Are realized volatility models good candidates for alternative Value at Risk prediction strategies?

Louzis, Dimitrios P. and Xanthopoulos-Sisinis, Spyros and Refenes, Apostolos P.

Athens University of Economics and Business, Bank of Greece

18 April 2011

Are realized volatility models good candidates for alternative Value at Risk prediction strategies?

Dimitrios P. Louzis[a,b,*], Spyros Xanthopoulos-Sisinis[a] and Apostolos P. Refenes[a]

This version: April 2011

Abstract

In this paper, we assess the Value at Risk (VaR) prediction accuracy and efficiency of six ARCH-type models, six realized volatility models and two GARCH models augmented with realized volatility regressors. The a^{th} quantile of the innovation's distribution is estimated with the fully parametric method using either the normal or the skewed student distributions and also with the Filtered Historical Simulation (FHS), or the Extreme Value Theory (EVT) methods. Our analysis is based on two S&P 500 cash index out-of-sample forecasting periods, one of which covers exclusively the recent 2007-2009 financial crisis. Using an extensive array of statistical and regulatory risk management loss functions, we find that the realized volatility and the augmented GARCH models with the FHS or the EVT quantile estimation methods produce superior VaR forecasts and allow for more efficient regulatory capital allocations. The skewed student distribution is also an attractive alternative, especially during periods of high market volatility.

Keywords: High frequency intraday data; Filtered Historical Simulation; Extreme Value Theory; Value-at-Risk forecasting; Financial crisis

JEL classifications codes: C13; C53; C58; G17; G21; G32

[a] Financial Engineering Research Unit, Department of Management Science and Technology, Athens University of Economics and Business, 47A Evelpidon Str., 11362 Athens, Greece.

[b] Bank of Greece, Financial Stability Department, 3 Amerikis Str., 105 64, Athens, Greece.

[*] Corresponding Author: E-mail address: dlouzis@aueb.gr; dlouzis@bankofgreece.gr

1. Introduction

The recent 2007 – 2009 financial crisis demonstrated, if nothing else, that the financial institutions' risk management systems were not as adept as previously thought in tracking and anticipating the extreme price movements witnessed during that highly volatile period. Nearly all financial institutions recorded multiple consecutive exceptions, i.e. days in which the trading book losses exceeded the prescribed Value-at-Risk (VaR)[1]. In several instances, the total number of exceptions during the previous trading year exceeded the threshold of ten violations which is the set regulatory maximum (Campel and Chen, 2008)[2]. Consequently, much doubt was cast and many questions were raised about the reliability and accuracy of the implemented VaR models, systems and procedures.

However, the criticisms faced by the risk management departments can hardly be attributed to a lack of allocated resources or research efforts. VaR measurement and forecasting has been one of the most vigorously researched areas in quantitative risk management and financial econometrics. It has also enjoyed significant investments both in terms of capex and in human capital within banks and financial institutions. In this context, the evaluation of some recently proposed volatility models which make use of the informational content in high frequency data could reveal some attractive alternative VaR modelling specifications.

The foundations of modern risk management were laid with the seminal work of Engle (1982) who introduced the AutoRegressive Conditional Heteroscedasticity (ARCH) model for modeling the conditional heteroscedasticity in financial assets returns. Since then, a plethora of ARCH–type models have been proposed in the open literature (see Bollerslev, 2010 for a short description for almost all ARCH–type models) and most of them have been included in VaR studies. Giot and Laurent (2003a and 2003b) for example, showed that flexible ARCH specifications combined with fat tailed distributions can provide accurate VaR forecasts for a wide range of assets.

More recently, Andersen and Bollerslev (1998), Andersen et al. (2001a), Andersen et al. (2001b) and Barndorff-Nielsen and Shephard (2002) introduced and promoted the realized volatility as a non-parametric approach for measuring the unobserved volatility. In Andersen et

[1] Value-at-Risk is the most common measure of downside market risk and is widely adopted by both the financial services industry and the regulators. It reflects an asset's market value loss not be exceeded over a specified holding period, with a specified confidence level (see also Section 4).
[2] A. Campel and X. L. Chen are the authors of a VaR survey article in the "Risk" magazine on July 2008.

al. (2003), the authors also suggested that standard time series techniques can be used in order to model the "observable" realized volatility. These concrete theoretical foundations coupled with the increased availability of high quality intraday data for a wider range of assets, fuelled the research interest on the use of high frequency data for measuring and forecasting the volatility of financial assets. Several authors demonstrated the superiority of realized volatility models over ARCH models for volatility forecasting (see Koopman et al., 2005; Martens et al., 2009; Martens, 2002 among others), while Giot and Laurent (2004) first utilized high frequency intraday data in a VaR forecasting context.

In Table 1, a concise literature review on the use of intraday data for VaR modeling is presented. Nine out of ten studies therein perform a direct or indirect comparison between ARCH-type and realized volatility models (except from Clements et al., 2008 who considered only realized volatility models). The findings are mixed as five out of nine studies give evidence in favor of the use of high frequency data, while the remaining four provide evidence in favor of ARCH-type models. Almost all of the researchers implement a fully parametric approach for the estimation of the VaR quantiles, i.e. they adopt a specific distributional assumption (e.g. the normal or the skewed student distribution) for the innovation process. The use of alternative assumptions is quite limited (see Kuester et al., 2006 for an excellent review of alternative VaR methods). Finally, the VaR models are almost invariably evaluated in terms of the statistical accuracy of their VaR forecasts (implementing Christoffersen's, 1998 tests for example) and less so by their efficiency with respect to specific regulatory provisions.

In this study, we contribute to this growing literature by examining the day-ahead 10%, 5%, 1% and 0.5% VaR forecasting performance of fourteen volatility models for three different quantile estimation methods and under eight statistical and regulatory evaluation criteria. We implement six ARCH – type specifications that include short memory, long memory and asymmetric GARCH models, six realized volatility models including two AR(FI)MA and four of the latest Heterogeneous Autoregressive (HAR) models and finally, two augmented GARCH models that incorporate the realized volatility and the realized power variation introduced by Barndoff-Nielsen and Shephard (2004) as explanatory variables.

Moreover, for each volatility model we implement three VaR quantile estimation methods: the fully parametric method using both the normal and the skewed student distributions, the semi-parametric filtered historical simulation (FHS) method and the conditional extreme value theory (EVT) method. For the first time, the FHS and the EVT quantile estimation methods are combined with the realized volatility models using Giot and Laurent's (2004) two step procedure. Hence, we have in total fifty-six unique VaR models, each estimated for four VaR quantiles on an approximately thirteen years (from 1.1.1997 to 09.30.2009) of daily and intra-daily returns for the S&P 500 cash index. The eight years out-of-sample period includes the latest financial crisis, while we also repeat our analysis adjusting the out-of-sample period to cover exclusively the 2007-2009 period.

The performance of the alternative VaR models is assessed using an extensive and diverse range of evaluation measures. On top of the usual statistical accuracy tests (e.g. failure rates, conditional and unconditional coverage tests, dynamic quantile test and quadratic loss functions), we lay particular emphasis in implementing efficiency measures in the form of regulatory oriented loss functions, including the one implied by the Market Risk Amendment (MRA) to the Basel Accord and a loss function which considers the opportunity cost of capital. On the latter two measures we additionally run equal and superior predictive ability tests (Diebold and Mariano, 1995; Hansen, 2005) in order to identify the models that satisfy both the statistical accuracy and efficiency conditions set herein.

The remaining of the paper is organized as follows: In section 2 we present the volatility models used in this paper. In section 3 we describe the VaR methods while in section 4 the VaR evaluation measures are presented. The empirical analysis is presented in Section 5. Section 6 summarizes and concludes this article.

[Insert Table 1 about here]

2. Volatility modelling and forecasting

The daily conditional heteroskedastic logarithmic returns of a financial asset, $r_t = p(t) - p(t-1)$, where $p(t)$ is the logarithmic asset price observed at day t, can be described by the following process:

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \tag{1}$$

where $\mu_t = E\left(r_t \left| I_{t-1}\right.\right)$ is the conditional mean (I_{t-1} is the available information available until $t-1$), $\sigma_t^2 = Var\left(r_t \left| I_{t-1}\right.\right)$ is the conditional variance of the return process and z_t is a zero mean unit variance independently and identically distributed (i.i.d.) process. In order to account for the often inherent serial autocorrelation in the financial assets returns process, we model the latter using an $AR(k)$ specification:

$$r_t = c + \phi(L) r_t + \varepsilon_t \tag{2}$$

where L is the lag operator $(Ly_t = y_{t-1})$ and $\phi(L) = \phi_1 L + \phi_2 L^2 + \ldots + \phi_k L^k$ is a polynomial of order *k*. Here, we fit an AR(1) model as in Giot and Laurent (2004) and Kuester et al. (2006). The conditional variance of the returns process can be modeled with one of the many volatility models that have been proposed in the open literature. In this paper, we adopt fourteen volatility models from three broad model classes: (i) ARCH – type models, (ii) Realized Volatility models and (iii) Augmented GARCH models. The volatility models implemented here are briefly presented in the following subsections.

2.1. ARCH – type models

In his seminal paper, Engle (1982) proposed the AutoRegressive Conditional Heteroscedasticity (ARCH) model as a feasible approach for modeling the conditional heteroscedasticity in financial assets return series. Since then, the ARCH based literature has being growing fast (see for example the "Glossary to ARCH (GARCH)" by Bollerslev, 2010), encompassing today a plethora of generalizations and extensions of the original ARCH model. The implementation of (G)ARCH based models in financial asset price volatility forecasting applications is considered today common practice among professionals and a benchmark in academic research. The ARCH models employed here fall into three broad categories: (i) symmetric GARCH models:

– The exponentially weighted moving average (EWMA) model:

$$\sigma_t^2 = (1-\lambda)\sum_{i=1}^{\infty}\lambda^{i-1}\varepsilon_{t-i}^2 = (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2 \quad (3)$$

where $\lambda = 0.94$ as in RiskMetrics^{TM} (JP Morgan, 1996) and Giot and Laurent (2004). The value of λ determines the persistence in the volatility process.

– The Bollerslev (1986) GARCH(1,1) model:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (4)$$

with $\omega > 0$, $\alpha, \beta \geq 0$ and $\alpha + \beta < 1$

(ii) asymmetric GARCH models that capture the asymmetric impact[3] of market news on the volatility process:

– The Glosten et al. (1993) GJR-GARCH, or in short GJR(1,1), model:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma I(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (5)$$

where $\mathrm{I}(\cdot)$ is an indicator function. For $\gamma > 0$ the impact of past negative returns on the conditional variance is greater than the impact of past positive returns.

– The Nelson (1991) Exponential GARCH, or in short EGARCH(1,1), model:

$$\log(\sigma_t^2) = \omega + \alpha(|z_{t-1}| - E(|z_{t-1}|)) + \gamma z_{t-1} + \beta\log(\sigma_{t-1}^2) \quad (6)$$

[3] The asymmetric impact of bad (good) news, or equivalently of negative (positive) returns on the volatility process has been well documented in empirical studies (Bekaert and Wu, 2000; Engle and Ng, 1993; Glosten, Jagannathan and Runkle, 1993; Nelson, 1991; Wu, 2001; Zakoian, 1994): negative shocks tend to increase volatility more than positive shocks due to leverage effects (see Black, 1976), or volatility feedback (e.g. see Bekaert and Wu, 2000; Campbell and Hentchel, 1992; French, Schwert and Stambaugh, 1987; Pindyc, 1984; Wu 2001).

where $z_t \equiv \varepsilon_t / \sigma_t$ are the standardized errors. The term $\alpha\left(\left|z_{t-1}\right| - E\left(\left|z_{t-1}\right|\right)\right)$ is referred to as the "size effect" of past shocks while the term γz_{t-1} is referred to as the "sign effect" of past shocks on current conditional variance. For $\gamma < 0$ a negative surprise would generate more volatility than a same magnitude positive one.

– The Asymmetric Power ARCH (APARCH) model proposed by Ding et al. (1993). The APARCH(1,1) is defined as:

$$\sigma_t^{\delta} = \omega + \alpha\left(\left|\varepsilon_{t-1}\right| - \gamma\varepsilon_{t-1}\right)^{\delta} + \beta\sigma_{t-1}^{\delta} \tag{7}$$

where δ $(\delta > 0)$ is the Box-Cox transformation of the conditional standard deviation, while the parameter γ, with $-1 < \gamma < 1$, captures the leverage effects.

