THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 10-07

Roxana Chiriac
University of Konstanz, CoFE

Winfried Pohlmeier
University of Konstanz, CoFE, ZEW, RCEA

HOW RISKY IS THE VALUE AT RISK?

How Risky is the Value at Risk?

Roxana Chiriac*
University of Konstanz,
CoFE

Winfried Pohlmeier
University of Konstanz,
CoFE, ZEW, RCEA

March 16, 2010

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 present empirical evidence from assessing the out-of-sample performance and robustness of VaR before and during the recent financial crisis with respect to the choice of sampling window, return distributional assumptions and stochastic properties of the underlying financial assets. Moreover we develop a new data driven approach that is based on the principle of optimal combination and that provides robust and precise VaR forecasts for periods when they are needed most, such as the recent financial crisis.

JEL classification: $C21, C5, G01, G17, G28, G32$

Keywords: Value at Risk, model risk, optimal forecast combination

*Corresponding author, Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone +49-7531-88-2204, Fax -4450, email: Roxana.Chiriac@uni-konstanz.de. Earlier versions of the paper have been presented at 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 Königsfeld, the risk seminar at Humboldt University and faculty seminar at the University of Duisburg. For helpful comments on earlier drafts of the paper we would like to thank Karim Abadir, Nikolaus Hautsch and Wolfgang Härdle. Financial support by the German Research Foundation within the research group "Price-, Liquidity- and Credit Risks: Measurement and Allocation" is gratefully acknowledged. All remaining errors are ours.

1 Introduction

The current worldwide financial crisis has revealed major deficiencies in the existent financial risk measures.[1] Although considerable progress has been made over the last decade to quantify financial risks by means of elaborate econometric tools, there is little evidence on their performance during turbulent times similar to the recent financial crisis. In general, 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. Although comparative studies on the performance of the risk models used in practice or proposed in academic journals are not new, a systematic analysis their limitations such that users are able to assess the model risk of using different approaches is still missing.

In this paper we provide evidence on some dimensions of the riskiness of risk models. The focus of our interest lies on the Value at Risk (VaR) as a major tool to assess financial-market risks. As pointed out by Jorion (2007) in the preface of the latest edition of his monograph on the Value at Risk, the VaR methodology was developed as a response to the financial disasters of the early 1990s. In particular, following the 1995 amendment of the Basel accord, whereby banks are allowed to use their own internal models, the search for the "best" VaR approach has become of practical importance for quantitative risk management. Initially developed to assess market risks, the "*VaR is now being used to control and manage risk actively, well beyond derivatives*" (Jorion (2007), p.vii) including credit risk and operational risk. Yet, experience with the performance of the VaR methodology in rough times was missing until the subprime crisis started to shake the world financial markets. Obviously the problem of lack of empirical experience is most severe for portfolios consisting of new financial products, for which by definition the number of observations is limited.

[1] For example, JP Chase reported 5, Credit Suisse 7 and UBS 16 exceedances based on one-day-ahead VaR forecasts in the 3rd quarter of 2007, which requires a maximum of 0.63 exceedances for a probability level of 1% (Jorion (2009)).

The Basel rules require that risk models are to be estimated on historical data sampled over a period of at least 250 trading days. No explicit rules are given on how the sampling period should be chosen. However, the choice of the sampling period strongly affects the mean squared prediction error of econometric estimators in the presence of structural breaks. It is well established that post-break sampling periods are suboptimal in terms of out-of-sample forecasting performance for mean predictions. As shown by Pesaran and Timmermann (2007) for the linear model, pre-break information creates a bias in prediction but improves the efficiency of estimates, so that the optimal sampling window generally includes also pre-break information. This argument is likely to be even more valid for VaR predictions. By definition small sampling windows are less likely to capture extreme outcomes, such that the valuable information on the tails of the return distribution is missing.

Driven by the suspicion that the VaR could potentially embody high model risk, we provide here what Summers (1991) calls some "*successful pieces of pragmatic empirical work*". We show for a fairly large range of assets, how unstable VaR estimates are, not only with respect to the pure choice of the risk model, but also concerning the choice of the underlying econometric model and the choice of the appropriate sampling period. In particular, we show that 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.

Many of the earlier studies on alternative econometric approaches to estimate the VaR perform horse races. They differ in many dimensions related to the choice of the models under investigation, the assets or portfolios analyzed, evaluation criteria and sample size. Kuester et al. (2006) consider the out-of-sample performance of GARCH based models using the test of unconditional and conditional coverage proposed by Christoffersen (1998) as well as a dynamic quantile test. For their horse race they use the daily data of *NASDAQ* Composite Index for a period of more than 30 years, but they restrict their analysis to a rolling sampling window of size T=1000 days for each estimate, corresponding to roughly four years of trading.

McAleer and da Veiga (2008) concentrate on VaR predictions of an equally weighted portfolio containing S&P500, FTSE100, CAC40 and SMI with a rolling window of T = 2000 days, where the variance of the portfolio is computed from estimates of different multivariate GARCH models. Besides the number of exceedances, their benchmark criteria for model comparison is the mean daily capital charge implied by each model.

The question whether information on intraday variation in returns is informative for VaR estimation is taken up in Giot and Laurent (2004), who compare one-day-ahead forecasts of VaR's based on a long memory skewed Student model for the realized volatility estimates with those from a skewed Student APARCH on daily returns and the standard RiskMetrics approach. Their comparison is based on two exchange rates (YEN-USD and DEM-USD) and two stock indices (S&P500, CAC40). They explicitly distinguish between the forecasting performance of VaR of long and short positions. Unfortunately, the authors use different sample sizes and sample periods, so that their empirical findings cannot be compared across assets.

Most of the horse races are applied to popular assets (large-cap stock, major indices or currencies) and data stemming from calm periods or developed markets. A notable exception is Gençay and Selçuk (2004) who look at daily stock price indices from eight different emerging markets. They argue that emerging market economies differ fundamentally from developed economies with respect to changes in the dynamics of the economy. The relative performance of different approaches compared in their study depends among others on the quantile of interest and the tail of the distribution.

Very typical for the last fraction of horse races reported in the academic literature is the concentration on one day forecasting horizons although Basel II requires backtesting for a horizon of ten trading days. Thus potential misspecification of the dynamics of the risk models is likely to play a larger role, when it comes to forecast over a ten day horizon.

In this paper we present a horse race in a different, but very simple dimension. We provide a list of stylized facts on how well the VaR worked in various time periods, especially before and during the recent financial crisis, for different securities and popular models. In the spirit of a meta study we summarize systematic patterns of VaR predictions based on different model specifications and estimation windows. The goal of this exercise is to provide more insights into the applicability of these approaches in different world settings. We show that the performance of standard VaR estimates differs among types of stocks (with small, middle and large market capitalization), subject to the choice of volatility model or distributional assumption.

We provide empirical evidence that using a sampling window of the last two years of data is sufficient to asses the downward risks before the crisis, while during the crisis the estimation window should incorporate also information on past extreme events (e.g. the market crash from 1987) in order to improve the accuracy of the standard VaR methods. Particulary, VaR estimates computed from General Extreme Value (GEV) distributed residuals, standardized by GARCH conditional volatilities estimated on windows starting before the crash of 1987, generate during the current crisis independent failure rates, which according to the Basel II rules are situated in the "green zone".

As a result of the performance variability of standard VaR estimates with respect to the model choice, distributional assumptions or the choice of estimation windows, we propose two new procedures of estimating financial losses based on the principle of optimal forecast combinations, which already has been applied to improve the accuracy of mean forecasts (Timmermann (2005)). Giacomini and Komunjer (2005) make reference to this principle and empirically apply it to VaR estimates in the context of deriving encompassing tests for comparing conditional quantile forecasts. Different from their study, 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. The first procedure finds the optimal combination by minimizing the conditional coverage tests on the failure rates stemming from the VaR combinations. The second procedure minimizes the distance between population and empirical quantiles, by means of quantile regression approaches. We show that optimal combinations of VaR estimates improve the performance of stand-alone estimates during the periods when they are needed most, such as the recent financial crisis.

