

Improving the value at risk forecasts: Theory and evidence from the financial crisis [☆]

Roxana Halbleib ^{a,*}, Winfried Pohlmeier ^b

^a Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany

^b University of Konstanz, CoFE, ZEW, RCEA, Germany

ARTICLE INFO

JEL classification:

C21
C5
G01
G17
G28
G32

Keywords:

Value-at-risk
Optimal forecast combination
Quantile regression
Method of moments
Financial crisis

ABSTRACT

The recent financial crisis has raised numerous questions about the accuracy of value at risk (VaR) as a tool to quantify extreme losses. In this paper we develop data driven VaR approaches that are based on the principle of optimal combination and that provide robust and precise VaR forecasts for periods when they are needed most, such as the recent financial crisis. Within a comprehensive comparative study we provide the latest piece of empirical evidence on the performance of a wide range of standard VaR approaches and highlight the overall outperformance of the newly developed methods.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The current worldwide financial crisis has revealed major deficiencies in the existent financial risk measures. For example, based on one day ahead value at risk (VaR) forecasts, which constitutes the focus of interest of the present paper, JPMorgan Chase reported 5, Credit Suisse 7 and UBS 16 exceedances in the 3rd quarter of 2007, which required a maximum of 0.63 exceedances for a probability level of 1% (Jorion, forthcoming). Considerable progress has been made over the last decade to quantify financial risks by means of elaborate econometric tools. However, experience with the performance of these methodologies in rough times was missing until the subprime crisis started to shake the world financial markets. Most of the risk models are based on a set of assumptions which may be more or less carefully validated most of the time, but may contribute to an increase in systemic risk if the model assumptions fail to be true for other (out of sample) time periods.

[☆] This project is supported by the European Social Fund and by the Ministry of Science, Research and the Arts Baden-Württemberg. Financial support from the German Science Foundation, the Belgian National Bank and the IAP P6/07 contract from the IAP programme (Belgian Scientific Policy), 'Economic policy and finance in the global economy' is also acknowledged. Previous versions of the paper circulated under the maiden name of the author, Roxana Chiriac. A previous version of the paper circulated under the title "How risky is the value at risk?".

* Corresponding author. Tel.: +49 7531 885373; fax +49 7531 4450.

E-mail address: Roxana.Halbleib@uni-konstanz.de (R. Halbleib).

In this paper we propose a new methodology of computing VaR's based on the principle of optimal combination that accurately and robustly forecasts losses during periods of high risk. In particular, we propose to compute the optimally combined VaR's by two methods: (1) by assessing the performance of the combined measure over a time window in accordance with the Basel II rules and (2) by assessing the performance of the combined measure sequentially (day by day). The optimal weights of the first method are obtained from minimizing the distance between the population and empirical moments derived from Basel II rules and other VaR evaluation techniques (e.g., independence of hits). The weights of the second method stem from a simple quantile regression on stand alone VaR forecasts. Thus, we aim at developing a new VaR method for forecasting future losses, which profits from the gains of forecast combination (Timmermann, 2005) such as: diversification gains, robustness to structural breaks, robustness to model misspecification risk and correction for bias of individual VaR's, and which consequently remains flexible to various market conditions.

In order to assess the performance of the new developed methods, we consider a comparison type study in the spirit of Kuester et al. (2006), where we first test the forecast performance of a wide range of standard VaR measures before and during the recent financial crisis and then evaluate the performance of their optimal combinations stemming from the methods developed in this paper. More precisely, the first part of the empirical study provides the most recent piece of empirical evidence on the performance of standard VaR measures during periods when they were needed most, such as the previous financial crisis for stock investments of different capitalization. We show that the performance of standard VaR estimates substantially declined during these turbulent financial times. However, their performance varies substantially depending on the type of stock, choice of volatility model, distributional assumption and estimation window. We further show that, while simple models are preferred during calm periods, a more flexible parametrization is necessary to forecast losses during crisis. Moreover, we show that standard VaR estimates generate too many exceedances if the sampling period is too short and that the inclusion of sample information from previous crises is absolutely necessary during turbulent financial times. In the second part of the empirical study we show that the deficient and unstable performance of the stand alone standard VaR forecasts during financial turmoil can be considerably improved simply by optimally combining them according to the new proposed methods, regardless of the stock type, model or distributional choice.

Thus, the contribution of the paper is twofold: (1) we introduce two simple methodologies to accurately and robustly forecast VaR by optimally combining standard approaches according to relevant statistical and economic requirements and, (2) in order to outline the predictive ability of the new method, we provide the most recent piece of comprehensive empirical evidence on the performance of most popular VaR measures during a turbulent financial time, such as the previous financial crisis.

While the idea of combining quantile forecasts is not new, it still has been exploited very little in the existing literature. To the best of our knowledge, there are only a few papers that make reference to this topic. Timmermann (2005) describes the optimal combination of mean forecasts at lengths and solely mentions the optimal combination of conditional quantile forecasts as a further application of the optimal combination principle. Giacomini and Komunjer (2005) use the same optimal combination as our first approach, but with the aim of deriving encompassing tests for comparing conditional quantile forecasts. Different from their approach, we directly apply the optimal combination procedure to improve the accuracy of individual standard VaR estimates, according to the Basel II rules and to the independence principle. Another paper that considers combinations of VaR forecasts is the paper of Jiménez Martín et al. (2011) that uses simple deterministic combinations of individual VaR forecasts, such as mean or median, to minimize the daily capital charge and violation penalties under the Basel II accord. Although very straightforward, these methods do not improve the input stand alone forecasts if each of them is performing poorly (the violations ratio is larger than the VaR probability) because, by construction, the new "combined" forecast will be between the worst and the least worst stand alone forecasts. The relatively good performance of the optimal combination of VaR's introduced in our paper emerges from the fact that we impose no boundaries on the combination weights and allow for an intercept. In regards to our second combination method, to the best of our knowledge, there are so far no previous references in the related literature.

The comprehensive comparative study described in the empirical application contributes to the existing literature in many directions. Most of the existing comparative studies (e.g., Kuester et al., 2006; McAleer and da Veiga, 2008) are applied to data stemming from calm periods or developed markets, popular assets (large cap stock, major indices or currencies) and restrict their analysis to rolling sampling windows. Driven by the suspicion that the VaR could potentially embody a high model risk, we provide here what Summers (1991) calls some "*successful pieces of pragmatic empirical work*" for the latest financial turmoil. Moreover, for the first time in the literature, our study manages to highlight the difference in the performance of standard VaR approaches from calm to crisis periods across stocks of companies with different market capitalization.¹ We also depart from the classic horse races and consider recursive sampling windows with different origins which allow to identify the past shocks that may have important predictive power for the recent turbulence. The parallel study by Schmidt (2011), which uses our data, but increases the basket of model specification with the CAViaR model of Engle and Manganelli (2004) and asymmetric type GARCH's, provides similar empirical results to ours: e.g., the performance of standard VaR measures drastically declines from a calm to a turbulent financial period and the optimal combinations based on the methods proposed in our paper improve their performance during the turbulent times.

The outline of the paper is as follows: Section 2 makes a brief introduction to the VaR measure and presents theoretically the new VaR methods. Section 3 describes the setup and the empirical results of the comprehensive empirical study for

¹ The comparison of how different model specifications and sampling schemes influence the standard VaR performance for other types of financial assets, such as currencies, commodities, bonds, etc. is left to further research.

evaluating the standard and newly developed VaR measures before and during the financial crisis of 2007. Finally, Section 4 summarizes the paper and provides an outlook on future research.

2. Method

2.1. Theoretical settings

The VaR is defined to be the worst possible loss from an investment over a target horizon and for a given probability level (Crouhy et al., 2001). The VaR at time t for a long position over time horizon h with probability p , $VaR_t(p, h)$, is formally defined to be given by:

$$p = \Pr[\Delta V_{t+h}(h) \leq VaR_t(p, h)] = F(VaR_t(p, h)),$$

where $\Delta V_{t+h}(h)$ is the change in the value of a long position V_t from t to $t+h$ and $F(x)$ is the cumulative distribution function (cdf) of $\Delta V_{t+h}(h)$. Let $\Delta V_{t+h}(h)$ be defined as a location scale process conditional on an information set \mathcal{F}_t :

$$\Delta V_{t+h}(h) = E[\Delta V_{t+h}(h) | \mathcal{F}_t] + \varepsilon_{t+h} = \mu_{t+h|t} + \sigma_{t+h|t} z_{t+h}, \quad (2.1)$$

where $\mu_{t+h|t}$ is the expected change in the value V_t from t to $t+h$ given the information setup to t , $\sigma_{t+h|t}$ is the conditional standard deviation in the value V_t at time $t+h$ given the information setup to t . ε_{t+h} is the error term (shock, surprise) and z_t has a zero location, unit scale probability density $f_Z(\cdot)$. Then VaR conditional on the information set \mathcal{F}_t can be derived to be:

$$VaR_{t+h|t}(p) = \mu_{t+h|t} + Q_p(Z) \sigma_{t+h|t}, \quad (2.2)$$

where $Q_p(Z)$ is the p th quantile of z_t .

According to the Basel II rules, financial institutions are required to hold capital that is sufficient to cover losses on the trading book over a 10 day holding period for 99% of the times ($h=10$ and $p=0.01$). Furthermore, the rules of Basel Committee (1996) allow the financial institutions to freely choose the model specification for estimating $\mu_{t+h|t}$, $\sigma_{t+h|t}$ and $f_Z(\cdot)$ as long as the resulting VaR measures have good predictive performance according to some backtesting rules. These rules imply the computation of a hit sequence over S days, defined by:

$$H_{t+h} = \mathbb{1}(\Delta V_{t+h}(h) < \hat{VaR}_{t+h|t}(p)), \quad t = T+1, T+2, \dots, T+S, \quad (2.3)$$

which follows a binomial distribution with parameter p .

