-values
FRM	0.48
GT	0.01
DFRM	0.01

Table 9: p values of ADF test for stationarity for FRM and GT

Model	AIC	HQ	SC	FPE
FRM and GT	20	20	20	20
DFRM and GT	20	20	20	20

Table 10: Suggested order for VAR process by different criteria

Model	Order	PT (asymptotic)	PT (adjusted)	BG	ES
FRM and GT	20	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$
DFRM and GT	20	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

Table 11: p values of model selection tests

Cause	Effect	p-values
FRM	GT	1.1×10^{-10}
GT	FRM	2.1×10^{-12}
DFRM	GT	6.8×10^{-11}
GT	DFRM	4.1×10^{-10}

Table 12: p values of Granger causality test

FRM_GT

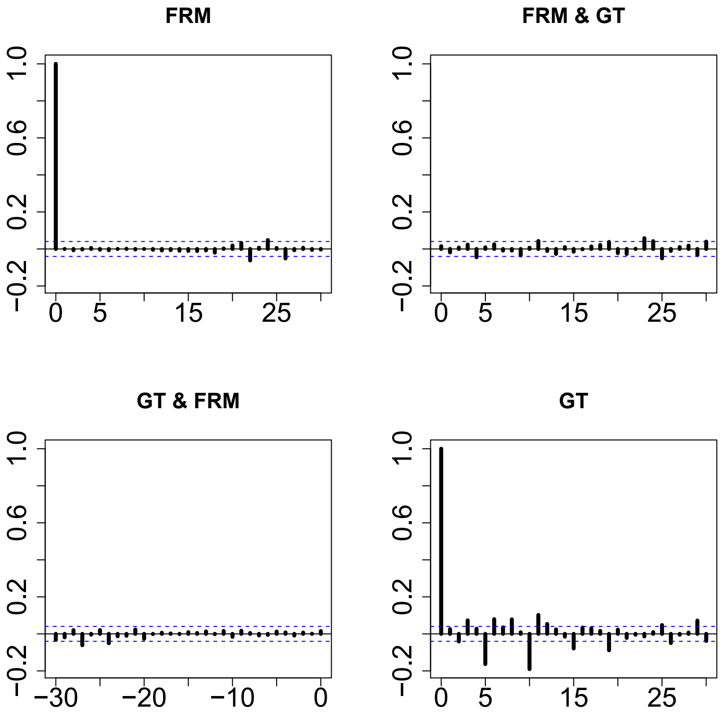

Figure 13: Autoregression functions of FRM and GT

FRM_GT

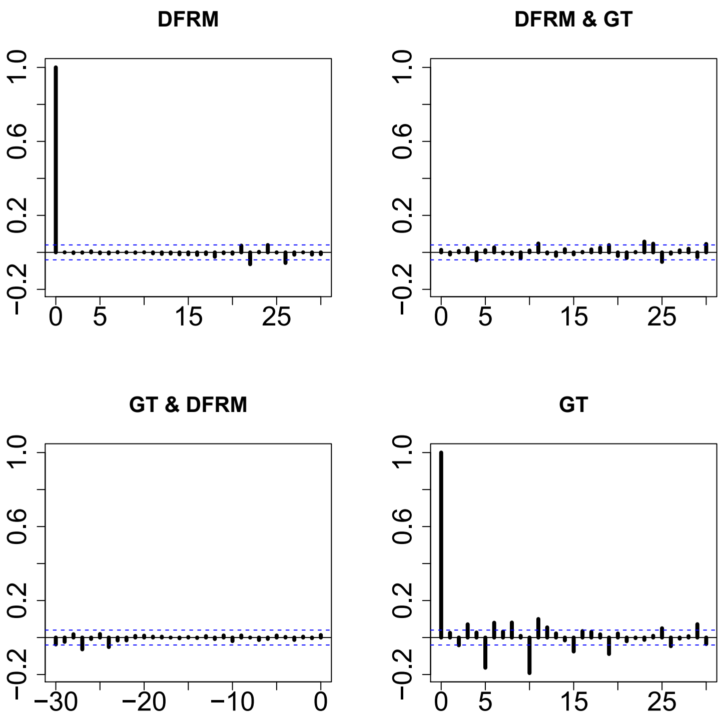

Figure 14: Autoregression functions of DFRM and GT

FRM_GT

5 Software implementation in R

Koenker and Mizera (2014) survey some recent developments of convex optimization and describe some implementations of these methods in R. Quadratic programming (QP), as part of convex optimization, involves the minimization of a positive semi-definite quadratic objective function subject to polyhedral constraints. There are many applications of QP in statistics, typically involving Gaussian likelihoods constrained by some form of linear inequalities. Shape constrained regression examples have gained recent attention, and the introduction of sparse regularization methods like lasso, has greatly stimulated interest in computational methods for such problems. One of the most familiar statistical QP applications in recent times has been the lasso estimator of Tibshirani (1996).

Standard quantile regression (QR) models can be estimated with the `rq()` function of the **quantreg** package (Koenker, 2016). However, software implementations for computing solution paths of lasso penalized QR are rare. **hqreg** (Yi, 2016) is such an example. This R package is relatively new (it was published for the first time on 21 June 2015), its version history is trackable via `https://github.com/cran/hqreg/commits/master`. The main advantage is its C optimization: `https://github.com/cran/hqreg/blob/master/src/hqreg.c`. Yi and Huang (2015) demonstrate both the convergence properties of the proposed algorithm and the numerical experiments, showing that their package implementation is very efficient and scalable to ultra-high dimensions.

Another available R implementation is the supplementary code of Li and Zhu (2008). At the time of the early stage development of the FRM project, only the latter code was known and available. Therefore, the current lasso penalized QR implementation of FRM relies on the idea of Li and Zhu (2008). In the following, numerical experiments and benchmarks will be provided in order to evaluate the speed and efficiency of the current FRM version.

5.1 RiskAnalytics package

In order to integrate and facilitate the research, calculation and analysis methods around the FRM project, the R package **RiskAnalytics** (Borke, 2017) has been developed. Its main goal is to provide data processing and parallelized quantile lasso regression methods for risk analysis based on NASDAQ data, Yahoo Finance data and the macro variables as described in Section 3.1. The derived "Risk Analytics" can help to forecast and evaluate the systemic risk for the corresponding markets.

As member of the Research Data Center (`https://rdc.hu-berlin.de`) Lukas Borke was involved into the development of the FRM project from the very beginning, having the main tasks: automation of data collection, optimization and parallelization of code, and data visualization. Based on this experience, the functionality of the **RiskAnalytics** package is subdivided into 4 major software components:
1) data processing (*get_data.R*);
2) parallel computing (*parallel_calculation.R*);

3) QR methods (*qrL1.R*);
4) “Risk Analytics” (*analytics.R*);

Every software component contains several related functions. Their interaction is presented in Listings 1, 2, 3, 4 and 5.

5.1.1 RiskAnalytics package: data extraction and analysis part

Listing 1 demonstrates the data extraction and analysis part of the **RiskAnalytics**. The functions `get.nasdaq.companies`, `get.yahoo.data` and `get.macro.data` are responsible for real time processing of NASDAQ, Yahoo Finance and Federal Reserve Bank of St. Louis data. `get.nasdaq.companies` extracts the top NASDAQ companies (sorted by their market capitalization) from the web resource `http://www.nasdaq.com/screening/companies-by-industry.aspx?industry=Finance` by means of the package **RCurl** (Lang and the CRAN team, 2016). `get.yahoo.data` provides daily log returns of the selected NASDAQ companies by use of the package **quantmod** (Ryan, 2016). `get.macro.data`, in its turn, employs both approaches: Yahoo Finance via **quantmod** for the download of the VIX, GSPC (S&P500) and IYR (iShares Dow Jones US Real Estate) macro variables, and direct downloads of the other 3 macro variables from the corresponding web resources on `https://fred.stlouisfed.org/`.