The outline of the paper is as follows: Section 2 introduces the standard VaR approaches and outlines the set-up of the meta study, whose results are presented in Section 3. Section 4 describes theoretically the new procedures of optimally combining VaR estimates and reports their empirical results. Finally, Section 5 summarizes the main results and concludes the paper.

2 Set-Up of the Meta Study

The Value at Risk 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). We focus in this paper on estimating the VaR at time t for a long position over time horizon h with probability p ($VaR_t(p,h)$), which is formally defined to be given by:

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

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 $\Delta V_{t+h}(h)$. Furthermore, given that $\Delta V_{t+h}(h)$ can be defined as a location scale process conditional on an information set $\mathcal{F}_t$:

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

where $\mu_{t+h|t}$ is the expected change in the value V_t from t to $t+h$ given the information set up to t, ε_{t+h} is the error term (shock, surprise) and z_t has a zero-location, unit-scale probability density $f_Z(\cdot)$, we can then derive the VaR conditional on the information set $\mathcal{F}_t$ to be:

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

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 (the Value at Risk) that is sufficient to cover losses on the trading book over a ten days holding period for 99% of the times ($h = 10$ and $p = 0.01$). Interestingly, the horse races presented in academic journals usually refer to a one day forecasting horizon. However, the choice of the forecasting horizon may seriously bias the results if the dynamics of the underlying econometric model is misspecified. Early empirical evidence is provided by Christoffersen and Diebold (2000), who show that volatility forecast ability varies with horizons and asset classes.

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 a good predictive performance according to some backtesting rules. These rules imply the computation of a hit sequence, defined by:

$$H_{t+h} = \mathbb{1}\left(\Delta V_{t+h}(h) < \hat{V}aR_{t+h|t}(p)\right) \qquad t = T+1, T+2, \ldots T+S, \tag{2.3}$$

which follows a binomial distribution with parameter p.

The rules imposed by 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)). Based on the results of this test and accounting for the Type

I error rate, the Basel Committee (1996) came 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 and S 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 testing simultaneously the two null hypotheses from the unconditional coverage and independence tests.

Our meta-study is designed to assess the quality of standard VaR estimates at different dimensions: asset choice, model choice, distributional assumption and estimation window, during the recent financial crisis. Given that the evaluation period implies a series of negative extreme events, we find proper to focus on assessing the performance of standard VaR estimates at $p = 0.01$, which is in line with the Basel II requirements. Regarding the asset choice, we consider three indices built on 30 randomly chosen stocks from the Dow Jones U.S. Small, Middle and respectively Large cap-indices[2]. 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

[2]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, ITG, 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, GWW, 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, SYY, TGT, WFT, XOM

thus complementing the results of previous evaluation researches done only on popular large-cap stocks, or major indices.

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 1^{st} up to July 17^{th}, 2007. July 17^{th}, 2007 represents in our study the beginning of the current financial crisis: on this day the FED signaled the first troubles on the subprime loan markets and announced their support and supervision to the subprime mortgage lenders. We consider this first evaluation period to be a calm one, representative for the years prior to the financial crisis; (2) the crisis period starting on July 18^{th}, 2007 up to July 2^{nd}, 2009 (the end of our sampling window); and (3) the crash period, which starts in September 2008, when the crisis hits the financial markets following the bankruptcy of Lehman Brothers on September 15^{th}. Starting with this date, major stock markets all over the world experience the largest decline since the black Monday crash on October 19^{th}, 1987. By dividing the whole evaluation period in these three subperiods, we intend to check the robustness of standard 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 parametrical distributional assumptions such as normal or Student-t distribution or semi-parametric approaches such as extreme value theory (EVT) for the standardized residuals. This leads to the following model specifications:

1. ARMA(1,0) for the conditional mean $\mu_{t+1|t}$ and GARCH(1,1) for the conditional variance $\sigma^2_{t+1|t}$ (Bollerslev (1986)):

$$\mu_{t+1|t} = c + \phi r_t, \tag{2.4}$$

where r_t is the log-return from $t-1$ to t of holding a stock and

$$\sigma^2_{t+1|t} = \omega + \alpha_1 \varepsilon^2_t + \beta_1 \sigma^2_t, \quad \omega, \alpha, \beta > 0, \quad \alpha + \beta < 1, \tag{2.5}$$

where $\varepsilon_t = r_t - \mu_{t|t-1}$. We further consider normal, Student-t and EVT functional forms for the probability density function of the standardized residuals. To estimate the EVT quantiles, we adapt the maximum likelihood procedure developed by Roncalli (2001) to the peak over the threshold (POT) approach. Given that there is no straightforward method of choosing the optimal threshold level, we follow the results of McNeil and Frey (2000), who find that a threshold of approximatively 10% minimizes the MSE of estimated $Q_p(Z)$, and choose a threshold of 9% from the largest losses, which assures in our case the stability of quantile estimates for all types of stocks and sampling windows.

2. RiskMetrics for the conditional covariance with $\mu_{t+1|t} = 0$,

$$\sigma^2_{t+1|t} = (1-\lambda)\varepsilon^2_t + \lambda \sigma^2_t, \quad 0 < \lambda < 1. \tag{2.6}$$

This approach, introduced by JP Morgan in 1995, aims at describing the conditional variance as an exponentially weighted moving average and is mostly applied in practice for a fix parameter $\lambda = 0.94$. In this case we apply the normal and Student-t with 7 degrees of freedom distribution on the standardized residuals. Along with this approach (*RM-fix*), we also proceed at estimating the parameter λ based on the underlying data (*RM-est*), assumed to be normal or Student-t distributed, for which we proceed at estimating the degrees of freedom.

3. ARMA(1,0) for the conditional mean $\mu_{t+1|t}$ and FIGARCH(1,d,0) for the conditional variance $\sigma^2_{t+1|t}$ (Baillie et al., 1996), which accounts for the long memory property of financial volatilities:

$$(1-L)^d \varepsilon^2_{t+1} = \omega + \nu_{t+1} - \beta \nu_t, \quad 0 < d < 1, \quad \nu_t \equiv \varepsilon^2_t - \sigma^2_t, \quad 0 < \beta < 1 \tag{2.7}$$

In this case, we apply the normal and Student-t distribution on the standardized residuals.

Subject to this dimension, 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 to forecast the risk of a certain type of stock in calm vis-á-vis 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.

Another important dimension we focus on in our meta-study is the size of the estimation window. Contrary to other studies, which apply a rolling window in the forecasting procedure, we use the augmented 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. This strategy is particulary useful in forecasting extreme losses, by incorporating in all in-sample windows the shock of the market crash from 1987, which seems to be the most sever extreme event on the stock markets until today. Through this strategy we aim at identifying the extreme events from the past, which contain valuable information in order to forecast losses from the current crisis. For this purpose, we consider four strategies of choosing the starting date of the augmented windows: January 1^{st}, 1987, for which all in-sample data include the Black Monday effect from October 19^{th}, 1987; January 1^{st}, 1996, for which the in-sample windows include the dot-com bubble crash from March up to October, 2002 and the terrorist events from

September 11^{th}, 2001; January 1^{st}, 2001, which includes just the September 11^{th}, 2001 event; and January 1^{st}, 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.

Alternatively 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, which 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 augmented windows starting in 1987, 1996, 2001 and respectively 2005. The FHS estimation is applied only on augmented windows.