The rules imposed by the Basel Committee (1996) for assessing the performance of the internal models are designed only for holding periods of one day ($h=1$) and imply testing the null hypothesis $H_0 : E[H_t] = p$, known in the literature as the *unconditional coverage test* (Christoffersen, 2003). Therefore, in this paper, we focus on $h=1$, leaving the analysis for $h > 1$ for further research. Based on the results of the unconditional coverage test and accounting for the Type I error rate, the Basel Committee (1996) comes with further regulations regarding the failure rate acceptance for a bank before being penalized: banks with failure rates in the 95% quantile of the Bernoulli distribution with parameters $p=0.01$ and $S=250$ are not penalized and are said to be in the “green zone”; banks with failure rates between the 95% and 99% quantiles lie in the “yellow zone” and are progressively penalized; and banks with failure rates outside the 99% quantile are automatically penalized, and are classified as being in the “red zone”.

Furthermore, Christoffersen (1998) develops an approach which tests the degree of “clustering” within the hit sequence as a result of time variation in the data. This approach, known in the literature as the *independence test*, exploits the martingale property of H_{t+h} and has as null hypothesis the statement that the probability of incurring a failure in time $t+1$ is independent of incurring a failure in time t . One can assess the ability of a VaR model to provide the correct *conditional coverage* probability by simultaneously testing the two null hypotheses from the unconditional coverage and independence tests.

2.2. Combining VaR forecasts

In the following we propose two methods of obtaining robust conditional VaR forecasts based on an adapting mechanism of optimally combining conditional VaR forecasts computed from a given set of estimates based on different models and/or samples. For expositional reasons, we restrict our attention to the case of combining two VaR predictions. The generalization to a large set of estimates to be combined is, however, straightforward. Ex ante, the superiority of forecast combinations is not compelling because they involve additional parameter estimation.

Consider S samples of two alternative conditional VaR predictions for the periods $T+1, \dots, T+S$, with $0 < T$ and $0 < S$ and let a combined VaR prediction for the out of sample day $T+s$, with $s=1, \dots, S$, be given by the linear combination of the two one period out of sample VaR predictions based on two different models, i.e., model 1 and model 2:

$$\hat{VaR}_{T+s}(\lambda_T) = \lambda_{T,0} + \lambda_{T,1} \hat{VaR}_{T+s}^1 + \lambda_{T,2} \hat{VaR}_{T+s}^2, \quad s = 1, \dots, S, \quad (2.4)$$

where $\lambda_T = (\lambda_{T,0}, \lambda_{T,1}, \lambda_{T,2})'$ is the vector of loading factors and $\hat{VaR}_{T+s}^j \equiv \hat{VaR}_{T+s|T+s-1}^j(p)$ is the one period ahead VaR prediction from model $j = \{1, 2\}$ with the same probability level p .

For example, let the VaR prediction from model 1 be based on a standard ARMA GARCH approach with normally distributed errors and the VaR prediction from model 2 be based on the pre filtering method using the same conditional mean and

conditional variance function; then, the combined VaR prediction simply uses the linear combination of the quantile of the standard normal, $\hat{F}_1^{-1}(p)$, and the estimated quantile from the distribution of the standardized residuals, $\hat{F}_2^{-1}(p)$:

$$\hat{VaR}_{T+s}(\lambda_T) = \lambda_{T,0} + \{\lambda_{T,1} + \lambda_{T,2}\} \hat{\mu}_{T+s|T+s-1} + \{\lambda_{T,1} \hat{F}_1^{-1}(p) + \lambda_{T,2} \hat{F}_2^{-1}(p)\} \hat{\sigma}_{T+s|T+s-1}. \quad (2.5)$$

Although Eq. (2.4) has a general structure, we opt in this paper for restricting the loadings of VaR^j 's, $j = \{1, 2\}$, to sum up to one. In this way, we have a more intuitive and structured interpretation of the results: e.g., a larger absolute loading on the first forecast than on the other, indicates that the first forecast has a larger impact on the optimal combination, while the sign indicates the direction. However, this restriction does not constrain the flexibility of the model: i.e., contrary to Jiménez Martín et al. (2011), we do not impose any boundary constraints on the weights and, therefore, we allow the combined estimator to take values beyond the values of the stand alone estimators. Thus Eq. (2.4) becomes:

$$\hat{VaR}_{T+s}(\lambda_T) = \lambda_{T,0} + \lambda_{T,1} \hat{VaR}_{T+s}^1 + (1 - \lambda_{T,1}) \hat{VaR}_{T+s}^2, \quad s = 1, \dots, S. \quad (2.6)$$

The role of the constant in the optimization is to correct for the biases in the VaR's forecasts, as recommended by Timmermann (2005). Therefore, we refrain from setting any constraint on this parameter.

We choose to combine the VaR estimates from Eq. (2.6) optimally: (1) by assessing the performance of the resulting risk estimator over the evaluation period starting at $T+1$ and ending at $T+S$ or (2) by sequentially assessing the daily performance of \hat{VaR}_{T+s} at each $s = 1, \dots, S$.² The first method, which we call Conditional Coverage Optimization Method (CCOM), consists in finding the optimal combination of VaR predictions, which maximizes the conditional coverage rate over an evaluation window, while the second method, called Conditional Quantile Optimization Method (CQOM), finds the optimal loadings by minimizing the distance between the population daily quantiles and the VaR's combinations.

2.2.1. Conditional coverage optimization method (CCOM)

The CCOM is optimal in the sense that it chooses the weights of the forecast combination such that the combined forecast implies a sequence of independent hits with moments close to the ones of the hit sequence of the true underlying process and thus satisfies the Basel II criteria. It consists of computing a sequence of S binary exceedance indicators $H_{T+s}(\lambda_T)$ based on the combined estimator $\hat{VaR}_{T+s}(\lambda_T)$ for any arbitrary vector of weighting parameters:

$$\underline{H}(\lambda_T) = \{H_{T+1}(\lambda_T), H_{T+2}(\lambda_T), \dots, H_{T+S}(\lambda_T)\},$$

where $H_{T+s}(\lambda_T)$ is given by

$$H_{T+s}(\lambda_T) = \mathbb{1}(r_{T+s} < \hat{VaR}_{T+s}(\lambda_T)), \quad s = 1, 2, \dots, S$$

and $\hat{VaR}_{T+s}(\lambda_T)$ is given in Eq. (2.6). The optimal weights $\lambda_T^* = (\lambda_{T,0}^*, \lambda_{T,1}^*, 1 - \lambda_{T,1}^*)$ minimize the conditional coverage test, which has the following two null hypotheses:

- (1) $E[H_{T+s}(\lambda_T^*)] = p$, which assures the unconditional coverage and
- (2) $E[H_{T+s}(\lambda_T^*) | H_{T+s-1}(\lambda_T^*)] = E[H_{T+s}(\lambda_T^*)] (1 - H_{T+s-1}(\lambda_T^*)) = E[H_{T+s}(\lambda_T^*)]$, which assures the independence of the hits.

The main advantage of this method is that it exclusively aims at fulfilling the Basel II criteria by minimizing the failure rate and optimizing the forecast performance with respect to backtesting rules. In addition, the optimal combination of VaR's aims at eliminating the violation clustering and reduces the probability of having consecutive violations. To the best of our knowledge, the paper of Giacomini and Komunjer (2005) is the only other study that considers this type of optimal combination, however within a different context. They use this framework in order to derive and implement a test for out-of-sample conditional quantile forecast comparison based on the principle of encompassing. Their empirical results provide favorable evidence for using combinations instead of stand alone forecasts. In contrast to Giacomini and Komunjer (2005), our aim is to use this type of optimal combination in order to improve the poor performance of stand alone conditional VaR forecasts according to the Basel II rules. In Section 3 we show that this approach outperforms a very wide range of stand alone forecasts regardless of their degree of common predictive ability.

Estimates of the two parameters $\lambda_{T,0}^*$ and $\lambda_{T,1}^*$ can be obtained by the Method of Moment (MM) approach using the following moment restrictions resulting from the martingale property of H_t :

- (1) $E[H_{T+s}(\lambda_T^*) - p] = 0$,
- (2) $E[(H_{T+s}(\lambda_T^*) - p)H_{T+s-1}(\lambda_T^*)] = 0$.

We can opt for further moments, such as $E[(H_{T+s}(\lambda_T^*) - p)VaR_{T+s}^j] = 0$, $j=1,2$ or $E[(H_{T+s}(\lambda_T^*) - p)r_{T+s-1}] = 0$ which may increase the efficiency of the estimated parameters (see Giacomini and Komunjer, 2005). Moreover, if we choose to combine more than two conditional VaR forecasts, the number of moments should be strictly larger than the number of

² Other objective functions based on monetary rather than statistical criteria may also be meaningful. For instance, one may combine two forecasts such that the penalty costs are minimized. This is left for further research.

VaR forecasts. The MM estimator $\hat{\lambda}_T^*$ is the solution to the following minimization problem:

$$\hat{\lambda}_T^* = \arg \min \psi(\lambda_T, X_{T+s})' \psi(\lambda_T, X_{T+s}), \quad (2.7)$$

where $\psi(\lambda_T, X_{T+s}) = (\psi^1(\lambda_T, X_{T+s}), \psi^2(\lambda_T, X_{T+s}, \dots))'$ with

$$\psi^1(\lambda_T, X_{T+s}) = \frac{\sum_{s=1}^S [H_{T+s}(\lambda_T) - p]}{S},$$

$$\psi^2(\lambda_T, X_{T+s}) = \frac{\sum_{s=1}^S [H_{T+s}(\lambda_T) - p] H_{T+s-1}(\lambda_T)}{S - 1},$$

and $X_{T+s} = (\hat{V}aR_{T+s}^1, \hat{V}aR_{T+s}^2, r_{T+s})'$. In the case of further moments, the estimation given in Eq. (2.7) is weighted by \hat{V}^{-1} , where \hat{V} is a consistent estimator of the asymptotic variance $V = E[\psi(\lambda_T, X_{T+s})\psi(\lambda_T, X_{T+s})']$.