And finally, (iii) long memory GARCH models:[4]

– Bailie et al. (1996) proposed the Fractionally Integrated GARCH (FIGARCH) model. The FIGARCH(1,*d*,1) is defined as:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \left[1 - \beta L - \left(1 - \alpha L - \beta L\right)\left(1 - L\right)^d\right]\varepsilon_t^2 \tag{8}$$

where $\omega > 0$, $\alpha + \beta < 1$ and d is the long memory, or fractional differencing parameter which captures the long range dependence in the conditional variance. For values of the differencing parameter d between 0 and 1, the shock to the volatility process exhibits a slow hyperbolic rate of decay. As the term $\left(1 - L\right)^d$ in (8) is an infinite summation, the FIGARCH obtains an infinite order specification which in practice is truncated at 1000 lags, as suggested in Baillie et al. (1996).

[4] In a short memory GARCH model, a shock to the volatility process would die out at a fast exponential rate. Nonetheless, many authors (see for example Andersen and Bollerslev, 1997; Ding et al., 1993) have argued that the impact of shocks on market volatility could persist for longer periods of time, before eventualy dying out.

2.2. Realized Volatility (RV) models

In this study, we implement the realized volatility models for Value at Risk forecasting following the two step procedure proposed in Giot and Laurent (2004) and later used in Angelidis and Degiannakis (2008) and in Shao et al. (2009). In the first step, the realized volatility of the return series is modelled using either standard time series AR(FI)MA models, or the recently proposed Heterogeneous Autoregressive (HAR) model (see Corsi, 2009; Andersen et al., 2007) and some of its extensions. In the second step, the dynamics of the conditional realized volatility are taken into account in the return process described in equation (1).

We model realized volatility as in Andersen et al. (2007) where the logarithmic asset price is assumed to follow a continuous time jump diffusion semi-martingale process of the form:

$$\mathrm{d}p(t) = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t) + \kappa(t)\mathrm{d}q(t), \; 0 \le t \le T \tag{9}$$

where $\mu(t)$ is a continuous and locally bounded (finite) variation process, $\sigma(t)$ is the strictly positive stochastic volatility process, $W(t)$ is a standard Brownian motion, $\kappa(t)$ is the jump size and $\mathrm{d}q(t)$ is the jump counting process which takes the value of one in the case of a jump and zero otherwise. The resulting one-period cumulative return is defined as:

$$r(t) = p(t) - p(t-1) = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s) + \sum_{t-1 \le s < t} \kappa(s) \tag{10}$$

while its corresponding quadratic variation is given by:

$$QV_t = \int_{t-1}^{t} \sigma^2(s)ds + \sum_{t-1 < s \le t} \kappa^2(s) \tag{11}$$

The first part of the summation in equation (11) is the continuous path component, or integrated variance (IV_t) and the second part is the sum of squared jumps. Note that in the absence of the discrete jump path component $QV_t \equiv IV_t$.

The quadratic variation can be consistently estimated as the sum of intraday squared returns, the so called realized volatility (RV) (see Andersen and Bollerslev, 1998; Andersen et al., 2001a; Barndorff-Nielsen and Shephard, 2002).[5] Here, if M is the total number of intraday returns for each day, we define the j^{th} continuous compounded intraday return of day t, as $r_{t,j} = p\left(t-1+\frac{j}{M}\right) - p\left(t-1+\frac{(j-1)}{M}\right)$, with $j = 1,...,M$. Hence, the realized volatility for day t is given by:

$$RV_t = \sum_{j=1}^{M}\left(r_{t,j}\right)^2 \tag{12}$$

Since the close-to-open price levels are often in practice quite different and the overnight returns could bias the realized variance estimation, we scale the realized volatility calculated in equation (12) as follows: $RV_t = \left[\left(\sigma_{oc}^2 + \sigma_{co}^2\right)/\sigma_{oc}^2\right]\sum_{j=1}^{M}\left(r_{t,j}\right)^2$, where σ_{oc}^2 and σ_{co}^2 are the "open-to-close" and "close-to-open" sample variances respectively (see Martens, 2002; Koopman et al., 2005). As $M \to \infty$, the realized volatility converges in probability to the quadratic variation which is "by construction … a prime candidate for formal volatility measure" (Andersen et al., 2006, p. 830) and hence, the daily unobservable volatility can be consistently estimated by realized volatility. The latter can be treated as an observable variable and thus standard time series techniques can be applied for modeling and forecasting purposes.

Barndorff-Nielsen and Shephard (2004) generalized the quadratic variation process to the power variation process by defining the integrated power variation of order p as:

$$IPV_t\left(p\right) = \int_{t-1}^{t}\sigma^p(s)ds\,,\ 0 < p \leq 2 \tag{13}$$

By cumulating absolute intraday returns raised to the p^{th} power, the authors defined the realized power variation (RPV) of order p as:

[5] The sum of squared intraday returns is actually the realized variance. Realized volatility is defined as the square root of realized variance. However, the term realized volatility is used interchangeably with the term realized variance.

$$RPV_t(p) = \mu_z^{-1} M^{-(1-p/2)} \sum_{j=1}^{M} \left| r_{t,j} \right|^p \tag{14}$$

where $0 < p < 2$ and $\mu_z = E|z|^p = 2^{p/2}\Gamma\left(\frac{1}{2}(p+1)\right)/\Gamma\left(\frac{1}{2}\right)$ with $z \sim N(0,1)$. For values of p between 0 and 2 and as $M \to \infty$, it holds that: $RPV_t(p) \xrightarrow{p} IPV_t$. Since absolute intraday returns are less sensitive to large price movements and mitigate the impact of outliers, it has been shown that as long as $p \in (0,2)$, the realized power variation is robust to jumps. Note that when $p = 2$, the realized power variation reduces to the realized volatility as defined in equation (12), i.e. $RPV_t(2) \equiv RV_t$ [6].

The appealing properties of the realized power variation have encouraged its use in volatility forecasting applications. In particular, Forsberg and Ghysels (2007), Ghysels et al. (2006) and Ghysels and Sinko (2006) demonstrated the ability of realized absolute variation, i.e. $RPV_t(1)$, to produce superior volatility forecasts compared to the squared return volatility measures. They argued that the realized power variation is a better predictor of realized volatility because of its robustness to jumps, its smaller sampling error and its improved predictability. In Liu and Maheu (2009) and Fuertes et al. (2009), the authors showed that an $RPV_t(\cdot)$ of order other than one can significantly improve the accuracy of volatility forecasts. Next, the realized volatility models employed in this study are briefly presented.

2.2.1. AR(FI)MA models for realized volatility

In Andersen et al. (2001a) and Andersen et al. (2003), the authors proposed a long memory Autoregressive Fractionally Integrated Moving Average (ARFIMA) model in order to capture the long range dependence in the realized volatility process. They also showed that the logarithm of realized volatility is approximately normally distributed. This implies that one could model the logarithmic realized volatility instead of the realized volatility itself and conveniently assume

[6] In this case, the realized power variation is not robust to jumps and converges to the integrated volatility plus the jump component.

that the models' errors are normally distributed. At the same time, the positivity of conditional realized volatility estimates is reassured without imposing any nonegativity constraints on the model's parameters. The $\text{ARFIMA}\left(1,d_{RV},1\right)$ model for the logarithm of realized volatility, $lrv_t^{(d)}=\log\left(RV_t\right)$, in terms of deviations from the mean μ, is defined as:

$$\left(1-\psi_1 L\right)\left(1-L\right)^{d_{RV}}\left(lrv_t^{(d)}-\mu\right) = \left(1+\delta_1 L\right)u_t \tag{15}$$

where d_{RV} is the fractionally differencing parameter and u_t are the normally distributed errors as $N\left(0,\sigma_u^2\right)$. ARFIMA models have been extensively employed in volatility forecasting (e.g. see Andersen et al., 2003; Pong et al., 2004; Koopman et al., 2005) and VaR forecasting applications (e.g. see Giot and Laurent, 2004; Beltratti and Morana, 2005; Angelidis and Degiannakis, 2008).

2.2.2. ARMA models for realized volatility

In order to examine whether a short memory implementation can provide accurate volatility forecasts, we also include in our analysis an ARMA(2,1) model as in Pong et al. (2004). The authors justified the suitability of an ARMA model for capturing the realized volatility process based on the findings of Gallant et al. (1999) and Alizadeh et al. (2002). Therein, it was shown that the sum of a two AR(1) processes could capture the persistent behavior of realized volatility and thus describe the evolution of the volatility process better than a single AR(1) process. The summation of two AR(1) processes is equivalently an ARMA(2,1) implementation (Granger and Newbold, 1976), given by:

$$\left(1-\psi_1 L-\psi_2 L^2\right)\left(lrv_t^{(d)}-\mu\right) = \left(1+\delta_1 L\right)u_t \tag{16}$$

2.2.3. Heterogeneous Autoregressive (HAR) models for realized volatility

Recently, Corsi (2009) proposed an approximate long memory realized volatility model, the Heterogeneous Autoregressive (HAR) model. In contrast to the AR(FI)MA models, the HAR model is based on the Heterogeneous Market Hypothesis of Muller et al. (1993) and the HARCH model of Muller et al. (1997) and it approximates the persistence in realized volatility by aggregating daily, weekly and monthly volatility components in an autoregressive structure.[7] The logarithmic version of the HAR-RV model is defined as:

$$lrv_t^{(d)} = a_0 + a_{(d)} lrv_{t-1}^{(d)} + a_{(w)} lrv_{t-1}^{(w)} + a_{(m)} lrv_{t-1}^{(m)} + u_t \tag{17}$$

where $lrv_t^{(d)}$ is the daily logarithmic realized variance and $lrv_t^{(h)} = (1/h)\left(lrv_t + lrv_{t,t-1} + lrv_{t,t-2} + \ldots + lrv_{t,t-h+1}\right)$ with $h = w = 5$ and $h = m = 22$ being the weekly and monthly volatility components respectively. The embedded long lag structure, equivalent to a restricted AR(22), is capable of reproducing the long memory behavior of realized volatility, while its simple autoregressive functional form requires no more than OLS for the estimation of its parameters.

Similarly, the HAR-RPV model is defined as:

$$lrv_t^{(d)} = a_0 + a_{(d)} lrpv_{t-1}^{(d)} + a_{(w)} lrpv_{t-1}^{(w)} + a_{(m)} lrpv_{t-1}^{(m)} + u_t \tag{18}$$

where $lrpv_t^{(d)} = \log\left(RPV_t\right)$ is the logarithm of the daily realized power variation and $lrpv_t^{(h)} = (1/h)\left(lrpv_t + lrpv_{t,t-1} + lrpv_{t,t-2} + \ldots + lrpv_{t,t-h+1}\right)$ with $h = w = 5$ and $h = m = 22$ being the weekly and monthly realized power variation components respectively. Here, following Liu and Maheu (2009), we use an $RPV_t(\cdot)$ of order 1.5 as a regressor.

In Corsi et al. (2008), the authors accounted for the time varying conditional heteroscedasticity of the normally distributed HAR errors, i.e. the so called “volatility of realized volatility” by implementing a GARCH error process and thus improving the model’s fitting and its predictive ability. The HAR-RV-GARCH model is given by:

[7] The Heterogeneous Market Hypothesis (Muller et al., 1993) states that market agents differ with respect to their investment horizon, risk aversion, degree of available information, institutional constraints, transaction costs, etc. This diversity is identified as the root cause of asset volatility as market agents aim to settle at different asset valuations according to their individual market view, preferences and expectations.

$$lrv_t^{(d)} = a_0 + a_{(d)} lrv_{t-1}^{(d)} + a_{(w)} lrv_{t-1}^{(w)} + a_{(m)} lrv_{t-1}^{(m)} + u_t \quad (19)$$

$$u_t = \sigma_{u,t} \upsilon_t \text{ and } \sigma_{u,t}^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{u,t-1}^2 \quad (20)$$

where $\upsilon_t \mid I_{t-1} \sim N(0,1)$ with I_{t-1} being the information available until $t-1$.

