Graduate School of Decision Sciences

Universität Konstanz

GSDS
Working Paper
No. 2017-11

Combining Value-at-Risk Forecasts Using Penalized Quantile Regression

Sebastian Bayer

May 2017

Graduate School of Decision Sciences

All processes within our society are based on decisions – whether they are individual or collective decisions. Understanding how these decisions are made will provide the tools with which we can address the root causes of social science issues.

The GSDS offers an open and communicative academic environment for doctoral researchers who deal with issues of decision making and their application to important social science problems. It combines the perspectives of the various social science disciplines for a comprehensive understanding of human decision behavior and its economic and political consequences.

The GSDS primarily focuses on economics, political science and psychology, but also encompasses the complementary disciplines computer science, sociology and statistics. The GSDS is structured around four interdisciplinary research areas: (A) Behavioural Decision Making, (B) Intertemporal Choice and Markets, (C) Political Decisions and Institutions and (D) Information Processing and Statistical Analysis.

GSDS – Graduate School of Decision Sciences
University of Konstanz
Box 146
78457 Konstanz

Phone: +49 (0)7531 88 3633
Fax: +49 (0)7531 88 5193
E-mail: gsds.office@uni-konstanz.de

-gsds.uni-konstanz.de

ISSN: 2365-4120
May 2017

Combining Value-at-Risk Forecasts Using Penalized Quantile Regression

Sebastian Bayer[a]

[a] *University of Konstanz, Department of Economics, 78457 Konstanz, Germany.*

Abstract

Elastic net penalized quantile regression is proposed for the combination of Value-at-Risk forecasts. The main reason for this regularization is the multicollinearity among the standalone forecasts, which leads to a poor forecast performance of the unpenalized quantile regression estimator due to unstable combination weights. The new approach is applied to combining the Value-at-Risk forecasts of 17 frequently used standalone models for 30 stocks comprising the Dow Jones Industrial Average Index. Within a thorough comparison analysis, the elastic net quantile regression performs better in terms of backtesting than the standalone models and several competing combination approaches. This is particularly the case during the global financial crisis from 2007 / 2008.

Keywords: Value-at-Risk, Forecast Combination, Quantile Regression, Elastic Net
JEL: C51, C52, C53, G32

1. Introduction

It is both difficult and relevant to decide between alternative Value-at-Risk (VaR) modeling and forecasting strategies. A poorly selected risk model may have drastic effects on banks and the economy as a whole, as it could be seen, for example, during the previous financial crisis when many of the standard approaches predicted too low levels of risk. Einhorn (2008, p. 12) compares the VaR to "an airbag that works all the time, except when you have a car accident". The VaR is defined to be the worst possible loss over a target horizon that will not be exceeded with a given probability (Jorion, 2006). VaR is, thus, a quantile of the distribution of returns over a horizon (usually one or ten days) for a given probability level (usually 1%). A major reason for its popularity is that the Basel Committee on Banking Supervision (1996, 2006, 2011) uses the VaR for the calculation of the minimum capital requirements which banks need to keep as reserves to cover the market risk of their investments.

There is an extensive literature on how to estimate and predict VaR (see Kuester et al. (2006), Komunjer (2013) and Nieto and Ruiz (2016) for overviews). The major problem of VaR forecasting, however, is that the models' performance and reliability to accurately predict the risk highly depends on the data at hand. While a parsimonious model might perform well in calm times, it can fail tremendously in a volatile period. Likewise, highly parametrized models might be adequate during periods of high volatility, but can be easily outperformed by simpler approaches in less turbulent times. So far, no unique model or approach dominates throughout the existing VaR forecasting comparisons (see e.g. Kuester et al., 2006; Marinelli et al., 2007; Halbleib and Pohlmeier, 2012; Abad and Benito, 2013; Boucher et al., 2014; Louzis et al., 2014; Ergen, 2015; Nieto and Ruiz, 2016; Bernardi and Catania, 2016). The key reasons for this finding is that the applied models are prone to suffer from model misspecification (e.g. through the application of a too simple model) and estimation uncertainty (e.g. they imply a complicated estimation procedure). For a much more detailed discussion of the risks and uncertainties involved in VaR forecasting, we refer to Boucher et al. (2014).

If the best model is unknown or likely to change over time, a promising alternative to deciding on a single risk model is to combine the predictions stemming from several approaches. In an overview on

Email address: sebastian.bayer@uni-konstanz.de (Sebastian Bayer)

forecasting combinations, Timmermann (2006) provides three explanations in favour of combining forecasts in order to stabilize and improve the predictive performance upon the standalone models. First, there are diversification gains coming from the combination of forecasts computed from different assumptions, specifications or information sets. Second, combined forecasts tend to be robust against structural breaks. Third, the influence of potential misspecification of the individual models is reduced due to averaging over a set of forecasts stemming from several models.

Halbleib and Pohlmeier (2012) propose to combine VaR forecasts using quantile regression (QR), introduced by Koenker and Bassett (1978), as the QR estimator minimizes the tick loss function. This asymmetric and piecewise linear loss function is consistent for quantiles, which means that the true quantile prediction minimizes the expected tick loss (Gneiting, 2011a). Thus, it is reasonable to incorporate the tick loss function for estimation and evaluation of VaR forecasts. QR is further applied by Fuertes and Olmo (2013), who use it for the combination of VaR forecasts from intra- and inter-day models and for a conditional QR forecast encompassing test. QR forecast combination under a variety of restrictions is also explored by Jeon and Taylor (2013), who combine forecasts stemming from the CAViaR class (Engle and Manganelli, 2004) with the forecasts from an option implied volatility model.

If someone aims at combining a possibly large number of VaR forecasts, she likely faces the problem of multicollinearity among them, as they stem from the same data and similar mathematical approaches. This is the case in our empirical application: we observe high pairwise correlations among the forecasts (sometimes larger than 99%), which indicates he presence of severe multicollinearity. In this situation, the standard QR estimator is unstable: small variations in the data can lead to large changes in the estimated parameters. Moreover, it can overfit the data such that, for two highly correlated forecasts we obtain a large positive weight on the one, and a large negative weight on the other. From an in-sample point of view this is not problematic as the estimated coefficients still minimize the tick loss function. For out-of-sample purposes, however, such imprecisely estimated parameters can be harmful as the model fails to generalize well to new data (e.g. Hastie et al., 2011, p. 38). An obvious solution is to only combine forecasts with small to moderate cross-correlations. However, our aim is to avoid manually selecting the models over whose forecasts we average, and instead consider combination techniques that can cope with high correlations among the predictions.

In this paper we propose penalized QR as a novel VaR combination technique. In particular, we consider the regularization with the elastic net penalty of Zou and Hastie (2005), which is a convex combination of the well known least shrinkage and selection operator (lasso) of Tibshirani (1996) and the ridge penalization of Hoerl and Kennard (1970a,b). Due to the geometry of the penalty function, the elastic net restriction simultaneously induces coefficient shrinkage and variable selection. These two properties allow for the combination of a large number of possibly highly correlated VaR forecasts, since the parameter estimates are stable, overfitting is reduced and variables are automatically selected.

Elastic net (elanet) QR features a number of advantages over alternative combination techniques: (1) It performs a data-driven selection of forecasts: the coefficients of uninformative standalone models can be set to zero. This may help to improve the predictability, as only a subset of the models enters the combination. We explore this by comparing elanet QR with ridge QR, which shrinks the coefficients but selects no variables. (2) It can cope with almost collinear forecasts by the regularization of the estimated weights. (3) The elanet QR includes an intercept term with the purpose of bias correction (Halbleib and Pohlmeier, 2012). This is important when all standalone predictions systematically over- or underestimate the VaR: an intercept can shift the combined forecast outside the range of the standalone predictions. Simple averaging techniques do not include such an intercept term and are bound between the minimum and the maximum of the standalone forecasts. (4) The elanet QR minimizes the (penalized) tick loss function.

In the literature, a range of other quantile combination techniques is proposed. Giacomini and Komunjer (2005) introduce a generalized method of moments (GMM) estimator aiming at the minimization of the tick loss function, with the purpose of forecasting combination and encompassing tests. In a similar spirit, Halbleib and Pohlmeier (2012), in addition to QR, introduce a GMM estimator to find the optimal combination weights by minimizing the conditional coverage test (Christoffersen, 1998). The mean and median over the standalone forecasts is considered by Huang and Lee (2013), who combine VaR predictions from models using high frequency information. McAleer et al. (2013a,b) combine VaR forecasts by the percentiles of the predictions with the goal of minimizing capital charges imposed by the Basel Accord. Shan and Yang (2009) and Casarin et al. (2013) introduce sequential combination approaches, where the weights of in the past well performing models are increased and vice versa. While Shan and Yang (2009) assess the performance of the standalone models using the tick loss, Casarin et al. (2013) evaluate the models with respect to the capital requirements imposed by the Basel Accords. A different route is followed by Hamidi et al. (2015), who average the VaR forecasts stemming from conditional autoregressive expectile models and estimate the combination weights by optimizing the squared difference between the nominal probability and the hit rate, i.e. the share of times a VaR prediction is smaller than the realized return. This approach is closely linked to the unconditional coverage test (Kupiec, 1995) which is standard in VaR evaluation. Recently, Bernardi et al. (2017) suggest filtering the standalone models with the model confidence set (MCS) approach of Hansen et al. (2011) prior to averaging the forecasts.

In the empirical part, we assess the performance of the proposed penalized QR combination method for a data set comprising 30 constituents of the Dow Jones Industrial Average Index (DJIA) over a horizon of 8 years. We compare the performance of the elanet QR combined forecasts to a large variety of competing approaches. For the forecast evaluation, we use backtesting via the dynamic quantile backtest of Engle and Manganelli (2004) and the unconditional coverage backtest of Kupiec (1995). We furthermore compare the forecasts with the MCS of Hansen et al. (2011) to find the approach producing the most precise predictions. Apart from the general elastic net, our comparison includes two special cases, namely the ridge and the lasso QR. It turns out, that the penalized QR combined forecasts exhibit the lowest number of backtest rejections and realize comparably low tick losses. By splitting the evaluation sample into two subperiods, we furthermore find that during volatile times, lasso and elanet perform better than ridge QR. This relationship reverses during the calm period. Additionally, we do not face the “forecast combination puzzle” (Stock and Watson, 2004), which states that simple approaches are hard to beat: in the combination of VaR forecasts, complexity pays off.

The remainder of this paper is organized as follows. Section 2 introduces the methodology and provides details on elastic net penalized QR. Section 3 introduces the data set, evaluation horizons, the forecast evaluation method and the standalone models. Section 4 presents the results of the empirical application. Section 5 concludes and gives an outlook on potential future research.

2. Methodology

Let the price of a financial asset or a portfolio at time t be P_t such that the logarithmic return from time t to $t+h$ is $r_{t+h} = \log(P_{t+h}/P_t)$. We denote the VaR forecast from t to $t+h$, conditional on all available information up to time t, $\mathcal{F}_t$, as $q_{t+h|t}(\alpha)$. Formally, the VaR is defined to be,

$$\alpha = \Pr\left(r_{t+h} \leq q_{t+h|t}(\alpha)|\mathcal{F}_t\right), \tag{1}$$

where $\alpha \in (0,\, 1)$ is the probability level. Throughout the paper we focus on the probability $\alpha = 1\%$ and the forecast horizon $h = 1$ day.

In the following, $\widehat{q}_{m,t+1|t}(\alpha)$ is the VaR forecast for the day $t+1$ of model $m = 1, \dots, M$ based on the information available up to time t and $\widehat{\boldsymbol{q}}_{t+1|t}(\alpha) = (\widehat{q}_{1,t+1|t}(\alpha), \dots, \widehat{q}_{M,t+1|t}(\alpha))'$ is the vector of all forecasts. The linear combination of the M forecasts including an intercept term is given by,

$$\begin{aligned} q^c_{t+1|t}(\alpha) &= \beta_{0,t}(\alpha) + \beta_{1,t}(\alpha)\widehat{q}_{1,t+1|t}(\alpha) + \ldots + \beta_{M,t}(\alpha)\widehat{q}_{M,t+1|t}(\alpha) \\ &= \beta_{0,t}(\alpha) + \widehat{\boldsymbol{q}}'_{t+1|t}(\alpha)\boldsymbol{\beta}_t(\alpha), \end{aligned} \tag{2}$$

where $\boldsymbol{\beta}_t(\alpha)$ is the quantile-specific vector of slope coefficients, which we loosely call the combination weights, even though the sum of the coefficients is not necessarily one. We explicitly incorporate an intercept term $\beta_{0,t}(\alpha)$ to bias correct misspecified standalone forecasts: if all standalone predictions systematically over- or underestimate the VaR, an intercept can shift the combined forecast outside the range of the standalone predictions (Halbleib and Pohlmeier, 2012). The time index of the coefficients indicates that the weights are time-varying: in order to incorporate the most recent information into the model parameters, we re-estimate the combination weights every day.