Although straightforward in principle, the estimation suffers from a set of numerical complications. Since the objective function is non differentiable and not sufficiently smooth, gradient based methods may encounter difficulties in finding the global, rather than the local minimum (Giacomini and Komunjer, 2005). In such cases, we use the simulated annealing algorithm, which is more robust to the choice of the objective function and reaches the global minimum, regardless of the degree of smoothness of the objective function (Goffe et al., 1994). Furthermore, we increase the degree of smoothness of the function to be minimized in Eq. (2.7), by replacing $H_{T+s}(\lambda_T)$ with $F(\lambda_T, X_{T+s}, h_T) = 1/(1 + \exp[(r_{T+s} - \lambda_{T,0} - \lambda_{T,1} \hat{V}aR_{T+s}^1 - (1 - \lambda_{T,1})\hat{V}aR_{T+s}^2)/h_T])$ such that, for $s \rightarrow \infty$, $h_T \rightarrow 0$ and $sh_T \rightarrow \infty$, $\sum_{s=1}^S F(\lambda_T, X_{T+s}, h_T)/S \xrightarrow{p} E[H_{T+s}(\lambda_T)]$. The heteroscedasticity and autocorrelation robust estimator of Newey and West (1987) for the asymptotic variance covariance matrix V should be applied.

2.2.2. Conditional quantile optimization method (CQOM)

The second method we propose for optimally combining VaR's estimators is based on the quantile regression approach (Koenker and Basset, 1978) and models the conditional p quantile as a linear function of $\hat{V}aR_{T+s}^j$, with $j = \{1, 2\}$:

$$Q_p(r_{T+s}) = \lambda_{T,0} + \lambda_{T,1} \hat{V}aR_{T+s}^1 + (1 - \lambda_{T,1}) \hat{V}aR_{T+s}^2 \equiv \hat{V}aR'_{12,T+s} \lambda_T, \quad s = 1, 2, \dots, S, \quad (2.8)$$

where $\hat{V}aR'_{12,T+s} = (1, \hat{V}aR_{T+s}^1, \hat{V}aR_{T+s}^2)$ is of dimension 3×1 and $\lambda_T = (\lambda_{T,0}, \lambda_{T,1}, 1 - \lambda_{T,1})$ is also of dimension 3×1 . In our framework, the vector of optimal weights λ_T is given by the solution to the following minimization problem:

$$\hat{\lambda}_T = \underset{\lambda_T}{\operatorname{argmin}} \left\{ \sum_{r_{T+s} \geq \hat{V}aR'_{12,T+s} \lambda_T} p |r_{T+s} - \hat{V}aR'_{12,T+s} \lambda_T| + \sum_{r_{T+s} < \hat{V}aR'_{12,T+s} \lambda_T} (1 - p) |r_{T+s} - \hat{V}aR'_{12,T+s} \lambda_T| \right\}$$

The main advantage of the quantile regression approach is that it requires no explicit distributional assumptions for the return series. This method allows for a flexible framework of combining VaR forecasts such that the distance between the population quantiles and the VaR combination is minimized. Although very straightforward to implement, there is no previous reference to this type of combination method to be found in the related literature. In contrast to the CCOM method, the construction of the CQOM does not directly aim at fulfilling some Basel II or independence criteria. Therefore, its predictive ability with respect to these rules is ultimately an empirical question. The empirical results from applying this method are reported in Section 3.

3. Empirical results

In this section we present results from evaluating stand alone VaR forecasts (Section 3.2) and optimal combinations (Section 3.3) following the theoretical settings presented in Section 2. The VaR evaluation is done by means of backtesting before and during the financial crisis from 2007.

3.1. Setup of the comparative study

This comparative study is designed to assess the quality of standard and new VaR estimates at different dimensions: asset choice, model choice, distributional assumption and estimation window, before and during the recent financial crisis. Given that the evaluation period implies a series of negative extreme events, we find it proper to focus on assessing the performance of the VaR estimates at $p=0.01$, which is in line with the Basel II requirements. Regarding the asset choice, we consider three equally weighted indices built on 30 randomly chosen stocks from the Dow Jones U.S. Small, Middle and respectively Large cap indices.³ In this way, we aim at verifying the stability of standard VaR methods with respect to the degree of capitalization of the underlying asset before and during the recent financial crisis and thus complementing the results of previous evaluation research (e.g., Kuester et al., 2006; McAleer and da Veiga, 2008) done only on popular large cap stocks or major indices. The data consists of daily log returns computed from closing dividend and split adjusted prices provided by Thomson Reuters Datastream.

³ The small cap index contain the stocks with the following symbols: AGL, AIR, AMR, ASH, BDN, BEZ, BIG, BIO, BRE, BXS, CBRL, CBT, COO, CTX, CW, DLX, ESL, GAS, HXL, ITC, LIZ, LPX, MDP, NEU, PBY, PCH, PPD, RLI, TXI, UNS; the mid cap index contains: ACV, ADSK, AMD, BCR, BDK, BMS, CBE, CCK, CEG, CSC, DBD, DOV, DTE, EK, DPL, GMGMQ, GR, GVVW, HOT, MAS, MDC, MWV, NAV, NI, ROST, RSH, SWK, UNM, VFC, WEC; and the large cap index contains: ABT, ADBE, AMAT, APC, APD, AVP, BAC, BEN, BK, CA, CL, D, DD, EMR, FPL, ITW, JCI, JPM, LOW, MMM, MRK, OXY, PCAR, PEP, PFE, SO, SY, TGT, WFT, XOM.

For our evaluation purposes, we divide the out of sample period in three subperiods around the beginning of the sub prime crisis: (1) the period before the crisis starting on January 1st up to July 17th, 2007. July 17th, 2007 represents in our study the beginning of the current financial crisis: on this day the FED signaled the first trouble on the subprime loan markets and announced their support and supervision to the subprime mortgage lenders. We consider the first evaluation period to be a calm one, representative for the years prior to the financial crisis; (2) the crisis period starting on July 18th, 2007 up to July 1st, 2009; and (3) the crash period, which starts on September 1st, 2008, which marks the first day on the month in which the crisis hits the financial markets following the bankruptcy of Lehman Brothers on September 15th, 2008. Starting on this day, major stock markets all over the world experienced the largest decline since the Black Monday crash on October 19th, 1987. The crash period ends on July 1st, 2009. By dividing the whole evaluation period into these three subperiods, we intend to check the robustness of VaR estimates across different financial settings and outline which methods perform “best” during calm periods as opposed to turbulent ones.

Regarding the model choice, we focus mainly on fully parametric location scale models for $\mu_{t+1|t}$ and $\sigma_{t+1|t}$, which account for the empirical properties of financial returns: volatility clustering, serial dependence, fat tailed distributions, etc. and parametric distributional assumptions, such as normal (ND), Student t (SD) and skewed Student t (skewed SD) distribution or semi parametric approaches, such as, extreme value theory (EVT) for the standardized residuals. We consider the following model specifications: for the conditional mean we estimate an ARMA(1,0) model with intercept and for the conditional variance we estimate the GARCH(1,1) model, the RiskMetrics model introduced by JP Morgan in 1995 with estimated (RM est) and fixed (RM fix) parameters and the FIGARCH (1,d,0) of Baillie et al., 1996. Within a parallel study to ours, Schmidt (2011) also estimates using the same data set a GARCH type model that accounts for the leverage effects, such as EGARCH(1,1). His empirical results show that the VaR forecasts computed on EGARCH(1,1) and GARCH(1,1) conditional variance forecasts have a similar performance during the evaluation periods considered. As already mentioned, in order to compute the quantile, we consider the normal, Student t and skewed Student t distributions for the standardized residuals stemming from the above mentioned conditional models.⁴ Finally, we also consider the EVT for the standardized residuals stemming from the ARMA(1,0) GARCH(1,1) specification.

Furthermore, given the nature of the VaR measure, we also consider VaR forecasts stemming from quantile regressions. More specifically, we consider the approach introduced by Coroneo and Veredas (2011) (QR), who compute VaR forecasts from quantile regressions on the absolute values of the previous day return. The parallel study by Schmidt (2011) opts for the CAViaR model of Engle and Manganelli (2004). His empirical results indicate a good performance of the approach during the calm financial period, but a relatively poor performance during the crisis. As described below, this type of behavior is similar for almost all standard approaches considered in our paper.

Subject to the model choice, we aim at assessing the model risk component in estimating financial risks based on standard approaches. By choosing different degrees of parametrization and tail distributional assumptions, we aim at outlining the approaches, which are most appropriate for forecasting the risk of a certain type of stock in calm versus turbulent periods. This analysis will allow us to identify the trade off between the estimation risk and model misspecification risk involved in estimating VaR: while the RM fix approach involves no estimation risk, its simple and fixed structure might face difficulties in correctly capturing the dynamics of conditional volatilities (e.g., long memory) and remains inflexible to recent informational content, which could be relevant for forecasting future risks. Contrary to the RM fix approach, the ARMA FIGARCH model easily adapts to the new information arrivals, but might significantly suffer from estimation risk, given that the precision of the degrees of fractional integration estimator increases with the window size of the underlying data.

As an alternative to the fully parametric models described above, we consider one nonparametric approach of estimating VaR, namely the Historical Simulation (HS), which is very popular among practitioners and consists in estimating VaR simply by the sample quantile of a rolling window of historical data. In the HS estimation, the size of the sampling window plays an important role for the accuracy of VaR estimates and is mostly set to be between 250 and 750 observations (Jorion, 2007). Although very easy to implement, this approach ignores the conditional dependencies among returns as well as the relevant information on extreme past events that situate outside the sampling window. Subject to these drawbacks, we also implement the filtered HS method (FHS) proposed by Barone Adesi et al. (1999), which applies the HS approach on residuals standardized by the parametric ARMA GARCH method. Furthermore, we implement the HS method on rolling windows of 250, 500, 750 and 1000 historical observations, but also on recursive windows starting in 1987, 1996, 2001 and 2005, respectively. The FHS estimation is applied only on recursive windows.