Extending the work in Corsi and Reno (2009), Louzis et al. (2010) proposed the Asymmetric HAR-RPV model allowing for heterogeneous leverage or asymmetric effects modeled as lagged standardized returns and absolute standardized returns (analogous to an EGARCH-type structure) occurring at distinct time horizons: daily, weekly and monthly. Moreover, in order to capture any remaining long range dependence in the volatility of realized volatility, they proposed a FIGARCH implementation for the conditional heteroscedasticity of the residuals, while at the same time utilizing the realized power variation as a regressor. Based on their proposal, we define the AHAR-RPV-GARCH model as:

$$\begin{aligned} lrv_t^{(d)} = {} & a_0 + a_{(d)} lrpv_{t-1}^{(d)} + a_{(w)} lrpv_{t-1}^{(w)} + a_{(m)} lrpv_{t-1}^{(m)} + \vartheta_{(d)} z_{t-1}^{(d)} + \vartheta_{(w)} z_{t-1}^{(w)} + \vartheta_{(m)} z_{t-1}^{(m)} + \\ & + \gamma_{(d)} \left| z_{t-1}^{(d)} \right| + \gamma_{(w)} \left| z_{t-1}^{(w)} \right| + \gamma_{(m)} \left| z_{t-1}^{(m)} \right| + u_t \end{aligned} \quad (21)$$

where $z_t^{(h)} = \sum_{i=1}^{h} r_{t-i+1} \Big/ \sqrt{\sum_{i=1}^{h} RV_{t-i+1}}$ are the daily ($h = d = 1$), weekly ($h = w = 5$) and monthly ($h = m = 22$) standardized returns, while the conditional variance of the errors, u_t, is modeled as in equation (20). The leverage effects are captured by the coefficient $\vartheta_{(\cdot)}$ which is expected to be negative and statistically different from zero, should past negative shocks yield a greater impact on future volatility.

2.2.4. Incorporating the conditional realized volatility into the return process

As previously mentioned, we use the two step procedure on the realized volatility estimates from the ARFIMA$(1, d_{RV}, 1)$, ARMA(2,1), HAR-RV, HAR-RPV, HAR-RV-GARCH and AHAR-RPV-GARCH models in order to integrate the conditional realized volatility into the return process. In the first step, the conditional realized volatility estimates are deduced as described in Sections 2.2.1-2.2.3 for each of the *t=1,…,T* in-sample data points using the estimated model parameters and the following transformation (see Beltratti and Morana, 2005; Giot and Laurent, 2004):

$$\widehat{RV}^{j}_{t/t-1} = \exp\left(lrv_t - \hat{u}^j_t + 0.5\hat{\sigma}^{2(j)}_{u(t)} \right) \tag{22}$$

where $\hat{u}^j_t$ are the estimated residuals, j denotes henceforth the j^{th} realized volatility model and $\hat{\sigma}^{2(j)}_{u(t)}$ is the residuals variance. The t in the subscript parenthesis denotes the time varying conditional variance of the residuals in the HAR-RV-GARCH and AHAR-RPV-GARCH models.

In the second step, the conditional variance in the return process of equation (1) is modeled as a fraction of the estimated conditional realized volatilities i.e.:

$$\sigma^2_{t,j} = g\widehat{RV}^{j}_{t/t-1} \tag{23}$$

Given the distributional assumption for the innovation process z_t, the scaling parameter g and the parameters of the conditional mean process specified in equation (2) are estimated via maximum likelihood (see also Section 2.4). This implementation allows for the different dynamics of the realized volatility models to be incorporated in the conditional variance of the return process, whilst we are able to assess their forecasting ability by ensuring that z_t is a unit variance process. In order to obtain the one step ahead conditional volatility forecast, the j^{th} model's day ahead realized volatility forecast , $\widehat{RV}^{j}_{t+1/t}$, is multiplied by the estimated scaling factor $\hat{g}$.

2.3. *Augmented GARCH-R(P)V models*

An alternative approach for accessing the informational content of realized volatility and realized power variation in VaR forecasting is to use them as explanatory variables in a GARCH model as in Fuertes et al. (2009), Grané and Veiga (2007) and Koopman et al. (2005) i.e.:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + bX_{t-1} \tag{24}$$

where X_{t-1} is either the realized volatility, or the realized power variation at $t-1$. Again, all coefficients in equation (24) are estimated by maximizing the likelihood function. Empirical evidence have shown that the GARCH model's volatility forecasting performance can be improved when realized volatility measures are used as additional explanatory variables (e.g. see Fuertes et al., 2009; Koopman et al. 2005). However, there is limited empirical evidence on the performance of the Augmented GARCH model in VaR forecasting applications (Grané and Veiga, 2007).

2.4. Estimation of the models

The scaling factor g, the conditional mean equation's coefficients vector and the coefficients vector for all the ARCH–type and Augmented GARCH models are estimated with the Quasi Maximum Likelihood (QML) method. Here, we consider two distributional forms for the innovation process z_t: the *Normal* (*N* in short) distribution and the *skewed student* (*skst* in short) distribution (Lambert and Laurent, 2001).

When $z_t \sim iid\ N(0,1)$, the QML estimates are deduced by maximizing the following logarithmic likelihood function with respect to the coefficients vector:

$$L_N = -\frac{1}{2}\sum_{t=1}^{T}\left[\ln(2\pi) + \ln\left(\sigma_t^2\right) + z_t^2\right] \tag{25}$$

However, the normality assumption has been shown to be inappropriate for the majority of financial assets returns (see for example Giot and Laurent, 2004, Giot, 2005 and Ferreira and Lopez (2005). Giot and Laurent (2003a, 2004) considered the skewed student distribution which takes into account the asymmetry and the excess kurtosis usually observed in the returns series. In this case, z_t follows a standardized (zero mean and unit variance) skewed student distribution, i.e. $z_t \sim iid\ \mathrm{skst}(0,1,\xi,\nu)$, where ν (with $\nu > 2$) and ξ are the degrees of freedom and the asymmetry coefficient respectively and both are estimated along with the coefficients vector. The respective logarithmic likelihood function is then defined as:

$$
\begin{aligned}
L_{skst} = T\left\{\ln\Gamma\left(\frac{\nu}{2}\right)-\ln\Gamma\left(\frac{\nu}{2}\right)-\frac{1}{2}\ln\left[\pi\left(\nu-1\right)\right]+\ln\left(\frac{2}{\xi+\frac{1}{\xi}}\right)+\ln\left(S\right)\right\}- \\
-\frac{1}{2}\sum_{t=1}^{T}\left\{\ln\left(\sigma_t^2\right)+\left(1+\nu\right)\ln\left[1+\frac{\left(sz_t+m\right)^2}{\nu-2}\xi^{-2I_t}\right]\right\}
\end{aligned}
\qquad (26)
$$

where m and s are the mean and the standard deviation of the non-standardized skewed student distribution and I_t equals 1 if $z_t \geq -m/s$ and -1 if $z_t < -m/s$. The estimated coefficients are used to compute day ahead forecasts for the conditional mean and variance.[8]

3. Value at Risk estimation methods

Value-at-Risk (VaR) has been adopted by practitioners and regulators as the standard method of measurement of the market risk of financial assets. It encapsulates in a single quantity the potential market value loss of a financial asset over a time horizon *h*, at a significance or coverage level α. Alternatively, it reflects the asset's market value loss over the time horizon *h*, that is not expected to be exceeded with probability $1-\alpha$, i.e. $\Pr\left(r_{t+h} \leq VaR_{t+h}^{\alpha} \middle| I_t\right) = 1-\alpha$, where r_{t+h} is the asset's return over the period *h* and I_t is the available information until time t. Hence, VaR is the α^{th} quantile of the conditional returns distribution defined as:

$$
VaR_{t+h}^{\alpha} = F_{t+h}^{-1}\left(\alpha\right) \qquad (27)
$$

[8] m and s are defined as $m = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\xi-\frac{1}{\xi}\right)$ and $s = \sqrt{\left(\xi^2+\frac{1}{\xi^2}-1\right)-m^2}$ respectively.

where F is the returns cumulative distribution function (cdf) and F^{-1} denotes the inverse cdf. For the returns process of equation (1), the next day's VaR is given by:

$$\widehat{VaR}_{t+1,j}^{\alpha} = \hat{\mu}_{t+1,j} + \hat{\sigma}_{t+1,j} F_z^{-1}(\alpha) \tag{28}$$

where $\hat{\mu}_{t+1,j}$ and $\hat{\sigma}_{t+1,j}$ are the j^{th} model's day ahead conditional mean and conditional volatility forecasts respectively and F_z^{-1} is the inverse cdf of the standardized returns, or innovations, i.e. $z_t = (r_t - \mu_t)/\sigma_t$.

A pivotal decision in VaR forecasting is the assumed conditional distribution and several authors have underlined the inappropriateness of the often used normal distribution (see for example Giot and Laurent, 2004; Giot, 2005). Here, we estimate the α^{th} quantile of the z_t process using three alternative methods: the fully parametric method utilizing either the normal or the skewed student distribution, the semi-parametric Filtered Historical Simulation (FHS) method and the conditional Extreme Value Theory (EVT) method. In the following sections the three methods are briefly discussed.

3.1. Fully Parametric method

In the fully parametric method, the risk manager makes an explicit distributional assumption for the innovation process. The conditional distribution of the standardized returns is assumed to have a specific functional form and its shape parameters are estimated along with the parameters of the conditional mean and of the volatility models, as described in Section 2.4. When normally distributed innovations, i.e., $z_t \sim iid\ N(0,1)$, are assumed, given a data sample of t=1,…,T daily returns tomorrow's VaR is deduced from:

$$\widehat{VaR}_{T+1,j}^{\alpha} = \hat{\mu}_{T+1,j} + \hat{\sigma}_{T+1,j} c_{\alpha}^{N} \tag{29}$$

where $c_\alpha^N = \Phi^{-1}(\alpha)$ is the standard normal α^{th} quantile which is readily available from statistical tables, Φ is the standard normal cdf, while $\hat{\mu}_{T+1,j}$ and $\hat{\sigma}_{T+1,j}$ are the j^{th} model's conditional mean and volatility forecasts respectively. Since the normal distribution is fully characterized by its mean and its variance, it does not require estimation of any additional parameters. Due to its simplicity, the normality assumption is widely adopted by practitioners for risk management purposes (see McMillan and Kambouroudis (2009)[9]). Nonetheless, empirical evidence suggests that it misspecifies the true conditional distribution of returns, especially in volatile periods where extreme price variations are more frequently observed.

The alternative specification of skewed student distributed innovations, i.e. $z_t \sim iid\ s\text{kst}(0,1,\xi,\nu)$, is more attractive as it captures the asymmetry and the fat tails of the returns process:

$$\widehat{VaR}_{T+1,j}^{\alpha} = \hat{\mu}_{T+1,j} + \hat{\sigma}_{T+1,j} c_{\alpha,\nu,\xi}^{skst} \tag{30}$$

$$\text{with } c_{\alpha,\nu,\xi}^{skst} = \begin{cases} \left\{ \frac{1}{\xi} c_{\alpha,\nu}^{st} \left[\frac{\alpha}{2}\left(1+\xi^2\right) \right] - m \right\} / \text{s} \ \text{ if } \ \alpha < \dfrac{1}{1+\xi^2} \\ \left\{ -\xi c_{\alpha,\nu}^{st} \left[\frac{1-\alpha}{2}\left(1+\xi^{-2}\right) \right] - m \right\} / \text{s} \ \text{ if } \ \alpha \geq \dfrac{1}{1+\xi^2} \end{cases}$$

where $c_{\alpha,\nu,\xi}^{skst}$ is the α^{th} quantile of the unit variance skewed student distribution with ν degrees of freedom and asymmetry parameter ξ, while $c_{\alpha,\nu}^{st}$ denotes the quantile function of the standardized Student-t density function (see Lambert and Laurent, 2001 and Giot and Laurent, 2003a).

3.2. Filtered Historical Simulation

The Filtered Historical Simulation (FHS) method proposed by Hull and White (1998), Barone-Alsesi et al. (1998) and Barone-Alsesi et al. (1999) combines the fully parametric

[9] The authors provide empirical evidence from 31 stock markets for the RiskMetrics model which assumes a normal distribution for the returns. They found that the RiskMetrics model produces adequate volatility forecasts for small emerging markets and for broader VaR measures.

approach described in the previous section and the non-parametric method of Historical Simulation (HS). In the HS method, no assumptions are made for the returns distribution, nor any parameter estimation is required. Assuming that asset returns are i.i.d., future VaR can be well approximated by the empirical distribution of historical returns. In this case, the VaR is calculated as the α^{th} empirical quantile of the unconditional return distribution of a moving window of w historical observations (see also Christoffersen, 2003 and 2009):

$$\widehat{VaR}_{t+1}^{\alpha} = Quantile\left(\left\{r_{t+1-\tau}\right\}_{\tau=1}^{w}, \alpha\right) \tag{31}$$

Despite that HS is a model-free VaR method used extensively by financial institutions[10], it suffers from several drawbacks such as the choice of the window length and the underlying i.i.d. returns assumption.