For quantiles, a consistent loss function is the tick loss function (Giacomini and Komunjer, 2005; Gneiting, 2011a) given by,

$$\rho_\alpha(u) = \left(\alpha - \mathbb{1}_{\{u \leq 0\}}\right) u. \tag{3}$$

Consistency of the tick loss implies that the true quantile prediction minimizes the expected tick loss, a concept directly linked to the fact that the VaR is elicitable (Gneiting, 2011a). More generally, a class of consistent loss functions for quantiles is the generalized piecewise linear (GPL) family (Gneiting, 2011b) which is of the form $L_\alpha(r, q) = (\alpha - \mathbb{1}_{\{r \leq q\}})(g(r) - g(q))$ for a non-decreasing function g. The tick loss given in eq. (3) is a special case when g is the identity function, by far the most widely used specification of the GPL family.

Equipped with a consistent loss function, the optimal forecast combination weights consequently minimize the expected loss of the forecast error,

$$\left(\beta^*_{0,t}(\alpha),\, \boldsymbol{\beta}^*_t(\alpha)\right) = \underset{\beta_{0,t}(\alpha),\, \boldsymbol{\beta}_t(\alpha)}{\arg\min}\ \mathrm{E}\left[\rho_\alpha\left(r_{t+1} - \beta_{0,t}(\alpha) - \widehat{\boldsymbol{q}}'_{t+1|t}(\alpha)\,\boldsymbol{\beta}_t(\alpha)\right) \mid \mathcal{F}_t\right], \tag{4}$$

which can be estimated by performing a linear QR of the standalone forecasts on the returns, as the tick loss given in eq. (3) lies at the heart of quantile regression (Koenker and Bassett, 1978; Halbleib and Pohlmeier, 2012). We therefore obtain a consistent and asymptotically normal estimator of the combination weights by minimizing the average tick loss (Koenker and Bassett, 1978),

$$\left(\widehat{\beta}_{0,t}(\alpha),\, \widehat{\boldsymbol{\beta}}_t(\alpha)\right) = \underset{\beta_{0,t}(\alpha),\, \boldsymbol{\beta}_t(\alpha)}{\arg\min}\ \frac{1}{t}\sum_{\tau=0}^{t-1} \rho_\alpha\left(r_{\tau+1} - \beta_{0,t}(\alpha) - \widehat{\boldsymbol{q}}'_{\tau+1|\tau}(\alpha)\boldsymbol{\beta}_t(\alpha)\right), \tag{5}$$

which can then be used to form the combined forecast for the next day via $\widehat{q}^{c}_{t+1|t}(\alpha) = \widehat{\beta}_{0,t}(\alpha) + \widehat{\boldsymbol{q}}'_{t+1|t}\widehat{\boldsymbol{\beta}}_t(\alpha)$.

2.1. *The Effect of Multicollinearity on Forecast Combinations*

Although combination weights estimated via QR are optimal for an in-sample combination of the standalone forecasts (they minimize the tick loss), they might be not optimal for out-of-sample purposes. This is due to the multicollinearity among the forecasts, which increases the variance of the estimated parameters and through that increases the expected tick loss of the prediction error.

To see why multicollinearity among the forecasts increases the variance of the combination weights estimated by QR, note that the asymptotic distribution of the QR estimator is given by $\sqrt{n}(\widehat{\beta}(\alpha) - \beta(\alpha)) \xrightarrow{d} \mathrm{N}\left(0,\, \alpha(1-\alpha)G^{-1}MG^{-1}\right)$ with $M = \mathrm{E}\left[XX'\right]$ and $G = \mathrm{E}\left[XX' f_{u|X}(0)\right]$ where X are the covariates, i.e. here the standalone forecasts. We can see that the term G^{-1} depends on the inverse of the XX' matrix, which gets very large when the data is highly correlated. Consequently, the variance of $\widehat{\beta}(\alpha)$ increases with the degree of correlation among the covariates.

In order to understand why a high variance of the combination weights can be problematic when combining forecasts, assume for the moment that we forecast the mean instead of quantiles. Consider an additive model of the form $Y = f(X) + \varepsilon$ with $\mathrm{E}[\varepsilon] = 0$ and $\mathrm{V}[\varepsilon] = \sigma^2$, where X are the covariates, Y is the dependent variable, f is an unknown function of the data and $\widehat{f}$ is an estimate of f. Then, the expected prediction error (under squared error loss) on data not used for the estimation of f, can be expressed as,

$$\mathrm{E}\left[(Y - \widehat{f}(X))^2\right] = \mathrm{E}\left[f(X) - \widehat{f}(X)\right]^2 + \mathrm{E}\left[(f(X) - \widehat{f}(X))^2\right] + \sigma^2, \tag{6}$$

which is the usual bias-variance trade off (e.g. Hastie et al., 2011, p. 223). Thus, we see that an increase in the variance of $\widehat{f}$ (e.g. through multicollinearity of the covariates) increases the expected squared prediction error.

Such straightforward and general calculations in terms of mean and variance are available only for the mean squared error, but not for the tick loss function. James (2003) generalizes the bias-variance trade off to general symmetric loss functions, but the case of asymmetric loss functions, such as the tick loss, is to the best of our knowledge unexplored. Nonetheless, the same logic intuitively carries over to the tick loss: an increase of the variance the estimated model parameters increases the expected tick loss of the prediction error, although the exact relation is unknown and is likely not so simple as in eq. (6).

We thus conclude that forecast combination is mainly beneficial if we can estimate the combination weights with a reasonable precision, which is, however, negatively correlated with the degree of multicollinearity among the forecasts. Consequently, reducing the estimation error of the combination weights can help to improve the forecasts.

2.2. Elastic Net Penalized Quantile Regression

In this paper we suggest the penalization of the QR estimator with the elastic net of Zou and Hastie (2005) for the combination of VaR forecasts. The elastic net is a linear combination of the well known ridge penalty of Hoerl and Kennard (1970a,b) and the lasso of Tibshirani (1996). While the ridge term shrinks the coefficients towards zero, the lasso shrinks the coefficients and additionally selects variables. Automatic variable selection through the lasso is attractive as the weights of uninformative models can be set to zero. With highly correlated variables, however, the lasso tends to select one of the coefficients of the correlated variables randomly, whereas the ridge shrinks them towards each other (Zou and Hastie, 2005). Under such a scenario, the elastic net offers a compromise between group wise shrinkage with no selection and random selection of variables: similar to ridge, the elastic net shrinks the variables in groups and similar to lasso, it sets some coefficients to zero. The elastic net, thus, combines the strengths of both approaches so that Zou and Hastie (2005) interpret the elastic net as a stabilized version of the lasso penalization. The QR estimator under elastic net penalization is given by

$$\begin{aligned}\left(\widehat{\beta}_{0,t}(\lambda,\,\delta),\,\widehat{\boldsymbol{\beta}}_t(\lambda,\,\delta)\right) = \underset{\beta_{0,t},\,\boldsymbol{\beta}_t}{\arg\min}\,\frac{1}{t}\sum_{\tau=0}^{t-1}\rho_\alpha\left(r_{\tau+1}-\beta_{0,t}-\widehat{\boldsymbol{q}}'_{\tau+1|\tau}\boldsymbol{\beta}_t\right)+ \\ +\lambda\left(\delta||\boldsymbol{\beta}_t||_1+(1-\delta)||\boldsymbol{\beta}_t||_2^2/2\right),\end{aligned} \tag{7}$$

where λ is the regularization parameter and $\delta \in [0,\,1]$ balances the ridge and the lasso term, given by the sum of the absolute, respectively the sum of the squared parameters. Note that to simplify notation, we mostly suppress the quantile level α in what follows.

If $\lambda \to \infty$, eq. (7) simplifies to the intercept as it remains unpenalized. In this case, we simply estimate the empirical quantile of the returns. For $\lambda = 0$, eq. (7) reduces to unpenalized QR. The value of λ, thus, determines the influence of the standalone predictions on the combined forecast. Considering the parameter δ, we obtain lasso QR for $\delta = 1$ and a pure ridge penalization for $\delta = 0$. It is common in the literature (Hastie et al., 2015, p. 57), to only estimate the λ parameter and consider preselected values of δ. In particular, we consider the three cases of $\delta = 0$ (ridge), $\delta = 1$ (lasso) and $\delta = 0.5$ (elanet). This choice allows to identify if improvements in the predictions come from the shrinkage part alone or whether the variable selection is also essential in terms of predictive performance.

We estimate the elastic net penalized QR with the semismooth Newton coordinate descent algorithm proposed by Yi and Huang (2017). An implementation for R (R Core Team, 2016) is available in the hqreg library of Yi (2017).

Relation to Convex Weights

In the forecast combination literature, convexity is frequently imposed on the combination weights which typically improves the predictive performance (see e.g. Timmermann, 2006; Hansen, 2008). Convex weights are non-negative and they sum to one, i.e. $0 \le \beta_m \le 1$, for $m = 1, \ldots, M$ and $\sum_{m=1}^{M}\beta_m = 1$. This particular restriction bears an interesting similarity to the elastic net penalty, which we can see by writing the Lagrangian form in eq. (7) as a restricted estimator,

$$\left(\widetilde{\beta}_{0,t}(\xi,\,\delta),\,\widetilde{\boldsymbol{\beta}}_t(\xi,\,\delta)\right) = \underset{\beta_{0,t},\,\boldsymbol{\beta}_t}{\arg\min}\,\frac{1}{t}\sum_{\tau=0}^{t-1}\rho_\alpha\left(r_{\tau+1}-\beta_{0,t}-\widehat{\boldsymbol{q}}'_{\tau+1|\tau}\boldsymbol{\beta}_t\right) \quad \text{s.t.}\ \left(\delta||\boldsymbol{\beta}_t||_1+(1-\delta)||\boldsymbol{\beta}_t||_2^2/2\right)\leq\xi. \tag{8}$$

As usual, there is a one-to-one mapping between λ in eq. (7) and ξ in eq. (8). If we consider the case of lasso QR ($\delta = 1$) and assume all slope coefficients to be non-negative, then eq. (8) collapses to convex QR for $\xi = 1$. As a result, we interpret the convexity constraint as a heavily restricted version of the elastic net penalty, which is more general due to many reasons: (1) the combination weights are allowed to be negative; (2) the weights must not sum to one, as on can choose the regularization parameter λ (or ξ) and (3) one can select the degree of sparsity that the model is enforcing by varying the δ parameter.

2.3. Selection of the Regularization Parameter

The optimal shrinkage parameter for forecasting purposes is the value that minimizes the expected prediction error over the out-of-sample data. The in-sample tick loss, $\frac{1}{t}\sum_{\tau=0}^{t-1}\rho_\alpha(r_{\tau+1}-\widehat{\beta}_{0,t}(\lambda,\,\delta)-\widehat{\boldsymbol{q}}'_{\tau+1|\tau}\widehat{\boldsymbol{\beta}}_t(\lambda,\,\delta))$, can not be used as the loss decreases in λ. Standard ways of estimating the parameter λ include information criteria and cross validation, which we discuss below. In addition, we propose a computationally convenient heuristic rule based on the sum of absolute combination weights.

Bayesian Information Criterion

The simplest approach for estimating the regularization parameter λ is via the Bayesian Information Criterion (BIC), which penalizes the in-sample loss. For the application of the BIC, we require a measure of the effective degrees of freedom of the model. In the case of lasso QR, Li and Zhu (2008) show that the effective degrees of freedom can be estimated by the number of non-zero coefficients, i.e. by $\mathrm{df} = \sum_{m=1}^{M}\mathbb{1}_{\{\widehat{\beta}_{m,t}(\lambda,\delta=1)\neq 0\}}$. Consequently, the BIC for lasso QR regression is given by (Li and Zhu, 2008),

$$\mathrm{BIC}(\lambda) = \ln\left(\frac{1}{t}\sum_{\tau=0}^{t-1}\rho_\alpha\left(r_{\tau+1}-\widehat{\beta}_{0,t}(\lambda,\,\delta=1)-\widehat{\boldsymbol{q}}'_{\tau+1|\tau}\widehat{\boldsymbol{\beta}}_t(\lambda,\,\delta=1)\right)\right)+\frac{\ln t}{2t}\mathrm{df}, \tag{9}$$

and we find $\widehat{\lambda}$ as the value that minimizes the BIC. This approach is implemented only for lasso QR, given that there is no similar approach available for elanet and ridge QR.

Time Series Cross Validation

More appropriate for out-of-sample purpose is the cross validation (CV) as it aims at minimizing the out-of-sample prediction error by evaluating a model on data that was not part of the estimation process. The two most common approaches, leave-v-out and K-fold CV, are, however, not applicable in the present context. The reason is a violation of the fundamental assumption of CV that the estimation and evaluation samples are independent (Arlot and Celisse, 2010). Financial returns may be assumed to be at least uncorrelated, but they are neither independent nor identically distributed. VaR predictions exhibit high positive autocorrelation: in our application the autocorrelations decrease only slowly, i.e. even after 250 days several autocorrelations are well above 50%.