Another important dimension we focus on in our study is the size of the estimation window. Contrary to other studies, which apply a rolling window in the forecasting procedure, we use the recursive sampling window approach, which is able to preserve the valuable information issued by past extreme shocks, regardless of how far we move forward with the forecasting window. Through this strategy, we aim at identifying the extreme events from the past, which contain valuable information in order to better forecast extreme losses in the recent financial crisis. For this purpose, we consider four strategies of choosing the starting date of the recursive windows: January 1st, 1987, for which all in sample data include the Black Monday effect from October 19th, 1987; January 1st, 1996, for which the in sample windows include the dot com bubble crash from March until October, 2000 and the terrorist events from September 11th, 2001; January 1st, 2001, which includes just the September 11th,

⁴ The fact that the RM-fix model is not linked with any skewed distribution, and consequently no fix skewness parameter is recommended by JP Morgan, we apply the skewed Student-t distribution only to GARCH, RM-est and FIGARCH standardized residuals.

2001 event; and January 1st, 2005, for which the in sample windows contain no exceptional extreme event and represent the standard estimation windows (last two years of data) used in practice before the beginning of the recent financial crisis.

The dimensions of the model implementation described so far lead to a total of 192 forecasts given by three indexes (small, middle, large cap), two parametric model specifications (RM est, ARMA FIGARCH) with three distributional assumptions (normal, Student t and skewed Student t), the RM fix model with normal and Student t distribution, the ARMA GARCH model with four distributional assumptions (normal, Student t , skewed Student t and EVT), the quantile regression approach of Coroneo and Veredas (2011), four strategies of choosing the origin of the recursive windows (starting in 1987, 1996, 2001 and 2005, respectively) and two nonparametric models based on rolling (HS) and recursive (HS, FHS) sampling windows of sizes 250, 500, 750 and 1000 and, respectively, starting at 1987, 1996, 2001 and 2005. Furthermore, we apply the unconditional and conditional coverage tests described in Section 2 to evaluate the VaR estimates on horizons of one day.

3.2. Empirical results from evaluating stand alone VaR forecasts

In this subsection we present the results from evaluating the performance of the stand alone VaR estimators described in Section 3.1. The percentage rate of violations as well as the results related to unconditional and conditional coverage tests for the periods considered are reported in Tables A1–A3 in Appendix B. Table A1 reports the VaR failure rates prior to the recent sub prime crisis, from January 1st up to July 17th, 2007. All in all, the standard VaR approaches perform well during this calm period. While the normal distribution produces most of the “yellow zone” type violations, applying a fat tailed distribution to financial returns such as Student t or skewed Student t improves the overall unconditional performance of VaR. In support of the results from previous horse races (e.g., Kuester et al., 2006), our study provides evidence on the frailty of the RiskMetrics estimated on normal distributed losses. However, its performance significantly improves when applied together with fat tailed distributions. Nevertheless, during this calm period, simple approaches, such as the quantile regression approach of Coroneo and Veredas (2011) generally suffices to capture the data dynamics and produces independent sequences of violations.

Contrary to the calm period, standard VAR estimators generally perform poorly during the recent financial crisis (Table A2). Estimators using the normality assumption exhibit the worst performance and yield only time dependent “red zone” type violations. The poor performance of the normal distribution is explained by its inadequacy in consistently estimating the tail behavior of returns facing numerous extreme negative shocks during periods of crisis and crashes (e.g., September 2008). A slightly better performance than the normal distribution exhibits the Student t distribution, which, in the absence of estimation noise (fixed degrees of freedom), delivers a smaller, but still disputable number of violations in the “yellow zone” for small and large cap stocks. However, the skewed Student t distribution proves to perform very well during the crisis period across all types of conditional variance models. This result is in line with the findings of Kuester et al. (2006) and Giot and Laurent (2004) and is due to the fact that this type of distribution accommodates both tail thickness and asymmetry, which are typical features of the financial return distributions.

During the crisis period, the EVT estimated on historical samples which incorporate information from previous crisis, also has a very good performance and results in all failure rates being independent and in the “green zone”. Although not new (e.g., Kuester et al., 2006), these results highlight the fact that by implementing EVT or skewed Student t distribution on large and informative historical samples, one can increase the predictive ability of standard VaR measures for stock investments during turbulent financial times.

Further evidence that the sampling window plays an important role in correctly forecasting losses during crisis times is provided by the QR approach, that gives yellow and even green type violations for small and middle cap stocks, respectively, when estimated on samples incorporating the financial crisis from 1987.

Aside from the estimation window, the degree of parametrization plays a further important role: the ARMA FIGARCH approach in the presence of fat tailed distributions outperforms (for the Student t) or performs comparatively well (for the skewed Student t) as the other models for all types of stocks.

Table A3 reports the backtest results for the third period, which we call “crash period”. Although in general better, the results from the crash period emphasize the findings from the crisis time: regardless of the asset type, higher parametrization increases the performance of most of the VaR estimates, including the ones built on the normal distribution.

The Student t distribution plays an important role in enhancing the performance of VaR measures, especially of the ones which involve little estimation noise, such as the fixed parameter RiskMetrics approach. However, similar to the crisis period, this popular method manages to correctly forecast the losses only for small and large cap stocks. Nevertheless, the skewed Student t distribution in general and the EVT approach estimated on large historical samples clearly outperform all other methods by generating “green zone” independent violations for all types of stocks.

By comparing the results of Tables A2 and A3, we notice that during the crash time there are more VaR estimates, which perform well according to the Basel II rules compared to the overall crisis period. Contrary to the estimates from the beginning of the crisis, whose performance could only be enhanced by accounting for very old shocks (e.g., Black Monday), the sampling windows used to estimate the risk models during the crash period already incorporate sufficient relevant information on extreme events from the crisis period. This result confirms the fact that including information from previous crisis enhances the performance of VaR estimates, especially of the ones based on distribution assumptions, which properly exploit this type of information, such as (skewed) Student t or EVT.

Tables A4–A6 in the Appendix report the results from backtesting nonparametric VaR estimates during the three periods. In the calm period, the HS method always generates independent “green zone” type violations, regardless of the sampling strategy.

This result justifies, once again, the popularity of this method among practitioners. Prior filtering of data brings no additional gains (last column of Table A4) to the VaR estimates. However, the standard HS methods fail entirely to forecast future losses and to capture their characteristic time dependence during turbulent periods. In this context, their performance is substantially improved by the ARMA GARCH filtration especially during the crash time, when regardless of the asset type, the FHS method applied on short windows generate “green zone” independent hits. This confirms once again the good performance of highly parameterized models during turbulent periods. However, the nonparametric methods built on short window samples have, by construction, a good performance during the crash time. Because they are sample quantiles of historical windows, the nonparametric VaR estimates are better when estimated on shorter windows, which incorporate mainly information from the past year of crisis, than on longer horizons, which contain also information from the calm period prior to the crisis.

Aside from the contribution to the performance of EVT and Student t distribution during the crisis, further evidence on the importance of using large samples when estimating the tail of return distribution is given by the results from estimating the degrees of freedom (d.f.) of the Student t distribution throughout the whole evaluation period. Based on our empirical results,⁵ the d.f. estimated on samples from 1987 or 1996 exhibit a stable behavior prior to and during the crisis. Their values vary around six, which is typical for a fat tailed distribution. Estimated on recent data, the d.f.’s seem to be very volatile and, thus, more flexible to the market conditions.

The trade off between using large samples to estimate stable parameters and using recent samples to estimate parameters which easily adapt to the market conditions translates into the following: financial institutions using stable d.f. specific to fat tailed distributions for stock investments follow strategies of holding constant and large reserves over a long period. By contrast, others using distributions with d.f. estimated on recent samples follow strategies of continuously adapting their reserve requirements according to the market settings mirrored in the estimated parameters. However, in the latter case, the risk is that if everybody follows the same strategy on the days when the VaR measures signal the need for larger reserves, the banks face huge liquidity problems.

Although this might seem to be the case of what happened during the recent financial crisis, we have to point out that our analysis and conclusions are confined to stock portfolio investments. Obviously, one severe problem during the crisis of 2007 was the lack of empirical experience with portfolios consisting of new financial products, for which, by definition, the number of observations is limited. Although a systemic analysis similar to the one presented above for such portfolios or portfolios of other financial assets, such as currencies, commodities, bonds is considered as very interesting and necessary, it is nevertheless beyond the aim of this paper and is left for future research.

3.3. Empirical results from evaluating optimal combinations of VaR forecasts

In the following we present the empirical performance on correctly estimating the failure probability of 1% for the combination methods described in Section 2. In particular, we focus on analyzing the relative performance of the combination approach compared to the stand alone forecasts presented above. Moreover, in order to emphasize the performance of the new methods, we concentrate on reporting results from combining mainly stand alone forecasts that deliver a poor prediction performance following the description from Section 3.2.

We make the assessment at two different stages:

- (1) Firstly, we estimate one pair of optimal weights for each of the three evaluation periods and assess the performance of the ex post optimal VaR combination by means of the coverage tests described in Section 2. This assessment is similar to an in sample evaluation of the fitted models with $S=510$ for the crisis period from July 18th, 2007 until July 1st, 2009 and with $S=217$ for the crash evaluation period from September 1st, 2008 until July 1st, 2009.
- (2) Secondly, we evaluate the optimal combination of VaR’s in a recursive manner: we divide the entire evaluation period of $S=653$ observations (from January 1st, 2007 to July 1st, 2009) into an in sample and an out of sample period and re estimate the optimal weights at each out of sample point with all in sample available data. At the end of the out of sample period, we assess the quality of the optimal forecasts by means of coverage tests. The first in sample period contains the first $T+S'$ observations, the second in sample period contains the first $T+S'+1$ and the last one contains $T+S-1$ observations, where $1 < S' < S$.

Through these two evaluation exercises we aim at assessing the ability of optimal combinations to improve the performance within sample (first assessment), but also out of sample (second assessment) of stand alone VaR estimates. For both stages, we report results from the crisis and crash period, given that during the calm period almost all models, with very few exceptions, perform well according to standard criteria (see Table A1). Moreover, we only present results here from combining parametric VaR estimates, while the assessment of optimal combinations based on individual nonparametric measures is left for further research. Based on the choice of the combined VaR predictors, we sort the results in five groups and report them in Tables 1 and 2 for each of the optimization method.