The dimensions described so far lead to a meta-study which comprises 120 forecasts given by three indexes (small-, middle-, large-cap), three parametric model specifications (RM-fix, RM-est, ARMA-FIGARCH) with two distributional assumptions (normal and Student-t), one parametric model (ARMA-GARCH) with three distributional assumptions (normal, Student-t and EVT), four strategies of choosing the in-sample windows (starting in 1987, 1996, 2001 and respectively 2005) and two nonparametric models based on rolling (HS) and augmented (HS, FHS) sampling windows of sizes 250, 500, 750 and 1000 and respectively starting in 1987, 1996, 2001 and 2005. Furthermore, we apply the unconditional and conditional coverage tests described above to evaluate the VaR estimates on horizons of one day. It remains for further research to carry out this meta-study on estimates for longer horizons, such as ten days.

3 Meta Study Results

In this section we present the results of evaluating the VaR estimators by means of backtesting before and during the current financial crisis. Tables 1, 2 and 3 report the percentage rate of violations as well as results related to unconditional and conditional coverage tests for the periods under consideration.

Table 1 reports the VaR failure rates prior to the recent sub-prime crisis, from January 1^{st} up to July 17^{th}, 2007. All in all, the standard VaR approaches seem to perform well during this calm period. While the normal distribution produces most of the "yellow zone"-type violations, applying a fat-tailed distributional assumption characteristic to financial returns such as 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 returns. However, its performance significantly improves when applied together with the Student-t distribution assumption. Nevertheless, during this calm period, the ARMA-GARCH approach generally suffices to capture the data dynamics and produces independent sequences of violations.

Table 1 also provides insights into the performance of the VaR estimators with respect to different sampling windows. Sample size plays no significant role in the presence of the normal and Student-t distribution, but it definitively affects the precision of VaR estimates based on EVT distributional assumption: incorporating in the sampling window major past extreme events, such as the ones from 1987 and 1996, leads to better EVT quantile estimates, which enhance the ability of VAR measure to correctly forecast potential losses.

Contrary to the calm period, standard VAR estimators perform generally poorly during the recent financial crisis (Table 2). Estimators using the normality assumption exhibit the worst performance and yield only time dependent "red zone"-type violations. The only exception shows the portfolio of large-cap stocks, for which the ARMA-FIGARCH approach seems to correctly capture the time dependence among the extreme losses. 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, exhibits the Student-t distribution, which in the absence of estimation noise (fixed degrees of freedom), delivers less, but still disputable number of violations in the "yellow zone" for small- and large- cap stocks.

During the crisis period, the EVT approach estimated on historical samples which incorporate information from previous crisis, has the best performance and results in all failure rates being independent and in the "green zone". Although not new (e.g. Kuester et al., 2006), this result highlights the fact that by simply implementing EVT-based VaR estimates, the financial institutions could have better handled the extreme market risks provoked by the recent financial crisis. This finding attenuates some of recent opinions, which state that the VaR is a faulty measure during hard times. Our results show that this is the case if VaR is built on standard approaches such as normal, but it becomes a sound method simply by implementing the EVT approach on large and informative historical samples.

As already seen, the sampling window plays an important role in correctly forecasting losses during crisis times. A further important role is played by the degree of parametrization of the underlying models: the ARMA-FIGARCH approach in the presence of Student-t distribution and estimated on samples from 1996 or 2001 outperforms other models for all type of stocks.

Table 1: Backtest results for the calm period. Percentage rate of violations for VaR at p = 1% for the period from January 1^{st}, 2007 to July 17^{th}, 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".

Distribution		ND				TD				EVT
Stock Type	Start date	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH
small	1987	*2.79*	*2.79*	*2.79*	2.09	2.09	2.09	2.09	2.09	0.69
	1996	*2.79*	*2.79*	*2.79*	2.09	2.09	2.09	2.09	2.09	0.69
	2001	2.09	2.09	*2.79*	2.09	2.09	2.09	2.09	2.09	0.69
	2005	2.09	2.09	*2.79*	2.09	2.09	2.09	2.09	2.09	2.09
middle	1987	2.09**	*2.79***	*2.79***	**3.49****	1.39	1.39	2.09	1.39	0.69
	1996	1.39	*2.79***	*2.79***	1.39	1.39	2.09**	2.09**	1.39	0.69
	2001	2.09**	*2.79***	*2.79***	2.79**	1.39	*2.79***	2.09**	2.09**	1.39
	2005	*2.79***	**3.49****	*2.79***	2.09	1.39	*2.79***	2.09**	1.39	1.39
large	1987	2.09	**3.49***	**3.49***	**3.49***	2.09	2.09	2.09	*2.79*	0.69
	1996	*2.79*	**3.49***	**3.49***	*2.79*	2.09	*2.79*	2.09	2.09	0.69
	2001	*2.79*	**3.49***	**3.49***	2.09	2.09	*2.79*	2.09	2.09	1.39
	2005	**3.49***	**4.19****	**3.49***	*2.79*	*2.79*	2.09	2.09	*2.79*	2.09

Table 3 reports the backtest results for the third period, which we call "crash period". Although in general better, these results emphasize the findings from the crisis period: regardless the asset type, higher parametrization increases the performance of most of the VaR estimates, including here of the ones built on the normal distributional assumption.

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 EVT approch estimated on large historical samples clearly outperforms all other methods by generating "green-zone" independent violations for all types of stocks.

By comparing the results of tables 2 and 3, 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. A substantial improvement brings the Student-t distribution, which seems to quickly assimilate and exploit the informational content related to extreme events preceding the crash period.

Table 2: Backtest results for the crisis period. Percentage rate of violations for VaR at p = 1% for the period from July 18^{th}, 2007 to July 2^{nd}, 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".

Distribution		ND				TD				EVT
Stock Type	Start date	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH
small	1987	**2.94****	**2.54****	**2.54****	**3.13****	**2.35****	*1.96*	*1.76*	*1.96*	0.39
	1996	**3.13****	**2.54****	**2.54****	**3.13****	**2.54****	*1.96*	*1.76*	**2.15***	1.17
	2001	**3.33****	**2.54****	**2.54****	**2.15****	**2.94****	**2.15**	*1.76*	1.56	1.56
	2005	**2.94****	**2.35****	**2.54****	**3.13****	**2.74****	**2.35**	*1.76*	1.56	**2.74****
middle	1987	**3.52****	**2.74****	**2.74****	**3.52****	**2.54****	**2.15***	**2.15***	**2.15***	1.17
	1996	**3.52****	**2.74****	**2.74****	**2.94****	**2.54****	**2.15***	**2.15***	*1.96*	1.37
	2001	**3.52****	**2.74****	**2.74****	**2.54****	**2.74****	**2.35****	**2.15***	**2.15***	*1.96*
	2005	**4.11****	**2.54****	**2.74****	**3.13****	**2.74****	**2.35****	**2.15***	*1.96*	*1.96*
large	1987	**3.72****	**3.13****	**2.74****	**3.72**	*1.96*	*1.96*	*1.76*	**2.15***	0.98
	1996	**3.72****	**3.13****	**2.74****	**2.74**	*1.96*	*1.96*	*1.76*	*1.76*	0.98
	2001	**4.11****	**3.13****	**2.74****	**2.74****	**2.35***	**2.15**	*1.76*	*1.76*	1.17
	2005	**4.51****	**9.21****	**2.74****	**5.68****	**2.35***	**2.15**	*1.76*	**3.13****	1.17

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 incorporate already 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 (Student-t or EVT).

Table 3: Backtest results for the crash period. Percentage rate of violations for VaR at p = 1% for the period from September 1^{st}, 2008 to July 2^{nd}, 2009 (total of 129 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".