To account for the dependence in the data, we therefore employ a time series CV method described in Hart (1994) which takes the form,

$$\mathrm{CV}(\lambda,\,\delta) = \frac{1}{t-n_{\min}}\sum_{\tau=n_{\min}}^{t-1}\rho_\alpha\left(r_{\tau+1}-\widehat{q}^{c}_{\tau+1|\tau}(\lambda,\,\delta)\right), \tag{10}$$

where $\widehat{q}^{c}_{\tau+1|\tau}(\lambda,\,\delta) = \widehat{\beta}_{0,\tau}(\lambda,\,\delta)+\widehat{\boldsymbol{q}}'_{\tau+1|\tau}\widehat{\boldsymbol{\beta}}_\tau(\lambda,\,\delta)$ is the combined VaR prediction for $\tau+1$ based on the information available up to τ and $n_{\min}$ is the minimum number of observations required to initially fit the model, which we set to four years in the empirical application. In contrast to leave-v-out or K-fold CV, this approach only employs past information to predict future values and is more robust towards autocorrelation in the data (Hart and Lee, 2005). Eventually, for a given value of δ, we select λ by the value that minimizes the CV loss given in eq. (10).

Heuristic Rule

Apart from the BIC for lasso QR and time series CV for the more general elastic net penalty, we propose an heuristic rule for selecting the regularization parameter λ. In particular, we suggest to choose the most restricted model such that for a given value of δ, the sum of the absolute estimated weights (i.e. the L_1-norm) is smaller than some value s,

$$\widehat{\lambda} = \max \lambda, \qquad \text{s.t.} \sum_{m=1}^{M} |\widehat{\beta}_{m,t}(\lambda, \delta)| \leq s. \tag{11}$$

The intuition and origin for this rule stems from the fact that the elastic net penalty generalizes the convexity restriction, for which $s = 1$, given that $\beta_{m,t} \geq 0$ for $m = 1, \ldots, M$. The suggestion above therefore connects a generalized variant of the convexity constraint (the L_1-norm) with the regularization parameter λ.

To find reasonable values of s, we compute $\sum_{m=1}^{M} |\widehat{\beta}_{m,t}(\lambda, \delta)|$ when λ is estimated with the time series CV procedure. It turns out that while the estimates of λ vary greatly depending on δ and the data, the values of $\sum_{m=1}^{M} |\widehat{\beta}_{m,t}(\lambda, \delta)|$ are remarkably stable across the time, the asset space and even across the different values of δ. We find that most of the values are in the range (0.7, 1.1) with the majority of values around 0.8. Consequently, in the empirical application, we include predictions formed with the above rule and set $s = 0.8$. We also provide a robustness check on the choice of s, where we show that a range of values of s yields precise predictions. This rule for selecting λ might not be optimal from a theoretical point of view, but on the practical side, it performs well empirically, it is robust, easy to implement and requires no computationally expensive CV procedure.

3. Empirical Application: Setup

In the empirical application, we compare the predictions of the penalized QR with forecasts of the standalone models and some competing combination approaches. This section outlines the data, the models to be combined, some competing combination techniques and the forecast evaluation methodology. The presentation of the results is deferred to Section 4.

3.1. Data and Evaluation Horizon

The dataset under consideration are the daily closing (dividend and split adjusted) prices of 30 constituents of the DJIA for a time horizon from January 2, 1996 to December 31, 2014, a total of 4784 days, which we obtained from Thomson Reuters Eikon. Note that the DJIA composition as of March 19, 2015 includes Goldman Sachs (GS) and Visa (V), which were only listed after 1996. Consequently, we replace these two stocks with two immediate predecessors, AT&T (T) and Hewlett Packard (HPQ). The symbols of the assets are thus: AAPL, AXP, BA, CAT, CSCO, CVX, DD, DIS, GE, HD, HPQ, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, NKE, PFE, PG, T, TRV, UNH, UTX, VZ, WMT, XOM.

Figure A.5 in the appendix shows the log return series of the stocks and Table A.2 presents the corresponding summary statistics, together with the ticker symbols and company names. The return series show volatility clustering, especially in the time around the dot-com bubble and in the time of the previous global financial crisis. Moreover, the returns exhibit excess kurtosis and non-zero skewness, the Jarque-Bera test strongly rejects normality of the log returns.

As we require data to estimate the standalone models, to combine the forecasts with penalized QR and to estimate the regularization parameter, λ, our evaluation horizon spans the time from January 3, 2007 to December 31, 2014 (2014 days). Besides the full 8 years of data, we split the sample into two equally sized windows of 4 years as the first half of the overall sample is mainly driven by the financial crisis and is, thus, much more volatile compared to the second subperiod. The goal of this split is to evaluate the models under different volatility regimes, which we term the crisis and calm period. For an illustration of the data and sample split, consider Figure 1, which shows the log returns of the equally weighted portfolio of the 30 returns series. Whereas the two gray areas together represent the overall evaluation sample, the light and dark gray areas depict the subperiods from January 3, 2007 to December 31, 2010 (1008 days), respectively from January 3, 2011 to December 31, 2014 (1006 days).

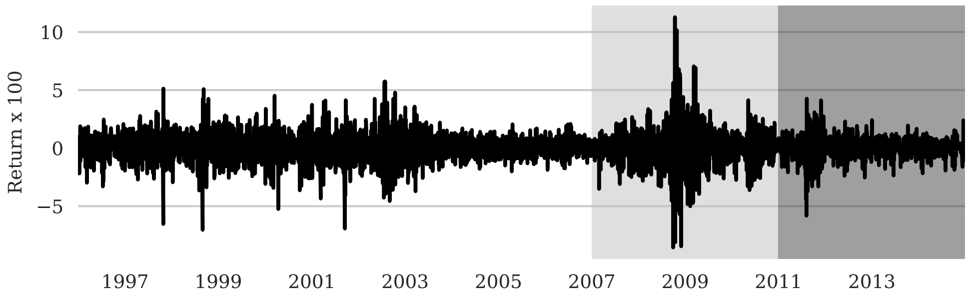


Figure 1: Equally weighted portfolio of the 30 assets considered in the empirical comparison. The two gray shaded areas indicate the forecast evaluation horizons January 3, 2007 to December 31, 2010 (1008 days) and January 3, 2011 to December 31, 2014 (1006 days). In addition, we consider the overall period from January 3, 2007 to December 31, 2014 (2014 days).

3.2. Standalone Models

Our pool of models, which we use to form the combined forecasts, contains a total of 17 approaches. The selected models cover a wide range of frequently applied parametric, semi-parametric and non-parametric techniques. While some of them are parsimonious, others are highly parametrized and can account for rich dynamics in the data.

Static Normal Distribution

This approach assumes that the returns are normally distributed with mean μ and variance σ^2, which we estimate based on a rolling window of 250 observations. For this technique, the quantile prediction for the next day is $\widehat{q}_{t+1|t}(\alpha) = \widehat{\mu} + \widehat{\sigma}\Phi(\alpha)^{-1}$ where $\Phi(\cdot)^{-1}$ is the inverse of the standard Normal distribution.

Weighted Historical Simulation

The popular historical simulation (HS) approach predicts the next day's VaR by the empirical α-quantile of the past returns. While the standard HS weights all past days equally, the weighted HS technique of Boudoukh et al. (1998) uses a geometrically declining weighting scheme: more recent data points receive a higher weight and, consequently, become more relevant for the prediction. The weight of the day $\tau = t-w+1,\ldots,t$ is $\eta_\tau = \eta^{\tau-1}(1-\eta)/(1-\eta^w)$, where w is the window length and we set $\eta = 0.99$. Similar to the Normal Distribution, we estimate the empirical quantile of the HS, respectively the weighted HS approach using a rolling window of 250 observations.

RiskMetrics

The popular exponential smoothing RiskMetrics method (RiskMetrics Group, 1996) assumes the VaR of day $t+1$ to be $\widehat{q}_{t+1|t}(\alpha) = \sigma_{t+1}\Phi^{-1}(\alpha)$, where $\sigma_t^2 = (1-\nu)r_{t-1}^2 + \nu\sigma_{t-1}^2$ with $\nu = 0.94$ for daily data. For RiskMetrics no parameters need be estimated.

CAViaR Models

The CAViaR class introduced by Engle and Manganelli (2004) assumes the VaR forecast to be a function of lagged VaR predictions and other explanatory variables. They propose the following four specifications,

Symmetric absolute value	(SAV)	$q_{t+1\|t}(\alpha) = \beta_0 + \beta_1 q_{t\|t-1}(\alpha) + \beta_2 \|r_t\|$,
Asymmetric slope	(AS)	$q_{t+1\|t}(\alpha) = \beta_0 + \beta_1 q_{t\|t-1}(\alpha) + \beta_2 (r_t)^+ + \beta_3 (r_t)^-$,
Indirect GARCH(1, 1)	(IG)	$q_{t+1\|t}(\alpha) = (\beta_0 + \beta_1 q^2_{t\|t-1}(\alpha) + \beta_2 r_t^2)^{1/2}$,
Adaptive	(AD)	$q_{t+1\|t}(\alpha) = q_{t\|t-1}(\alpha) + \beta_1 \{[1 + \exp(G[r_t - q_{t\|t-1}(\alpha)])]^{-1} - \alpha\}$,

where $(x)^+ = \max(x, 0)$, $(x)^- = -\min(x, 0)$ and we set $G = 10$ as in Engle and Manganelli (2004). The estimation of the CAViaR models follows the procedure described in Engle and Manganelli (2004) using a rolling window of 1000 days.

Location-Scale Models

The remaining 9 methods all assume the location-scale family for the return process such that returns can be decomposed into $r_t = \mu_t + \sigma_t z_t$. The component μ_t is the conditional mean of the distribution of r_t, σ_t is a volatility process and the innovation term z_t is independent and identically distributed with mean zero and unit variance. VaR forecasts are obtained via $\widehat{q}_{t+1|t}(\alpha) = \widehat{\mu}_{t+1|t} + \widehat{\sigma}_{t+1|t} Q_\alpha(z_t)$ where $\widehat{\mu}_{t+1|t}$ is a one-step-ahead forecasts of the mean, $\widehat{\sigma}_{t+1|t}$ is a one-step-ahead forecast of the volatility and $Q_\alpha(z_t)$ is the unconditional α quantile of the innovations.

We assume that returns are not predictable and consequently set the conditional mean to zero. For the volatility process, we assume either the standard GARCH(1, 1) of Bollerslev (1986), the exponential GARCH(1, 1) of Nelson (1991) or the asymmetric power ARCH(1, 1) of Ding et al. (1993), subsequently denoted by GARCH, EGARCH and APARCH. They are given by:

GARCH(1, 1)	$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$,
EGARCH(1, 1)	$\log\left(\sigma_t^2\right) = \omega + \alpha z_{t-1} + \gamma\left(\|z_{t-1}\| - \mathrm{E}\left[\|z_{t-1}\|\right]\right) + \beta \log\left(\sigma_{t-1}^2\right)$,
APARCH(1, 1)	$\sigma_t^\delta = \omega + \alpha\left(\|\varepsilon_{t-1}\| - \gamma \varepsilon_{t-1}\right)^\delta + \beta \sigma_{t-1}^\delta$.

In contrast to the standard GARCH, the EGARCH and APARCH specifications can respond asymmetrically with respect to positive and negative returns. The APARCH additionally accounts for the Taylor effect: the finding that the autocorrelation of absolute returns is often larger than that of squared returns (Taylor, 1986), which might improve the predictive power of the model. For the innovations z_t we assume the Normal distribution (N), the Student-t distribution (t) to account for fat tails and the filtered historical simulation (FHS) method of Barone-Adesi et al. (1999), which estimates $Q_\alpha(z_t)$ by the empirical quantile of the standardized returns. Combining the three variance processes with the three assumptions on the innovations, we obtain a total of 9 additional models. For estimation of the GARCH models we employ the rugarch library for R by Ghalanos (2015) and a rolling window of 1000 days.

3.3. *Competing Combination Approaches*

This section introduces a collection of 9 competing quantile combination approaches.

Unpenalized Quantile Regression

For a comparison with its penalized variants, we include the unpenalized QR estimator given by,

$$\widehat{\boldsymbol{\beta}}_t = \underset{\beta_{0,t},\, \boldsymbol{\beta}_t}{\arg\min} \frac{1}{t} \sum_{\tau=0}^{t-1} \rho_\alpha \left(r_{\tau+1} - \beta_{0,t} - \widehat{\boldsymbol{q}}'_{\tau+1|\tau} \boldsymbol{\beta}_t \right). \tag{12}$$

Convex Quantile Regression

Given that we interpret the convexity constraint as a heavily restricted version of the elastic net penalty, we include convex QR defined as,

$$\widehat{\boldsymbol{\beta}}_t = \underset{\beta_{0,t},\, \boldsymbol{\beta}_t}{\arg\min} \frac{1}{t} \sum_{\tau=0}^{t-1} \rho_\alpha \left(r_{\tau+1} - \beta_{0,t} - \widehat{\boldsymbol{q}}'_{\tau+1|\tau} \boldsymbol{\beta}_t \right), \text{ s.t. } \beta_{m,t} \geq 0 \text{ for } m = 1, \ldots, M \text{ and } \sum_{m=1}^{M} \beta_{m,t} = 1. \tag{13}$$

For the estimation of the unpenalized and convex QR, we employ the quantreg package for R by Koenker (2016).