The first group of results (Part A of both Tables 1 and 2) presents the performance of optimally combining VaR estimates based on different distributional assumptions, given a certain pre filtering method. It illustrates the forecasting

⁵ The results from estimating the models implemented in this paper are available from the authors upon request.

power of a linear quantile combination for the same conditional mean and conditional variance specification (see e.g., Eq. (2.5)). More specifically, we combine VaR predictors based on the normal and Student t distribution, and we report the results for different pre filtering methods: from ARMA GARCH estimated on the sample starting in 1987 to ARMA FIGARCH estimated on the sample starting in 2005. The results from both tables show that, independent of the pre filtering and combination method, optimal combinations of normal and Student t distribution quantiles significantly reduce the failure rates and their time dependencies with very small differences between the two methods. While the CQOM method (Table 2) delivers only “green zone” type violations with all p values of conditional coverage tests larger than 0.1, the CCOM method (Table 1)

Table 1

Percentage rate of violations based on CCOM, in-sample assessment. Percentage rate of violations at $p = 1\%$. ** refers to p -values of conditional coverage test smaller than 0.05, * refers to p -values between 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold entries mark “red zone”-type violations, italic ones mark “yellow zone”-type violations and no typeface entries mark “green zone”-type violations.

Comb\Stock type	Crisis period			Crash period		
	Small	Medium	Large	Small	Medium	Large
Part A (Normal and Student-t distribution)						
ARMA-GARCH						
1987	0.98	1.56	0.98	0.92	0.92	0.92
1996	0.98	0.98	0.98	0.92	0.92	0.92
2001	2.54*	0.98	0.98	0.92	0.92	0.92
2005	1.76	1.76	0.98	0.92	0.92	0.92
RM-est						
1987	0.98	0.98	0.98	0.92	0.92	0.92
1996	0.98	0.98	0.98	0.92	0.92	0.92
2001	0.98	0.98	0.98	0.92	0.92	0.92
2005	0.98	0.98	2.15*	0.92	0.92	1.38
RM-fix						
1987	0.98	0.98	0.98	0.92	0.92	0.92
1996	0.98	0.98	0.98	0.92	0.92	0.92
2001	0.98	0.98	0.98	0.92	0.92	0.92
2005	0.98	0.98	0.98	0.92	0.92	0.92
ARMA-FIGARCH						
1987	0.98	0.98	0.98	0.92	0.92	0.92
1996	7.25**	0.98	0.98	0.92	0.92	0.92
2001	0.98	0.98	6.47**	0.92	0.92	0.92
2005	2.94**	1.56	1.17	0.92	4.14**	0.92
Part B (ARMA-GARCH and RM-fix)						
ND						
1987	1.37	0.98	1.37	0.92	0.92	0.92
1996	1.37	0.98	0.98	0.92	0.92	0.92
2001	0.98	1.17	0.98	0.92	4.14**	0.92
2005	1.17	0.98	0.98	0.92	0.92	0.92
TD						
1987	1.56	0.98	0.98	0.92	0.92	0.92
1996	1.17	1.56	0.98	0.92	0.92	0.92
2001	0.98	0.98	1.17	0.92	0.92	0.92
2005	6.07**	0.98	0.98	0.92	0.92	0.92
Part C (ARMA-FIGARCH and RM-fix)						
ND						
1987	0.78	0.98	0.98	0.92	0.92	0.92
1996	0.98	0.78	0.98	0.92	1.84	5.99**
2001	0.98	0.98	1.17	2.30	0.92	0.92
2005	1.56	0.98	0.98	0.92	0.92	0.92
TD						
1987	0.98	0.98	0.98	0.92	0.92	0.92
1996	0.98	1.56	1.56	0.92	0.92	0.92
2001	0.98	0.98	0.98	0.92	0.92	0.92
2005	1.96	6.47**	0.98	0.92	0.92	0.92
Part D (samples starting in 1987 and 2005)						
ND						
ARMA-GARCH	0.98	1.76	0.98	0.92	0.92	0.92
RM-est	1.17	1.96	0.98	0.92	0.92	0.92
ARMA-FIGARCH	1.17	2.94**	1.96	0.92	0.92	0.92
TD						
ARMA-GARCH	0.98	1.17	0.98	0.92	0.92	0.92
RMEST	2.74**	0.98	0.98	0.92	0.92	0.92
ARMA-FIGARCH	0.98	0.98	1.56	0.92	0.92	0.92

Table 2

Percentage rate of violations based on CQOM, in-sample assessment. Percentage rate of violations at $p = 1\%$. ** refers to p -values of conditional coverage test smaller than 0.05, * refers to p -values between 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold entries mark "red zone"-type violations, italic ones mark "yellow zone"-type violations and no typeface entries mark "green zone"-type violations.

Comb\Stock type	Crisis period			Crash period		
	Small	Medium	Large	Small	Medium	Large
Part A (Normal and Student-t distribution)						
ARMA-GARCH						
1987	0.98	1.17	1.17	0.92	1.38	1.38
1996	1.17	1.17	1.17	0.92	1.84	1.38
2001	1.17	0.98	1.17	1.38	1.38	1.38
2005	1.17	1.17	1.17	1.38	1.38	1.38
RM-est						
1987	1.17	1.17	0.98	1.38	1.84	1.38
1996	0.78	1.17	1.17	1.38	1.38	1.38
2001	0.98	0.98	1.17	1.84	0.46	1.38
2005	1.17	1.17	0.98	1.38	1.38	1.38
RM-fix						
1987	0.98	1.17	0.98	0.46	1.84	1.38
1996	0.98	1.17	0.98	0.46	1.84	1.38
2001	0.98	1.17	0.98	0.46	1.84	1.38
2005	0.98	1.17	0.98	0.46	1.84	1.38
ARMA-FIGARCH						
1987	1.17	1.17	1.17	1.38	1.38	1.38
1996	1.17	1.17	0.98	1.84	0.92	0.92
2001	0.98	1.17	1.17	1.38	1.38	1.38
2005	1.17	0.98	0.98	1.38	0.46	1.38
Part B (ARMA-GARCH and RM-fix)						
ND						
1987	0.98	0.98	1.17	0.92	1.38	1.84
1996	1.17	0.98	0.98	0.92	1.38	1.38
2001	1.17	0.98	1.17	1.38	1.38	1.38
2005	1.17	1.17	1.17	1.38	1.38	1.38
TD						
1987	0.98	0.98	1.17	0.92	1.38	1.38
1996	1.17	1.17	1.17	0.92	1.38	1.38
2001	1.17	1.17	0.98	1.38	1.38	1.38
2005	1.17	1.17	0.98	1.38	1.38	1.38
Part C (ARMA-FIGARCH and RM-fix)						
ND						
1987	1.17	1.17	1.17	0.92	1.38	1.38
1996	1.17	0.78	1.17	0.92	1.38	1.38
2001	0.98	1.17	0.98	1.38	1.38	1.38
2005	1.17	1.17	1.17	0.92	1.38	1.38
T						
1987	1.17	1.17	1.17	1.84	1.38	1.38
1996	1.17	1.17	1.17	0.92	0.92	1.38
2001	1.17	1.17	1.17	1.38	1.38	1.38
2005	0.98	1.17	1.17	0.92	1.38	1.38
Part D (samples starting in 1987 and 2005)						
ND						
ARMA-GARCH	1.37	1.17	1.17	0.92	1.38	1.38
RM-est	0.98	1.17	0.98	0.92	1.38	1.38
ARMA-FIGARCH	1.17	1.17	1.17	1.38	1.84	0.92
TD						
ARMA-GARCH	1.17	1.17	0.98	0.92	0.92	1.38
RM-est	1.17	0.98	0.98	1.38	1.38	1.38
ARMA-FIGARCH	1.17	1.17	1.17	1.38	1.84	1.38

reveals a few exceptions, especially when combining VaR estimates based on recent data. However, both methods deliver a clear improvement in the performance of the combination components (see Tables A2 and A3).

Similar good results are obtained when combining two VaR estimates with different conditional mean and conditional variance estimators: Part B and C of both Tables 1 and 2 report the results from optimally combining VaR predictors based on the fixed parameter RiskMetrics approach, which is widely used in practice, parsimonious and exhibits no estimation noise and two other approaches, which are more flexible to the volatility properties, but involve large estimation risk: ARMA GARCH (Part B) and ARMA FIGARCH (Part C). For example, Table A2 reports the failure rate of the estimators built on ARMA GARCH and RM fix

Table 3

Percentage rate of violations, out-of-sample assessment. Percentage rate of violations at $p = 1\%$. ** refers to p -values of conditional coverage test smaller than 0.05, * refers to p -values between 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold entries mark "red zone"-type violations, italic ones mark "yellow zone"-type violations and no typeface entries mark "green zone"-type violations.