Distribution		ND				TD				EVT
Stock Type	Start date	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH	RM-est	RM-fix	ARMA-FIGARCH	ARMA-GARCH
small	1987	**3.22****	**2.76***	**2.76***	**2.76***	*2.30*	*2.30*	1.84	*2.30*	0.00
	1996	**3.22****	**2.76***	**2.76***	**3.22****	*2.30*	*2.30*	1.84	*2.30*	0.92
	2001	**3.22****	**2.76***	**2.76***	1.84	**3.22****	**2.76***	1.84	1.38	1.84
	2005	**3.22****	**2.76***	**2.76***	**3.68****	**3.22****	**2.76***	1.84	0.92	**3.22****
middle	1987	**3.68****	*2.30*	*2.30*	**3.68****	**2.76***	*2.30*	*2.30*	1.84	1.38
	1996	**3.68****	*2.30*	*2.30*	**2.76***	**2.76***	*2.30*	*2.30*	1.84	1.38
	2001	**3.68****	*2.30*	*2.30*	*2.30*	**2.76**	*2.30*	*2.30*	1.38	1.84
	2005	**3.68****	*2.30*	*2.30*	*2.30*	*2.30*	*2.30*	*2.30*	1.38	1.84
large	1987	**4.60****	**2.76***	**2.76***	**4.60****	1.38	*2.30*	1.38	1.84	0.92
	1996	**4.60****	**2.76***	**2.76***	**2.76***	1.38	1.84	1.38	1.84	0.92
	2001	**4.60****	**2.76***	**2.76***	*2.30*	*2.30*	1.84	1.38	1.84	0.92
	2005	**4.60****	**12.90****	**2.76***	**6.45****	1.38	1.84	1.38	**2.76***	0.92

The main finding of the analysis so far is that standard VaR methods perform totally differently from calm to crisis periods. 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 highly-parameterized methods estimated on larger samples perform well for all type of stocks. Finally, large sampling windows, incorporating valuable information on past shocks and properly exploited by suitable distributional assumptions (Student-t with estimated degrees of freedom and EVT) deliver better quantile estimates and thus better VaR forecasts. When the normal or fixed-degrees of freedom Student-t distribution are applied, which provide fixed quantile estimates, the sampling window plays a minimal role (it affects only the estimation of conditional mean and volatilities) and thus fails to properly forecast potential losses during turbulent times.

Tables A.9, A.10 and A.11 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 the sampling strategy. This result justifies once again the popularity of this method among practitioners. Prior filtering the data brings no additional gains (last column of Table A.9) 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 preceding the crisis.

Tables 4 and 5 summarize our findings in terms of meta regressions. Using OLS regressions, we quantify the impact that different model specifications and implementations have on the violation rate (Table 4) and on the p-value of conditional coverage tests (Table 5). Distributional choice always matters regardless of the evaluation period, but differently in sign from calm to crisis times: while less parametrization seems to be appropriate for forecasting financial risks during calm periods (column 1), choosing the EVT and Student-t over the normal or HS significantly increase the probability of correct coverage during turbulent times (columns 2 and 3).

Table 4: Results of OLS Meta Regressions. Dependent variable: Percentage number of violations. Explanatory variables: *Market Cap*, which equals 0, if the underlying asset is the small-cap index; 1 if middle-cap index; and 2 if large-cap index; *Model*, which equals 0, if there is no parametric model in the VaR estimation; 1, if the model is RM-fix; 2, if RM-est.; 3, if ARMA-GARCH; and 4 if ARMA-FIGARCH; *Distr.*, which equals 0, if the underlying distribution is HS; 1 if FHS; 2 if normal distribution; 3 if Student-t distribution; and 4, if EVT; *Years*, which equals 0, if the estimation sample starts in 2005; 1, if the estimation sample starts in 2001; 2, if the estimation sample starts in 1996; 3, if the estimation sample starts in 1987. *** significant at 1%; ** at 5%; * at 10%. Total number of observations: 120.

Variables	**Calm Period**	**Crisis Period**	**Crash Period**
Constant	1.549***	5.747***	8.647***
Market Cap	0.263***	0.084	-0.023
Model	0.046	-0.287***	-0.735***
Distr.	0.129**	-1.003***	-1.707***
Years	-0.061	0.024	0.181
$Adj.R^2$	0.083	0.452	0.479

Interestingly, the model choice does not matter for the calm evaluation period, but it contributes to better results for the crisis and the crash period. The results on the regressor *Model*, which indicates the parametric richness of the underlying model, show that higher model parametrization decreases the number of violations during turbulent times and this effect intensifies during the crash periods. Although they suffer from a large estimation risk, they outperform the low parameterized models, which slowly adapt to extreme information arrival.

Contrary to the model choice, the asset choice does matter for the calm evaluation period, but it has no contribution during turbulent times. Investing in large-cap stocks may lead to higher violation rates during calm periods, but this effect totally disappears in difficult times. Although not significant, the negative sign on the regressor *Market Cap* in column 3 of Table 4 indicates to a certain degree that during very turbulent times, investing in small cap stocks might have been less risky than investing in portfolios of large cap stocks.[3]

[3] In a pervious analysis for single stocks rather than portfolios this effect turned out to be even more pronounced.

Regardless of the evaluation period, the impact of the sampling window on the percentage rate of violations is not significant at 5% level. However, this result should be carefully discussed: while the nonparametric methods (e.g. FHS) manage to generate "green zone"-type violations based on short samples only for certain type of stocks, the EVT distribution estimated on larger historical datasets correctly forecast the risk for *all* types of stocks.

Table 5: Results of OLS Meta Regressions. Dependent variable: ln(p-value of conditional coverage test). Explanatory variables: *Market Cap*, which equals 0, if the underlying asset is the small-cap index; 1 if middle-cap index; and 2 if large-cap index; *Model*, which equals 0, if there is no parametric model in the VaR estimation; 1, if the model is RM-fix; 2, if RM-est.; 3, if ARMA-GARCH; and 4 if ARMA-FIGARCH; *Distr.*, which equals 0, if the underlying distribution is HS; 1 if FHS; 2 if normal distribution; 3 if Student-t distribution; and 4, if EVT; *Years*, which equals 0, if the estimation sample starts in 2005; 1, if the estimation sample starts in 2001; 2, if the estimation sample starts in 1996; 3, if the estimation sample starts in 1987. *** significant at 1%; ** at 5%; * at 10%. Total number of observations: 120.

Variables	**Calm Period**	**Crisis Period**	**Crash Period**
Constant	-0.934***	-31.158***	-28.398***
Market Cap	-0.347***	-0.193	0.139
Model	0.135	2.879***	3.465***
Distr.	-0.180	7.555***	7.025***
Years	0.049	-0.774	-1.175
$Adj.R^2$	0.051	0.390	0.438

The results of Table 5 show that during calm periods, the hits stemming from investing in small-cap stocks exhibit less dynamical correlation than the ones of other stock types. This might be explained by the fact that the returns of stocks with less capitalization show less dynamical dependency than the ones of large cap stocks. Moreover, besides reducing the number of violations during turbulent periods, richer parametrization also contributes to the elimination of time dependencies among exceedances. This result provides more evidence of the importance of using highly parameterized risk models to forecast losses in the periods when they are needed most, such as the recent financial crisis.

An interesting result reported by both meta regressions regards the values of the adjusted R^2 before and during the recent financial crisis. Its small value during calm periods indicates that none of the model specification and implementation factors plays a significant role on improving the VaR performance. However, during turbulent times, these factors reveal a major explanatory power in the variation of backtest results, which emphasizes the importance of our analysis carried out so far.