FHS retains the simplicity of the HS approach regarding the estimation of the innovations' quantiles. It uses however filtered instead of raw returns in order to account for the dynamic structure and the volatility clustering asset returns, whilst capturing any non-normalities of the innovation process. Given a data sample of t=1,…,T daily historical returns, the FHS requires two steps: In the first step, Gaussian QMLE is employed in order to estimate the coefficients of the j^{th} model as well as its conditional mean, $\hat{\mu}_{t,j}$ and variance, $\hat{\sigma}^2_{t,j}$. Then, the historical returns are filtered so as to obtain the standardized returns, i.e. $\hat{z}_{t,j} = \left(r_t - \hat{\mu}_{t,j}\right)/\hat{\sigma}_{t,j}$. In the second step, the resulting sequence of past estimated standardized returns is used for calculating their α^{th} quantile. Thus, tomorrow's VaR for model j is given by:

$$\widehat{VaR}_{T+1,j}^{\alpha} = \hat{\mu}_{T+1,j} + \hat{\sigma}_{T+1,j} Quantile\left(\left\{\hat{z}_{t,j}\right\}_{t=1}^{T}, \alpha\right) \tag{32}$$

[10] Perignon and Smith (2010) reported that from the 64,9% of the banks that revealed their VaR methodology in their survey, 73% declare the use of HS.

3.3. *Extreme Value Theory (EVT)*

As Extreme Value Theory (EVT) is concerned with the tails of a distribution, its application in a VaR context especially in periods characterized by abrupt and extreme price movements is to say the least, appealing. Following McNeil and Frey (2000), we use the filtered returns $\hat{z}_{t,j}$, $t=1,\ldots,T$ with a distribution function $F_z(\hat{z}_{t,j}) = \Pr\{\hat{z}_{t,j} \leq z\}$ as derived for each model *j* in order to implement the Peaks Over Threshold (POT) EVT method (for details see McNeil and Frey, 2000; Embrechts et al., 1997). The method effectively models the filtered returns which exceed a prespecified threshold U (for the choice of U see section 3.3.1). The use of the filtered instead of the raw returns validates the inherent EVT method i.i.d. assumption, while at the same time it accounts for the conditional heteroscedasticity of the returns (Byström, 2004).

If the magnitude of exceedence of z [11] over U is defined as $y_i = z_i - U$, where $i = 1,\ldots,T_U$ and T_U being the total number of exceedences for a given threshold U, then the distribution of y given that $z > U$ is defined as:

$$F_U(y) = \Pr\{z - U \leq y | z > U\} = \frac{F_z(y+U) - F_z(U)}{1 - F_z(U)}, \text{ for } y \geq 0 \tag{33}$$

Since $z = y + U$, and for $z > U$, equation (33) can be re-written as $F_z(z) = F_U(y)\left[1 - F_z(U)\right] + F_z(U)$ and as $F_z(U)$ is equal to $(T - T_U)/T$, after some algebra, it can be shown that:

$$F_z(z) = 1 + \frac{T_U}{T}\left(F_U(y) - 1\right) \tag{34}$$

A key result in EVT is that for a sufficiently high threshold U, the conditional distribution $F_U(y)$ in equation (34) converges to the Generalized Pareto Distribution (GPD) (see Balkema and de Han, 1974; Pickands, 1975) which is defined as:

[11] We drop the subscripts and the hat from $\hat{z}_{t,j}$ in order to simplify the notation.

$$G_{\zeta,\beta}(y)=\begin{cases}1-\left(1+\dfrac{\zeta}{\beta}y\right)^{-1/\zeta} & \text{if } \zeta \neq 0\\ 1\text{-}\exp(-y/\beta) & \text{if } \zeta = 0\end{cases}\text{, for } 0 \le y \le z-U \tag{35}$$

where ζ and $\beta > 0$ are the shape and scale parameters respectively. The GPD covers a variety of distributions depending on the value of the shape parameter ζ. Heavy tailed distributions such as the Pareto, Student *t*, Cauchy, and the Frechet with power tails, correspond to $\zeta > 0$, while the normal and other thin tailed distributions with exponential tails correspond to $\zeta = 0$. Short tailed distributions are also accounted for by the GPD when $\zeta < 0$. As most financial time series exhibit fat tails, we expect a positive ζ and for $\zeta \neq 0$ and $1+\zeta\frac{y}{\beta} > 0$, the probability density function (pdf) of the GPD in equation (35) is given by (Smith, 1987):

$$g_{\zeta,\beta}(y)=\frac{1}{\beta}\left(1+\zeta\tfrac{y}{\beta}\right)^{-(1+1/\zeta)} \tag{36}$$

Hence, the tail estimator in equation (34) reduces to:

$$\begin{aligned} F_z(z) &= 1+\frac{T_U}{T}\left(G_{\zeta,\beta}(y)-1\right)\\ &= 1-\frac{T_U}{T}\left(1+\frac{\zeta}{\beta}y\right)^{-1/\zeta} = 1-\frac{T_U}{T}\left(1+\frac{\zeta}{\beta}(z-U)\right)^{-1/\zeta} \end{aligned} \tag{37}$$

Solving equation (37) for z, the α^{th} quantile is defined as:

$$F_z^{-1}(\alpha)=U+\frac{\beta}{\zeta}\left[\left(\frac{T}{T_U}\alpha\right)^{-\zeta}-1\right] \tag{38}$$

Both shape and scale parameters can be deduced with maximum likelihood estimation and as Hosking and Wallis (1987) showed, for $\zeta > -0.5$ the maximum likelihood regularity conditions are fulfilled and the estimates are asymptotically normal. Hence, the estimates for ζ and β ($\hat{\zeta}$ and $\hat{\beta}$) are obtained by maximizing the corresponding log-likelihood function:

$$L_g\left(\zeta,\beta\right) = -T_U \log\left(\beta\right) - \left(1+\frac{1}{\zeta}\right)\sum_{i=1}^{T_U}\log\left(1+\frac{\zeta}{\beta}y_i\right) \tag{39}$$

Once substituted in equation (38) and for an α - significance level < $1-F_z\left(U\right)$ or T_U/T , the $\alpha\%$ VaR for model *j* is deduced from:

$$\widehat{VaR}_{T+1,j}^{\alpha} = \hat{\mu}_{T+1,j} + \hat{\sigma}_{T+1,j}F_z^{-1}\left(\alpha\right) \tag{40}$$

3.3.1. Choosing the threshold level

The threshold level U should be carefully selected as it has been characterized the “Achilles’ heel” of the POT – EVT method (Christoffersen, 2003, p. 83). The threshold should be sufficiently high so that the asymptotic results of Balkema and de Han (1974) and Pickands (1975) are valid, while at the same time it should be sufficiently low so that there are enough observations for the ML estimations. This is effectively interpreted as a tradeoff between unbiasness and efficiency. For a high threshold and subsequently a small number of observations, unbiased but volatile ML estimates are produced, while a lower threshold increases the number of observations leading to more efficient (less volatile) but biased estimates (Embrechts et al., 1997).

Here, we follow Gencay and Selcuk (2004) and Chan and Gray (2006) who jointly employed the Mean Excess Function (MEF) and the Hill plots (Hill, 1975) in order to deduce U (for a detailed discussion see Embrechts et al., 1997). The empirical MEF is defined as the average value of exceedances given a threshold U :

$$MEF\left(U\right) = \frac{1}{T_U}\sum_{i=1}^{T_U}\left(z_i - U\right) \tag{41}$$

Based on the mean excess plot from equation (41), we choose a threshold level U for which the resulting graph is an approximately positive straight line, suggesting that the data are generated by a GPD with $\zeta > 0$. The Hill estimator is a simple estimator of ζ when $\zeta > 0$ (Hill, 1975):

$$\hat{\zeta}_H = \frac{1}{T_U}\sum_{i=1}^{T_U}\log\left(z_i / U\right) \quad \text{for} \quad z > U . \tag{42}$$

In the Hill plot, the estimated $\hat{\zeta}_H$ values are plotted against different values for U. The threshold level is the lowest value of U from which and over the value of $\hat{\zeta}_H$ remains almost constant.

4. VaR evaluation methods

The 1996 Market Risk Amendment (MRA) to the Basel Capital Accord states that financial institutions can use their own internal VaR models in order to calculate the daily regulatory reserved capital. Should however they record too many VaR exceptions, then they are penalized with reserving more capital, or even worse, they have to re-evaluate with a view to revise their VaR models and systems. On the other hand, the banks profit maximization mandate dictates the use of capital in an efficient and productive way which is in effect interpreted as minimization of the idle capital. Thus, VaR models are inevitably driven towards achieving a balance between accuracy and efficiency. A VaR model should be conservative enough so as to produce the prespecified number of violations (in line with the chosen coverage level), whilst keeping VaR capital reserves at a minimum. With this practical risk management view in mind, we evaluate the VaR models forecasting performance with a variety of statistical accuracy and efficiency tests. The evaluation measures are briefly presented in the subsequent sections.

4.1. Failure Rate

A standard VaR model validation approach is the backtesting procedure where the estimated exante VaR is compared with the expost realized returns in order to calculate the number of VaR exceptions, i.e. the number of times the losses exceed the VaR estimates. The result of the backtesting procedure is the so called "failure, or hit process" and is described by the following indicator function:

$$H_t = \begin{cases} 1 \text{ if } r_t < VaR_t^{\alpha} \\ 0 \text{ if } r_t \geq VaR_t^{\alpha} \end{cases} \tag{43}$$

If n is the sample size, we define the number of exceptions as $n_1 = \sum_{t=1}^{n} H_t$, the number of non-exceptions as $n_0 = n - n_1$ and the proportion of failures, or Failure Rate (FR), as $\hat{\alpha} = n_1 / n$. The closer the FR is to the predetermined coverage level, the more accurate the VaR model is considered to be.

4.2. Unconditional and Conditional Coverage

From a statistical point of view, an accurate VaR model must exhibit correct conditional coverage, meaning that the hit process $\{H_t\}_{t=1}^{n}$ must be an i.i.d. Bernoulli process with parameter α. This is equivalent to testing whether the VaR model generates a Proportion of Failures (PF) $\hat{\alpha}$, equal to the required coverage level α, conditional on information set Ω_{t-1} for every point in time t, i.e.:

$$E\left(H_t \middle| \Omega_{t-1}\right) = \alpha \ \forall \ t . \tag{44}$$

Christoffersen's (1998) unconditional coverage test examines whether the VaR model's failure rate is statistically equal to the predetermined coverage level α, ignoring the history of the failure process. Under the null hypothesis of accurate unconditional coverage, i.e. $E\left(H_t\right) = \alpha$ and given the assumption of independence, the Likelihood Ratio (LR) test for verifying that $\hat{\alpha} = \alpha$ is:

$$LR^{uc} = 2\left[\log\left(\left(1-\hat{\alpha}\right)^{n_0} \hat{\alpha}^{n_1}\right) - \log\left(\left(1-\alpha\right)^{n_0} \alpha^{n_1}\right)\right] \tag{45}$$

which follows an asymptotic $\chi^2(1)$ distribution. Note that the null hypothesis is rejected if the VaR model generates too many or too few exceptions while the test ignores the possible dependence in the failure process structure. However, the robustness of the LR^{uc} test statistic is somehow limited since it is actually an unconditional test, i.e. exceptions are not taken into consideration as they occur, rather they are considered over the whole sample period. Although the unconditional coverage is accurately evaluated, the inherent variance dynamics are ignored and the conditional coverage at specific points in time is ignored. Hence, an otherwise inadequate VaR model could be accepted.