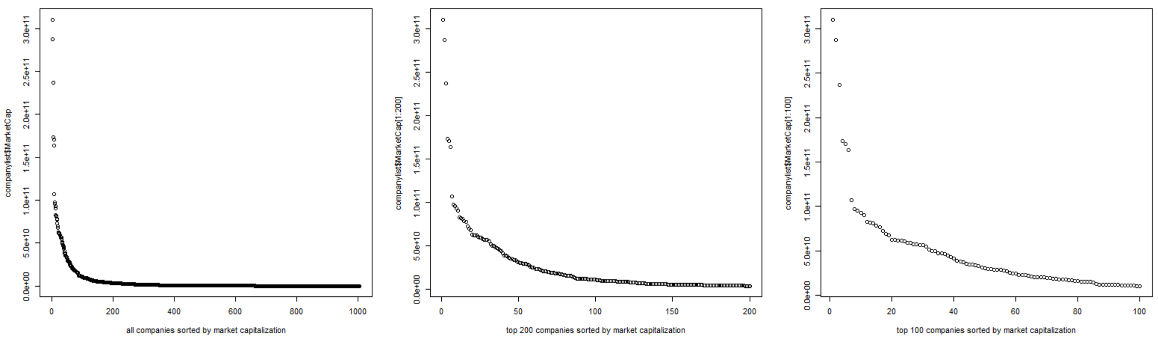

Figure 15: NASDAQ companies sorted by the market capitalization: all (left), top 200 (middle) and top 100 (right), produced via *get.nasdaq.companies*

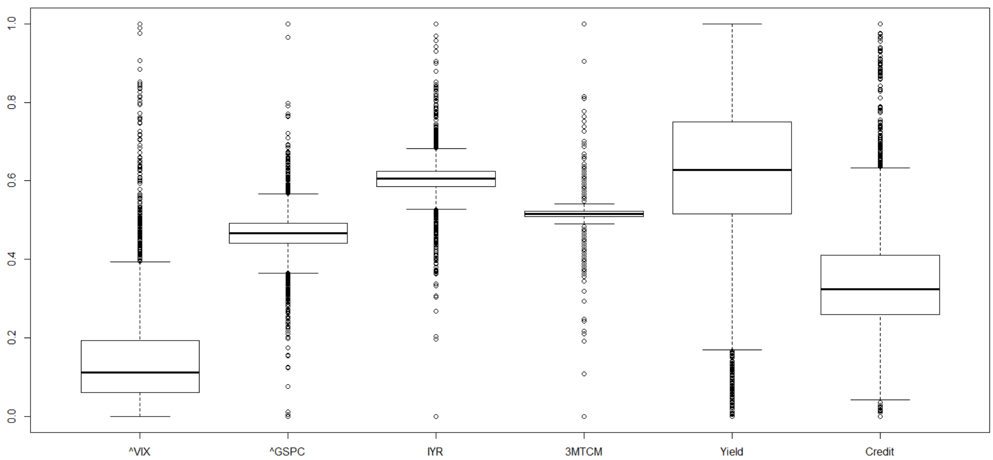

Figure 16: Box plots of macro variables produced via *get.macro.data*

The helping function `combine.data` combines all previously obtained time series in an appropriate time and date format. Additionally, the dimension and the preview of a sub sample of the resulting data frame object is displayed. The latter can be controlled by the parameter `summary_dim`, see also Listing 1. All aforementioned functions provide additional information and, where appropriate, graphical plots for better audit and validation checks of the extracted data, see e.g., Figure 15, 16.

```
#------------------------------------------------
# Initialization
#------------------------------------------------
library(snow)
library(RiskAnalytics)
work_dir = "c:/r/frm/2017"
max_companies = 100

#------------------------------------------------
# Load data
#------------------------------------------------
companylist = get.nasdaq.companies()

system.time( yahoo_data <- get.yahoo.data(companylist, max_comp_num = max_companies,
    from_date = "2006-12-29") )
# truncated output for illustration
[1] "97 : SBNY"
[1] "98 : ZION"
[1] "diff length : CIT"
[1] "diff length : APO"
[1] "99 : WRB"
[1] "100 : SEIC"
      User     System    Elapsed
      7.85       1.44      58.20

system.time( macro_data <- get.macro.data(from_date = "2006-12-28") )
      User     System    Elapsed
      1.05       0.06       5.13

final_data = combine.data(yahoo_data, macro_data, summary_dim = c(1:3, 102:107))
[1] "Dimension of the final data: 2534 * 107"
        Date         JPM         WFC      ^VIX     ^GSPC       IYR     3MTCM
1  03/01/2007 0.002290948 0.005049110 0.02353107 0.4414408 0.5978950 0.4904459
2  04/01/2007 0.002493227 0.001677339 0.03029449 0.4577132 0.6204483 0.5159236
3  05/01/2007 -0.008335091 -0.005602276 0.02282655 0.4695943 0.6035441 0.5222930
4  08/01/2007 0.003342404 -0.002812915 0.03170354 0.4337065 0.5632682 0.5286624
5  09/01/2007 -0.004179761 0.002532009 0.02973087 0.4744425 0.6035341 0.4840764

data.analytics(yahoo_data, macro_data)
# truncated output for illustration, correlation matrix of the macro var's
       ^VIX ^GSPC  IYR 3MTCM Yield Credit
^VIX    1.00 -0.14 -0.11 -0.06 0.26 0.55
^GSPC  -0.14  1.00  0.81 -0.02 0.00  0.01
IYR    -0.11  0.81  1.00 -0.04 0.01  0.02
3MTCM  -0.06 -0.02 -0.04  1.00 0.00  0.00
Yield   0.26  0.00  0.01  0.00 1.00  0.36
Credit  0.55  0.01  0.02  0.00 0.36  1.00
```

Listing 1: RiskAnalytics application example: data extraction and analysis part

The function `data.analytics` from Listing 1, which is actually a part of the "Risk Analytics" software component, provides descriptive statistics for both the NASDAQ companies and the macro variables. All statistical information vital for the subsequent QR methods is summarized in a brief overview. For instance, it becomes immediately obvious that the macro variables will be dominant regressors due to their larger Euclidean norms, compared to those of the NASDAQ companies (see the box plots in Figure 17). Together with the output in Figure 16 and 18, one can easily conclude that the macro variables VIX (1), "Yield spread" (3) and "Credit spread" (4) (see Section 3.1 for the enumeration assignment) will be "driving factors" in the QR process because of their high variances. Furthermore, `data.analytics` returns also the correlation matrix of all six macro variables, revealing that the aforementioned variables VIX, "Yield spread" and "Credit spread" have positive correlations among each other, see Listing 1. In the light of this technical analysis, it is hardly surprising that both the FRM and VIX time series reveal a similar behavior, see Section 4.1 and Figure 8.