Besides the contribution to the performance of EVT and Student-t distribution during crisis, further evidence of the importance of using large samples when estimating the tail of return distribution is given by results from estimating the degrees of freedom (d.f) of the Student-t distribution throughout the whole evaluation period. As Figure B.1 in the Appendix shows, the d.f estimated on samples from 1987 or 1996 exhibit a steady 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. seem to be very volatile and thus more flexible to the market conditions.

The trade off between using large samples and estimating stable parameters and using recent samples and estimating parameters which easily adapt to the market conditions translates into following: financial institutions using stable d.f. specific to fat-tailed distributions, follow strategies of holding constant and large reserves for a large period. Contrary, others using distributions with d.f. estimated on recent sample, 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 that if everybody follows the same strategy, on the days when the VaR measures signal the need of larger reserves, the banks face huge liquidity problems.

This last scenario seems to be more realistic in illustrating what happened during the recent financial crisis. According to our findings, if banks used to apply recent data to estimate their VaR models, they obtained flexible d.f estimators, which in general adapt fast to the new market conditions. Thus, when the crisis started, the estimators also started to signal higher risks, which require higher reserves. Unfortunately, given the high correlation among the financial institutions, the signal simultaneously spread and led to a liquidity trap: banks faced difficulties to borrow from other banks with similar strategy and thus encountered large losses. Interestingly, the results of our study show that this effect could have been mitigated by simply using large estimation samples starting in 1996 or earlier together with fat-tailed or extreme value distributional assumptions.

4 Combining VaR Forecasts

As the results from Section 3 show, the performance of individual VaR estimates is in general instable with respect to model, estimation window or asset choice. Therefore, in the following we propose two new methods of obtaining more robust VaR out of sample forecasts, which are based on an adapting mechanism of optimally combining VaR forecasts 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 straightforward. For ex-ante forecasts, the superiority of forecast combinations is not compelling because they involve additional parameter estimation.

Consider S' rolling samples of two alternative conditional VaR predictions for the periods $\tau+1,\dots\tau+S'$, respectively, with $0<\tau\leq T$ and let a combined VaR prediction for the out of sample period $\tau+s$, be given by the linear combination of the two one-period out-of-sample VaR predictions based on models 1 and 2:

$$\hat{V}aR_{\tau+s}(\lambda_\tau)=\lambda_{\tau,0}+\lambda_{\tau,1}\hat{V}aR^1_{\tau+s}+\lambda_{\tau,2}\hat{V}aR^2_{\tau+s},\quad s=1,\dots,S' \tag{4.1}$$

where $\lambda_\tau=(\lambda_{\tau,0},\lambda_{\tau,1},\lambda_{\tau,2})'$ is the vector of loading factors and $\hat{V}aR^j_{\tau+s}$ is the one-period ahead VaR prediction from model $j=\{1,2\}$.

For example, let the VaR prediction of model 1 be based on a standard ARMA-GARCH approach with normally distributed errors and the VaR prediction of 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 combinations of the quantiles of the standard normal, $\hat{F}_1^{-1}(p)$, and estimated quantile from the distribution of the standardized residuals, $\hat{F}_2^{-1}(p)$:

$$\hat{V}aR_{\tau+s}(\lambda_\tau) = \lambda_{\tau,0} + \big\{\lambda_{\tau,1} + \lambda_{\tau,2}\big\}\hat{\mu}_{\tau+s|\tau+s-1} + \big\{\lambda_{\tau,1}\hat{F}_1^{-1}(p) + \lambda_{\tau,2}\hat{F}_2^{-1}(p)\big\}\hat{\sigma}_{\tau+s|\tau+s-1}. \tag{4.2}$$

Although Equation (4.1) 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 in the optimal combination, while the sign indicates the direction. Thus Equation (4.1) becomes:

$$\hat{V}aR_{\tau+s}(\lambda_\tau) = \lambda_{\tau,0} + \lambda_{\tau,1}\hat{V}aR^1_{\tau+s} + (1 - \lambda_{\tau,1})\hat{V}aR^2_{\tau+s}, \quad s = 1, \dots, S' \tag{4.3}$$

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 Equation (4.3) optimally: by assessing the performance of the resulting risk estimator over the evaluation period starting at $\tau + 1$ and ending at $\tau + S'$ or by sequentially assessing the daily performance of $\hat{V}aR_{\tau+s}$, at each $s = 1, \dots, S'$.[4] The first method, which we call Conditional Coverage Optimization

[4]Other objective functions based on monetary rather 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.

Method (CCOM) consists in finding the optimal combination of VaR predictions, which maximizes the conditional coverage rate, while the second method, called Conditional Quantile Optimization Method (CQOM) finds the optimal loadings by minimizing the distance between the population quantiles and the VaR's combinations.

In what follows we present the combination methods and their empirical performance on correctly estimating the failure probability of 1%.

4.1 Conditional Coverage Optimization Method (CCOM)

The CCMO 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 Basle II criteria. It consists of computing a sequence of S' binary exceedance indicators $H_{\tau+s}(\lambda_\tau)$ based on the combined estimator $\hat{V}aR_{\tau+s}(\lambda_\tau)$ for any arbitrary vector of weighting parameter:

$$\underline{H}(\lambda_\tau) = \{H_{\tau+1}(\lambda_\tau), H_{\tau+2}(\lambda_\tau), \ldots, H_{\tau+S'}(\lambda_\tau)\}$$

where $H_{\tau+s}(\lambda_\tau)$ is given by

$$H_{\tau+s}(\lambda_\tau) = \mathbb{1}\left(r_{\tau+s} < \hat{V}aR_{\tau+s}(\lambda_\tau)\right), \quad s = 1, 2, \ldots S'.$$

and $\hat{V}aR_{\tau+s}(\lambda_\tau)$ is given in Equation (4.3). The optimal weights $\lambda^*_\tau = (\lambda^*_{\tau,0}, \lambda^*_{\tau,1})$ minimize the conditional coverage test, which has the following two null hypotheses (see Section 2 for a detailed discussion):

1) $E[H_{\tau+s}(\lambda^*_\tau)] = p$, which assures the unconditional coverage and

2) $E[H_{\tau+s}(\lambda^*_\tau)|H_{\tau+s-1}(\lambda^*_\tau)] = E[H_{\tau+s}(\lambda^*_\tau)|(1 - H_{\tau+s-1}(\lambda^*_\tau))] = E[H_{\tau+s}(\lambda^*_\tau)]$, 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. In addition, the optimal combination of VaR's aim at eliminating the violation clustering, which, if it is ignored, increases the probability of having a violation after an occurrence. Estimates of the two parameters $\lambda^*_{\tau,0}$ and $\lambda^*_{\tau,1}$, can be obtained by the Method of Moment approach using the following moment restrictions resulting from the martingale property of H_t:

1) $E[H_{\tau+s}(\lambda^*_\tau) - p] = 0$

2) $E[(H_{\tau+s}(\lambda^*_\tau) - p)H_{\tau+s-1}(\lambda^*_\tau)] = 0$

We can opt for further moments, such as $E[(H_{\tau+s}(\lambda^*_\tau) - p)VaR^j_{\tau+s}] = 0$, $j = 1, 2$ or $E[(H_{\tau+s}(\lambda^*_\tau) - p)r_{\tau+s-1}] = 0$ which may increase the efficiency of the estimated parameters (see Giacomini and Komunjer (2005)). Thus, the MM estimator $\hat{\lambda}^*_\tau$ is the solution to the following minimization problem:

$$\hat{\lambda}^*_\tau = \arg\min \qquad \psi(\lambda_\tau, X_{\tau+s})'\psi(\lambda_\tau, X_{\tau+s}), \tag{4.4}$$

where $\psi(\lambda_\tau, X_{\tau+s}) = (\psi^1(\lambda_\tau, X_{\tau+s}), \psi^2(\lambda_\tau, X_{\tau+s}, ...))'$ with

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

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

and $X_{\tau+s} = (\hat{V}aR^1_{\tau+s}, \hat{V}aR^2_{\tau+s}, r_{\tau+s})'$. In the case of further moments, the estimation given in Equation (4.4) is weighted by $\hat{V}^{-1}$, where $\hat{V}$ is a consistent estimator of the asymptotic variance $V = E[\psi(\lambda_\tau, X_{\tau+s})\psi(\lambda_\tau, X_{\tau+s})']$.