Comb Stock type	CCOM			CQOM		
	Small	Medium	Large	Small	Medium	Large
Part A (Normal and Student-t distribution)						
ARMA-GARCH						
1987	1.60	1.20	0.80	0.40	1.60	1.20
1996	1.20	1.20	2.00	0.40	1.60	2.00
2001	1.60	0.80	2.40	1.20	1.60	2.00
2005	0.40	2.00	1.60	2.80*	1.20	1.20
RM-est						
1987	1.60	1.60	0.80	1.20	2.80*	1.20
1996	1.20	1.20	0.80	1.60	1.20	1.60
2001	1.60	1.20	0.40	2.00	1.20	1.60
2005	1.20	1.20	1.20	2.40	1.20	1.60
RM-fix						
1987	1.20	1.20	0.40	2.80*	1.20	1.60
1996	1.20	1.20	0.40	2.80*	1.20	1.60
2001	1.20	1.20	0.40	2.80*	1.20	1.60
2005	1.20	1.20	0.40	2.80*	1.20	1.60
ARMA-FIGARCH						
1987	1.60	0.40	1.60	1.20	0.80	1.20
1996	2.00	1.20	1.20	2.00	2.00	1.60
2001	1.20	0.80	0.40	0.40	1.20	2.00
2005	0.40	1.20	2.00*	2.40	1.20	1.20
Part B (ARMA-GARCH and RM-fix)						
ND						
1987	1.60	1.20	1.20	1.20	1.60	1.20
1996	2.00	1.60	1.60	1.60	1.60	1.60
2001	2.00	1.60	1.60	2.00	1.60	1.60
2005	1.20	1.20	1.60	1.20	1.60	2.00
TD						
1987	0.80	2.00	1.60	1.20	1.60	1.60
1987	1.60	1.20	2.00	1.20	1.60	1.60
2001	1.60	1.20	2.40	1.60	1.60	2.00
2005	2.40	0.80	2.40	0.80	1.60	1.20
Part C (ARMA-FIGARCH and RM-fix)						
ND						
1987	1.60	2.00	1.20	1.60	1.60	1.60
1987	1.20	1.60	1.20	1.60	1.20	2.00
2001	1.20	2.00	1.20	0.80	1.20	1.20
2005	2.00	1.20	1.20	0.40	0.80	1.60
1987	0.00	0.80	1.20	0.80	0.80	1.20
1987	0.80	0.80	2.00*	0.80	0.80	0.80
2001	2.00	0.80	1.20	0.80	1.20	0.80
2005	1.20	0.40	1.20	0.80	1.60	1.60
Part D (samples starting in 1987 and 2005)						
ND						
ARMA-GARCH	1.60	1.60	0.40	1.20	0.80	0.80
RM-est	2.00	2.00	2.40	2.00	1.20	1.60
ARMA-FIGARCH	2.40	1.20	1.20	0.40	1.20	2.40
TD						
ARMA-GARCH	1.60	0.80	0.00	0.80	0.80	1.20
RM-est	1.60	1.20	2.00	2.00	2.00	0.80
ARMA-FIGARCH	2.00	1.20	2.00	0.40	1.20	2.00

given the normal distribution that ranges between 2.15 to 4.51. However, after combining the two approaches, the failure rate drastically decreases and ranges from 0.98 to 1.56 (see the first four rows and first six columns of Part B of Tables 1 and 2).

Again, combining VaR estimates based on recent data seems to be inadequate in forecasting future losses. However, combining estimates based on recent and older data (Part D), increases the performance of individual estimates in all cases when applying the CQOM method (Table 2) and, with a few exceptions, when applying the CCOM method (Table 1). Aside from improving the performance of the input stand alone forecasts, the optimal combination also performs equally well or better than the estimators not included in the combination. Thus, for example, combining estimates with different sampling windows improves the performance of the normal distribution during the crash period (see the first four rows and first four columns of

Part D of Tables 1 and 2) by delivering failure rates smaller than 1.4 and generally outperforms or performs equally well compared to the skewed Student t distribution or EVT (see Table A3) with the largest failure rate of about 3.22.

So far, we can conclude that although both combination methods perform well in sample, the CQOM approach yields the best results. However, of higher interest is assessing the ability of these methods to provide combinations of VaR which perform well during out of sample periods, especially during the recent financial crisis.

Table 3 reports the backtest results from the second stage assessment. Since our focus is on forecasting the risks during the crisis and crash period, we choose July 17th, 2008 to be the end date of the first in sample period, which marks one year since the beginning of the financial crisis ($S' = 250$). Thus, all in sample data entail at least one year of crisis and the out of sample period comprises the crash phase.

The results given in Table 3 support the previous results and show that combining VaR estimates performs well, not just within the sample, but also out of sample. Compared to the stand alone estimates, the combinations deliver very good results, which remain robust with respect to the choice of asset, model and evaluation period. In most of the cases, they produce independent failure rates, which are located in the “green zone”, according to the Basel II regulations.

Finally, the results from the tables above reveal an overall stability of the backtesting performance of the combination methods according to the Basel II and independence criteria. Further evidence on the robustness of the new methods with respect to the asset choice or model specification provides the sequence of combination weight estimates for the out of sample evaluation period. For illustration purposes, we plot the sequence of estimated optimal weights from combining normal and Student t distribution with ARMA GARCH pre filtration, subject to different sample windows.⁶

Except for a short inherently volatile phase, beginning around the events from September 15th, 2008 (45th observation) and lasting until December 1st, 2008, all sequences of estimates stemming from CQOM (see Figs. B1 and B2) exhibit stable behavior during the whole crisis period, which emphasizes the robustness of the new method. The CCOM weights exhibit a more volatile behavior (see Figs. B3 and B4), which indicates that estimators stemming from unconditional type methods are less robust to market changes than the ones stemming from a conditional computation, such as CQOM.

4. Conclusions

In this paper we propose data driven methodologies of accurately forecasting VaR based on the principle of forecast combination. The optimal loadings of VaR measures are driven by the maximization of conditional coverage rates or by the minimization of the distance between the population quantiles and VaR's combinations. Allowing for data driven combination weights that are optimally estimated with respect to statistic or economic requirements, the new methods feature an increased degree of flexibility in adapting to various market conditions and simultaneously delivering accurate VaR forecasts.

The empirical performance of the new methods is depicted within a comprehensive comparative study that includes a large range of standard VaR approaches evaluated before and during the recent financial crisis on stock investments. We show that popular VaR methods perform very differently from calm to crisis periods, subject to stock type, model or estimation window. While in calm periods VaR estimates based on normal distributional assumptions and parsimonious models estimated on recent windows of data are appropriate to forecast the potential losses, this is not the case for turbulent times. In these cases, fat tailed distributions estimated on larger samples perform well for all types of stocks. Finally, large sampling windows, incorporating valuable information on past shocks (e.g., market crash from 1987) and properly exploited by suitable distributional assumptions (Student t with estimated degrees of freedom, skewed Student t and EVT), deliver better quantile estimates and, thus, better VaR forecasts.

Contrary to standard simple methods, the combination forecasts exhibit a stable in sample and out of sample performance across the types of asset or evaluation period. Overall, they outperform the stand alone estimates by generating independent exceedances within the limits imposed by Basel II rules. The relatively excellent performance of the new VaR forecast methodologies benefit from the general advantages of forecast combinations, such as robustness to structural breaks and misspecification risk and from its flexibility with respect to market conditions and type of assets.

Future research should aim at assessing and developing robust risk measures in real time settings, which are essential in the field of risk management, where investors face the continuous challenge of making spontaneous decisions. The new risk measures should be able to instantaneously incorporate all relevant new information related to the underlying asset, market conditions, or other economical and financial variables, which affect the market price risk, such as, shortages in market liquidity.

Acknowledgments

We are grateful to the editor David Veredas and the two anonymous referees for insightful remarks that have significantly improved the paper. For further helpful comments we would like to thank Karim Abadir, Kasia Bień, Nikolaus Hautsch, Wolfgang Härdle, Lidan Li, Peter Schmidt, participants to the Quantitative Finance Workshop on ‘Financial Crisis, Market Microstructure and Asset Pricing’, Rimini Center of Economic Analysis (RCEA), in Rimini, May 2009, the CoFE workshop in Königfeld, the Risk Seminar at Humboldt University and the faculty seminar at the University of Duisburg. All remaining errors are ours.

⁶ Similar graphs stemming from other combinations are available from the authors upon request.

Appendix A

See Tables A1–A6.

Table A1

Backtest results for the calm period: parametric methods. Percentage rate of violations for VaR at $p = 1\%$ for the period from January 1st, 2007 to July 17th, 2007 (total of 143 days). ** refers to p -values of conditional coverage test smaller than 0.05, * to p -values between and 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold type entries are in the “red zone”, italic type entries are in the “yellow zone” and no typeface entries are in the “green zone”.

Stock type	Start date	ND				TD				skewed TD			QR	EVT
		ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	ARMA-FIGARCH		
Small	1987	2.79	2.79	2.79	2.09	2.09	2.09	2.09	2.09	0.70	1.40	1.40	0.69	0.69
	1996	2.79	2.79	2.79	2.09	2.09	2.09	2.09	2.09	1.40	2.10	1.40	0.69	0.69
	2001	2.09	2.09	2.79	2.09	2.09	2.09	2.09	2.09	2.10	2.10	1.40	0.69	0.69
	2005	2.09	2.09	2.79	2.09	2.09	2.09	2.09	2.09	2.10	2.10	2.10	0.69	2.09
Middle	1987	2.09**	2.79**	2.79**	3.49**	1.39	1.39	2.09	1.39	1.40	1.40	1.40	0.00	0.69
	1996	1.39	2.79**	2.79**	1.39	1.39	2.09**	2.09**	1.39	1.40	2.10*	0.70	0.69	0.69
	2001	2.09**	2.79**	2.79**	2.79**	1.39	2.79**	2.09**	2.09**	1.40	2.80*	3.50*	0.69	1.39
	2005	2.79**	3.49**	2.79**	2.09	1.39	2.79**	2.09**	1.39	2.10	2.80*	1.40	0.69	1.39
Large	1987	2.09	3.49*	3.49*	3.49*	2.09	2.09	2.09	2.79	2.10	2.10	2.80	0.00	0.69
	1996	2.79	3.49*	3.49*	2.79	2.09	2.79	2.09	2.09	2.10	2.10	2.80	0.69	0.69
	2001	2.79	3.49*	3.49*	2.09	2.09	2.79	2.09	2.09	2.10	2.10	2.10	0.69	1.39
	2005	3.49*	4.19**	3.49*	2.79	2.79	2.09	2.09	2.79	2.10	1.40	2.80	2.09	2.09

Table A2

Backtest results for the crisis period: parametric methods. Percentage rate of violations for VaR at $p = 1\%$ for the period from July 18th, 2007 to July 1st, 2009 (total of 510 days). ** refers to p -values of conditional coverage test smaller than 0.05, * to p -values between and 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold type entries are in the “red zone”, italic type entries are in the “yellow zone” and no typeface entries are in the “green zone”.