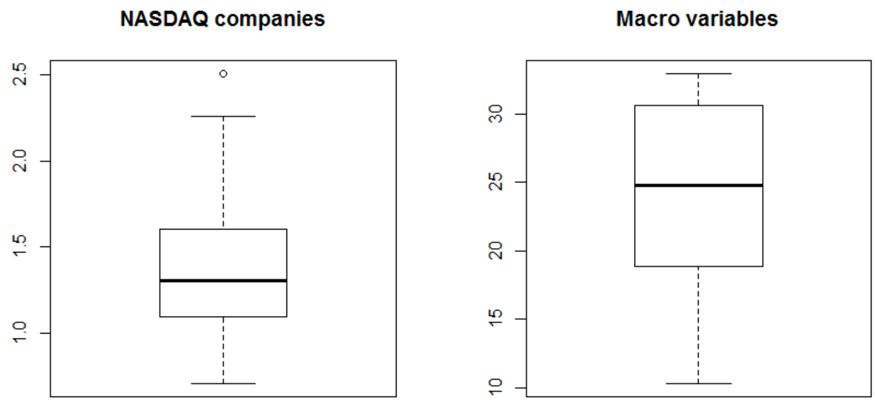

Figure 17: Box plots of the euclidean norms of the Yahoo Finance data/companies (left) and the macro variables (right), produced via *data.analytics*

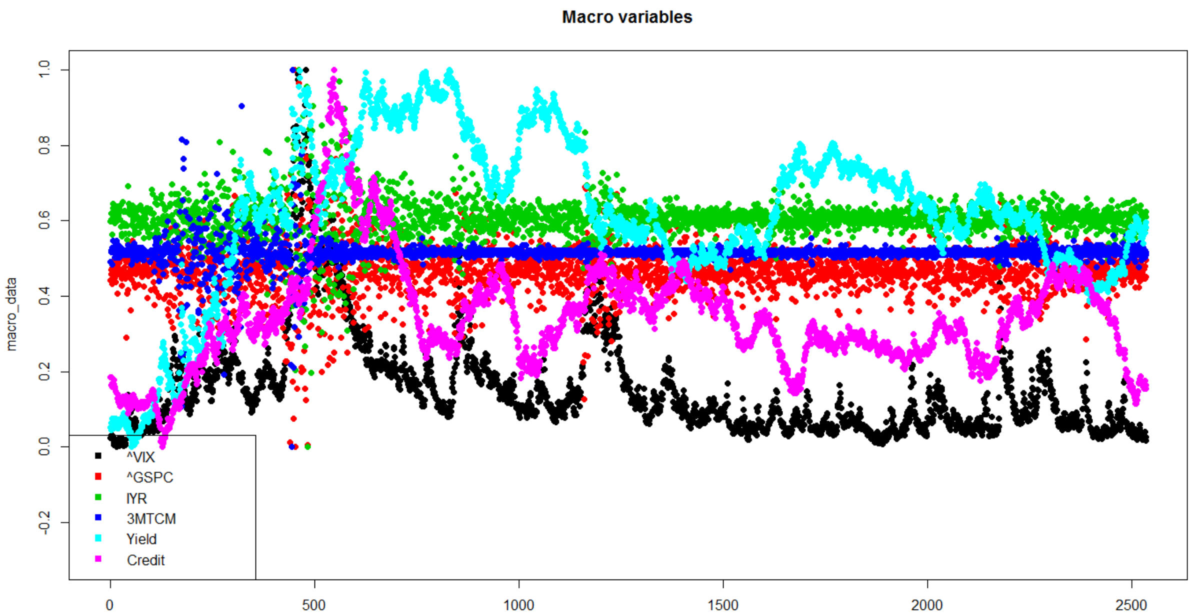

Figure 18: Plot of the macro variables, produced via *data.analytics*

According to Listing 1, the data processing component extracts all needed data in around one minute. Additionally, the `data.analytics` function provides statistical information for the further QR process. All obtained data are stored in the RAM, hence no further write or storage operations are required, and the real time data can be passed over to the next component: parallel computing.

5.1.2 RiskAnalytics package: parallel computing part

Listing 2 shows the execution and benchmark results of the parallel computing component of **RiskAnalytics**. Based on the packages **snow** (Tierney et al., 2016) and **snowfall** (Knaus, 2015) and the lasso penalized QR implementation of Li and Zhu (2008), the calculation of the QR method is performed for all moving windows and all NASDAQ companies. The most important parameters of the function `parallel.lasso.computation` are `max_companies`, `new_days`, `parallel_cpu`, `p`, `winsize` meaning: 1) number of desired NASDAQ companies, the parallelization is performed along this dimension; 2) number of desired moving windows within the total data observation time frame; 3) number of available CPU's for the parallel computing via **snowfall**; 4) desired quantile value for the QR method; and 5) the length of the moving window. Most of these parameters have default values as displayed in Listing 2.

```
#-------------------------------------------------
# Calculate data
#-------------------------------------------------

# by default: new_days = 5, parallel_cpu = 4, p = 0.05, winsize = 60

# main calculation for the FRM visualization
parResult = parallel.lasso.computation(final_data, max_companies, work_dir = work_dir,
     new_days = 2469, parallel_cpu = 32, winsize = 63)
# R Version: R version 3.3.2 (2016-10-31)
# snowfall 1.84-6.1 initialized (using snow 0.4-2): parallel execution on 32 CPUs.
# Stopping cluster
#    user   system elapsed
#    1.45     3.43 54895.78

# test benchmark for 200 working days, ca. 10 months
parResult = parallel.lasso.computation(final_data, max_companies, work_dir = work_dir,
     new_days = 200, parallel_cpu = 32)
# Stopping cluster
#    user  system elapsed
#    0.14    0.14 4287.58

# test benchmark for 5 working days, 1 working week
parResult = parallel.lasso.computation(final_data, max_companies, work_dir = work_dir,
     parallel_cpu = 32)
# Stopping cluster
#    user  system elapsed
#    0.17    0.14  115.89
```

Listing 2: RiskAnalytics application example: parallel computing part

For each company and each moving window the QR results are stored in the data structure `parResult`. The latter is basically a list with elements corresponding to the companies. Every list element j contains the lambda values (λ_j) and beta coefficients (β_j) from the

QR procedure. For a given company j, the lambda values are a vector enumerated by the calculated days `new_days`, and the beta coefficients are a matrix, whose rows are the calculated days `new_days` and whose columns are the regressors/covariates.

According to Li and Zhu (2008), the computational complexity of the L_1-norm QR algorithm is $\mathcal{O}(p\min(n,p)^3)$, with n being the length of the moving window and p the number of covariates. The main calculation for the FRM visualization is performed with $n = 63$ and $p = 105$ (99 companies except the regressed one and 6 macro variables), see also Listing 2. In addition to the basic complexity, we have to deal with two further dimensions, i.e. n_c (`max_companies` or *number* of *companies*) and n_w (`new_days` or *number* of moving *windows*).

In summary, the `parallel.lasso.computation` function for the main calculation of the *FRM* lambda time series has a computational complexity of

$$\mathcal{O}(n_c n_w)\mathcal{O}(p\min(n,p)^3), \tag{5}$$

which results in approximately 6.5×10^{12} basic calculations, if we compute the *FRM* lambda for $n_w = 2469$ (around 10 years).