Although straightforward in principle, the estimation method suffers from a set 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 smoothness degree of the objective function (Goffe et al., 1994). Furthermore, we increase the degree of smoothness of the function to be minimized in Equation (4.4), by replacing $H_{\tau+s}(\lambda_\tau)$ with $F(\lambda_\tau, X_{\tau+s}, h_T) = 1/(1 + exp[(r_{\tau+s} - \lambda_{\tau,0} - \lambda_{\tau,1}\hat{V}aR^1_{\tau+s} - (1-\lambda_{\tau,1})\hat{V}aR^2_{\tau+s})/h_T])$ such that, for $s \to \infty$, $h_T \to 0$ and $sh_t \to \infty$, $\sum_{s=1}^{S'} \frac{F(\lambda_\tau, X_{\tau+s}, h_T)}{S'} \xrightarrow{p} E[H_{\tau+s}(\lambda_\tau)]$. The heteroscedasticity-and autocorrelation-robust estimator of Newey and West (1987) for the asymptotic variance-covariance matrix V should be applied.

4.2 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 $\hat{V}aR^j_{\tau+s}$, with $j = \{1, 2\}$:

$$Q_p(r_{\tau+s}) = \lambda_{\tau,0} + \lambda_{\tau,1}\hat{V}aR^1_{\tau+s} + (1-\lambda_{\tau,1})\hat{V}aR^2_{\tau+s} \equiv \hat{V}aR'_{12,\tau+s}\lambda_\tau, \tag{4.5}$$

where $\hat{V}aR_{12,\tau+s} = (1, \hat{V}aR^1_{\tau+s}, \hat{V}aR^2_{\tau+s})'$ and $\lambda_\tau = (\lambda_{\tau,0}, \lambda_{\tau,1})'$. In our framework, the optimal λ_τ is given by the solution to the following minimization problem:

$$\begin{aligned} \hat{\lambda}_\tau &= \arg\min_{\lambda_\tau} \Bigg\{ \sum_{r_{\tau+s} \geq \hat{V}aR'_{12,\tau+s}\lambda_\tau} p|r_{\tau+s} - \hat{V}aR'_{12,\tau+s}\lambda_\tau| \\ &+ \sum_{r_{\tau+s} < \hat{V}aR'_{12,\tau+s}\lambda_\tau} (1-p)|r_{\tau+s} - \hat{V}aR'_{12,\tau+s}\lambda_\tau| \Bigg\} \end{aligned}$$

The main advantage of the quantile regression approach is that it requires no explicit distributional assumptions for the return series. Given the difficulties in estimating the standard errors of quantile regression estimators, we apply the moving block bootstrap procedure as recommended by Fitzenberger (1997).

4.3 Empirical Results

In the following we present results from assessing the performance of the two methods in terms of Basel II and independence criteria. We make the assessment at two different stages:

(1) First, 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 model with $S' = 143, 510, 129$ for the first, second and respectively third evaluation period.

(2) Second, we evaluate the optimal combination of VaR's in a recursive manner: we divide the entire evaluation period of T observations (from January 1^{st}, 2007 to July 2^{nd}, 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 τ_1 observations, the second "in-sample" period contains the first $\tau_1 + 1$ and the last one contains $T - 1$ observations. $S' = T - \tau_1$.

Through these two evaluation exercises, we aim at assessing the ability of optimal combinations to improve the performance within the sample (first assessment), but also out-of-sample (second assessment) of single 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 1). Moreover, we present here results 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 6 and 7 for each of the optimization method.

The first group of results (Part A of both tables 6 and 7) present the performance of optimally combining VaR estimates based on different distributional assumptions, given a certain pre-filtering method. It illustrates the forecasting power of linear quantile combination, for the same conditional mean and conditional variance specification (see e.g. Equation (4.2)). 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 sample starting in 1987 to ARMA-FIGARCH estimated on sample starting in 2005. The results from both tables show that, independent of the pre-filtering and combination method, optimal combinations between normal and Student-t distribution quantiles reduce significantly the failure rates and their time dependencies, with very small differences between the two methods. While the CQOM method (Table 7) delivers only "green zone"-type violations with all p-values of conditional coverage tests larger than 0.1, the CCOM method (Table 6) reveals a few exceptions, especially when combining VaR estimates based on recent data.

Similar good results are obtained when combining two VaR estimates with different conditional mean and conditional variance estimations: Part B and C of both tables 6 and 7 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 involves large estimation risk: ARMA-GARCH (Part B) and ARMA-FIGARCH (Part C).

Again, combining VaR estimates based on recent data seems to be inadequate in forecasting eventual losses. However, combining estimates based on recent and old data (Part D), increases the performance of individual estimates in all cases when applying the CQOM method (Table 7) and with a few exceptions when applying the CCOM method (Table 6).

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 to assess the ability of these methods in providing combinations of VaR which perform well during "out-of-sample" periods, especially during the recent financial crisis.

Table 8 reports the backtest results from the second stage assessment. Because our focus is on forecasting the risks during the crisis and crash period, we choose July 17^{th}, 2008 to be the end date of the first "in-sample" period, which marks one year since the beginning of the financial crisis ($\tau_1 = 250$). Thus all in-sample data entail at least one year of crisis and the out-of-sample period comprises the crash phase.

The results of Table 8 complete the previous results and show that combining VaR estimates performs well not just within sample, but also out of sample when forecasting financial losses. In general the VaR predictors based on both combination methods considerably improve the performance of single estimates and remain robust with respect to the asset or model choice. They produce in most of the cases 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 combining 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[5].

[5]Similar graphs stemming from other combinations are available from the authors upon request.

Table 6: Percentage rate of violations based on CCOM, Assessment Stage 1. Percentage rate of violations at p = 1%. ** refers to p-values of conditional coverage test smaller than 0.05, * refers to p-values between and 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 Small	Crisis Period Middle	Crisis Period Large	Crash Period Small	Crash Period Middle	Crash Period 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
T	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
T	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
T	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 7: Percentage rate of violations based on CQOM, Assessment Stage 1. Percentage rate of violations at p = 1%. ** refers to p-values of conditional coverage test smaller than 0.05, * refers to p-values between and 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.

		Crisis Period			Crash Period		
Comb.	Stock Type	Small	Middle	Large	Small	Middle	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
T	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
T	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

Table 8: Percentage rate of violations, Assessment Stage 2. Percentage rate of violations at p = 1%. ** refers to p-values of conditional coverage test smaller than 0.05, * refers to p-values between and 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*
T	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
T	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*
T	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*

Except for a short inherently volatile phase, starting around the events from September 15^{th}, 2008 (45^{th} observation) and lasting until December 1^{st}, 2008, all sequences of estimates stemming from CQOM (see figures B.2 and B.3), exhibit a stable behavior during the whole crisis period, which emphasizes the robustness of this new method. The CCOM weights exhibit a more volatile behavior (see figures B.4 and B.5)), which indicates that estimators stemming from unconditional-type methods are less robust to market changes than the ones based on conditional approaches, such as CQOM.