Stock type	Start date	ND				TD				skewed TD			QR	EVT
		ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	ARMA-FIGARCH		
Small	1987	2.94**	2.54**	2.54**	3.13**	2.35**	1.96	1.76	1.96	0.78	0.59	1.57	3.33**	0.39
	1996	3.13**	2.54**	2.54**	3.13**	2.54**	1.96	1.76	2.15*	1.37	0.78	1.76	4.31**	1.17
	2001	3.33**	2.54**	2.54**	2.15**	2.94**	2.15	1.76	1.56	1.57	1.57	1.76	6.66**	1.56
	2005	2.94**	2.35**	2.54**	3.13**	2.74**	2.35	1.76	1.56	1.18	1.76	2.16	6.66**	2.74**
Middle	1987	3.52**	2.74**	2.74**	3.52**	2.54**	2.15*	2.15*	2.15*	1.96	1.57	1.76	0.98	1.17
	1996	3.52**	2.74**	2.74**	2.94**	2.54**	2.15*	2.15*	1.96	1.76	2.16	3.52**	1.37	
	2001	3.52**	2.74**	2.74**	2.54**	2.74**	2.35**	2.15*	2.15*	1.96	1.76	1.57	4.51**	1.96
	2005	4.11**	2.54**	2.74**	3.13**	2.74**	2.35**	2.15*	1.96	1.76	1.57	1.96	5.68**	1.96
Large	1987	3.72**	3.13**	2.74**	3.72	1.96	1.96	1.76	2.15*	1.37	1.37	1.96	1.76	0.98
	1996	3.72**	3.13**	2.74**	2.74	1.96	1.96	1.76	1.76	1.57	1.37	1.96	3.33**	0.98
	2001	4.11**	3.13**	2.74**	2.74**	2.35*	2.15	1.76	1.76	1.37	1.37	1.57	5.68**	1.17
	2005	4.51**	9.21**	2.74**	5.68**	2.35*	2.15	1.76	3.13**	1.18	1.37	1.76	5.68**	1.17

Table A3

Backtest results for the crash period: parametric methods. Percentage rate of violations for VaR at $p = 1\%$ for the period from September 1st, 2008 to July 1st, 2009 (total of 217 days). ** refers to p -values of conditional coverage test smaller than 0.05, * to p -values between and 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold type entries are in the “red zone”, italic type entries are in the “yellow zone” and no typeface entries are in the “green zone”.

Stock type	Start date	ND				TD				skewed TD			QR	EVT
		ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	ARMA-FIGARCH		
Small	1987	3.22**	2.76*	2.76*	2.76*	2.30	2.30	1.84	2.30	0.46	0.00	1.38	5.99**	0.00
	1996	3.22**	2.76*	2.76*	3.22**	2.30	2.30	1.84	2.30	1.38	0.46	1.84	7.37**	0.92
	2001	3.22**	2.76*	2.76*	1.84	3.22**	2.76*	1.84	1.38	1.84	1.38	1.84	8.29**	1.84
	2005	3.22**	2.76*	2.76*	3.68**	3.22**	2.76*	1.84	0.92	0.46	1.84	1.84	8.29**	3.22**
Middle	1987	3.68**	2.30	2.30	3.68**	2.76*	2.30	2.30	1.84	1.84	1.38	1.38	2.30	1.38
	1996	3.68**	2.30	2.30	2.76*	2.76*	2.30	2.30	1.84	1.84	1.38	2.30	7.83**	1.38
	2001	3.68**	2.30	2.30	2.30	2.76	2.30	2.30	1.38	1.84	1.38	1.38	6.91**	1.84
	2005	3.68**	2.30	2.30	2.30	2.30	2.30	2.30	1.38	1.38	1.38	1.84	7.37**	1.84
Large	1987	4.60**	2.76*	2.76*	4.60**	1.38	2.30	1.38	1.84	1.38	1.38	2.30	4.14**	0.92
	1996	4.60**	2.76*	2.76*	2.76*	1.38	1.84	1.38	1.84	1.38	1.38	2.30	6.91**	0.92
	2001	4.60**	2.76*	2.76*	2.30	2.30	1.84	1.38	1.84	0.92	1.38	1.38	8.75**	0.92
	2005	4.60**	12.90**	2.76*	6.45**	1.38	1.84	1.38	2.76*	0.92	1.38	1.84	7.37**	0.92

Table A4

Backtest results for the calm period: nonparametric methods. Percentage rate of violations for VaR at $p = 1\%$ for the period from January 1st, 2007 to July 17th, 2007 (total of 143 days). ** refers to p -values of conditional coverage test smaller than 0.05, * to p -values between 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold type entries are in the "red zone", italic type entries are in the "yellow zone" and no typeface entries are in the "green zone".

Stock type	Number of observations	HS	Start date	HS	FHS
Small	1000	0.69	1987	0.69	2.09
	750	0.69	1996	0.69	2.09
	500	0.69	2001	0.69	2.09
	250	0.69	2005	0.69	1.39
Middle	1000	1.39	1987	0.69	1.39
	750	1.39	1996	0.00	1.39
	500	1.39	2001	0.00	1.39
	250	1.39	2005	1.39	1.39
Large	1000	2.09	1987	0.69	2.09
	750	2.09	1996	0.69	1.39
	500	2.09	2001	0.69	2.09
	250	2.09	2005	2.09	2.09

Table A5

Backtest results for the crisis period: nonparametric methods. Percentage rate of violations for VaR at $p = 1\%$ for the period from July 18th, 2007 to July 1st, 2009 (total of 510 days). ** refers to p -values of conditional coverage test smaller than 0.05, * to p -values between 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold type entries are in the "red zone", italic type entries are in the "yellow zone" and no typeface entries are in the "green zone".

Stock type	Number of observations	HS	Start date	HS	FHS
Small	1000	2.94**	1987	12.15**	1.76
	750	4.90**	1996	10.39**	2.35**
	500	5.88**	2001	7.64**	2.54**
	250	7.25**	2005	6.86**	1.56
Middle	1000	2.54**	1987	7.64**	2.15*
	750	3.92**	1996	7.25**	2.35**
	500	5.09**	2001	5.49**	2.35**
	250	6.47**	2005	5.68**	2.15*
Large	1000	3.13**	1987	8.82**	1.76
	750	4.51**	1996	8.03**	1.56
	500	6.27**	2001	7.25**	1.96
	250	6.86**	2005	7.05**	1.96

Table A6

Backtest results for the crash period: nonparametric methods. Percentage rate of violations for VaR at $p = 1\%$ for the period from September 1st, 2008 to July 1st, 2009 (total of 217 days). ** refers to p -values of conditional coverage test smaller than 0.05, * to p -values between 0.05 and 0.10 and no mark refers to p -values larger than 0.10. Bold type entries are in the "red zone", italic type entries are in the "yellow zone" and no typeface entries are in the "green zone".

Stock type	Number of observations	HS	Start date	HS	FHS
Small	1000	4.14**	1987	19.35**	1.84
	750	6.91**	1996	16.59**	1.84
	500	8.29**	2001	11.98**	2.76*
	250	8.75**	2005	9.21**	1.38
Middle	1000	2.76**	1987	13.36**	1.84
	750	5.53**	1996	12.44**	2.30
	500	7.37**	2001	10.13**	2.30
	250	8.29**	2005	8.29**	1.84
Large	1000	4.14**	1987	15.23**	1.84
	750	5.53**	1996	14.28**	1.38
	500	7.83**	2001	12.44**	1.38
	250	8.75**	2005	9.67**	0.92

Appendix B

See Figs. B1 B4.

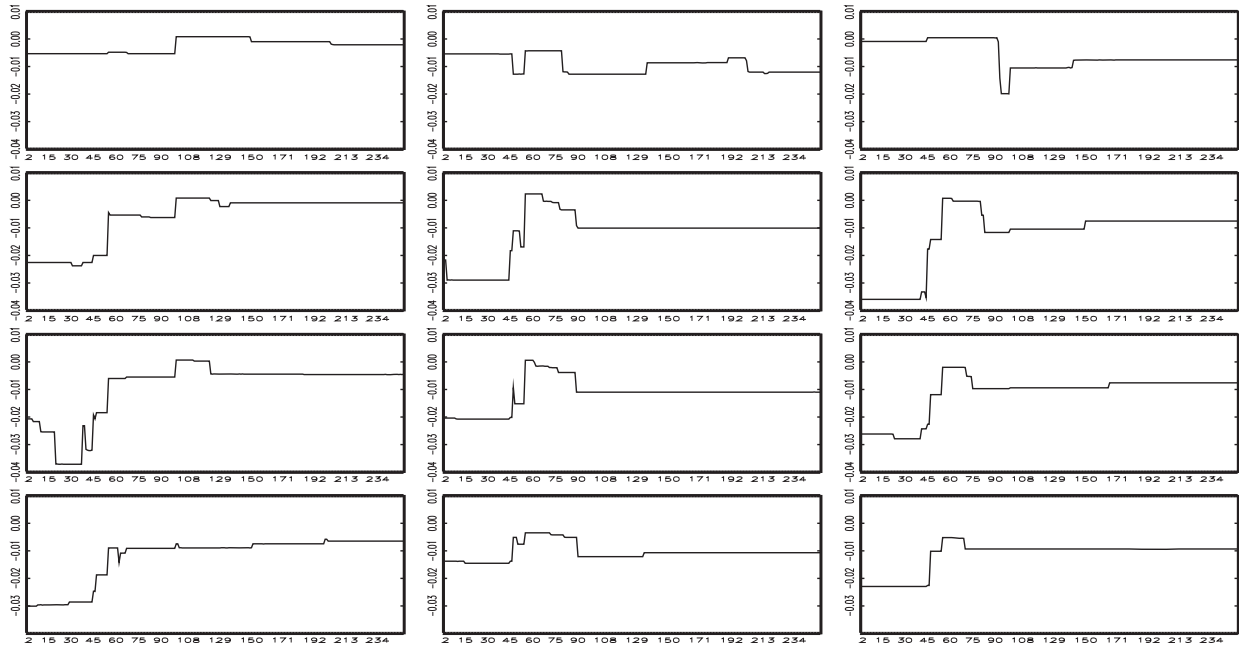


Fig. B1. Estimated $\lambda_{T,0}^*$ through the crisis, CQOM. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.