The time complexity benchmarks for four cases: $n_w = 2469$, $n_w = 200$, $n_w = 10$ and $n_w = 5$ are provided in Table 13, see Listing 2 for some examples. The tests were performed on a RDC (Research Data Center, `https://rdc.hu-berlin.de`) Windows server with 16 physical and 32 logical cores and Intel Xeon CPU E5-2690 0 @ 2.90 GHz. In each case `max_companies` was equal to 100, `parallel_cpu` = 32, `p` = 0.05. The corresponding length of the moving window n (`winsize`) and the number of moving windows n_w (`new_days`) are given in the table columns.

n (window size)	n_w	time in seconds	time in minutes	time in hours
60	5	116	2	0.03
60	10	222	4	0.06
60	200	4288	71	1.19
63	2469	54896	915	15.25

Table 13: Time complexity benchmarks for *parallel.lasso.computation* of the **RiskAnalytics** package

As can be expected from Formula 5, the running time of `parallel.lasso.computation` scales in proportion to n_w. For a better comparison, the same time measurements are displayed in seconds, minutes and hours, respectively. As main results of the time complexity benchmarks, we can conclude that:
I) The lasso penalized QR implementation in FRM can be performed within 2 minutes for the calculation of 5 working days and within 15 hours for a time period of 10 years, which shows that the QR calculation is feasible on a contemporary computer with 16 physical cores.
II) For the increase of the speed, only the physical CPU cores are relevant, what means that the calculations can be performed on a usual home PC with 4 CPU cores, like for

example Intel Core i5-2500 with 4 physical cores. In this case, the time demand must be multiplied by factor of 4 (16 cores ÷ 4 cores).
III) The memory demand for the storage of all necessary data and calculation results is very modest and is mainly dictated by the dimensions of the data matrices and frames and the data structure `parResult`. Saved as files, the data object `final_data` from Listing 1 and `parResult` from Listing 2 require around 2 MByte and 60 MByte, respectively.

The `parallel.lasso.computation` function accepts some additional optional parameters for minor validation outputs and allowing to save the calculation results as file outputs. By default, the parallel computing component operates as an "in-memory application" without requiring any disk Input/Output operations.

5.1.3 RiskAnalytics package: QR.analytics part

The code examples in Listing 3 demonstrate the *QR.analytics* part of **RiskAnalytics**, the former being a subset of the "Risk Analytics" software component. QR.analytics comprises 3 functions: `QR.regressors.stats`, `QR.beta.stats` and `QR.variance.vs.beta`. The output `parResult` from the parallel computing part serves as an "object of investigation".

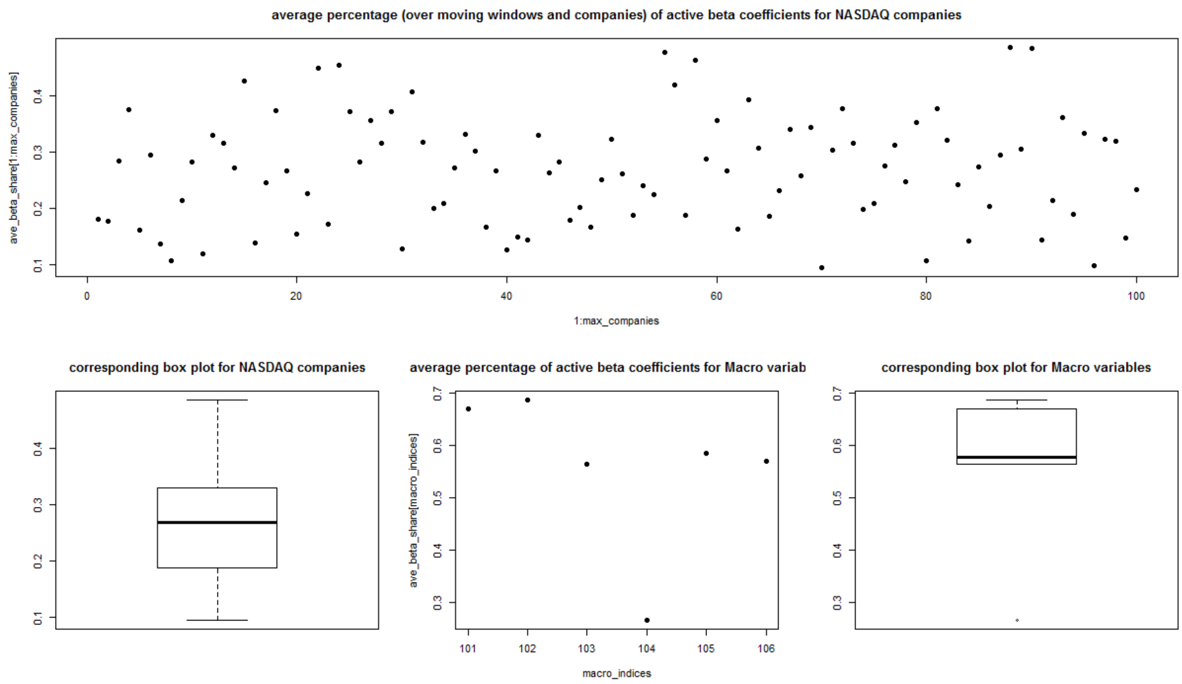

Figure 19: Average percentage (over moving windows and companies) of active beta coefficients for NASDAQ companies and macro variables and the corresponding box plots, produced via *QR.beta.stats*

The functions `QR.regressors.stats` and `QR.beta.stats` analyze the structure of the beta coefficients from the QR process. `QR.regressors.stats` provides the frequency of the covariates for a given percentage threshold `sel_threshold` and the filter value `min_regressed_comp`. For instance, for a given `sel_threshold` = 0.55 and `min_regressed_comp` = 10, we see in the first part of Listing 3 that only the following covariates: 88, 101, 102, 103, 105, 106 (88 is the number of a NASDAQ company, numbers higher than 100 are macro variables) have non-zero beta coefficients in the QR of a company. Additionally, we have the restrictions that the filtered and displayed covariates are active regressors in at least 55% of all moving windows (n_w) for at least 10 companies. For `sel_threshold` = 0.55 and `min_regressed_comp` = 10 we can conclude that the company with the number 88 is an active regressor for some 11 companies, being present in at least 55% of all moving windows for each of those 11 companies. The macro variable with the number 101 (VIX), on the other hand, is an active regressor with non-zero beta's in all 100 NASDAQ companies, being present in at least 55% of all moving windows for each of them. The second most influential regressor is the macro variable with the number 102 (S&P500), it is an active regressor for 99 NASDAQ companies (in at least 55% of all moving windows).

```
#-----------------------------------------------
# QR.analytics: QR.regressors.stats
#-----------------------------------------------
sapply( c(0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.72), function(x) { QR.regressors.stats(
    parResult, sel_threshold = x, min_regressed_comp = 10 )$R_tab_min_regressed_comp
    })

[[1]]
 15  22  24  31  55  56  58  63  67  81  88  90 101 102 103 105 106
 31  46  51  11  74  21  55  15  10  11  79  73 100 100 100 100 100

[[2]]
 22  24  55  58  88  90 101 102 103 105 106
 13  15  23  21  43  37 100 100 93 100 98

[[3]]                       [[4]]
 88 101 102 103 105 106      101 102 103 105 106
 11 100 99  67  86  72        98  99  15  30  16

[[5]]         [[6]]        [[7]]
101 102        101 102      102
 76  87         14  38       16

#-----------------------------------------------
# QR.analytics: QR.beta.stats
#-----------------------------------------------
ave_beta_share = QR.beta.stats(parResult)
which(ave_beta_share > 0.5)
[1] 101 102 103 105 106
which(ave_beta_share > 0.666)
[1] 101 102

#-----------------------------------------------
# QR.analytics: QR.variance.vs.beta
#-----------------------------------------------
variance_vs_beta = QR.variance.vs.beta(final_data, ave_beta_share)
# truncated output for illustration
$corr_comp_vars_beta
[1] 0.5752723
$corr_macro_vars_beta
[1] 0.3024359
```

Listing 3: RiskAnalytics application example: QR.analytics part

Iterating through different `sel_threshold` values (c(0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.72)) via the `sapply` function, we can observe that with the increasing threshold only the macro variables remain as active regressors for the NASDAQ companies. For `sel_threshold` = 0.7 only the macro variables 101 and 102 serve as regressors for 14 and 38 companies, respectively. Reaching `sel_threshold` = 0.72, the macro variable 101 (VIX) vanishes, what means that it is a regressor of maximally 9 companies, whereas the macro variable 102 (S&P500) is still an active regressor for some 16 companies.