5 Conclusions

In this paper we provide empirical evidence on the riskiness of the VaR. Given its performance during the recent financial crisis, we feel that there is more need of taking stock and communicating its potential shortcomings before adding another VaR approach to the literature.

Our goal in this paper is to study the robustness of the standard VaR with respect to different risk factors when applied to forecast losses in periods when it is needed most, such as the current financial crisis. By means of a meta-study approach, we show that the performance of VaR estimates differs across type of stocks subject to model choice, distributional assumptions or the choice of estimation window. Thus we show that popular VaR measures manage to accurately forecast the risk in calm periods or when applied to large-cap stocks, but exhibit a relatively low power when applied in crisis times or to stocks with lower capitalization. Higher parametric approaches (e.g. GARCH) which account for past shocks (e.g. market crash from 1987) within extreme value distributional settings lead to accurate loss forecasts according to Basel II rules during the recent financial crisis.

In order to improve the performance of standard risk measures, we propose a data-driven methodology of accurately estimating VaR based on the principle of forecast combination.

The optimal loadings of VaR measures are driven by the maximization of conditional coverage rates (CCOM) or by the minimization of the distance between the population quantiles and the VaR's combinations (CQOM). Using an empirical example, we show that optimal combinations of VaR forecasts radically improve the performance of stand-alone estimates during the recent crisis. All combinations exhibit very good "out-of-sample" performance by generating independent exceedances within the limits imposed by Basel II rules.

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 taking 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 market liquidity shortages.

References

Baillie, R., T. Bollerslev, and H. Mikkelsen (1996): "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper (1999): "VaR without Correlations for Nonlinear Portfolios," *Journal of Futures Markets*, 19, 583 – 602.

Basel Committee (1996): "Overview over the Amendment to the Capital Accord to Incorporate Market Risks," Discussion paper, Basel Committee on Banking Supervision.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Christoffersen, P. F. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

——— (2003): *Elements of Financial Risk Management.* Elsevier Science, USA.

Christoffersen, P. F., and F. X. Diebold (2000): "How Relevant is Volatility Forecasting for Financial Risk Management?," *The Review of Economics and Statistics*, 82, 12–22.

Crouhy, M., D. Galai, and R. Mark (2001): *Risk Management.* McGraw - Hill, New York.

Fitzenberger, B. (1997): "The Moving Blocks Bootstrap and Robust Inference in Linear Regressions," *Journal of Econometrics*, 82, 235–287.

Gençay, R., and F. Selçuk (2004): "Extreme value theory and Value-at-Risk: Relative performance in emerging markets," *International Journal of Forecasting*, 20, 287–303.

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

Giot, P., and S. Laurent (2004): "Modelling daily Value-at-Risk using realized volatility and ARCH type models," *Journal of Empirical Finance*, 11, 379 – 398.

Goffe, W., G. D. Ferrier, and J. Rogers (1994): "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics*, 60, 65–100.

Jorion, P. (2007): *Value At Risk*. McGraw-Hill, New York, 3rd edn.

——— (2009): "Risk Management Lessons from the Credit Crisis," *European Financial Management*, forthcomming.

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

Kuester, K., S. Mitinik, and M. S. Paolella (2006): "Value-at-Risk Prediction: A Comparison of Alternative Strategies," *Journal of Financial Econometrics*, 4(1), 5389.

McAleer, M., and B. da Veiga (2008): "Forecasting value-at-risk with a parsimonious portfolio spillover GARCH (PS-GARCH) model," *Journal of Forecasting*, 27, 1–19.

McNeil, A. J., and R. Frey (2000): "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, 7, 271300.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pesaran, H., and A. Timmermann (2007): "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics*, 137(1), 134–161.

Roncalli, T. (2001): "La Théorie des Extreêmes et La Gestion des Risques de Marcheé," Cours ENSAI.

Summers, L. (1991): "The Scientific Illusion in Empirical Macroeconomics," *Scandinavian Journal of Economics*, 93(2), 129–148.

Timmermann, A. (2005): "Forecast Combinations," Working Paper.

Appendix A: Tables

Table A.9: Backtest results for the calm period: nonparametric methods. Percentage rate of violations for VaR at p = 1% for the period from January 1^{st}, 2007 to July 17^{th}, 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	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 A.10: Backtest results for the crisis period: nonparametric methods. Percentage rate of violations for VaR at p = 1% for the period from July 18^{th}, 2007 to July 2^{nd}, 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	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 A.11: Backtest results for the crash period: nonparametric methods. Percentage rate of violations for VaR at p = 1% for the period from September 1^{st}, 2008 to July 2^{nd}, 2009 (total of 129 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	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: Figures

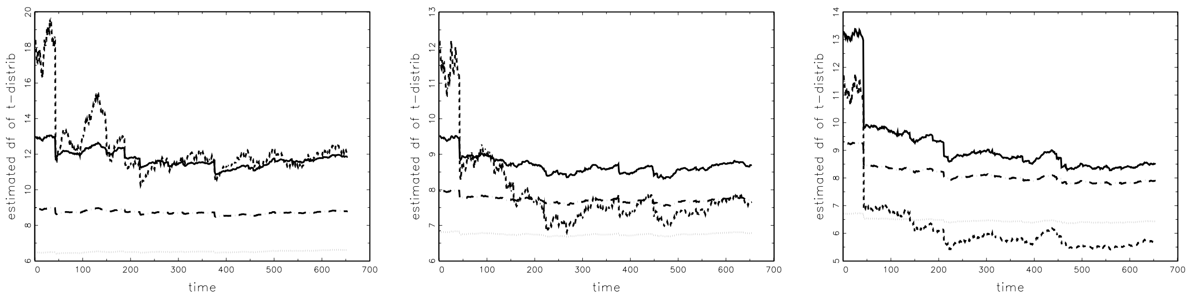

Figure B.1: Estimated d.f. of t-distribution. Left panel: small cap index, middle: middle cap index, right: large cap index; dotted lines correspond to d.f estimated on samples starting in 1987, dotted and dashed lines correspond to d.f estimated on samples starting in 1996, solid lines correspond to d.f estimated on samples starting in 2001 and dashed lines correspond to d.f estimated on samples starting in 2005.

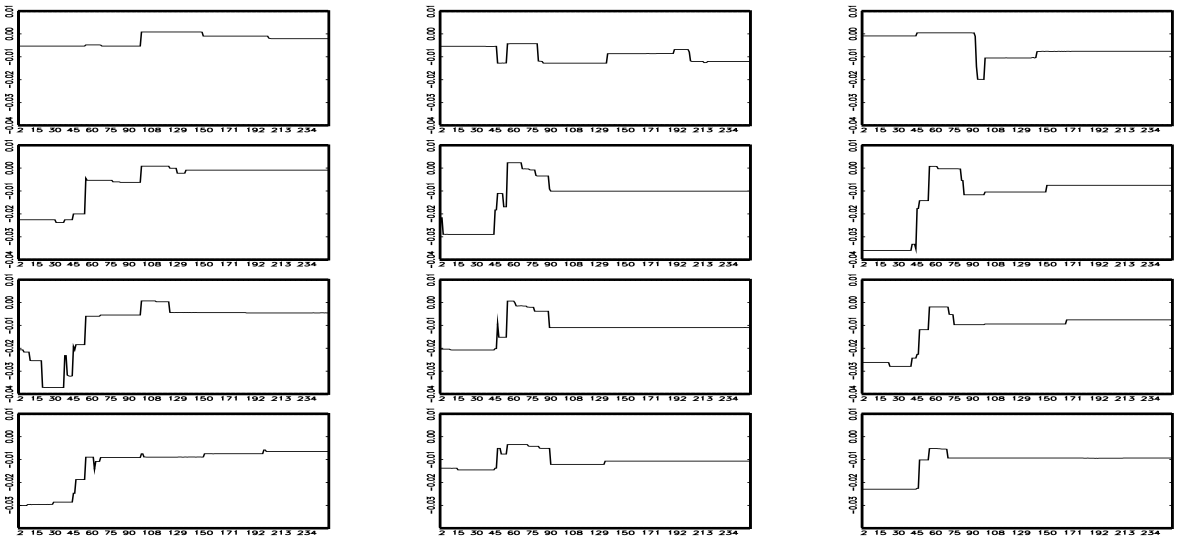

Figure B.2: Estimated $\lambda^*_{\tau,0}$ through the crisis, CQOM, Part A: ARMA-GARCH. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.