Simple Mean

Due to simplicity and empirical success in the mean forecasting literature (Timmermann, 2006), we also consider the simple average over all forecasts,

$$\widehat{\beta}_{m,t} = \frac{1}{M}, \quad \text{for all } m = 1, \dots, M. \tag{14}$$

Note that the simple mean, and all following combination approaches, do not include an intercept term, such that $\beta_{0,t} = 0$.

Trimmed Mean

A trimmed version of the simple mean combination is proposed by Timmermann (2006), which uses the relative rankings of the models to set the weight of some models to zero. This method is supposed to be more robust than the simple mean, as only the forecasts of the best models enter the combination. The weights are given by

$$\widehat{\beta}_{m,t} = \begin{cases} \frac{1}{\lfloor \eta M \rfloor}, & \text{if } R_t^m \leq \eta M \\ 0, & \text{else,} \end{cases} \quad \text{for all } m = 1, \dots, M, \tag{15}$$

where R_t^m is the rank of model m at time t with respect to the average tick losses up to time t, L_t^m. As we set $\eta = 0.25$ and our pool of standalone models includes 17 approaches, we average over the forecasts of the four best models.

Single Best

This approach assigns all weight to the in the past best performing standalone forecast, so that this method is rather a forecast selection than forecast combination,

$$\widehat{\beta}_{m,t} = \begin{cases} 1, & \text{if } R_t^m = 1 \\ 0, & \text{else.} \end{cases} \quad \text{for all } m = 1, \dots, M. \tag{16}$$

Inverse Loss

Also common in the mean forecasting literature is to weight the forecasts inversely proportional with respect to their historical performance measured by the loss (Timmermann, 2006),

$$\widehat{\beta}_{m,t} = \frac{\left(L_t^m\right)^{-1}}{\sum_{n=1}^{M} \left(L_t^n\right)^{-1}} \quad \text{for all } m = 1, \dots, M. \tag{17}$$

Inverse Rank

Timmermann (2006) suggests to weight the forecasts inversely proportional to their rank instead of the losses directly, as ranks are less sensitive to outliers than ranking by losses directly,

$$\widehat{\beta}_{m,t} = \frac{\left(R_t^m\right)^{-1}}{\sum_{n=1}^{M} \left(R_t^n\right)^{-1}} \quad \text{for all } m = 1, \dots, M. \tag{18}$$

Optimizing the Hit Rate

Hamidi et al. (2015) propose to estimate the combination weights by minimizing Mallows's C_p (Mallows, 1973) on the squared difference between the nominal and empirical hit rates subject to convexity of the weights, i.e.

$$\widehat{\boldsymbol{\beta}}_t = \arg\min C(\boldsymbol{\beta}_t), \quad \text{s.t. } \beta_{m,t} \geq 0 \text{ for } m = 1, \dots, M \text{ and } \sum_{m=1}^{M} \beta_{m,t} = 1, \tag{19}$$

where $C(\boldsymbol{\beta}_t) = \left(\alpha - \widehat{\alpha}_t(\boldsymbol{\beta}_t)\right)^2 + 2M \sum_{m=1}^{M} \beta_{m,t} s_m^2$. The term $\widehat{\alpha}_t(\boldsymbol{\beta}_t)$ is the in-sample hit rate of the combination when using the weight $\boldsymbol{\beta}_t$, $\widehat{\alpha}_{m,t}$ is the hit rate of the mth model and $s_m^2 = (\alpha - \widehat{\alpha}_{m,t})^2/(t - M)$ is an estimate of the error variance.

Sequential Relative Performance Approach

Shan and Yang (2009) propose a sequential method that is based on the relative historical performance of the standalone forecasts. Their approach takes the form,

$$\widehat{\beta}_{m,t} = \frac{\widehat{\beta}_{m,t-1} \exp\left(-\phi\rho_\alpha\left(r_t - \widehat{q}^{m}_{t|t-1}\right)\right)}{\sum_{n=1}^{M} \widehat{\beta}_{n,t-1} \exp\left(-\phi\rho_\alpha\left(r_t - \widehat{q}^{n}_{t|t-1}\right)\right)} \quad \text{for } m = 1, \ldots, M, \tag{20}$$

where the initial weights are $\beta_{m,0} = 1/M$ for all $m = 1, \ldots, M$. Each period, this technique increases the weight of the models that produced low losses in the past and vice versa. We set the tuning parameter $\phi = 1$ as this value performs best in the empirical application of Shan and Yang (2009).

3.4. Forecast Evaluation

We evaluate the VaR forecasts via backtesting and by comparing the realized tick losses. While backtesting allows us to individually assess whether a VaR forecast is valid or not, a comparison of the tick losses allows us to find the statistically most precise predictions.

Christoffersen (1998) terms a VaR forecast efficient with respect to the information set $\mathcal{F}_t$ if the hit variable

$$H_{t+1}(\alpha) = \begin{cases} 1, & \text{if } r_{t+1} \leq \widehat{q}_{t+1|t}(\alpha), \\ 0, & \text{else,} \end{cases} \tag{21}$$

satisfies the property of correct conditional coverage given by $\mathrm{E}\left[H_{t+1}(\alpha)|\mathcal{F}_t\right] = \alpha$. Similarly, the criterion of unconditional coverage is satisfied if $\mathrm{E}\left[H_{t+1}(\alpha)\right] = \alpha$.

As the likelihood ratio test of Christoffersen (1998) shows inferior size and power properties compared to more recent alternatives (e.g. Berkowitz et al., 2011), we test the hypothesis of correct conditional coverage of a VaR forecast with the dynamic quantile (DQ) backtest of Engle and Manganelli (2004). For application of the DQ backtest we estimate the equation

$$H_{t+1}(\alpha) - \alpha = \gamma_0 + \gamma_1 H_t(\alpha) + \gamma_2 \widehat{q}_{t+1|t}(\alpha) + u_{t+1}, \tag{22}$$

with least squares, where the specification of the regressors follows Berkowitz et al. (2011). The actual backtest is a Wald test for the hypothesis $\gamma_0 = \gamma_1 = \gamma_2 = 0$, which is asymptotically χ^2_3 distributed. In addition to the DQ test, we also provide results for the unconditional coverage hypothesis using the likelihood ratio test of Kupiec (1995), which tests whether $H_{t+1}(\alpha)$ is unconditionally Bernoulli distributed with probability α.

As McAleer et al. (2013a) and Bernardi and Catania (2016), we formally compare the realized tick losses by applying the MCS procedure of Hansen et al. (2011). The MCS repeatedly evaluates the hypothesis $\mathrm{E}\left[d_{ij}\right] = 0$ for all $i, j = 1, \ldots, M$, where d_{ij} is the loss differential between the predictions of model i and model j. Whenever it is possible to reject the hypothesis of equal predictive ability among all forecasts, the worst performing model (with respect to the losses) is eliminated and the procedure starts anew. Eventually, this technique produces a set of models that can statistically not be further distinguished at a certain significance level. Thus, the more often a model is included within this superior set of models (SSM), the better is its predictive power.

For computation of the MCS we use the ARCH package for Python by Sheppard (2017). We report results for the T_R statistic of Hansen et al. (2011), based on 100,000 repetitions of the moving block bootstrap with a block size of 10 days to account for the possibility of clustered VaR hits. We also check the results for block sizes of 5 and 20 days and find the results to be robust with respect to the choice of the block length.

4. Empirical Results

4.1. Estimation Window of Elastic Net Quantile Regression

Apart from the regularization parameter, λ, and the balance between lasso and ridge, δ, we need to decide on the length of the estimation window. To determine the optimal estimation window for lasso, elanet and ridge QR, we compare the out-of-sample predictive performance depending on the window length used for the estimation of the parameters when we hold the regularization level λ fixed. For each stock $i = 1, \ldots, N$, rolling window sizes w = 250, 500, 1000, 1500 and a recursively extending window starting in January 3 2000, we compute the average tick loss,

$$\mathrm{TL}_{i,w}(\lambda,\, \delta) = 1/R \sum_{t=T}^{T+R-1} \rho_\alpha \left(r^i_{t+1} - \widehat{q}^i_{t+1|w}(\lambda,\, \delta) \right), \tag{23}$$

where r^i_{t+1} is the return of stock i at time $t+1$ and $\widehat{q}^i_{t+1|w}(\lambda,\, \delta)$ is the elastic net combined VaR forecast of stock i for day $t+1$ based on a window w. Here, we compute the tick loss for the overall evaluation horizon spanning 8 years. For an easy interpretation of eq. (23), we average over the assets to get a single number per shrinkage value and window length,

$$\overline{\mathrm{TL}}_w(\lambda,\, \delta) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{TL}_{i,w}(\lambda,\, \delta), \tag{24}$$

and thereby obtain a measure for the average precision. Figure 2 shows the average tick loss for ridge ($\delta = 0$), the elastic net ($\delta = 0.5$) and lasso ($\delta = 1$) QR for fixed regularization levels between 10^{-5} and 10^2. We can see that all loss curves reach their minimum within the considered grid of shrinkage values, which implies that neither the empirical quantile of the data ($\lambda \to \infty$) nor the unpenalized quantile regression estimator ($\lambda \to 0$) is optimal. Considering these two extreme cases, we see that the empirical quantile is best estimated with short windows, while less penalized models profit from longer estimation samples. When we consider the minimum of the respective loss curves, we see that the average loss is decreasing in the length of the estimation window: if we obtain a good estimate of λ it is beneficial to use all available data for the estimation of the parameters. Consequently, in what follows we estimate the elastic net penalized QR with the recursively extending window approach. For a fair comparison, we apply the competing combination approaches based on the same window of data.

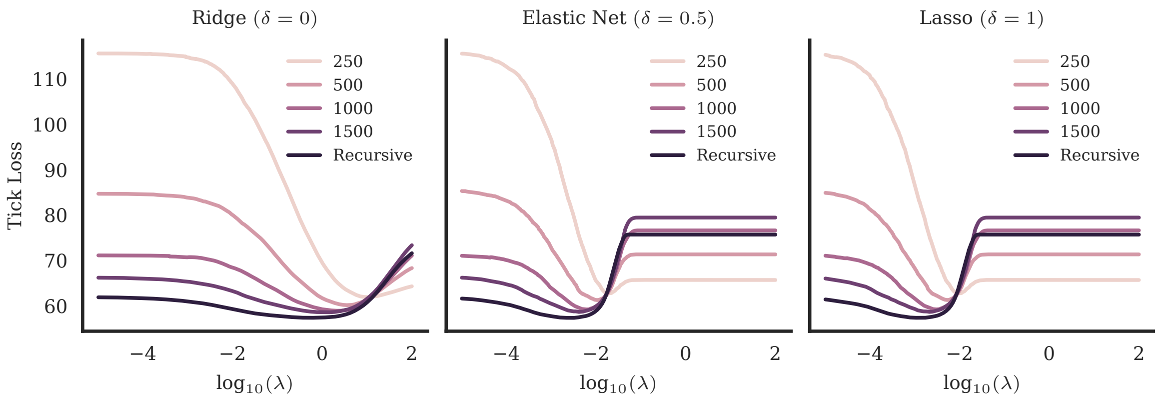


Figure 2: Average tick loss x 1e5 over all 30 assets for ridge, elanet and lasso QR for a grid values of the regularization parameter, λ, between 10^{-5} and 10^2. Each of the three panels shows the tick losses for a variety of rolling window sizes and a recursively extending window.

4.2. Conditional Coverage Backtesting

We start the discussion of the forecast comparison results by evaluating the standalone and the combined predictions via conditional coverage backtesting. As presenting tables with detailed results for all 30 assets is not feasible, we instead discuss the DQ backtest results based on stacked bar plots with the number of backtest rejections at two different significance levels. The results for all horizons are presented in Figure 3; one panel for each of the three evaluation horizons. In this figure, the red bars indicate the number of times (out of a maximum of 30) a VaR forecast is rejected at the 1% significance level, which indicates severe evidence against the validity of the forecast. Similarly, the orange bar denotes a p-value of the DQ backtest between 1% and 10%, which can indicate both, a valid or a non-valid prediction.

Figure 3a shows the backtest results for the predictions during the overall evaluation sample from January 2007 to December 2014. In this figure, we can see that the standalone models are often and highly rejected. Especially the HS and RiskMetrics, which are particularly popular in practice, are among the models with the most rejections of the conditional coverage hypothesis. The best performing standalone models are the GARCH specifications using the t-distribution, FHS and the CAViaR-SAV model.

Considering the combination approaches, we note that with the exception of the unpenalized QR and lasso QR with BIC estimated regularization parameters, the combined forecasts are less often rejected than the standalone predictions. Thus, implementing forecast combination is generally beneficial for the VaR prediction and it is a great improvement on the standalone models.