While `QR.regressors.stats` provides the frequency of the covariates based on the active set of the beta coefficients, `QR.beta.stats` analyzes the beta coefficients themselves. Basically, `QR.beta.stats` calculates the average percentage (over all moving windows and companies) of active beta coefficients for the covariates. The average percentage of active beta coefficients with the value 0.2, for instance, would mean that the covariate, which has this percentage, acts as an active regressor in exactly 20% of all moving windows (n_w) averaged over all companies. The vector of the average percentage of active beta coefficients is stored in the variable `ave_beta_share` (each element corresponds to a covariate), see Listing 3. Besides the corresponding plots and box plots for the NASDAQ companies and macro variables, which are provided by `QR.beta.stats` based on `ave_beta_share` (see also Figure 19), `ave_beta_share` can be subjected to further statistical analysis. For example, `ave_beta_share` is minimal (= 0.095) for the company with the number 70 (Loews Corporation (L)) and is maximal (= 0.484) for the company with the number 88 (CBRE Group, Inc. (CBG)). The distribution of the `ave_beta_share` values of the macro variables is provided in Figure 19 and Table 14.

The statistical analysis provided by the functions `QR.regressors.stats` and `QR.beta.stats` reveals that the macro variables have a dominant effect on the regressed companies. Except the macro variable with the number 104 (3MTCM: the changes in the three-month Treasury bill rate) all other macro variables have an average percentage of active beta coefficients of at least 56%. The two most influential regressors are the variables 101 and 102 (VIX and S&P500). Averaged over all moving windows and regressed companies, VIX and S&P500 are present in around two thirds of the performed quantile regressions.

	^VIX	^GSPC	IYR	3MTCM	Yield	Credit
Variance	0.0187	0.0042	0.0032	0.0013	0.0447	0.0227
Beta_share	0.6695	0.6865	0.5635	0.2651	0.5840	0.5698

Table 14: Variances versus average percentage of active beta coefficients of the macro variables, produced via *QR.variance.vs.beta*

An interesting observation is the relationship between the variances of the covariates and the average percentages of active beta coefficients as calculated in `ave_beta_share`. The function `QR.variance.vs.beta` examines this issue. Among other details, this function delivers the correlations between the variances and the `ave_beta_share` values (0.575 for the companies and 0.302 for macro variables, see the last part of Listing 3), the corresponding plots and scatter plots in Figure 20, and the output for Table 14. In particular, the scatter plot in Figure 20 illustrates the positive correlation between the variances of the NASDAQ companies and the corresponding average percentages of active beta coefficients, i.e. companies with higher volatility are tendentially more often active regressors with non-zero beta coefficients.

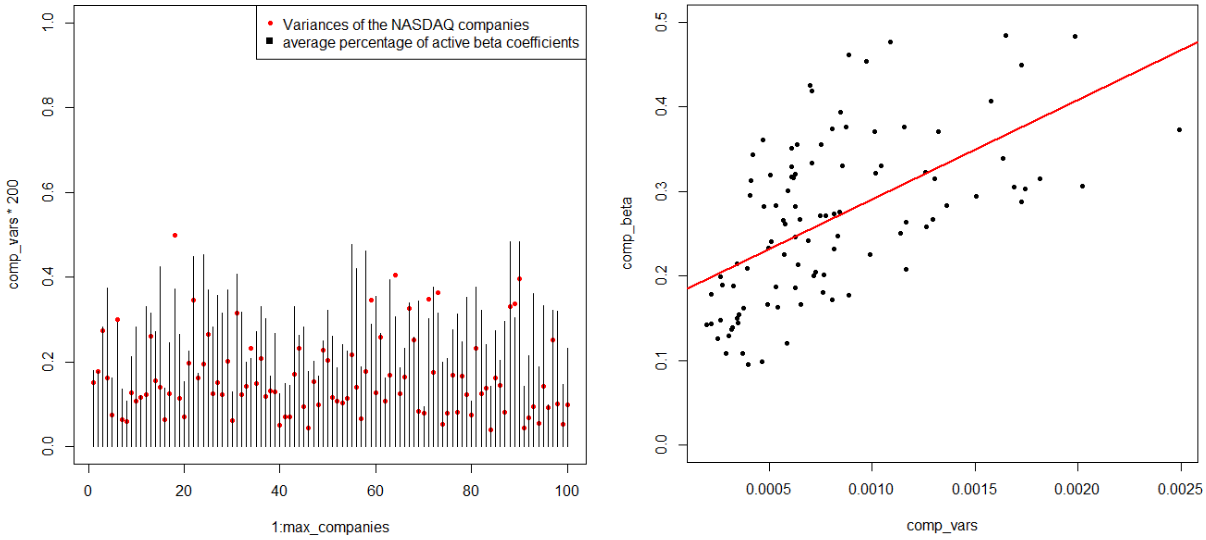

Figure 20: Variances versus average percentage of active beta coefficients of the NASDAQ companies: as a multiple plot with rescaled variances by factor of 200 on the left, and as a scatter plot with linear regression on the right, produced via *QR.variance.vs.beta*

5.1.4 RiskAnalytics package: "Risk Analytics" part

```
#-------------------------------------------------
# Aggregate data
#-------------------------------------------------
last_lambda = aggregate.parallel.results(final_data, max_companies, parResult,
     work_dir = work_dir, new_days = 2469, winsize = 63)

#-------------------------------------------------
# Risk Analytics
#-------------------------------------------------
lambda.analytics(last_lambda, final_data, max_companies)
# "Lambda Analytics Summary"
#     Min.  1st Qu.  Median    Mean  3rd Qu.    Max.
# 0.004418 0.006351 0.007255 0.009206 0.009903 0.032530
# "Lambda quantiles corresponding to the given probabilities:"
#         0%         20%         40%         60%         80%        100%
# 0.004418067 0.006227005 0.006791834 0.008091999 0.010884704 0.032527493
# "last lambda value is the quantile for this probability: 0.384927066450567"
# "last lambda value: 0.00673865282581167"
# "Correlation between Lambda and macro variables"
#      ^VIX       ^GSPC        IYR      3MTCM     Yield    Credit
# 0.8196245 -0.01427044 -0.01283302 -0.0365718 0.2497622 0.6803068
```

Listing 4: RiskAnalytics application example: "Risk Analytics" part

Listing 4 shows how the QR calculation results from the parallel computing component of **RiskAnalytics**, which are saved in the `parResult` object, are aggregated and the *FRM* risk measure as proposed in Section 2.2 is constructed. It is recalled that the *FRM* risk measure is defined as the averaged lambda over all k NASDAQ companies:

$$FRM(t) \stackrel{def}{=} \frac{1}{k}\sum_{j=1}^{k}\lambda_j^*(t), \quad t \in \{t_0, \dots, T\}. \tag{6}$$

The function `aggregate.parallel.results` serves the purpose of combining the λ_j^*-values from each company and applying Formula 6. Additionally, a previous lambda time series can be read in from a CSV file and concatenated with the new lambda values counting `new_days` entries. Finally, the current lambda time series is saved as a CSV file and returned as the vector `last_lambda` for further analysis.