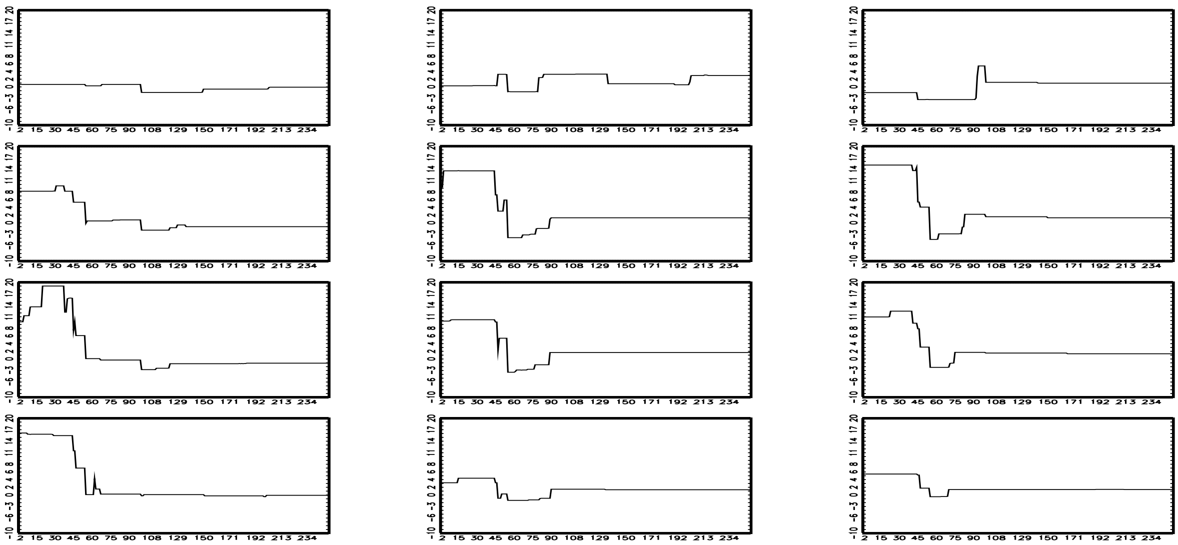

Figure B.3: Estimated $\lambda^*_{\tau,1}$ through the crisis, CQOM, Part A: ARMA-GARCH. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.

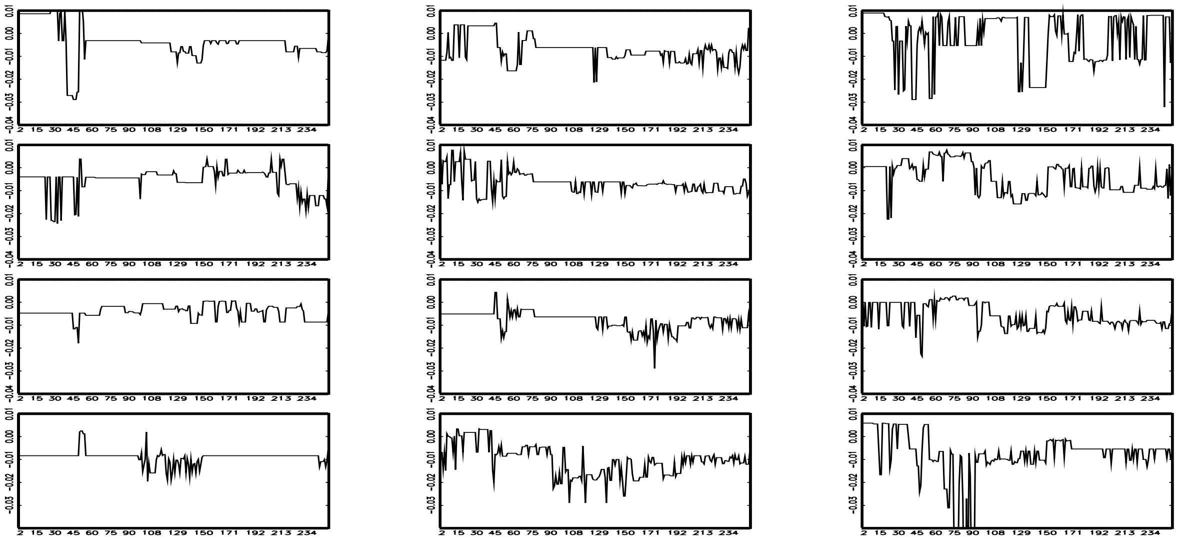

Figure B.4: Estimated $\lambda^*_{\tau,0}$ through the crisis, CCOM, Part A: ARMA-GARCH. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.

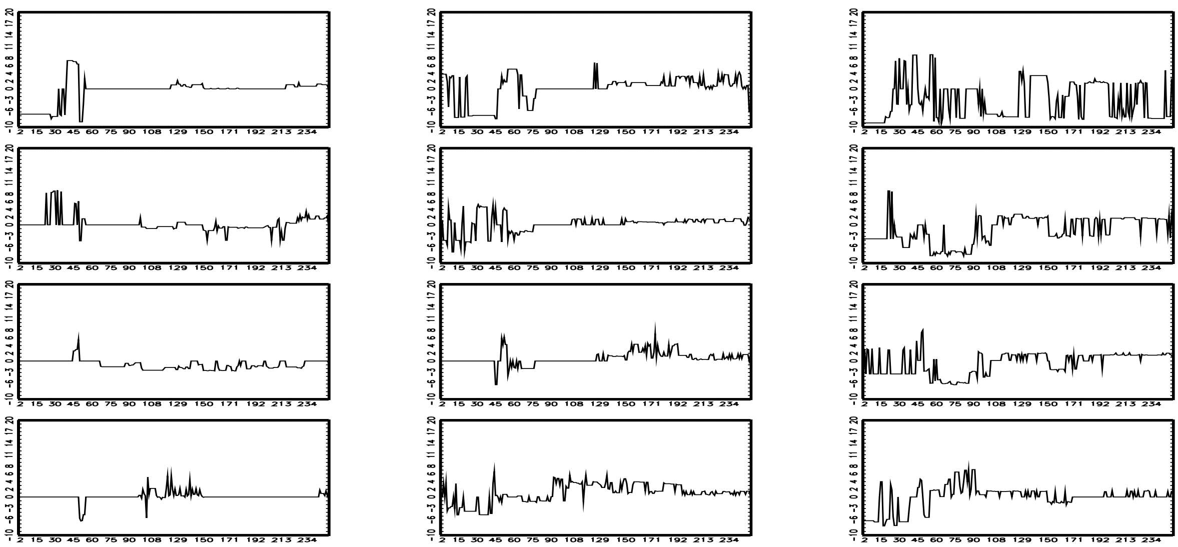

Figure B.5: Estimated $\lambda^*_{\tau,1}$ through the crisis, CCOM, Part A: ARMA-GARCH. Left panels: small cap index, middle panels: middle cap index, right panels: large cap index; first row: sampling starting in 1987, second row: sampling starting in 1996, third row: sampling starting in 2001 and fourth row: sampling starting in 2005.