Evaluating the different combination approaches, we find that the unpenalized QR fails to produce valid VaR forecasts as it is the most often rejected technique. The reason for its poor performance is the previously discussed multicollinearity among the forecasts: the unpenalized QR overfits the data. From the figure, we can see that imposing the convexity restriction improves the predictions and as a matter of fact, the convex QR is among the least often rejected models. We furthermore note that trimming the models before averaging them leads to more backtest rejections than the simple mean, even though the trimming is supposed to improve on its simpler variant. Also averaging over the inverse of the ranks of the models, in contrast to the inverse of the tick losses directly, does not improve the predictive performance. Thus, two conclusions from the mean forecasting literature (Timmermann, 2006), namely that trimming and averaging based on ranks improves upon the simpler variants, do not apply in our comparison. Moreover, selecting a single model on a day-by-day basis performs worse than the averaging techniques and even worse than many of the standalone models, which is in line with Aiolfi and Timmermann (2006). The quantile-specific suggestions of Shan and Yang (2009) and Hamidi et al. (2015) exhibit roughly the same number of rejections as the other competing combination approaches. A potential reason why these techniques are not able to improve on the simpler averaging approaches is the lack of an intercept term, which corrects the biased standalone predictions.

Considering the penalized QR, we find that the forecasts of lasso with BIC selected shrinkage values are nearly as often rejected as the unpenalized QR. The reason is that the BIC induces an insufficient amount of shrinkage such that the predictions are too similar with those of the unpenalized QR, which is in line with the findings in Koenker (2011). We furthermore find that lasso and elanet QR are less often rejected than ridged QR, independent of the approach of selecting the shrinkage parameter. Thus, the sparsity enforcing property of the lasso operator is crucial for good predictions. When we compare the proposed heuristic rule (denoted by fix) and the time series CV, we see that the heuristic rule leads to less rejections. In fact, lasso and elanet QR with the heuristic rule are just once rejected with a DQ backtest p-value between 1% and 10%, which is only very weak evidence against the validity of this strategy, in particular in comparison to the rejection rates of the alternatives.

Next, we split the overall evaluation sample into two equally sized windows of 4 years. Considering the first half of the evaluation sample in Figure 3b, which represents a period of high volatility around the 2007 / 2008 global financial crisis, we find that the number of rejections generally increases, in comparison to the overall horizon. The number of backtest rejections raises especially for the standalone models and with the exception of the best performing approach (GARCH with t distributed innovations), the forecasts of all standalone models are now at least 5 times rejected at the 1% level. Also among the combination approaches, the number of rejections is larger than during the overall evaluation period. Nevertheless, lasso

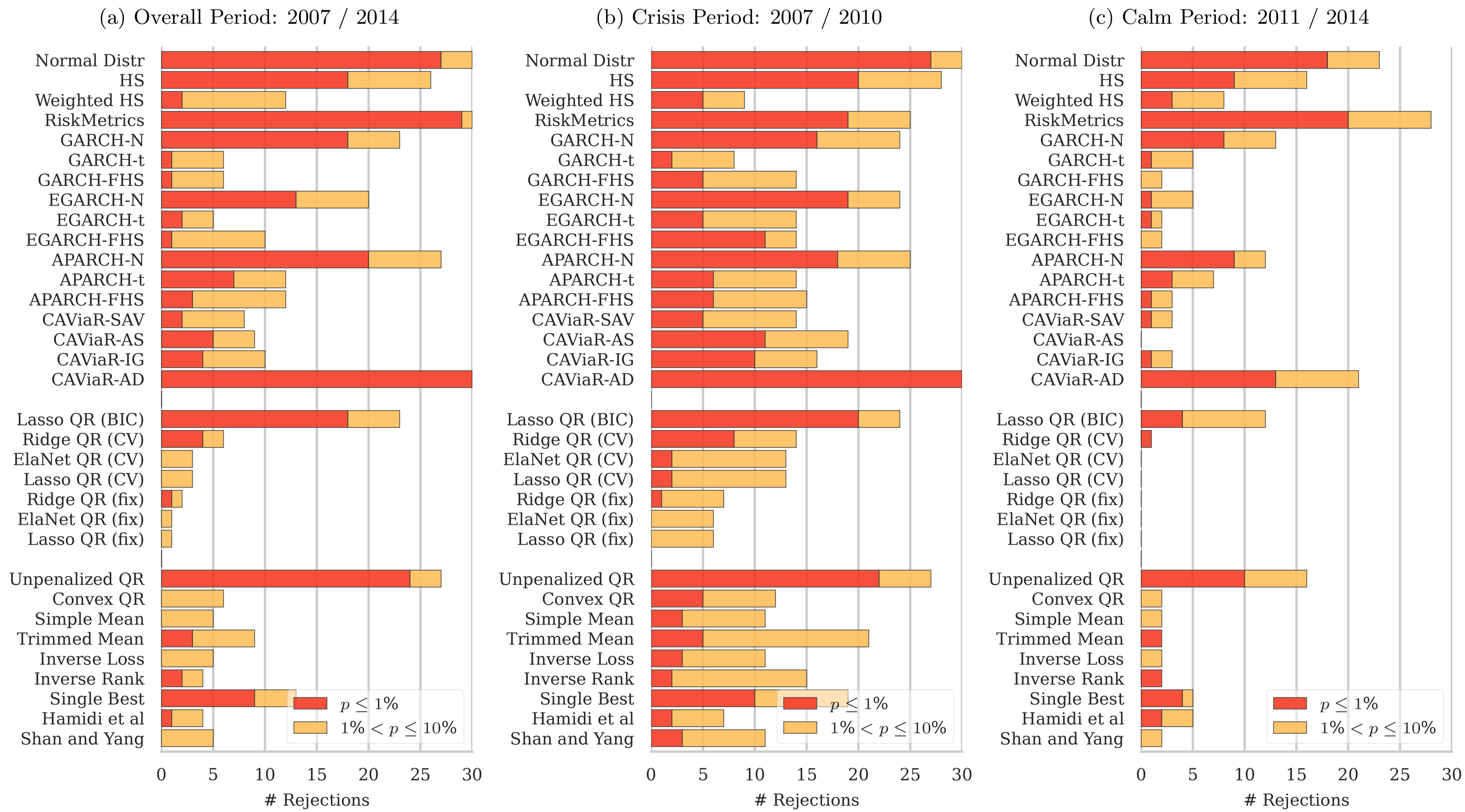


Figure 3: Number of dynamic quantile backtest rejections for all approaches at two different significant levels indicated by the colored bars. The three panels show the results for the overall / crisis / calm period, respectively The empty lines separate the standalone models, the penalized QR and the competing combination techniques.

and elanet QR, with the amount of regularization estimated by the heuristic rule, are never rejected at the 1% level and only a few times we find a p-value between 1% and 10%.

Last, we evaluate the VaR forecasts during the calm time from January 2011 to December 2014. During this time, the volatility is lower and, thus, the VaR is easier to predict. We can see this in Figure 3c, which shows that several standalone models are now hardly rejected, and the CAViaR-AS is never rejected at the 1% level. Nevertheless, the techniques involving the Normal distribution are still often rejected, raising doubt against the validity of this assumption. Regarding the combination approaches, we note that the penalized QR performs quite well: with the exception of lasso QR (BIC) and ridge QR with CV selected shrinkage values, there is no single violation of the conditional coverage hypothesis at the 1% level. We also observe that this does not apply for the competing combination approaches, which are all at least twice rejected at the 10% level.

4.3. Hit Ratios: Unconditional Coverage

In addition to the conditional coverage tests, we also provide details on the unconditional coverage properties of the predictions. Thus, in Figure 4 we provide the empirical hit rates, which should be close to the nominal value. The figure is split into three panels for the evaluation horizons. In each panel, the gray dots represent the empirical hit rates in percentage points. The gray line shows the 1% value and the two black lines depict the 99% confidence interval for the unconditional coverage backtest of Kupiec (1995). Note that the hit rates beyond 4%, which are occasionally produced by the CAViaR-AD model, are not shown to improve presentation of the figure. In comparison to many other approaches, penalized QR performs well with respect to the unconditional coverage test: almost always the hit ratios are within the 99% confidence interval. We also note that there is a tendency of the combination approaches not involving a bias correcting intercept term to produce hit rates larger than 1%, whereas the penalized QR through an intercept correction are more centered around the 1% value.

4.4. Relative Evaluation of all Forecast Approaches

Besides the individual evaluation via backtesting we now present the result of the MCS procedure by Hansen et al. (2011). The application of the MCS on the tick losses yields a p-value for each of the models, which can be used to decide whether some model is in or out of the superior set of models (SSM). Note that Hansen et al. (2011) express concerns about the validity of the crucial assumption of stationarity loss differentials d_{ij} when the model parameters are recursively estimated. In order to account for this concern, we perform unit-root tests on the loss differentials and do not find evidence against stationarity.

The results for all three evaluation horizons are presented in Table 1; one panel for each of the three horizons. In each, the first and second column are the number of times a model is included in the 90%, respectively 75% SSM. In the third column, similar to Grigoryeva et al. (2017), we present the average of the 30 MCS p-values per model: the larger the average p-value, the higher a model is ranked by the MCS procedure.

Overall, the results are less decisive than the backtesting, as typically a larger number of models is included in the 90% and 75% SSM. This is similar to the findings of Bernardi and Catania (2016) who can only eliminate a small number of models from the full set of models using the MCS. During the crisis time, the forecasts of lasso and elanet QR are more often included in the 75% and 90% SSM than the predictions of ridged QR and their average p-values are larger. That indicates again that penalized QR performs better when the model is allowed to set some coefficients to zero. During the calm time, this relation reverses: ridging performs slightly better as its forecasts are more often included in the MCS. We can conclude that the variable selection property of the elastic net and lasso penalization is especially important in volatile times when many of the standalone models fail. When all models perform well (e.g. during the calm time), this property is less relevant and simply shrinking the coefficients suffices to obtain precise predictions. Comparing the heuristic rule and the cross validation, we find that the number of times the models are within the SSM are comparable. Nonetheless, the average p-values are larger for the heuristic rule and, thus, the forecasts are more precise.

Regarding the competing combination approaches, we find that during the calm period, many approaches (e.g. the simple mean or the inverse ranking approach) perform well. The reason is that during the calm time,

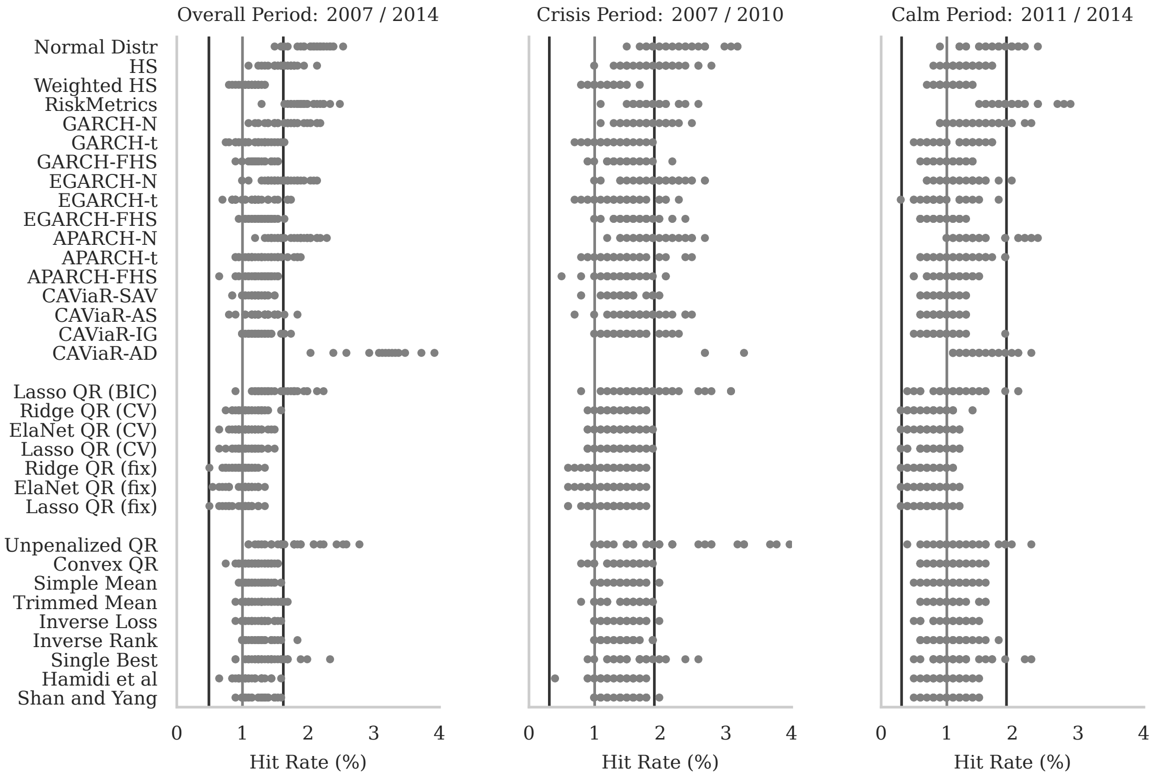


Figure 4: Empirical hit rates. The figure is split into panels for the the overall / crisis / calm period. In each, every dot represents the empirical hit rate in percentage points for one of the assets. The gray line denotes the nominal value of 1% and the black lines indicate the 99% confidence interval for the unconditional coverage test.

almost all standalone approaches exhibit a good performance and, thus, also simple averaging techniques perform well as they average over the forecasts of reasonable models only. When we consider the crisis and the overall period, however, none of these approaches can outperform the lasso and elanet QR as these two techniques can set the weight of poorly performing models to zero and thereby increase the predictive performance.