The complementary conditional coverage test proposed by Christoffersen (1998) is a joint test of correct unconditional coverage and first order independence of the failure process. Under the null hypothesis that the exceptions are independently distributed through time and the failure rate is equal to the prespecified one, the corresponding LR test is:

$$LR^{cc} = 2\left[\log\left(\left(1-\hat{p}_{01}\right)^{n_{00}}\hat{p}_{01}^{n_{01}}\left(1-\hat{p}_{11}\right)^{n_{10}}\hat{p}_{11}^{n_{11}}\right)-\log\left(\left(1-\alpha\right)^{n_0}\alpha^{n_1}\right)\right] \sim \chi^2(2) \tag{46}$$

where p_{ij} is the transition probability between two consecutive observations from state *i* to state *j* assuming a first-order Markov chain probability transition matrix between the two possible states (a successful VaR estimation, or an exception), n_{ij} is the number of all occurrences of transitions from state *i* to state *j*, with $i, j = 0,1$ and $\hat{p}_{ij} = n_{ij} / \sum_{j=0}^{1} n_{ij}$ are the maximum likelihood estimates for p_{ij}. Christoffersen's test can also reveal information about the conditional probability of occurrence of two consecutive exceptions in the VaR model, as well as the average time interval between exceptions. A VaR model may be rejected if it generates too many, too few, or too clustered exceptions.

4.3. Dynamic Quantile test

Engle and Manganelli (2004) argued that given the sequence of returns, r_t, it is straightforward to generate an i.i.d. failure process H_t and they proposed a more powerful test.

Specifically, they defined $Hit_t = H_t - \alpha$, where α is the significance level and they suggested a regression based approach to test whether $E\left(Hit_t\right)=0$ and also if Hit_t is uncorrelated with the variables included in the information set Ω_{t-1} in equation (44). In matrix notation the regression equation can be written as: $\mathbf{Hit} = \mathbf{X\beta}$ where $\mathbf{X}$ is the explanatory variables vector and $\boldsymbol{\beta}$ is the coefficients vector. The authors emphasized on the use of the contemporaneous value of VaR, VaR_t^{α}, in the explanatory variables set, as well as the use of lagged values of Hit_t, i.e. $Hit_{t-1}, Hit_{t-2}, ...Hit_{t-q}$, with $q=5$ in our case. Under the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$, the regressors, i.e. the five lags of Hit_t and the VaR_t^{α}, should have no explanatory power. The corresponding test statistic is:

$$DQ = \frac{\boldsymbol{\beta}_{LS}^{'}\mathbf{X}^{'}\mathbf{X}\boldsymbol{\beta}_{LS}^{'}}{\alpha(1-\alpha)} \tag{47}$$

which follows an asymptotic $\chi^2\left(p+1\right)$ distribution, where p is the total number of explanatory variables used in the regression.

4.4. Quadratic loss Function

Adhering to the Basel Committee's guidelines, supervisors are not only concerned with the number of failures of a VaR model, but also with the magnitude of these failures. Based on this provision, Lopez (1999) proposed the Quadratic Loss Function (QLF) which considers both the number of exceptions and their magnitude, calculated as the square of the distance between the VaR estimate and the exception:

$$QLF_t = \begin{cases} 1+\left(r_t - VaR_t^{\alpha}\right)^2 & \text{if } r_t < VaR_t^{\alpha} \\ 0 & \text{if } r_t \geq VaR_t^{\alpha} \end{cases} \tag{48}$$

Hence, large exceptions are penalized more heavily compared to the linear, or binary loss function of equation (43). Note that although the model with the smallest average QLF is considered to be the most accurate one, the metric tends to favor models that are too conservative and their number of violations is well below the prescribed coverage level. Therefore, we also calculate the QLF relative to the number of violations as in Martens et al. (2009):

$$\overline{QLF} = \sum_{i=1}^{n} QLF_t \Big/ n_1 \tag{49}$$

4.5. Mean Relative Scaled Bias

Hendricks (1996) proposed the Mean Relative Scaled Bias (MRSB) as a measure of relative efficiency of alternative VaR models (see also Engel and Gizyck, 1999). The purpose of the test is to identify which of the competing VaR models, once properly scaled to obtain the prespecified coverage level, generates the lowest average level of VaR. The VaR forecasts for each VaR model j are multiplied with a scaling factor SF_j in order to obtain the desired coverage level α. Then, the resulting scaled VaR, i.e. $VaR^*_{t,j} = SF_j \cdot VaR_{t,j}$, is used for calculating the mean relative bias of the j^{th} model as follows:

$$MRSB_j = \frac{1}{n} \sum_{t=1}^{n} \frac{VaR^*_{t,j} - \overline{VaR^*_t}}{\overline{VaR^*_t}} \tag{50}$$

where $\overline{VaR^*_t} = \frac{1}{J} \sum_{j=1}^{J} VaR^*_{t,j}$ is the average VaR forecast across the different *J* models for day *t*. The smaller the respective $MRSB_j$ measure is, the more efficient the VaR model is considered to be.

4.6. Regulatory loss Function

The regulatory Market Risk Capital (MRC) loss function implied in the 1996 MRA to the Basel Capital Accord is also a widely accepted method for evaluating alternative VaR models (see Ferreira and Lopez, 2005; Lopez, 1999 and the discussion therein) and it is given by:

$$MRC_t = \max\left[VaR_t^{0.01}(10), \frac{k}{60}\sum_{i=1}^{60} VaR_{t-i}^{0.01}(10)\right] \quad (51)$$

where $VaR_t^{0.01}(10)$ denotes the 1% VaR estimate of day *t* for a holding period of ten days, while k is a multiplier set by the MRA's traffic light system[12]. Specifically, the value of k is based on the number of 1% daily VaR exceptions over the previous 250 trading days. If the model produces 4 or less violations, then it is considered sufficiently accurate and the multiplier k takes its minimum value of 3. These are the so-called green zone or green light models. If the model generates between 5 and 9 violations over the previous trading year then it is placed in the yellow zone, or it is given a yellow light. It is also considered acceptable for regulatory purposes, with k being set to 3.4, 3.5, 3.65, 3.75 or 3.85, for the corresponding exceptions in the interval $[5,9]$. A red zone or red light model is one which generates 10 or more exceptions and then k takes its maximum value of 4. In this case, the regulators can reject the VaR model and put a request to the financial institution to revise their risk management systems.

4.7. Firm Loss Function

In the Firm Loss Function (FLF) proposed by Sarma et.al (2003), the non-exception days are penalized according to the opportunity cost of the reserved capital held by the firm for risk management purposes:

$$FLF_t = \begin{cases} 1+\left(r_t - VaR_t^{\alpha}\right)^2 & \text{if } r_t < VaR_t^{\alpha} \\ -cVaR_t^{\alpha} & \text{if } r_t \geq VaR_t^{\alpha} \end{cases} \quad (52)$$

where c is the firm's opportunity cost for its reserved capital. Thus, an otherwise accurate model producing a limited number of small magnitude violations may be highly inefficient as high daily VaR estimates entail additional opportunity costs. Sarma et al. (2003) proposed performing the FLF test only to VaR models which satisfied specific statistical test conditions, such as the

[12] For the calculation of the MRC, daily VaR is expressed in dollars: $VaR\$_t^{0.01} = P_{t-1}\left[1-\exp\left(VaR_t^{0.01}\right)\right]$, where P is the asset's price and is multiplied by $\sqrt{10}$ to get the 10 day VaR estimates as in Ferreira and Lopez (2005) .

Christoffersen's conditional coverage test and the DQ test. Here, we additionally require the fulfillment of the MRA regulatory rules, i.e. a VaR model should not have ten or more violations during the previous year as in Sajjad et al. (2008). Thus, the VaR models assessed in terms of the FLF are those which fulfilled both the statistical accuracy and the regulatory expectations.

4.8. Equal and Superior Predictive Ability Test for the MRC and FLF loss functions

The Equal Predictive Ability (EPA) test of Diebold and Mariano (1995) and the more general Superior Predictive Ability (SPA) test of Hansen (2005) are used in order to evaluate whether the performance differences between different VaR models in the regulatory oriented Market Risk Capital and Firm Loss Function tests are statistically significant.

The EPA test is a pairwise test based on the loss function differential between two competing models i and j, $d_t = l_{i,t} - l_{j,t}$, where $l_{i,t} = MRC_{i,t}$ or $l_{i,t} = FLF_{i,t}$. The null hypothesis of equal forecasting performance between models i and j implies that the models' loss function average values are equal, i.e. $E\left(l_{i,t}\right) = E\left(l_{j,t}\right)$, or equivalently that the average of the differential is zero, i.e. $E\left(d_t\right) = 0$. As the sample mean of d_t, i.e. $\overline{d}_t = n^{-1}\sum_{t=1}^{n} d_t$, is asymptotical distributed as $\sqrt{n}\left(\overline{d} - \mu\right) \rightarrow N\left(0, \operatorname{var}\left(\overline{d}\right)\right)$, the corresponding test statistic is given by:

$$t_{DM} = \frac{\overline{d}}{\sqrt{\widehat{\operatorname{var}}\left(\overline{d}\right)}} \sim N\left(0,1\right) \tag{53}$$

where $\widehat{\operatorname{var}}\left(\overline{d}\right) = n^{-1}\left(\hat{\gamma}_0 + 2\sum_{\kappa=1}^{q} w_\kappa \hat{\gamma}_\kappa\right)$ is the sample variance, $w_\kappa = 1 - \kappa / \left(q+1\right)$ is the lag window, $\hat{\gamma}_\kappa$ is the κ^{th} order sample autocovariance of d_t estimated as $\hat{\gamma}_\kappa = n^{-1}\sum_{t=\kappa+1}^{n}\left(d_t - \overline{d}\right)\left(d_{t-\kappa} - \overline{d}\right)$ and $q = \left[4\left(n/100\right)^{2/9}\right]$ as in Marcucci (2005).

Diebold and Mariano (1995) proposed also the sign test which is an exact finite-sample EPA test. It differs from the EPA test as the null hypothesis is that of a zero median function loss differential: $H_0 : \operatorname{med}\left(l_{i,t} - l_{j,t}\right) = \operatorname{med}\left(d_t\right) = 0$ against the alternative of a negative median, where negative values for d_t indicate superiority of model i over model j.[13] This is equivalent to testing whether the respective loss function value for model *i* is less than that of model *j* in most of the *n* days of the forecasting period. The null hypothesis is tested via the sign statistic, which is actually the number of non-negative *d*'s. This is defined as $S_{ij} = \sum_{t=1}^{T} \psi_t$, where ψ_t takes the value of one if $d_t \geq 0$ and zero otherwise. Under the null, the sign statistic is binomially distributed with parameters $(T, 0.5)$, but for large samples it is asymptotically normally distributed implying that the standardized version of the test statistic, $\hat{S}_{ij} = \left(S_{ij} - 0.5T\right) / \sqrt{0.25T}$, is distributed as $N(0,1)$. The null hypothesis for a 5% significance level is rejected for values of $\hat{S}_{ij} < -1.645$, suggesting the superiority of model i over model j.

The predictive ability of the VaR models in terms of the MRC_t and the FLF_t is also assessed via Hansen's (2005) Superior Predictive Ability (SPA) test. The SPA test examines whether the null hypothesis that the benchmark model is not outperformed by any of its competitors is rejected or not. Thus, it is much more straightforward to come to a generic conclusion for the models tested. The forecasting performance of the benchmark model, model 0, with respect to model k is deduced from the loss function differential: $f_{t,k} = l_{0,t} - l_{k,t}$, where $k = 1 \ldots j$ is the total number of competing counterparts. Under the null hypothesis and assuming stationarity for $f_{t,k}$, we expect that on average the forecasting loss function of the benchmark model will be smaller, or at least equal to that of model k. Thus, the null hypothesis can be stated as: $H_0: \max\limits_{k=1\ldots l} \mu_k = E\left(f_{t,k}\right) \leq 0$ and is tested through the following test statistic:

$$T_n^{SPA} = \max_{k=1\ldots l} \frac{\sqrt{n}\bar{f}_k}{\sqrt{\mathrm{var}\left(\sqrt{n}\bar{f}_k\right)}} \tag{54}$$

where $\bar{f}_k = (1/n)\sum_{t=1}^{n} f_{t,k}$ and $\mathrm{var}\left(\sqrt{n}\bar{f}_k\right)$ is the variance of $\sqrt{n}\bar{f}_k$. Both $\mathrm{var}\left(\sqrt{n}\bar{f}_k\right)$ and the test statistic *p*-values are consistently estimated via stationary bootstrapping as in Politis and Romano (1994).

[13] Note that $\mathrm{med}\left(l_{i,t} - l_{j,t}\right) \neq \mathrm{med}\left(l_{i,t}\right) - \mathrm{med}\left(l_{j,t}\right)$.