Subsequently, the function `lambda.analytics` provides as part of the "Risk Analytics" software component descriptive statistics for the current lambda time series `last_lambda`, furthermore λ quantiles corresponding to the risk level probabilities as suggested in Section 3.3.2, the last λ with its quantile probability, and the correlations between λ and the macro variables are calculated. Finally, a simple plot preview of the FRM lambda time series is generated, see Figure 21.

5.1.5 RiskAnalytics package: full program run

```
library(snow)
library(RiskAnalytics)

work_dir = "c:/r/frm/2017"
max_companies = 100

# Load data
companylist = get.nasdaq.companies()
system.time( yahoo_data <- get.yahoo.data(companylist, max_comp_num = max_companies,
    from_date = "2006-12-29") )
system.time( macro_data <- get.macro.data(from_date = "2006-12-28") )
final_data = combine.data(yahoo_data, macro_data)

# Calculate data
parResult = parallel.lasso.computation(final_data, max_companies, work_dir = work_dir,
     new_days = 2469, parallel_cpu = 32, winsize = 63)
# Aggregate data
last_lambda = aggregate.parallel.results(final_data, max_companies, parResult,
     work_dir = work_dir, new_days = 2469, winsize = 63)

# Risk Analytics / QR.analytics
data.analytics(yahoo_data, macro_data)
QR.regressors.stats(parResult, sel_threshold = 0.5, min_regressed_comp = 10)
ave_beta_share = QR.beta.stats(parResult)
QR.variance.vs.beta(final_data, ave_beta_share)
lambda.analytics(last_lambda, final_data, max_companies)
```

Listing 5: RiskAnalytics application example: full program run

The full program run of the package **RiskAnalytics** is demonstrated in Listing 5. The *data processing component* extracts all needed data in real time, which are passed over to the *parallel computing component*. The latter performs the lasso penalized QR (*QR meth-*

ods component) via cluster computing (**snowfall** (Knaus, 2015)), thereby operating as an "in-memory application". That means that only the computational power of the physical CPU cores is needed and no disk Input/Output operations are required. Subsequently, the parallelization results are aggregated and the *FRM* risk measure is calculated. In conclusion, the *"Risk Analytics" component*, which comprises the tools *data.analytics*, *QR.analytics* and *lambda.analytics*, provides descriptive statistics of the data collected at different stages of the **RiskAnalytics** program run, hence helping to analyze, evaluate and forecast the systemic risk for the considered markets (Nasdaq Stock Market).

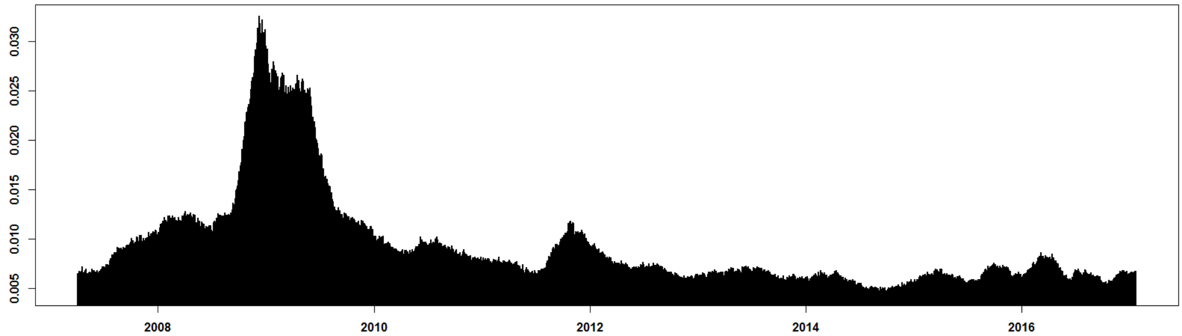

Figure 21: Simple plot preview of the FRM lambda time series, generated after the full program run of the **RiskAnalytics** package

5.2 RiskAnalytics (scientific IDE)

The ***RiskAnalytics*** *scientific IDE* is available under `http://borke.net/RiskAnalytics/`. IDE stands for "integrated development environment". This interactive and web based IDE has the purpose of combining and presenting the scientific, technical and visual materials, elements and sources around the topic "Risk Analytics and FRM". It provides different risk meter designs both for the risk indicators and for the time series visualizations, containing current but also previous risk measure calculations. Further, scientific references concerning the methodology but also software implementations can be found within the ***RiskAnalytics*** *scientific IDE*.

Interactive exploratory data analysis (EDA) can be conducted with the aid of the D3 based risk measure visualizations, current Google Trends statistics and real-time charts (encompassing VIX, S&P 500, Nasdaq etc.), see also Figure 22. The real-time charts are provided by TradingView, a social network for traders and investors on Stock and Futures and Forex markets (`https://www.tradingview.com/chart`).

5.3 Future Developments

5.3.1 Package namespace

The current **RiskAnalytics** package could be improved by using a namespace. Namespaces make a package self-contained in two ways: the **imports** and the **exports** behavior. The **imports** defines how a function in one package finds a function in another.

The **exports** helps to avoid conflicts with other packages by specifying which functions are available outside of the package (internal functions are available only within the own package and can't easily be used by another package). For more details, the following book is recommended (`http://r-pkgs.had.co.nz/namespace.html`) (Wickham, 2015). Furthermore, a package namespace could help to reduce redundant arguments, which are passed to several functions (see e.g. `parallel.lasso.computation`, `aggregate.parallel.results`), by storing the relevant variables in a namespace, from where they can be accessed from other functions without being explicitly provided as redundant arguments.

5.3.2 Incorporation of the hqreg package

The aforementioned **hqreg** (Yi, 2016) package, which provides efficient and C optimized algorithms for fitting regularization paths for lasso or elastic-net penalized regression models with Huber loss, quantile loss or squared loss, is a promising alternative for the time-intensive lasso penalized QR procedure, see Section 5.1.2. A further version of the **RiskAnalytics** package could provide different lasso penalized QR implementations, with **hqreg** as a possible option. But first the necessary studies and benchmarks should be carried out in order to compare the numerical consistency, reliability and time complexity with the former methods and results.

5.3.3 More risk measures involving the beta coefficients and the market volatility

The results from Section 5.1.3 indicate that there is a considerable relationship between the variances of the covariates and the average percentages of active beta coefficients, i.e. covariates with higher volatility are tendentially more often active regressors with non-zero beta coefficients. Because the L1-norm penalty in Formula 2 shrinks the fitted coefficients toward zero by $|\beta_1| + \ldots + |\beta_p| \leq s$, there is a duality between the λ value and the shrinkage parameter s of the β's L1-norm. Hence, the incorporation of the whole market volatility (in the given moving window or another time period) and some appropriate transformations of the β-coefficients into the new risk measure variants should be considered and examined. The ***RiskAnalytics*** *scientific IDE* is a good platform for further experiments.