Summing up the results, we find that: (1) The conditional coverage hypothesis is less often rejected for the penalized QR than for the standalone models and the competing combination forecasts and the hit rates are close to the nominal value of 1%. (2) The penalized QR is often included in SSM, which indicates a good relative performance. (3) The proposed heuristic rule performs well and even better than the time series CV. (4) The differences between lasso, elastic net and ridge QR are rather small, yet the former two perform slightly better during the crisis time and the overall period as they are able to set the coefficients of poorly performing models to zero. (5) Contradictory to the fact that elanet QR should be superior to ridge and lasso QR as it combines the strengths from both, we do not find that the elastic net performs better than lasso QR on its own.

4.5. Robustness Check: The Heuristic Rule

The empirical comparison reveals that the proposed heuristic rule performs well with respect to backtesting and tick losses. In order to demonstrate that the results are not simply due to a lucky choice of the maximum allowed absolute sum of the weights s, Figure B.6 in the appendix shows the number of dynamic quantile backtest rejections at the 1% level and the average tick loss over all assets for $0.5 \leq s \leq 1.5$. For s in (0.75, 1.0), we find just up to two rejections at the 1% significance level and relatively flat tick loss curves. Throughout the overall and calm period, i.e. excluding the crisis time, this range is even consider-

Table 1: Relative Comparison of all forecasting approaches

Approach	Overall Period			Crisis Period			Calm Period		
	$\#_{90\%}$	$\#_{75\%}$	$\overline{p}_{\mathrm{MCS}}$	$\#_{90\%}$	$\#_{75\%}$	$\overline{p}_{\mathrm{MCS}}$	$\#_{90\%}$	$\#_{75\%}$	$\overline{p}_{\mathrm{MCS}}$
Normal Distr	25	17	0.32	24	19	0.34	30	30	0.77
HS	20	9	0.27	24	12	0.31	29	27	0.63
Weighted HS	16	13	0.31	18	15	0.42	17	13	0.30
RiskMetrics	30	28	0.72	30	30	0.82	30	29	0.64
GARCH-N	30	27	0.68	30	26	0.70	30	29	0.75
GARCH-t	28	26	0.74	30	26	0.75	27	24	0.66
GARCH-FHS	28	25	0.62	26	25	0.63	27	23	0.61
EGARCH-N	30	30	0.83	30	29	0.76	30	30	0.92
EGARCH-t	30	30	0.90	30	30	0.87	30	29	0.88
EGARCH-FHS	29	28	0.77	29	28	0.73	29	29	0.81
APARCH-N	30	29	0.72	30	29	0.76	30	30	0.78
APARCH-t	30	29	0.85	29	29	0.84	29	29	0.79
APARCH-FHS	26	23	0.60	25	23	0.57	29	27	0.69
CAViaR-SAV	29	27	0.64	29	25	0.67	28	26	0.69
CAViaR-AS	30	29	0.67	30	27	0.62	29	29	0.84
CAViaR-IG	28	26	0.65	29	25	0.67	27	23	0.66
CAViaR-AD	9	4	0.10	11	4	0.12	30	29	0.71
Lasso QR (BIC)	27	24	0.54	27	24	0.60	28	23	0.59
Ridge QR (CV)	29	29	0.76	29	28	0.76	30	27	0.67
ElaNet QR (CV)	30	30	0.77	30	30	0.80	29	23	0.62
Lasso QR (CV)	30	30	0.79	30	30	0.80	27	23	0.63
Ridge QR (fix)	30	30	0.84	30	29	0.82	29	28	0.70
ElaNet QR (fix)	30	30	0.85	30	30	0.85	29	26	0.72
Lasso QR (fix)	30	30	0.86	30	30	0.88	27	25	0.68
Unpenalized QR	24	19	0.37	28	22	0.46	26	24	0.55
Convex QR	30	29	0.78	30	28	0.77	27	24	0.65
Simple Mean	30	29	0.84	29	27	0.75	30	29	0.84
Trimmed Mean	30	29	0.79	29	28	0.75	29	27	0.75
Inverse Loss	29	29	0.77	27	25	0.63	30	29	0.81
Inverse Rank	30	30	0.83	30	28	0.76	30	30	0.82
Single Best	30	27	0.71	30	29	0.69	29	28	0.76
Hamidi et al	24	22	0.58	25	22	0.65	26	25	0.59
Shan and Yang	30	30	0.84	29	27	0.74	30	29	0.83

This table presents the results of the model confidence set procedure across all 30 assets. $\#_{90\%}$ and $\#_{75\%}$ are the number of times a model is included in the 90%, respectively 75% superior set of models and $\overline{p}_{\mathrm{MCS}}$ is the average over the 30 individual MCS p-values based on the T_R statistics using 100,000 iterations of the moving block bootstrap with a block length of ten days.

ably larger. These findings thus confirm the robustness and performance of the proposed way of selecting the regularization parameter λ.

4.6. Combination Weights and Relative Importance of the Predictors

To get some insight into how the penalized QR estimates the weights and selects the models, Figure C.7 in the appendix shows the estimated combination weights for the VaR forecasts of the AT&T stock as an illustration. The three panels show the estimated weights and intercepts for lasso, elastic net and ridge QR over the time from January 2007 to December 2014. For that particular stock, the models RiskMetrics and APARCH-N dominate the estimated weights of lasso and elanet QR, the weights of the other standalone models are comparably small. The lasso and the elastic net penalty, thus, efficiently set the coefficients of 15 out of 17 variables to zero. When we look at the estimated weights of ridge QR, we find that the weights are very similar across the models, which underlines the grouping effect of the ridge penalty: the coefficients of highly correlated variables are shrunk towards each other. The combination weights are moreover relatively stable over time, there is not much variation in neither the choice of models nor in the value of the weights.

To conserve some space, in Figure D.8 we report the median of the estimated weights and intercepts over the time for each combination of asset and standalone model. Each cell entry contains the median estimated weight (over the time) for all standalone models and assets. A blank entry indicates that a coefficient is on average zero. We find that for lasso and elanet QR, the most important predictors are the GARCH models and RiskMetrics. Even though RiskMetrics is individually not a satisfactory predictor, the lasso and elastic net strongly opt for its inclusion in the combinations. A potential reason is that RiskMetrics' estimation error is zero as it is a calibrated model, i.e. it may serve as stabilizing component. From the median weights of ridge QR, we can again observe the strong grouping effect as almost all weights are between 0 and 0.1.

Regarding the number of active predictors, i.e. the number of non zero coefficients, we find that on average the lasso and elanet QR combine the predictions of up to 6 models. We can also see that there is often just one dominating approach, for example the GARCH-FHS model for the DD stock or RiskMetrics for WMT.

5. Conclusion

In this paper we propose the combination of VaR forecasts via elastic net penalized QR. The advantage of this technique is that it reduces overfitting due to the high multicollinearity of the standalone forecasts. Through the shrinkage and variable selection properties of the elastic net penalty, regularized QR stabilizes the estimates of the combination weights and thereby improves the predictions.

In the empirical application, we combine the VaR forecasts of 17 standalone models for 30 constituents of the DJIA and consider three evaluation horizons. We compare the penalized QR combined predictions with the standalone forecasts and a large variety of competing approaches. Apart from the elastic net QR, we consider the two corner cases, namely lasso and ridge QR. The penalized QR combined forecasts are less often rejected by backtests than the individual models and competing combination approaches and are most of the time included in the MCS proposed by Hansen et al. (2011). We find that in volatile times, the lasso and elanet QR perform better than the ridge QR, i.e. in times when many standalone models fail, so that the variable selection property is very relevant. We also observe that the elanet QR does not perform better than pure lasso QR, even though the elastic net is supposed to stabilize the lasso in case of highly correlated covariates (Zou and Hastie, 2005). The results of this paper also suggest that some approaches often working well in mean forecasting, like trimming the models prior to averaging or using ranks instead of losses, do not perform well for VaR combination.

For future research, a comparison between the elastic net and quantile regression boosting (Zheng, 2012) would be interesting. One could furthermore consider nonlinear forecast combination via quantile random forests introduced by Meinshausen (2006) or the post-lasso quantile regression estimator by Belloni and Chernozhukov (2011).

Acknowledgement

The author thanks Winfried Pohlmeier, Roxana Halbleib, Timo Dimitriadis, Lyudmila Grigoryeva and the participants of the International Symposium on Forecasting 2016, the German Statistical Week 2015 and the doctoral workshops in Lancaster, Hannover, Hamburg and Konstanz for comments and suggestions that helped to improve this paper. Financial support by the German Research Foundation (DFG) within the research group "Robust Risk Measures in Real Time Settings" and general support by the Graduate School of Decision Sciences (University of Konstanz) is gratefully acknowledged. This work was performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC. All remaining errors are ours.

Appendix A. Plots and Summary Statistics of the Return Series

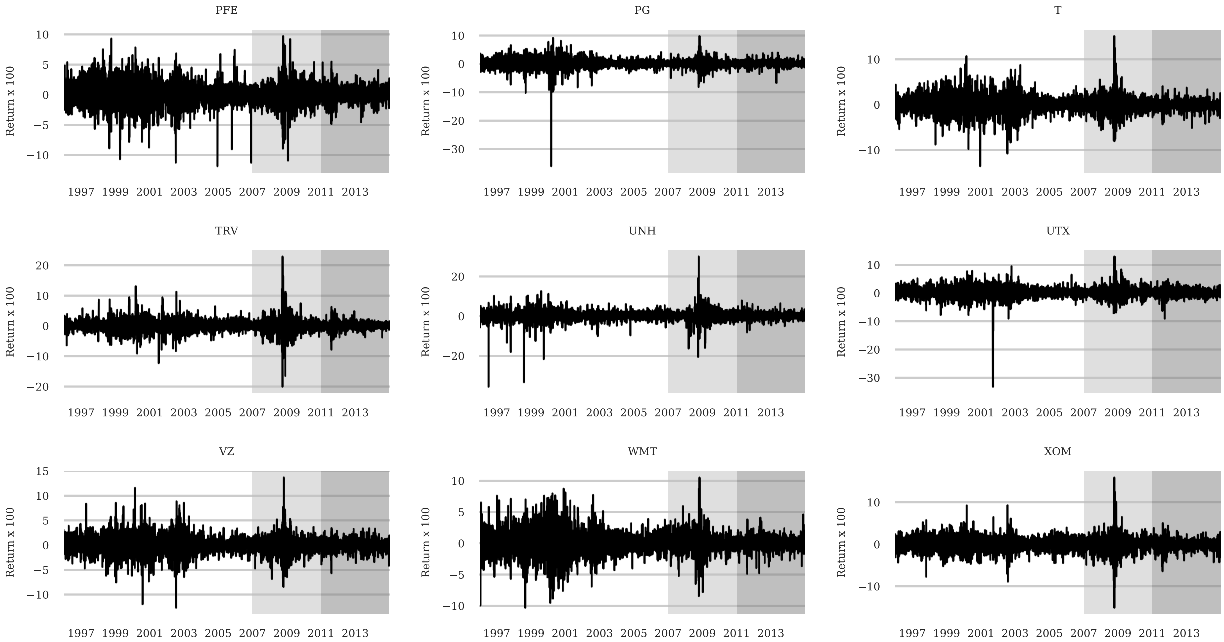
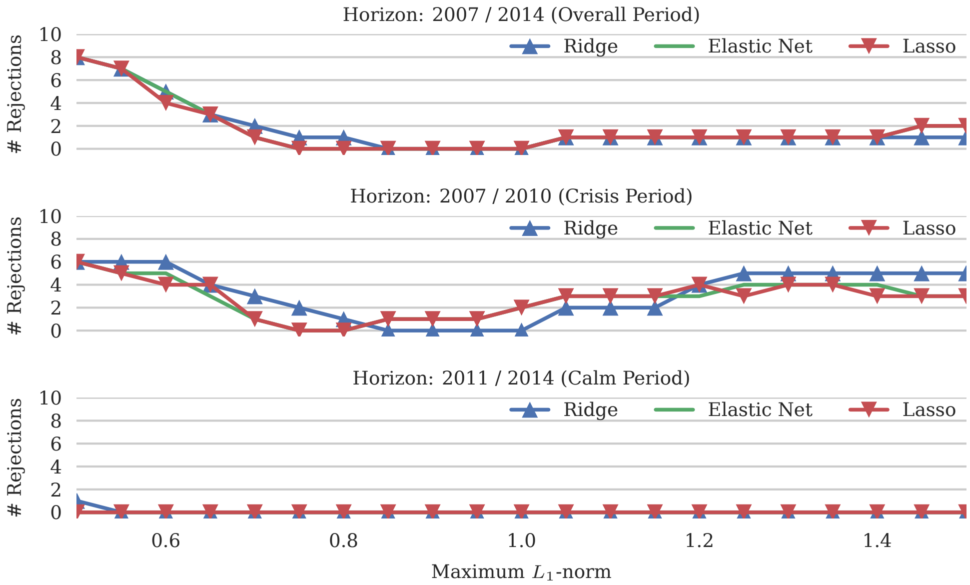


Figure A.5: Log return series from January 2, 1996 to December 31, 2014. The two gray shaded areas indicate the forecast evaluation horizons January 3, 2007 to December 31, 2010 (1008 days) and January 3, 2011 to December 31, 2014 (1006 days). In addition, we consider the overall period from January 3, 2007 to December 31, 2014 (2014 days).