5. Empirical results

5.1. The data set

The data set was obtained from Tick Data and consists of five minutes previous tick interpolated prices for the S&P 500 cash index[14] over an approximately thirteen year period, from 1.1.1997 to 09.30.2009. After adjustments for holidays and half-holidays, there were $T = 3{,}196$ trading days, with six and a half trading hours per day, interpreted as $M = 78$ intraday returns. The five minutes intraday sampling frequency used for the computation of the realized volatility and realized power variation has been found to be the highest sampling frequency with acceptable market microstructure bias, especially for liquid assets like the S&P 500 index (see Andersen *et al.*, 2001a; Koopman *et al.*, 2005; Corsi *et al.*, 2008 and Degiannakis, 2008).

The descriptive statistics for the daily returns, the realized volatility and the logarithmic realized variance are presented in Table 2. The return series exhibits negative skewness and fat tails, a departure from normality which can be attributed to mainly negative price shocks near the end of 1997 and 1998, all through 2000, towards the end of 2002 and obviously during the 2007 – 2009 financial turmoil. The skewness and kurtosis of the logarithmic variance series suggest that the respective distribution is approximately normal.

[Insert Table 2 about here]

5.2. VaR forecasting methodology and evaluation results

In this study, we evaluate the day ahead VaR forecasting performance of the following volatility models: (i) ARCH – type models: EWMA, GARCH, GJR, EGARCH, APARCH and FIGARCH, (ii) realized volatility models: ARFIMA, ARMA, HAR-RV, HAR-RPV, HAR-RV-GARCH and AHAR-RPV-GARCH and (iii) Augmented GARCH models: GARCH-RV and GARCH-RPV. For each of the fourteen (14) volatility models, the α^{th} quantile of the z_t process is estimated with all three aforementioned methods: the fully parametric method utilizing either the normal or the skewed student distributions, the FHS method and the EVT-POT method.

[14] In this paper we assume a long position on the index.

Hence, we implement in total fifty six (56) distinct VaR models for four different VaR coverage levels: 10%, 5%, 1% and 0.5%.[15]

The full data set was divided into $T^* = 1,250$ in-sample observations and $n = T - T^* = 1,946$ out-of-sample observations, from 12.20.2000 to 09.30.2009. Moreover, towards the end of our analysis we concentrate explicitly on the 2007 – 2009 period in order to evaluate the VaR models' forecasting performance during the recent financial crisis.

The day ahead out-of-sample VaR forecasts were obtained using a rolling window of approximately five years, or 1,250 trading days. Hence, the $T^* + 1$ day VaR forecast was estimated using the complete in-sample data set, the $T^* + 2$ day VaR forecast was estimated using observations $\{2, ..., T^* + 1\}$ and so on. For each iteration, all the realized volatility and ARCH-type models parameters, the conditional realized volatilities, the conditional mean parameters, the scaling parameter g and the quantile estimation as described in Section 3 were re-estimated anew. To our knowledge, this is the first time the two step procedure is implemented for realized volatility models with a rolling window forecasting methodology.

5.2.1. Failure Rates

The failure rates for the 10%, 5%, 1% and 0.5% quantiles for each of the fifty six VaR models are presented in Table 3 along with the respective models' rankings. The model rankings are deduced with respect to the minimization of the distance between the observed failure rate and the prespecified coverage level, i.e. the model ranking first has its observed failure rate closest to the required coverage level. Overall, the realized volatility and the augmented GARCH models failure rates are closest to the required coverage level.

A closer examination of the results reveals that the augmented GARCH models yield the smallest distance from the benchmark 0.5% quantile for the normal distribution and for the 1% and 0.5% quantiles for the FHS method, while the AHAR-RPV-GARCH model ranks consistently first for the skewed student distribution. In total, the realized volatility models rank first nine times (five of which are attributed to the AHAR-RPV-GARCH model), followed by

[15] For the EVT method the 10% VaR was not calculated as the threshold U deduced with the mean excess and Hill plot methods results in a value for T_U for which $T_U / T = 1 - F_z(U)$ is smaller than the 10% coverage level.

the augmented GARCH models which rank first seven times, while the alternative VaR models take the top ranking three times altogether.

The worst[16] 10% quantile performances for nearly all VaR models are noted for the FHS method where excessive VaR provisions are suggested, while the worst 5%, 1% and 0.5% quantile results are observed for the normal distribution, as expected. A final interesting note is that the VaR estimates subscription to the target coverage level tends to improve for lower quantiles.

[Insert Table 3 about here]

5.2.2. Unconditional and Conditional coverage test

Christoffersen's unconditional and conditional coverage tests results are presented in Tables 4 and 5 respectively, where the bold faced fonts indicate rejection of the respective null hypothesis at a 5% significance level. With the normal distribution, almost all of the VaR models fail to produce the correct unconditional and conditional coverage for the 1% and 0.5% quantiles. The only exceptions are the augmented GARCH models for the 0.5% quantile where the null hypothesis of correct unconditional and conditional coverage cannot be rejected and also the HAR-RV-GARCH model for the conditional coverage test. These results adhere to the aforementioned inappropriateness of the normal distribution to accurately represent actual asset returns, especially when low coverage levels are required. For higher coverage levels, i.e. for 10% and 5%, all models par the EWMA exhibit correct unconditional and conditional coverage.

When the skewed student distribution, or the FHS and the EVT methods are used, the conditional and unconditional coverage of the VaR estimates is significantly improved. The rejections of the null hypothesis are mainly concentrated in the unconditional coverage test and the skewed student distribution for the EWMA model for the 10% and 5% quantiles, for the GARCH-RV model for the 5% and 1% quantiles and for the GARCH-RPV for the 1% quantile. The Augmented GARCH models results were anticipated, since when they are combined with the skewed student distribution they tend to overestimate VaR producing very few exceptions

[16] Measured as the average deviation of the observed FR with respect to the prescribed quantile.

(see Table 3). Interestingly, the EGARCH null hypothesis of correct unconditional coverage is always rejected for the 1% quantile, irrespective of the VaR distribution estimation method.

These results confirm that the key driver for correct (un)conditional coverage of the VaR forecasts is the innovations' distributional assumption. This conclusion is consistent with the findings of other authors e.g. Giot and Laurent (2004), Giot (2005) and Ferreira and Lopez (2005). Nonetheless, the value of the informational content in intraday data should not be dismissed as the models relying on high frequency data tend to produce the highest p-values across all VaR methods and quantiles.

[Insert Table 4 about here]
[Insert Table 5 about here]

5.2.3. Dynamic Quantile test

Table 6 reports the Engle and Manganelli (2004) Dynamic Quantile test results. The 5% and 10% quantile results are qualitative similar to those in Table 5 for the conditional coverage test, as the null hypothesis of correct VaR estimates cannot be rejected at a 5% significance level for all models, irrespective of the distributional assumption. The 1% and 0.5% quantile results for the normal distribution also align with the respective conditional coverage outcomes. However, the picture is somewhat different with the skewed student distribution, the FHS and the EVT methods results for the 1% and 0.5% quantiles. For the 1% quantile, the skewed student EWMA and EGARCH VaR estimates and all the ARCH-type models estimates when combined with the FHS and EVT methods reject the null hypothesis. For the 0.5% quantile, the null hypothesis is rejected for the EGARCH and the APARCH models when combined with the skewed student distribution or the EVT method and similarly for the GARCH, EGARCH and FIGARCH models with the FHS method. However, all realized volatility and augmented GARCH models with the exception of the ARFIMA-FHS, do not reject the null hypothesis of correct VaR estimates for the 1% and 0.5% quantiles. These findings suggest that for low quantiles, the use of intraday data can help improve the statistical accuracy of the VaR forecasts, as this is evaluated by the DQ test.

[Insert Table 6 about here]

5.2.4. *Quadratic Loss Function*

Tables 7 and 8 summarize the Average QLF[17] (AQLF) and the QLF divided by the number of exceptions results respectively. The realized volatility and the augmented GARCH models almost invariably minimize both QLF metrics and rank in the first places irrespective of the VaR method or quantile. In particular, the GARCH-R(P)V and the ARMA models generate the smallest exceptions across all VaR methods and quantiles as they are nearly always in the top rankings for both metrics. On the contrary, the ARCH models usually rank in the last places with the only exceptions being the EWMA-EVT for the AQLF 5% quantile and the FIGARCH-skst 1% quantile for the QLF relative to the number of exceptions metric. Hence, even when accounting for the magnitude of the exceptions and not only for their number, it is clear that models utilizing the informational content of intraday returns provide the most accurate VaR estimates.

[Insert Table 7 about here]
[Insert Table 8 about here]

5.2.5. *Mean Relative Scaled Bias*

The VaR estimates' relative efficiency as measured by the MRSB metric are presented in Table 9. According to the MRSB values, the augmented GARCH models followed by the realized volatility models clearly outperform the ARCH-type models across all VaR methods and quantiles. Once the VaR estimates are appropriately scaled to obtain the prespecified frequency of exceptions, the high frequency data models typically produce the lowest average risk estimate which in terms of efficiency amounts to reduced reserved regulatory capital requirements (see also the next section) and improved resource allocation signalling.

[Insert Table 9 about here]

[17] In order to facilitate the reviewing of the results, all Average QLF values have been multipied by 100.

5.2.6. Regulatory loss Function

Table 10 shows the percentage of days that the 1% VaR out-of-sample estimates merit a green, yellow, or red light according to the Basel committee traffic light system. The average capital reserved for regulatory purposes and its respective standard deviation are also presented. The bold faced typing indicates that a model has been assigned a red light for at least one day and/or has failed one or all of the (un)conditional coverage and dynamic quantile tests.

All models relying on the normal distribution fail to pass Basel's traffic light test, highlighting the inappropriateness of the normal distribution when regulatory compliance is required. It is worth noting however that the realized volatility and the augmented GARCH models minimize the red light days. When the volatility models are combined with the asymmetric and fat tailed skewed student distribution, no red light days are recorded except for the EGARCH and the GJR models. The stand-out result in that column is the GARCH-RV model which has 100% of the days in the green zone, the single such occurrence in our tests and an outstanding performance considering that the out-of-sample period includes two periods of financial stress (2000-2002 and 2007-2009). The GARCH-RPV model also merits a yellow light for only 3% of the out-of-sample days which is an excellent result as well. However both have previously failed the correct unconditional coverage test for the 1% VaR coverage level.

When the FHS method is used, all ARCH-type models register a limited number of red light days except from the EWMA and FIGARCH models which record no red light days but are nonetheless rejected since they have previously failed the dynamic quantile test. However, when the GARCH specification is augmented with the R(P)V regressors, no red light days are observed whilst registering some of the higher overall percentages of green light days. Among the realized volatility models, only the HAR-RV-GARCH and the Asymmetric HAR-RPV-GARCH models satisfy the regulators loss function and register no red light days. It is also worth noting that after the normal distribution, the FHS method yields the most red and yellow light days.

The majority of the VaR models when coupled with the EVT method produce acceptable VaR forecasts with respect to the supervisors' requirements. Only the GARCH, EGARCH and GJR models violate at least once the threshold of the ten exceptions over the previous 250 trading days. Overall, the HAR-RV-GARCH and the Asymmetric HAR-RPV-GARCH models are the most consistent performers from a regulator's point of view. They generate zero red light VaR estimates when used with the skewed student distribution, the FHS and EVT methods, whilst satisfying all the statistical accuracy conditions set herein.

Turning to the regulatory reserved capital results, we note that irrespective of the α^{th} quantile estimation method used, the GARCH-R(P)V models followed by the realized volatility models minimize the regulatory capital requirements and its volatility. This can be attributed to the fact that realized volatility is a consistent and less noisy estimator of the daily unobserved volatility than the squared daily returns used by the ARCH-type models. Thus, when used in a market risk management context, it can produce VaR estimates that track more closely the actual asset returns volatility dynamics. These results also align with the MRSB findings presented in Table 9 where the augmented GARCH models produced the lowest VaR estimates for the correct coverage level.

[Insert Table 10 about here]

Table 11 summarizes the Market Risk Capital (MRC) loss function EPA test results for the average of the loss differential, the median sign test results of the loss differential and the loss function SPA test results. The tests were run only for models passing both the statistical and regulatory accuracy tests, i.e. the (un)conditional coverage and DQ tests with zero red light days. From the fifty-six models examined here, only twenty-one models passed both kinds of tests and none of them used the normal distribution. Moreover, only three are ARCH–type models: the GARCH, the FIGARCH and the APARCH models when combined with the skewed student distribution.