5.3.4 More D3/C3 visualizations based on the beta structure

The powerful capabilities and features of the D3.js framework but also the C3.js extension, a D3-based reusable chart library (`http://c3js.org/`), can be used to implement more interactive designs and visualizations of the risk measures. For instance, the rich structure of the QR-components, lambdas and beta coefficients as time dependent vectors and matrices, can be exploited for the generation of time-variant risk dependency graphs, where the beta coefficients serve as proxies for the adjacency matrix of the systemic risk. First steps within R can be easily done by means of the package **networkD3** (Gandrud et al., 2016), see also `https://github.com/Quantlet/forceNetwork`.

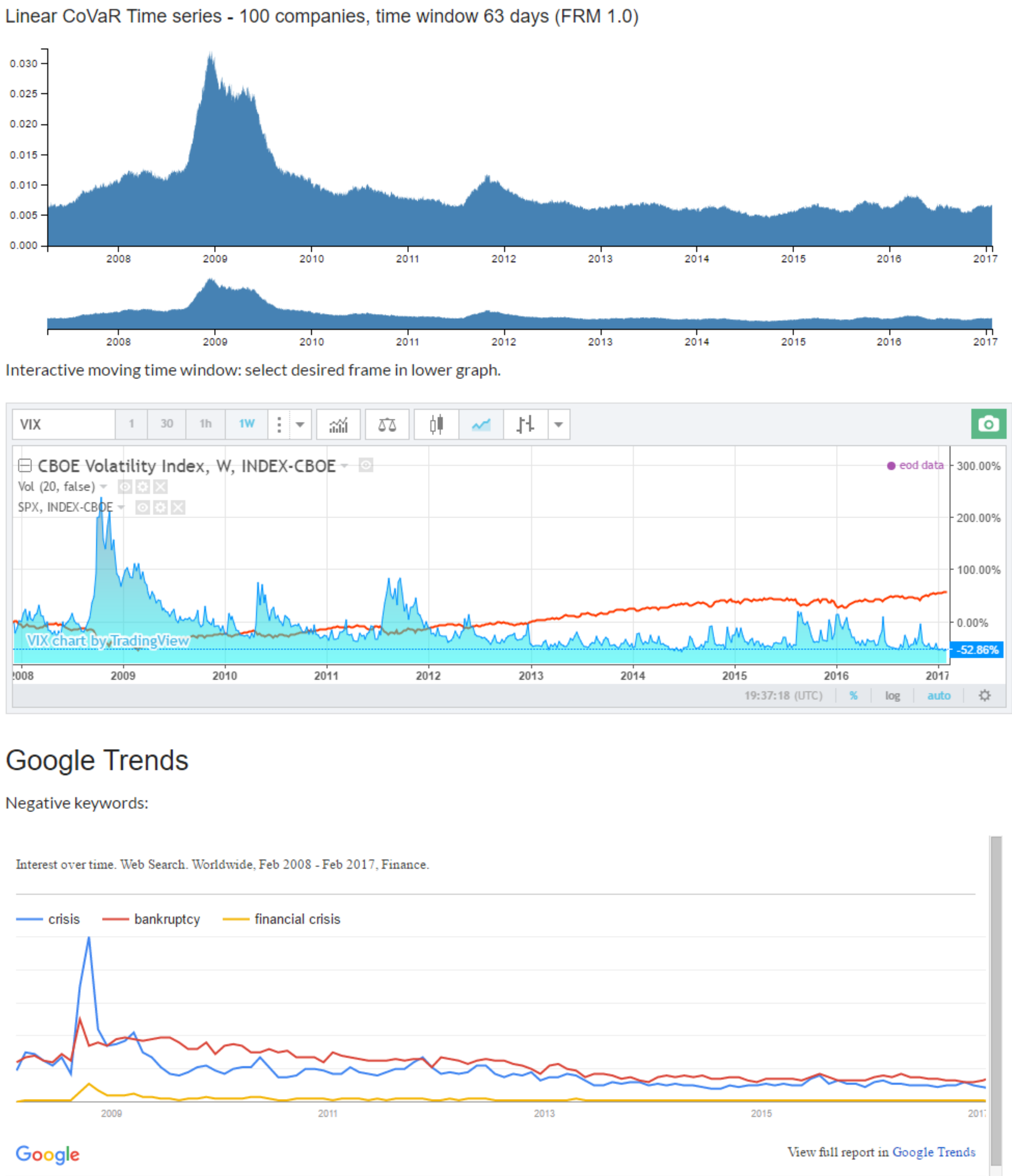

Figure 22: D3 based *FRM* risk measure visualization (created via the **RiskAnalytics** package), current Google Trends statistics and real-time charts (encompassing VIX and S&P 500), each of them covering the same time range; available for interactive exploratory data analysis on the ***RiskAnalytics** scientific IDE*

6 Conclusion

In this paper we propose and develop a measure for systemic risk in financial markets: the Financial Risk Meter (FRM). The FRM is a measure for systemic risk based on the penalty term λ of the linear quantile lasso regression, which is defined as the average of the λ series over the 100 largest US publicly traded financial institutions. The implementation is carried out by using parallel computing. The risk levels are classified by five levels. The empirical result shows that our Financial Risk Meter can be a good indicator for trends in systemic risk. Compared with other systemic risk measures, such as VIX, SRISK, Google Trends with the keyword "financial crisis", we find that the FRM and VIX, FRM and SRISK, FRM and GT mutually granger cause one another, which means that our FRM is a good measure of systemic risk for the US financial market. All the codes of FRM are published on `www.quantlet.de` with keyword FRM. The R package **RiskAnalytics** (Borke, 2017) is another tool with the purpose of integrating and facilitating the research, calculation and analysis methods around the FRM project. The up-to-date FRM can be found on `http://frm.wiwi.hu-berlin.de`.

7 Appendix: Financial Institutions

WFC	Wells Fargo & Company	AON	Aon plc
JPM	J P Morgan Chase & Co	ALL	Allstate Corporation
BAC	Bank of America Corporation	BEN	Franklin Resources, Inc.
C	Citigroup Inc.	STI	SunTrust Banks, Inc.
AIG	American International Group, Inc.	MCO	Moody's Corporation
GS	Goldman Sachs Group, Inc.	PGR	Progressive Corporation
USB	U.S. Bancorp	AMP	Ameriprise Financial Services, Inc.
AXP	American Express Company	AMTD	TD Ameritrade Holding Corporation
MS	Morgan Stanley	HIG	Hartford Financial Services Group, Inc.
BLK	BlackRock, Inc.	TROW	T. Rowe Price Group, Inc.
MET	MetLife, Inc.	NTRS	Northern Trust Corporation
PNC	PNC Financial Services Group, Inc. (The)	MTB	M&T Bank Corporation
BK	Bank Of New York Mellon Corporation (The)	FITB	Fifth Third Bancorp
SCHW	The Charles Schwab Corporation	IVZ	Invesco Plc
COF	Capital One Financial Corporation	L	Loews Corporation
PRU	Prudential Financial, Inc.	EFX	Equifax, Inc.
TRV	The Travelers Companies, Inc.	PFG	Principal Financial Group Inc
CME	CME Group Inc.	RF	Regions Financial Corporation
CB	Chubb Corporation	MKL	Markel Corporation
MMC	Marsh & McLennan Companies, Inc.	FNF	Fidelity National Financial, Inc.
BBT	BB&T Corporation	LNC	Lincoln National Corporation
ICE	Intercontinental Exchange Inc.	CBG	CBRE Group, Inc.
STT	State Street Corporation	KEY	KeyCorp
AFL	Aflac Incorporated	NDAQ	The NASDAQ OMX Group, Inc.
CINF	Cincinnati Financial Corporation	CACC	Credit Acceptance Corporation
CNA	CNA Financial Corporation	BRO	Brown & Brown, Inc.
HBAN	Huntington Bancshares Incorporated	ERIE	Erie Indemnity Company
SEIC	SEI Investments Company	OZRK	Bank of the Ozarks
ETFC	E*TRADE Financial Corporation	WTM	White Mountains Insurance Group, Ltd.
AMG	Affiliated Managers Group, Inc.	SNV	Synovus Financial Corp.