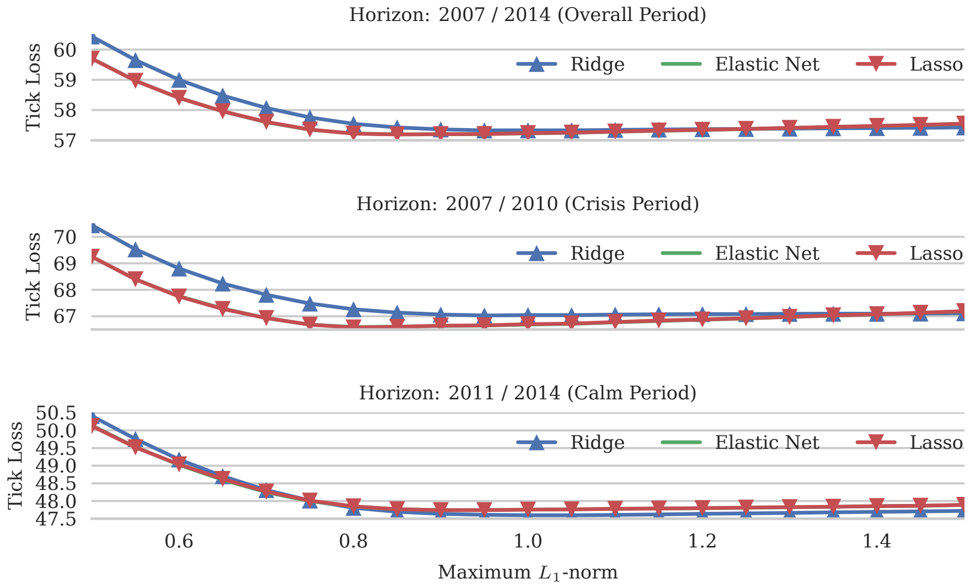
Table A.2: Ticker symbols, company names and summary statistics of the log returns x 100.

Symbol	Name	Min	Max	Mean	Var.	Kurt.	Skew.	JB	JB-p
AAPL	Apple	−73.12	28.69	0.10	9.54	73.62	−2.56	1.0e6	0.00
AXP	American Express Company	−19.35	18.77	0.04	5.66	10.97	0.01	1.3e4	0.00
BA	The Boeing Company	−19.39	14.38	0.02	4.07	9.78	−0.37	9.3e3	0.00
CAT	Caterpillar Inc.	−15.69	13.73	0.04	4.55	7.17	−0.09	3.5e3	0.00
CSCO	Cisco Systems, Inc.	−17.69	21.82	0.04	7.21	9.38	0.05	8.1e3	0.00
CVX	Chevron Corporation	−13.34	18.94	0.03	2.71	12.40	0.08	1.8e4	0.00
DD	E. I. du Pont de Nemours and Company	−12.03	10.86	0.02	3.51	7.17	−0.15	3.5e3	0.00
DIS	The Walt Disney Company	−20.29	14.82	0.03	4.07	10.69	−0.07	1.2e4	0.00
GE	General Electric Company	−13.68	17.98	0.02	3.79	10.52	0.01	1.1e4	0.00
HD	The Home Depot, Inc.	−33.88	13.16	0.05	4.47	20.33	−0.79	6.0e4	0.00
HPQ	Hewlett-Packard Company	−22.35	15.95	0.01	6.53	10.08	−0.30	1.0e4	0.00
IBM	International Business Machines	−16.89	12.37	0.04	3.36	10.50	−0.03	1.1e4	0.00
INTC	Intel Corporation	−24.89	18.33	0.03	6.41	9.79	−0.37	9.3e3	0.00
JNJ	Johnson & Johnson	−17.25	11.54	0.03	1.82	13.00	−0.22	2.0e4	0.00
JPM	JPMorgan Chase & Co.	−23.23	22.39	0.02	6.74	14.26	0.23	2.5e4	0.00
KO	The Coca-Cola Company	−11.07	13.00	0.02	2.20	9.57	0.00	8.6e3	0.00
MCD	McDonald's Corporation	−13.72	10.31	0.03	2.52	8.47	−0.04	6.0e3	0.00
MMM	3M Co	−10.08	10.50	0.03	2.41	7.30	−0.02	3.7e3	0.00
MRK	Merck & Co., Inc.	−31.17	12.25	0.01	3.37	25.90	−1.26	1.1e5	0.00
MSFT	Microsoft Corporation	−16.96	17.87	0.04	4.25	10.11	−0.07	1.0e4	0.00
NKE	Nike	−21.65	13.78	0.05	4.45	11.84	−0.15	1.6e4	0.00
PFE	Pfizer, Inc.	−11.82	9.69	0.02	3.17	6.84	−0.19	3.0e3	0.00
PG	Procter & Gamble	−36.01	9.73	0.03	2.31	73.03	−2.96	9.8e5	0.00
T	AT&T Inc.	−13.54	15.08	0.00	3.14	8.26	0.06	5.5e3	0.00
TRV	The Travelers Companies, Inc.	−20.07	22.76	0.03	3.70	16.55	0.35	3.7e4	0.00
UNH	UnitedHealth Group	−35.59	29.83	0.05	5.61	34.18	−1.36	2.0e5	0.00
UTX	United Technologies Corporation	−33.20	12.79	0.05	3.21	30.93	−1.30	1.6e5	0.00
VZ	Verizon Communications Inc.	−12.61	13.66	0.01	2.91	7.97	0.14	4.9e3	0.00
WMT	Wal-Mart Stores, Inc.	−10.26	10.50	0.04	2.87	7.14	0.08	3.4e3	0.00
XOM	Exxon Mobil Corporation	−15.03	15.86	0.03	2.53	11.61	0.02	1.5e4	0.00

Appendix B. Robustness Check



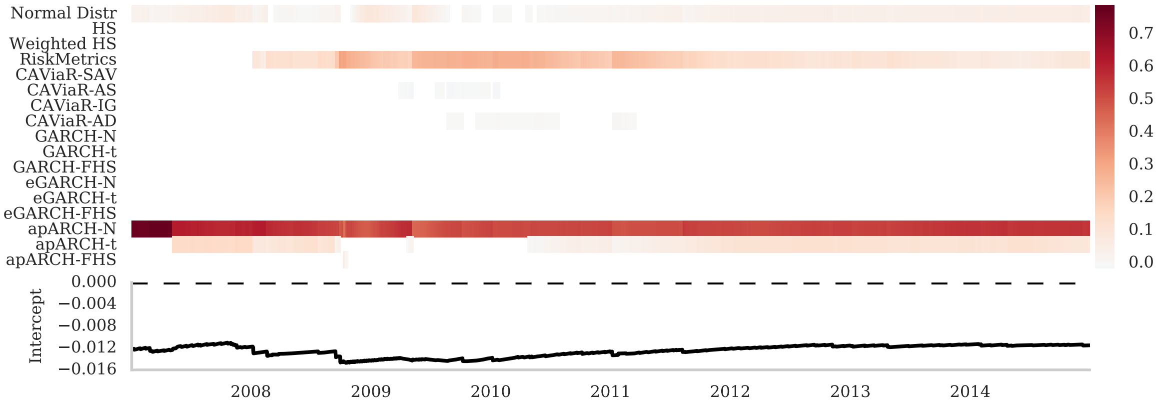
(a) Number of dynamic quantile backtest rejections at the 1% level.



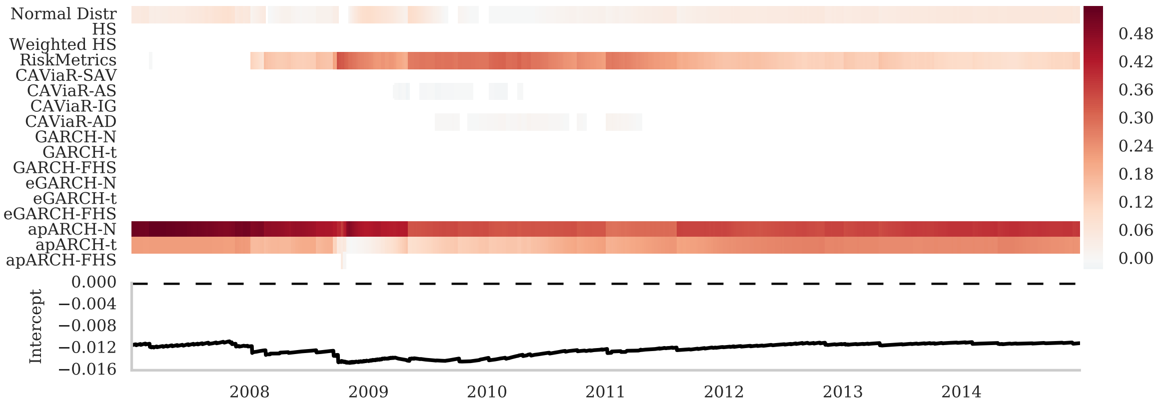
(b) Average tick loss over all 30 assets scaled by 10^5.

Figure B.6: Robustness check: in each of the two figures, the three panels depict all three evaluation horizons and all three varieties of the balance between lasso and ridge. For the quantities shown in the subfigures, see the subcaptions. The horizontal axis is the maximum allowed sum of the absolute weights for the heuristic rule in eq. (11).

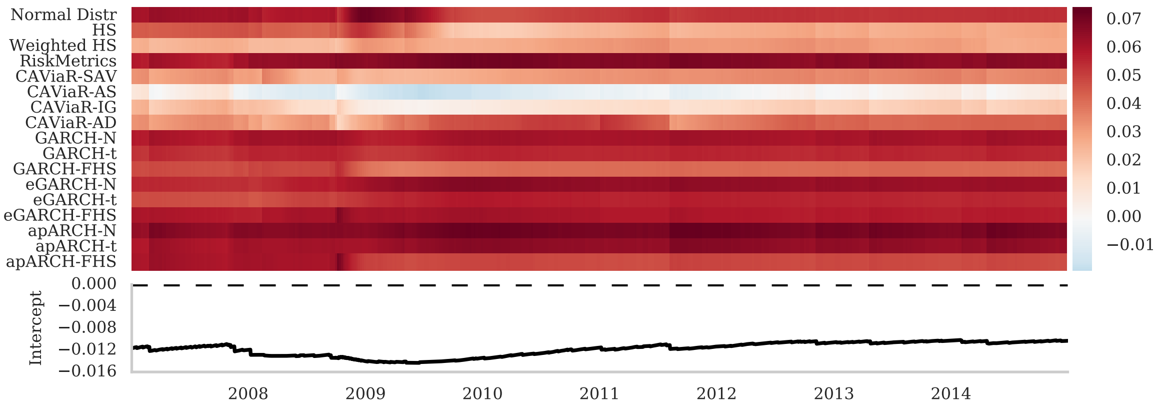
Appendix C. Estimated Combination Weights for AT&T