For the EPA test, we chose the AHAR-RPV-GARCH with the EVT method as the benchmark model since it passed both types of tests across all three quantile estimation methods, requiring also less overall reserved capital than the close competing HAR-RV-GARCH model. The first column of the table shows the percentage of days that the benchmark model requires less regulatory reserved capital than each one of its competitors. The results reveal that more than half of the days, the benchmark model demands less regulatory capital than that demanded by the alternative VaR specifications. The p-values of the average MRC loss function differentials show that the null of equal performance cannot be rejected for four realized volatility models and the two augmented GARCH models with the EVT and the FHS methods. However, the median sign-test p-values confirm that the benchmark model outperformed its competitors as the null hypothesis of equal performance is always rejected. For the SPA test, each of the Table 11 models is alternatively used as the benchmark model and the null hypothesis that it is not outperformed by any of its counterparts is tested. The p-values give evidence in favor only for the augmented GARCH-RV-FHS and GARCH-RPV-FHS models.

[Insert Table 11 about here]

5.2.7. Firm Loss Function

In Table 12 the Firm Loss Function (FLF) EPA test results for the average of the loss differential, the median sign test results of the loss differential and the loss function SPA test results are presented. The average FLF value is also shown in the first column. As before, only models satisfying both the statistical and regulatory tests are included. The GARCH-RPV-FHS/EVT followed by the GARCH-RV-FHS/EVT models produce the lowest average FLF, while the AHAR-RPV-GARCH-EVT and then the HAR-RV-GARCH-FHS models are lagging closely behind. The AHAR-RPV-GARCH-EVT model is the EPA benchmark model as before and more often than not, less opportunity costs are incurred for its reserved regulatory capital than that of its alternatives. Nonetheless, the EPA test for the average loss differential gives also evidence in favour of the augmented GARCH models and the HAR-RV-GARCH-FHS model, in line with the FLF value findings. The median sign-test results are qualitative equivalent to the ones presented for the MRC in Table 11. The SPA test results also support the aforementioned findings as the null hypothesis cannot be rejected for the augmented GARCH-RPV model at a 5% significance level. There is also evidence in favour of the AHAR-RPV-GARCH-EVT model for a 2% significance level.

[Insert Table 12 about here]

5.2.8. Evaluating the models during the 2007-2009 period

In order to explicitly evaluate the VaR models during a period of high market volatility, we repeated our analysis by setting the out-of-sample data range to span from 07.01.2007 to 30.09.2009. During this turbulent period, the latest global financial crisis gradually unfolded with the highlight event probably being the collapse of Lehman Brothers on the 15th of September 2008. We focused solely on the 1% VaR estimates as this quantile bears the greatest significance for practical applications. A synopsis of the statistical accuracy and efficiency evaluation measure results[18] is presented in Tables 13 and 14.

From Table 13 we note that the augmented GARCH models typically produce the most accurate empirical failure rates and the lowest Average QLF measures, followed by the realized volatility models. These are also the only two model classes that when combined with the skewed student distribution, or the EVT method, exhibit correct conditional and unconditional coverage and do not reject the null of the DQ test. The ARCH–type models performances in the (un)conditional and DQ tests deteriorate over the 2007-2009 period, as all of them fail either or both tests. The results for the FHS method are somewhat discouraging since the majority of the models fail the (un)conditional and DQ tests and eight out of fourteen models record red light days. Overall, the models that recorded red light days in the full out-of-sample period exhibit red light days in this sub-sample period as well, albeit to a greater extend.

In total, only three models satisfy all of the conditions set herein across the three VaR quantile methods: the two augmented GARCH models and the AHAR-RPV-GARCH model. In particular, the augmented GARCH-RV model performs consistently very well across all methods and tests. However, the best overall VaR forecasting performance is shared between the closely competing augmented GARCH models and the HAR-RPV model, when they are combined with the skewed student distribution.

[Insert Table 13 about here]

In Table 14, the models satisfying both the statistical and regulatory tests for the period in question are compared with respect to their MRC and FLF performances as in Tables 11 and 12.

[18] The results with the normal distribution are not presented as they have no material difference with those presented for the full sample. They are available from the authors upon request.

We note that all average MRC and the FLF values worsen during the financial crisis period, an outcome which was largely to be expected. The increased market volatility is reflected in the VaR estimates which in turn are translated into higher reserved capital (MRC) levels and hence higher opportunity costs (FLF). Overall, the augmented GARCH-RPV model when combined with the skewed student distribution and the EVT method produces the lowest overall MRC and FLF average values respectively. For comparison purposes, the benchmark model is again the AHAR-RPV-GARCH-EVT model, but this time alternative models demand less regulatory capital for most of the days. In particular, all augmented GARCH models almost invariably require less regulatory capital and most notably the GARCH-RPV-skst model always demands less capital. The EPA average MRC p-values indicate that the null hypothesis of equal performance cannot be rejected for all alternative models and the median sign-test p-values confirm that the benchmark model is outperformed by its competitors. The SPA p-values give evidence in favor only for the augmented GARCH-RV-skst model. Similar conclusions are drawn for the FLF results, but now, according to the SPA test p-values, the GARCH-RPV-EVT model outperforms its alternatives.

[Insert Table 14 about here]

6. Conclusions

Our empirical analysis yielded several interesting conclusions. For the volatility models, the evidence was unequivocal: volatility models utilizing high frequency data, i.e. the realized volatility and the augmented GARCH models, almost invariably produced superior VaR forecasts in terms of statistical accuracy and regulatory capital efficiency, irrespective of the VaR quantile or its estimation method. The ARCH-type models, including the popular RiskMetrics™ model, typically finished last in the statistical accuracy tests, while their performance usually deteriorated further for the more demanding 1% and 0.5% VaR quantiles. Would they have passed the 1% statistical accuracy conditions, they would have produced the most red light days and recorded the highest regulatory capital demands, whilst incurring the highest opportunity costs. Their performance in periods of high volatility confirmed and amplified the aforementioned observations. Although it is difficult to distinguish any one volatility model, we were particularly impressed with the performance of the augmented GARCH models, the Asymmetric HAR-RPV-GARCH and the HAR-RV-GARCH models. Based on the evidence collected here, in a practical application we would like to see these models partnered either with the FHS, or the EVT quantile estimation methods. In periods of high volatility we would also consider switching to the fully parametric method with the skewed student distribution.

The skewed student distribution helped achieve robust statistical accuracy results during the full sample period and better than those produced with the alternative quantile methods during the recent financial crisis. On average, it led to higher regulatory capital requirements, thus minimizing the red light days, however during periods of high volatility it yielded more conservative regulatory demands than the alternative methods. The FHS and the EVT methods also proved to be powerful VaR quantile estimators. We noted however a relative weakness of the FHS method with respect to the statistical accuracy conditions during highly volatile periods. It is also worth highlighting the solid performance of the EVT method in terms of the regulatory criteria, i.e. the small number of red light days with respect to the reserved regulatory capital levels (the lowest in the full sample) and consequently the low opportunity costs. Periods of high volatility extended significantly the regulatory capital provisions for both methods. We also confirmed previously published evidence that the normal distribution is a poor choice for the innovations distribution.

The results presented here should also be of interest to regulatory authorities and financial institutions. For financial institutions, apart from improving their confidence in their VaR forecasts which is beneficial for themselves and their clients, they could also improve their regulatory compliance profile, manage their regulatory reserves more effectively and even free tied up capital towards more productive and rewarding uses. From a regulator's point of view, enhancing the accuracy of risk management models could help mitigate systemic risks in periods of extraordinary market volatility and contribute towards the overall stability of the financial system. Off-course, practical issues, e.g. the availability of high frequency data for a broad range of assets classes, storage and the real-time processing requirements, will have to be addressed before deploying these VaR models to cover all the activities of a financial institution.

We have no reason to doubt that similar VaR forecasting results can also be realized for other liquid stock indices. However, further investigation is necessary into the VaR forecasting performance of the high frequency volatility models examined here for other asset classes such as futures, bonds, currencies and commodities. Finally, alternative realized volatility models such as the Multiplicative Error Models (MEM) (Engle, 2002; Engle and Gallo, 2006; Brownless and Gallo, 2009) and the recently proposed High-frEquency-bAsed VolatilitY (HEAVY) models (Shephard and Sheppard, 2010) could also be evaluated in combination with the FHS and the EVT quantile estimation methods.

References

Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. Journal of Finance 57, 1047-1091.

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance, 4, 115–158.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001a. The distribution of realized stock returns volatility. Journal of Financial Economics 6, 43-76.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001b. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42-55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579– 625.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F. Diebold, F.X., 2006. Volatility and correlation forecasting. In: Elliott, G., et al. (Eds.), Handbook of Economic Forecasting, Vol. I, North-Holland.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including Jump components in the measurement, modelling and forecasting of return volatility. The Review of Economics and Statistics 89, 701-720.

Angelidis, T., Degiannakis, S., 2008. Volatility forecasting: Intra-day versus Inter-day models. Journal of International Financial Markets, Institutions and Money 18, 449-465.

Balkema, A., De Haan, L., 1974. Residual life at great age. Annals of Probability 2, 792–804.

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74, 3-30.

Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. Don’t look back. Risk 11, 100-103.

Barone-Adesi, G., Giannopoulos, K., Vosper, L., 1999. VaR without correlations for nonlinear portfolios. Journal of Futures Markets 19, 583-602.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society. B 64, 253-280.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1–48.

Basle Committee on Banking Supervision (1996). Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements. Bank for International Settlements Publication No. 22.

Bekaert, G., Wu, G., 2000. Asymmetric volatility and risk in equity markets. Review of Financial Studies 13, 1 – 42.

Beltratti A., Morana C., 2005. Statistical Benefits of Value-at-risk with long memory. Journal of Risk 7, 21-45.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business and Economic Statistics, 19, 465-474.

Black, F., 1976. Studies of stock prices volatility changes. Proceedings of the 976 Meeting of the American Statistical Association, Business and Economic Statistics Section,177– 181.

Bollerslev, T., 1986. Generalized autoregressive heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., 2010. Glossary to ARCH (GARCH). In: Bollerslev, T., et al. (Eds.), Volatility and time series econometrics: Essays in honor of Robert F. Engle. Oxford University Press.

Brownless, C.T., Gallo, G.M., 2009. Comparison of volatility measures: A risk management perspective. Journal of Financial Econometrics 8, 29-56.

Byström, H., 2004. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. International Review of Financial Analysis 13, 133–152.

Campel, A., Chen, X.L., 2008. The year of living riskily. Risk July 2008, 28-32.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: An asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281– 318.

Chan, K.F., Gray, P., 2006. Using extreme value theory to measure value-at-risk for daily electricity spot prices. International Journal of Forecasting 22,283-300.

Clements, M., Galvao. A., Kim, J., 2008. Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. Journal of Empirical Finance 15, 729-750.

Corsi, F., 2009. A simple approximate long memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Reno, R., 2009. HAR volatility modelling with heterogeneous leverage and jumps. Working Paper.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 27, 46-78.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841-862.

Christofferesen, P., 2003. Elements of financial risk management. Academic Press, Boston.

Christofferesen, P., 2009. Value-at-Risk models. In: Andersen, T.G., et al., (Eds.), Handbook of financial time series, Berlin, Heildelberg.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253– 263.

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

Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. Modeling Extreme Events for Insurance and Finance, Springer, Berlin.

Engel, J., Gizyck, M., 1999. Conservatism, accuracy and efficiency: comparing Value-at-Risk models. Working Paper 2, Australian Prudential Regulatory Authority.

Engle, R.F., 1982. Autoregressive, conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1007.

Engle, R.F., 2002. New frontiers for ARCH models. Journal of Applied Econometrics 17, 425-446.

Engle, R.F., Ng, V., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749-1778.

Engle, R.F., Manganelli, S., 2004. CaViaR: conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22, 367-381.

Engle, R.F., Gallo J.P., 2006. A multiple indicator model for volatility using intra daily data. Journal of Econometrics 131, 3-27.

Ferreira, M.A., Lopez, J.A., 2005. Evaluating interest rate covariance models within a Value-at-Risk framework. Journal of Financial Econometrics 3, 126-168.

Forsberg, L., Ghysels, E., 2007. Why do absolute returns predict volatility so well? Journal of Financial Econometrics 5, 31-67.

French, K.R., Schwert, G.W., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3 –29.