RJF	Raymond James Financial, Inc.	ISBC	Investors Bancorp, Inc.
UNM	Unum Group	MKTX	MarketAxess Holdings, Inc.
NYCB	New York Community Bancorp, Inc.	LM	Legg Mason, Inc.
Y	Alleghany Corporation	CBSH	Commerce Bancshares, Inc.
SBNY	Signature Bank	BOKF	BOK Financial Corporation
CMA	Comerica Incorporated	EEFT	Euronet Worldwide, Inc.
AJG	Arthur J. Gallagher & Co.	DNB	Dun & Bradstreet Corporation
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
FCNCA	First Citizens BancShares, Inc.	MTG	MGIC Investment Corporation

Table 15: The list of 100 financial companies used to estimate FRM in our sample.

References

Acharya, V., Engle, R. and Richardson, M. (2012). Srisk as a macro-finance measure of systemic risk, https://bfi.uchicago.edu/sites/default/files/file_uploads/%234.pdf. Accessed: 2016-12-16.

Adrian, T. and Brunnermeier, M. K. (2016). CoVaR, *American Economic Review* **106**(7): 1705–1741.

Borke, L. (2017). *RiskAnalytics: Real time processing of Nasdaq and Yahoo finance data and parallelized quantile lasso regression methods*. R package version 0.2.0.
URL: *https://github.com/lborke/RiskAnalytics*

Borke, L. and Härdle, W. K. (2017a). GitHub API based QuantNet Mining infrastructure in R, *SFB 649 Discussion Paper (forthcoming)* . Humboldt Universität zu Berlin.

Borke, L. and Härdle, W. K. (2017b). Q3-D3-LSA, *in* W. K. Härdle, H. H. Lu and X. Shen (eds), *Handbook of Big data Analytics*, Springer.

Bostock, M., Ogievetsky, V. and Heer, J. (2011). D3 Data-Driven Documents, *IEEE Transactions on Visualization and Computer Graphics* **17**(12): 2301–2309.
URL: *http://dx.doi.org/10.1109/TVCG.2011.185*

Brooks, C. (2014). *Introductory Econometrics for Finance*, Cambridge University Press.

Brownlees, C. T. and Engle, R. F. (2016). SRISK: A Conditional Capital Shortfall Measure of Systemic Risk, *The Review of Financial Studies* .

Calaway, R., Weston, S. and Analytics, R. (2015). *foreach: Provides Foreach Looping Construct for R*. R package version 1.4.3.
URL: *https://CRAN.R-project.org/package=foreach*

Calaway, R., Weston, S., Tenenbaum, D. and Analytics, R. (2015). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.10.
URL: *https://CRAN.R-project.org/package=doParallel*

Fan, Y., Härdle, W. K., Wang, W. and Zhu, L. (2016). Composite Quantile Regression for the Single-Index Model, *Journal of Business economics and statistics* .

Gandrud, C., Allaire, J. and Russell, K. (2016). *networkD3: D3 JavaScript Network Graphs from R*. R package version 0.2.13.
URL: *https://CRAN.R-project.org/package=networkD3*

Granger, C. (1988). Some recent development in a concept of causality, *Journal of Econometrics* **39**(1): 199 – 211.
URL: *http://www.sciencedirect.com/science/article/pii/0304407688900450*

Granger, C. and Newbold, P. (1974). Spurious regressions in econometrics, *Journal of Econometrics* **2**(2): 111–120.

Hallett, N. (2009). *Taming Risk–A guide for Traders*, The Disciplined Trader.

Härdle, W., Wang, W. and Yu, L. (2016). TENET: Tail-Event driven NETwork risk, *Journal of Econometrics* **192**(2): 499–513.

Hautsch, N., Schaumburg, J. and Schienle, M. (2015). Financial network systemic risk contributions, *Review of Finance* **19**(2): 685–738.

Kane, M., Emerson, J. and Weston, S. (2013). Scalable strategies for computing with massive data, *Journal of Statistical Software* **55**(1): 1–19.
URL: *https://www.jstatsoft.org/index.php/jss/article/view/v055i14*

Knaus, J. (2015). *snowfall: Easier cluster computing (based on snow).* R package version 1.84-6.1.
URL: *https://CRAN.R-project.org/package=snowfall*

Koenker, R. (2016). *quantreg: Quantile Regression.* R package version 5.29.
URL: *https://CRAN.R-project.org/package=quantreg*

Koenker, R. and Mizera, I. (2014). Convex Optimization in R, *Journal of Statistical Software* **60**(1): 1–23.
URL: *https://www.jstatsoft.org/index.php/jss/article/view/v060i05*

Lang, D. T. and the CRAN team (2016). *RCurl: General Network (HTTP/FTP/...) Client Interface for R.* R package version 1.95-4.8.
URL: *https://CRAN.R-project.org/package=RCurl*

Li, Y. and Zhu, J. (2008). L1-Norm Quantile Regression, *Journal of Computational and Graphical Statistics* **17**(1).

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer.

Preis, T., Moat, H. S. and Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends, *Scientific Reports* **3**: 1684.

Ryan, J. A. (2016). *quantmod: Quantitative Financial Modelling Framework.* R package version 0.4-7.
URL: *https://CRAN.R-project.org/package=quantmod*

Schwarcz, S. L. (2008). Systemic Risk, *Georgetown Law Journal* **97**(1): 193–249.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.

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

Tierney, L., Rossini, A. J., Li, N. and Sevcikova, H. (2016). *snow: Simple Network of Workstations.* R package version 0.4-2.
URL: *https://CRAN.R-project.org/package=snow*

Wickham, H. (2015). *R Packages*, 1st edn, O'Reilly Media, Inc.
URL: *http://r-pkgs.had.co.nz*

Yi, C. (2016). *hqreg: Regularization Paths for Lasso or Elastic-Net Penalized Huber Loss Regression and Quantile Regression.* R package version 1.3.
URL: *https://CRAN.R-project.org/package=hqreg*

Yi, C. and Huang, J. (2015). Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss Regression and Quantile Regression, *ArXiv e-prints* . 1509.02957.

Yuan, M. (2006). Gacv for quantile smoothing splines, *Computational statistics and data analysis* **50**(3): 813–829.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(1): 49–67.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series?–A study in sampling and the nature of time-series, *Journal of the Royal Statistical Society* **89**(1): 1–63.

Zbonakova, L., Härdle, W. K. and Wang, W. (2016). Time Varying Quantile Lasso, *SFB 649 Discussion Paper 2016-047*, Sonderforschungsbereich 649, Humboldt Universität zu Berlin, Germany.

SFB 649 Discussion Paper Series 2017

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Fake Alpha" by Marcel Müller, Tobias Rosenberger and Marliese Uhrig-Homburg, January 2017.
002 "Estimating location values of agricultural land" by Georg Helbing, Zhiwei Shen, Martin Odening and Matthias Ritter, January 2017.
003 "FRM: a Financial Risk Meter based on penalizing tail events occurrence" by Lining Yu, Wolfgang Karl Härdle, Lukas Borke and Thijs Benschop, January 2017.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".