(a) Lasso Quantile Regression



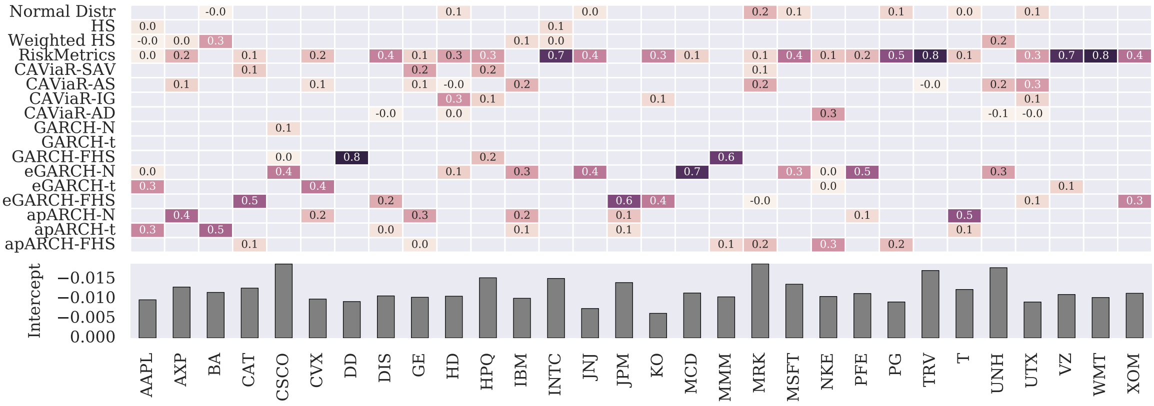
(b) Elastic Net Quantile Regression



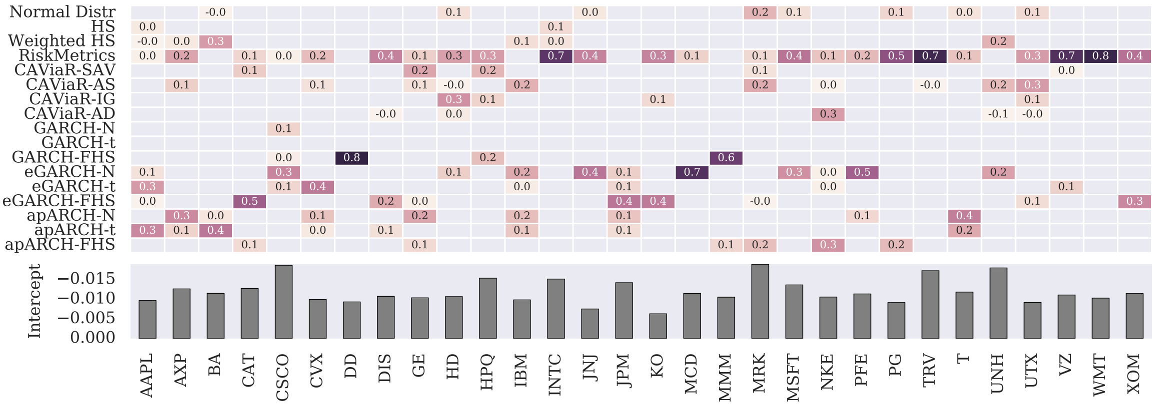
(c) Ridge Quantile Regression

Figure C.7: Values of the estimated combination weights and intercepts for the AT&T stock over the time from January 2007 to December 2014 for lasso, elastic net and ridge quantile regression.

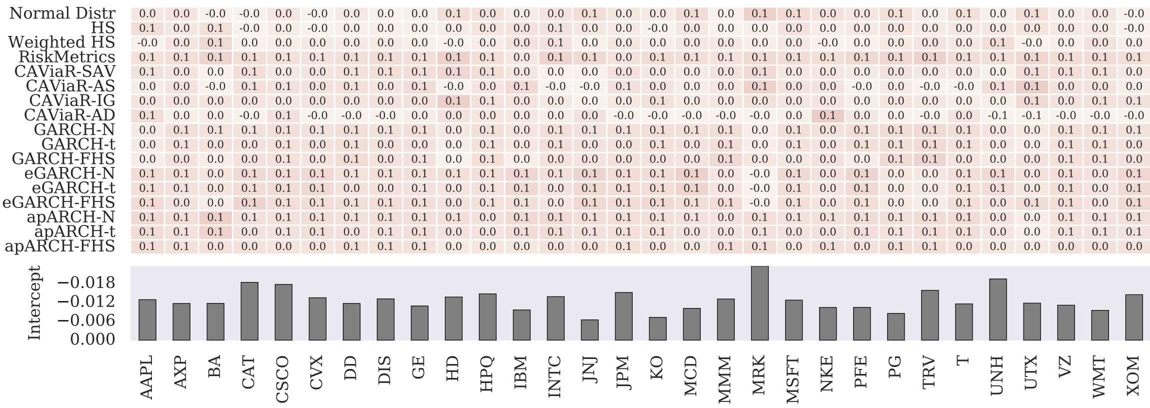
Appendix D. Median Estimated Combination Weights



(a) Lasso Quantile Regression



(b) Elastic Net Quantile Regression



(c) Ridge Quantile Regression

Figure D.8: Median of the estimated combination weights and intercepts for lasso, elastic net and ridge quantile regression over the time from January 2007 to December 2014; for all assets and standalone models. The values of the median weights are given in the cells, a blank entry indicates that a weight is on average zero. The cells are furthermore colored: the darker the cell, the larger is the coefficient. The bars show the value of the median intercepts.

References

Abad, P., Benito, S., 2013. A detailed comparison of value at risk estimates. Mathematics and Computers in Simulation 94, 258–276.

Aiolfi, M., Timmermann, A., 2006. Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics 135 (12), 31–53.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statistics Surveys 4, 40–79.

Barone-Adesi, G., Giannopoulos, K., Vosper, L., 1999. VaR without correlations for portfolios of derivative securities. Journal of Futures Markets 19 (5), 583–602.

Basel Committee on Banking Supervision, 1996. Overview of the Amendment to the Capital Accord to Incorporate Market Risks. Tech. rep., Bank for International Settlements.

Basel Committee on Banking Supervision, 2006. International Convergence of Capital Measurement and Capital Standards. Tech. rep., Bank for International Settlements.

Basel Committee on Banking Supervision, 2011. Basel III: A global regulatory framework for more resilient banks and banking systems. Tech. rep., Bank for International Settlements.

Belloni, A., Chernozhukov, V., 2011. ℓ_1-penalized quantile regression in high-dimensional sparse models. The Annals of Statistics 39 (1), 82–130.

Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating value-at-risk models with desk-level data. Management Science 57 (12), 2213–2227.

Bernardi, M., Catania, L., 2016. Comparison of Value-at-Risk models using the MCS approach. Computational Statistics 31 (2), 579–608.

Bernardi, M., Catania, L., Petrella, L., 2017. Are news important to predict the Value-at-Risk? The European Journal of Finance 23 (6), 535–572.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.

Boucher, C. M., Danielsson, J., Kouontchou, P. S., Maillet, B. B., 2014. Risk models-at-risk. Journal of Banking & Finance 44, 72–92.

Boudoukh, J., Richardson, M., Whitelaw, R. F., 1998. The best of both worlds: A hybrid approach to calculating value at risk. Risk 11 (5), 64–67.

Casarin, R., Chang, C.-L., Jimenez-Martin, J.-A., McAleer, M., Perez-Amaral, T., 2013. Risk management of risk under the Basel Accord: A Bayesian approach to forecasting Value-at-Risk of VIX futures. Mathematics and Computers in Simulation 94, 183–204.

Christoffersen, P., 1998. Evaluating Interval Forecasts. International Economic Review 39 (4), 841–862.

Ding, Z., Granger, C. W. J. G., Engle, R. F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1 (1), 83–106.

Einhorn, D., 2008. Private Profits and Socialized Risk. In: Einhorn, D., Brown, A. (Eds.), Global Association of Risk Professionals Risk Review (June/July 2008). Vol. 42. pp. 10–26.

Engle, R. F., Manganelli, S., 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. Journal of Business & Economic Statistics 22 (4), 367–381.

Ergen, I., 2015. Two-step methods in VaR prediction and the importance of fat tails. Quantitative Finance 15 (6), 1013–1030.

Fuertes, A.-M., Olmo, J., 2013. Optimally harnessing inter-day and intra-day information for daily value-at-risk prediction. International Journal of Forecasting 29 (1), 28–42.

Ghalanos, A., 2015. rugarch: Univariate GARCH models. R package version 1.3-6.

Giacomini, R., Komunjer, I., 2005. Evaluation and Combination of Conditional Quantile Forecasts. Journal of Business & Economic Statistics 23 (4), 416–431.

Gneiting, T., 2011a. Making and Evaluating Point Forecasts. Journal of the American Statistical Association 106 (494), 746–762.

Gneiting, T., 2011b. Quantiles as optimal point forecasts. International Journal of Forecasting 27 (2), 197–207.

Grigoryeva, L., Ortega, J.-P., Peresetsky, A., 2017. Volatility forecasting using global stochastic financial trends extracted from non-synchronous data. Forthcoming in Econometrics and Statistics. DOI: 10.1016/j.ecosta.2017.01.003.

Halbleib, R., Pohlmeier, W., 2012. Improving the Value at Risk Forecasts: Theory and Evidence from the Financial Crisis. Journal of Economic Dynamics and Control 36 (8), 1212–1228.

Hamidi, B., Hurlin, C., Kouontchou, P., Maillet, B., 2015. A DARE for VaR. Finance 36 (1), 7–38.

Hansen, B., 2008. Least-squares forecast averaging. Journal of Econometrics 146 (2), 342–350.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The Model Confidence Set. Econometrica 79 (2), 453–497.

Hart, J. D., 1994. Automated Kernel Smoothing of Dependent Data by Using Time Series Cross- Validation. Journal of the Royal Statistical Society. Series B (Methodological) 56 (3), 529–542.

Hart, J. D., Lee, C.-L., 2005. Robustness of one-sided cross-validation to autocorrelation. Journal of Multivariate Analysis 92 (1), 77–96.

Hastie, T., Tibshirani, R., Friedman, J., 2011. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition. Springer.

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC.

Hoerl, A. E., Kennard, R. W., 1970a. Ridge Regression: Applications to Nonorthogonal Problems. Technometrics 12 (1), 69–82.

Hoerl, A. E., Kennard, R. W., 1970b. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12 (1), 55–67.

Huang, H., Lee, T.-H., 2013. Forecasting Value-at-Risk Using High-Frequency Information. Econometrics 1 (1), 127–140.
James, G. M., 2003. Variance and Bias for General Loss Functions. Machine Learning 51 (2), 115–135.
Jeon, J., Taylor, J. W., 2013. Using CAViaR Models with Implied Volatility for Value-at-Risk Estimation. Journal of Forecasting 32 (1), 62–74.
Jorion, P., 2006. Value at Risk: The New Benchmark for Managing Financial Risk, 3rd Edition. McGraw-Hill.
Koenker, R., 2011. Additive models for quantile regression: Model selection and confidence bandaids. Brazilian Journal of Probability and Statistics 25 (3), 239–262.
Koenker, R., 2016. quantreg: Quantile Regression. R package version 5.29.
Koenker, R., Bassett, G., 1978. Regression Quantiles. Econometrica 46 (1), 33–50.
Komunjer, I., 2013. Quantile Prediction. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 2. Elsevier, Ch. 17, pp. 961–994.
Kuester, K., Mittnik, S., Paolella, M., 2006. Value-at-Risk Prediction: A Comparison of Alternative Strategies. Journal of Financial Econometrics 4 (1), 53–89.
Kupiec, P. H., 1995. Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives 3 (2), 73–84.
Li, Y., Zhu, J., 2008. L1-Norm Quantile Regression. Journal of Computational and Graphical Statistics 17 (1), 163–185.
Louzis, D. P., Xanthopoulos-Sisinis, S., Refenes, A. P., 2014. Realized volatility models and alternative Value-at-Risk prediction strategies. Economic Modelling 40, 101–116.
Mallows, C. L., 1973. Some Comments on Cp. Technometrics 15 (4), 661–675.
Marinelli, C., D'addona, S., Rachev, S. T., 2007. A Comparison Of Some Univariate Models For Value-at-risk And Expected Shortfall. International Journal of Theoretical and Applied Finance 10 (06), 1043–1075.
McAleer, M., Jimenez-Martin, J.-A., Teodosio, P.-A., 2013a. GFC-robust risk management strategies under the Basel Accord. International Review of Economics & Finance 27, 97–111.
McAleer, M., Jimenez-Martin, J.-A., Teodosio, P.-A., 2013b. International Evidence on GFC-Robust Forecasts for Risk Management under the Basel Accord. Journal of Forecasting 32 (3), 267–288.
Meinshausen, N., 2006. Quantile regression forests. Journal of Machine Learning Research 7, 983–999.
Nelson, D. B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59 (2), 347–370.
Nieto, M. R., Ruiz, E., 2016. Frontiers in VaR forecasting and backtesting. International Journal of Forecasting 32 (2), 475–501.
R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
RiskMetrics Group, 1996. RiskMetrics – Technical Document. J. P. Morgan and Reuters, New York.
Shan, K., Yang, Y., 2009. Combining Regression Quantile Estimators. Statistica Sinica 19, 1171–1191.
Sheppard, K., 2017. ARCH. Python package version 4.0.
Stock, J. H., Watson, M. W., 2004. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 23 (6), 405–430.
Taylor, S. J., 1986. Modelling Financial Time Series. World Scientific Publishing.
Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58 (1), 267–288.
Timmermann, A., 2006. Forecast Combinations. In: Elliott, G., Granger, C. W., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 1. Elsevier, Ch. 4, pp. 135–196.
Yi, C., 2017. hqreg: Regularization Paths for Lasso or Elastic-Net Penalized Huber Loss Regression and Quantile Regression. R package version 1.4.
Yi, C., Huang, J., 2017. Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss Regression and Quantile Regression. Forthcoming in Journal of Computational and Graphical Statistics. DOI: 10.1080/10618600.2016.1256816.
Zheng, S., 2012. QBoost: Predicting quantiles with boosting for regression and binary classification. Expert Systems with Applications 39 (2), 1687–1697.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society. Series B (Methodological) 67, 301–320.