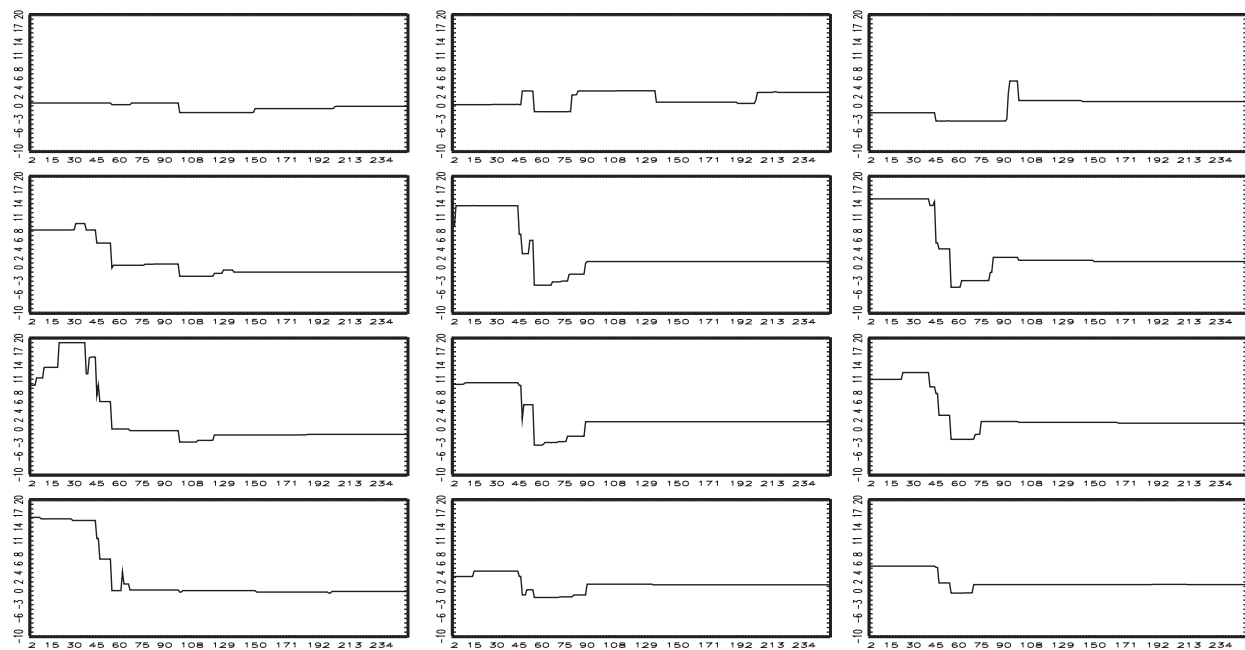


Fig. B2. Estimated $\lambda_{T,1}^*$ through the crisis, CQOM. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.

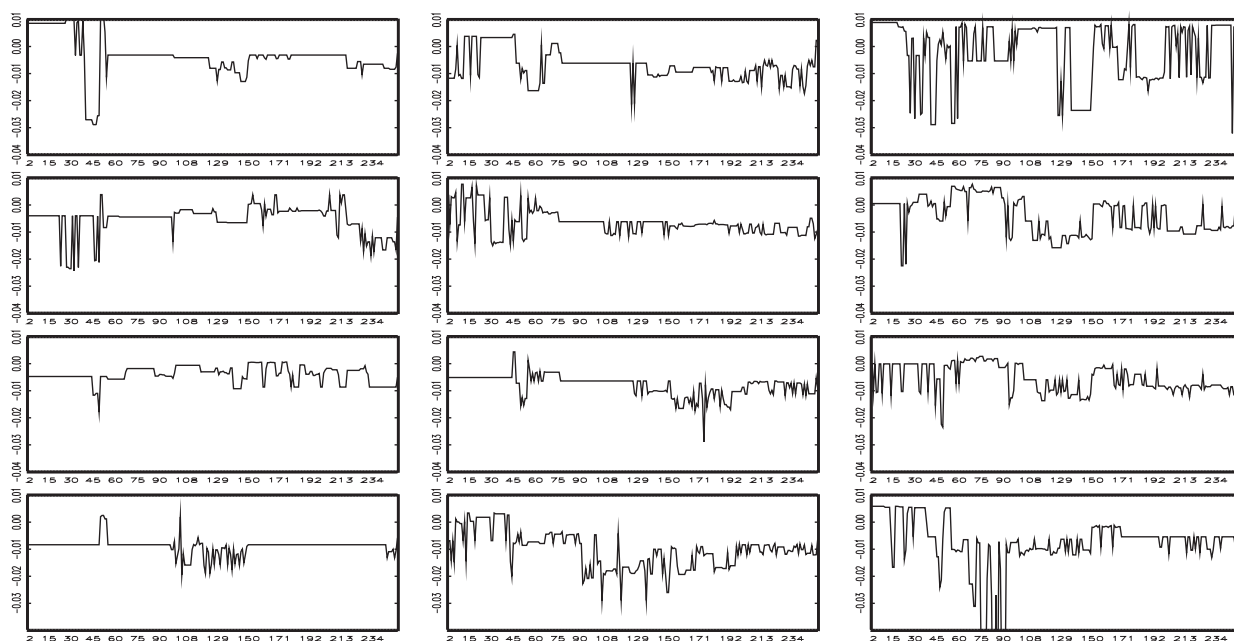


Fig. B3. Estimated $\lambda_{T,0}^*$ through the crisis, CCOM. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.

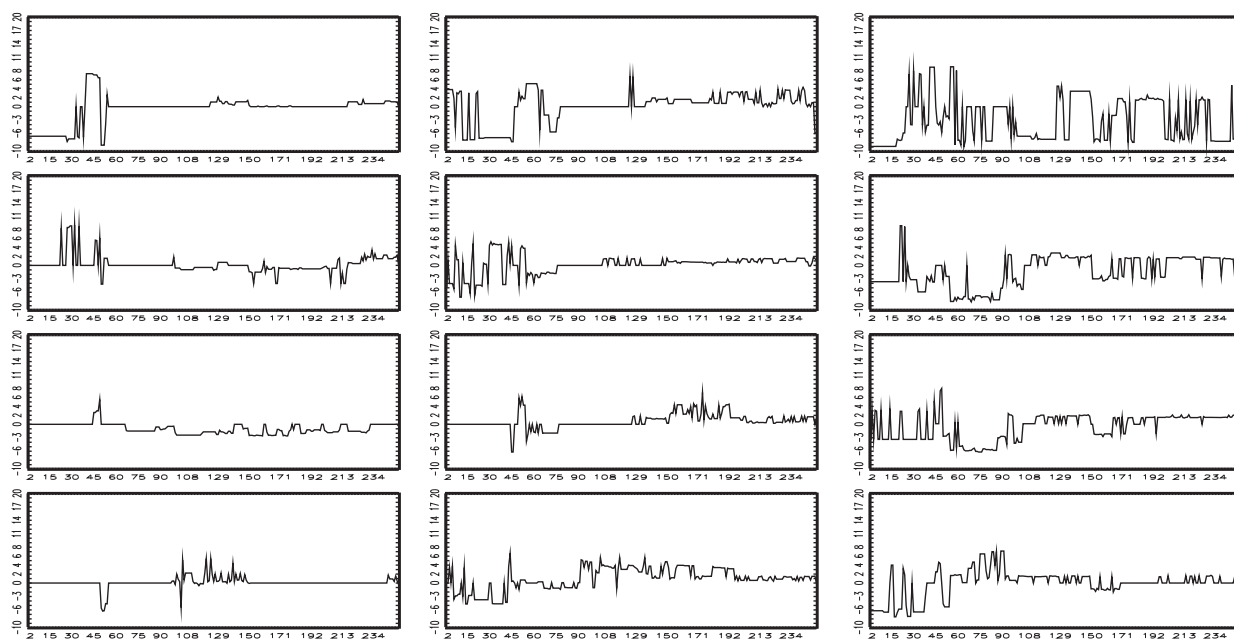


Fig. B4. Estimated $\lambda_{T,1}^*$ through the crisis, CCOM. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.

References

- Baillie, R., Bollerslev, T., Mikkelsen, H., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3–30.
- Barone-Adesi, G., Giannopoulos, K., Vosper, L., 1999. VaR without correlations for nonlinear portfolios. *Journal of Futures Markets* 19, 583–602.
- Basel Committee, 1996. Overview of the Amendment to the Capital Accord to Incorporate Market Risks. Discussion Paper, Basel Committee on Banking Supervision.
- Christoffersen, P.F., 1998. Evaluating interval forecasts. *International Economic Review* 39, 841–862.
- Christoffersen, P.F., 2003. *Elements of Financial Risk Management*. Elsevier Science, USA.

- Coroneo, L., Veredas, D., 2011. A simple two-component model for the distribution of intraday returns. *The European Journal of Finance*, forthcoming.
- Crouhy, M., Galai, D., Mark, R., 2001. *Risk Management*. McGraw-Hill, New York.
- Engle, R., Manganelli, S., 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22 (4), 367–381.
- Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics* 23 (4), 416–431.
- Giot, P., Laurent, A., 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance* 11, 379–398.
- Goffe, W., Ferrier, G.D., Rogers, J., 1994. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics* 60, 65–100.
- Jiménez-Martín, J.A., McAleer, M., Pérez-Amaral, T., 2011. GFC-Robust Risk Management Strategies under the Basel Accord. KIER Working Papers 757, Kyoto University.
- Jorion, P., 2007. *Value at Risk*, 3rd ed. McGraw-Hill, New York.
- Jorion, P., Risk management lessons from the credit crisis. *European Financial Management*, forthcoming.
- Koenker, R., Basset, G., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Kuester, K., Mittnik, S., Paolella, M.S., 2006. Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics* 4 (1), 53–89.
- McAleer, M., da Veiga, B., 2008. Forecasting value-at-risk with a parsimonious portfolio spillover GARCH (PS-GARCH) model. *Journal of Forecasting* 27, 1–19.
- Newey, W.K., West, K.D., 1987. A simple, positivesemi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Schmidt, P., 2011. Robustifying Value-at-Risk Forecasts. Master's Thesis, University of Konstanz, Germany.
- Summers, L., 1991. The scientific illusion in empirical macroeconomics. *Scandinavian Journal of Economics* 93 (2), 129–148.
- Timmermann, A., 2005. Forecast Combinations. Working Paper.