Fuertes, A.M., Izzeldin, M., Kalotychou, E., 2009. On forecasting daily stock volatility: The role of intraday information and market conditions. International Journal of Forecasting 25, 259-281.

Gallant, R., Hsu, C.T., Tauchen, G., 1999. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. The Review of Economics and Statistics 81, 617-631.

Gencay, R., Selcuk, F., 2004. Extreme value theory and value-at-risk: Relative performance in emerging markets. International Journal of Forecasting 20, 287– 303.

Ghysels, E., Sinko, A., 2006. Comment on Hansen and Lunde JBES paper, Journal of Business and Economic Forecasting 26,192-194.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: How to get most out of returns data sampled at different frequencies. Journal of Econometrics 131, 59-95.

Giot, P., Laurent, S., 2003a. Value-at-risk for long and short trading positions. Journal of Applied Econometrics 6, 641-663.

Giot, P., Laurent, S., 2003b. Market risk in commodity markets: A VaR approach. Energy Economics 25, 435-457.

Giot, P., Laurent, S., 2004. Modeling daily value-at-risk using realized volatility and ARCH type models. Journal of Empirical Finance 11, 379-398.

Giot, P., 2005. Implied volatility indexes and daily Value at Risk models. The Journal of derivatives 12, 54-64.

Glosten, L., Jagannathan, R., Runke, D., 1993. Relationship between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779-1801.

Grané, A., Veiga, H., 2007. The effect of realized volatility on stock return estimates, Working Paper, Universidad Carlos III de Madrid.

Granger, C.W.J., Newbold, P., 1976. Forecasting Economic Time Series. Academic Press, New York.

Hansen, P.R., 2005. A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365-380.

Hendricks, D., 1996. Evaluation of value-at-risk models using historical data. Economic Police Review 2, 39-70.

Hill, B., 1975. A simple general approach to inference about the tail of distribution. Annals of Statistics 46, 1163-1173.

Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29, 339-349.

Hull, J., White, A., 1998. Incorporating volatility updating into the historical simulation method of Value-at-Risk. Journal of Risk 1, 5-19.

JP Morgan, 1996. Risk Metrics Technical Document, 4th edition.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements. Journal of Empirical Finance 12, 445-475.

Kruse, R., 2006. Can realized volatility improve the accuracy of Value-at-Risk forecasts? Working Paper, Leibniz University of Hannover.

Kuester, K., Mittnik, S., Paolella M.S., 2006. Value-at-Risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 4, 53-89.

Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3, 73-84.

Lambert, P., Laurent, S., 2001. Modelling financial time series using GARCH-type models and skewed Student density. Mimeo, Universite de Liege.

Liu, C., Maheu, J.M., 2009. Forecasting realized volatility: A Bayesian model-averaging approach. Journal of Applied Econometrics 24, 709-733.

Lopez J.A., 1999. Methods for evaluating Value-at-Risk estimates. Federal Reserve Bank of San Francisco Economic Review 2, 3-17.

Louzis, D.P., Xanthopoulos Sisinis, S., Refenes, A.P., 2010. Stock index realized volatility forecasting in the presence of heterogeneous leverage effects and long range dependence in the volatility of realized volatility. Applied Economics, forthcoming.

Marcucci, J., 2005. Forecasting stock market volatility with regime-switching GARCH models. Studies in Nonlinear dynamics and Econometrics 9, 1-53.

Martens, M., 2002. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. Journal of Futures Markets 22, 497-518.

Martens, M, Dijk, D., Pooter, M., 2009. Forecasting S&P 500 volatility: Long memory, level shifts, leverage effects, day of the week seasonality and macroeconomic announcements. International Journal of Forecasting 25, 282-303.

McMillan, D.G., Speight A.E.H., Evans, K.P., 2008. How useful is intraday data for evaluating daily Value-at-Risk: Evidence from three Euro rates. Journal of multinational Financial Management 18,488-503.

McMillan, D.G., Kambouridis, D., 2009. Are RiskMetrics good enough? Evidence from 31 stock markets. International Review of Financial Analysis 18, 117-124.

McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance 7, 271-300.

Muller, U.A., Dacorogna, M.M., Davi, R.D., Olsen, R.B., Pictet, O.V., von Weizsacker, J.E., 1997. Volatilities of different time resolutions – analyzing the dynamics of market components. Journal of Empirical Finance 4, 213-239.

Muller, U.A, Dacorogna, M.M, Dav, R.D., Pictet, O.V., Olsen, R.B., Ward, J., 1993. Fractals and intrinsic time – A challenge to econometricians. 39th International AEA Conference on Real Time Econometrics, 14–15 October 1993, Luxembourg.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347-370.

Perignon, C., Smith, D.R., 2010. The level and quality of Value-at-Risk disclosure by commercial banks. Journal of Banking and Finance 34, 362-377.

Pickands, J., 1975. Statistical inference using extreme order statistics. Annals of Statistics 3, 119-131.

Pindyck, R.S., 1984. Risk, inflation and stock market. American Economic Review, 74, 334-351.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303-1313.

Pong, S., Shackleton, M.B., Taylor, S.J., Xu, X., 2004. Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. Journal of Banking and Finance 28, 2541-2563.

Sarma, M., Thomas, S., Shah, A., 2003. Selection of Value-at-Risk models. Journal of Forecasting 22, 337-358.

Sajjad, R., Coakley, J., Nankervis, J.C., 2008. Markov-Switching GARCH modeling of Value-at-Risk. Studies in Nonlinear dynamics and Econometrics 12, Article 7.

Shephard, N., Sheppard, K., 2010. Realising the future: Forecasting with high-frequency-based volatility (HEAVY) models. Journal of Applied Econometrics 25, 197-231.

Shao, X.D., Lian, Y.J, Yin L.Q., 2009. Forecasting Value-at-Risk using high frequency data: The realized range model. Global Finance Journal 20, 128-136.

Smith, R.L., 1987. Estimating tails of probability distributions. Annals of Statistics 15, 1174-1207.

Wu, G., 2001. The determinants of asymmetric volatility. Review of Financial Studies 14, 837-859.

Zakoian, J.M., 1994. Threshold heteroskedastic models. Journal of Economic Dynamics and Control, 18, 931-955.

Table 1 Literature review: Value at Risk and realized volatility

Author(s)	Methodology and VaR evaluation	Data set	Main conclusions
Giot and Laurent (2004)	The RiskMetrics and the skewed student APARCH model are compared with a realized volatility (RV) ARFIMAX-RV model combined with a normal and a skewed student distribution. A two step approach was used to relate the RV dynamics with the returns process. Evaluation: Kupiec's (1995) test and the Dynamic Quantile (DQ) test of Manganelli and Engle (2004).	CAC 40 (1995-1999), SP 500 futures (1989-2000), YEN/USD and DEM /USD (1989-2001)	The RV model did not improve the daily VaR forecasting performance of the APARCH-skst model.
Beltratti and Morana (2005)	An ARFIMA-RV model, an ARFIMA model with a FIGARCH specification for the heteroskedastic errors and a GARCH model were used to generate multi-step VaR forecasts.Evaluation: Christoffersen's (1998) (un)conditional coverage test and Berkowitz (2001) density forecast tests.	YEN and DEM against USD (1986-1999)	The RV ARFIMA-FIGARCH model provided superior VaR forecasts.
Kruse (2006)	GARCH type, RV and stochastic volatility models were used to forecast VaR based on the Normal, GED, skst errors distributions, the Filtered Historical Simulation (FHS) and the EVT methods. Evaluation: (un)conditional coverage tests and the Firm Loss Function (FLF) (Sarma et al., 2003).	SP 500 futures index (the time period is not reported)	The RiskMetrics and the GARCH models were not outperformed by any other model.
Grané and Veiga (2007)	GARCH, EGARCH and GJR-GARCH were augmented with realized volatility as an explanatory variable. Evaluation: the Minimum Capital Risk Requirements (MCRR).	American Express, Coca-Cola, Walt Disney, Pfizer (1997-2007)	RV enhanced the capacity of the models to calculate accurate MCRR.
Angelidis and Degiannakis (2008)	A normal TARCH, a FIAPARCH-skst model and an ARFIMAX-RV model combined with a skst distribution (as in Giot and Laurent (2004)) were used to forecast daily VaR. Evaluation: (un)conditional coverage tests.	CAC 40, DAX 30 (1995-2003), FTSE 100 (1998-2003).	The TARCH model was the overall best performing model.
Clements et al. (2008)	The authors used AR(5), MIDAS regressions and HAR models combined with normal, t-student (8 degrees of freedom) and the FHS method. Evaluation: A 'tick' or check function assessed with the Diebold and Mariano (1995) test.	AUD, CAD, EUR, GBP, YEN vs USD rates (1999-2003)	The HAR model provided superior forecasts for currencies with volatility shifts.
McMillan et al. (2008)	Intraday GARCH, Component GARCH and EGARCH models were compared with their daily counterparts and RV models (AR-structure) in volatility and VaR forecasting. Evaluation: Kupiec's (1995) and the DQ tests.	EUR against USD, GBP and YEN (2002-2003)	Intraday models provided improved performance wrt daily & RV models
Brownless and Gallo (2009)	RV, bipower RV, two scales RV, realized kernel as well as the daily range are modeled with a P-Spline Multiplicative Error Model. A t-student GARCH was also used. Evaluation: (un)conditional coverage tests, the DQ test and the probability deviation loss functions (Kuester et al., 2006).	3 NYSE stocks: Boeing, General Electric, Johnson and Johnson (2001-2006)	RV measures improved the VaR forecasts with respect to the GARCH model, but not so relative to the range.
Martens et al. (2009)	An AR(22), an ARFI, a HAR and a GJR-GARCH model were extended to incorporate level shifts, leverage effects, day-of the week seasonality and the effect of the macroeconomic announcements. Evaluation: (un)conditional coverage tests, the Quadratic Loss Function (QLF) (Lopez, 1999) and the Basel II Capital Requirements (CR).	SP 500 futures index (1994-2006)	All models failed the coverage tests. The RV models produce less volatile CRs and minimize the QLF.
Shao et al. (2009)	The Realized Range (RR) modelled with a Conditional Autoregressive Range (CARR) model (Chou, 2005) and combined with the skst distribution as in Giot and Laurent (2004). An ARFIMA-RV model, the RiskMetrics, t-student GARCH and the APARCH-skst models were also used. Evaluation: Kupiec's (1995) and the DQ test.	Shanghai Composite and Shenzhen Component Index (2005-2007)	The RV and RR models had similar performance and outperformed the daily ARCH-type models.

Table 2 Descriptive statistics for the S&P 500 cash index (01.01.1997-09.30.2009)

	r	RV	ln(RV)
Mean	0.01	1.36	-0.30
Median	0.06	0.71	-0.34
Maximum	10.71	76.80	4.34
Minimum	-9.79	0.05	-2.91
Std. Dev.	1.34	2.77	0.99
Skewness	-0.09	10.85	0.55
Kurtosis	9.75	210.62	3.66

Table 3 Failure rates (in percentage points)

	Normal				Skewed Student				FHS				EVT			
	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%
EWMA	11.51 *14*	6.27 *14*	1.95 *13*	1.13 *11*	11.97 *14*	6.12 *14*	1.28 *11*	0.57 *4*	**9.82** ***1***	5.24 *11*	1.18 *4*	0.72 *11*		4.73 *9*	1.13 *8*	0.31 *12*
GARCH	10.48 *13*	5.70 *12*	2.16 *14*	0.98 *4*	10.79 *12*	5.34 *9*	0.92 *2*	0.46 *1*	9.40 *8*	5.19 *7*	1.28 *10*	0.77 *13*		4.62 *12*	1.18 *10*	0.46 *2*
FIGARCH	10.38 *11*	5.70 *12*	1.90 *12*	1.08 *10*	10.84 *13*	5.60 *11*	1.13 *6*	0.41 *6*	9.71 *2*	5.24 *11*	1.18 *4*	0.57 *1*		5.24 *8*	1.08 *7*	0.41 *9*
EGARCH	10.12 *4*	5.60 *11*	1.75 *7*	1.03 *7*	10.69 *11*	5.76 *12*	1.49 *14*	0.77 *14*	10.53 *6*	5.45 *14*	1.75 *14*	0.77 *13*		5.70 *14*	1.49 *14*	0.77 *14*
APARCH	9.61 *12*	5.45 *8*	1.80 *10*	1.13 *11*	9.76 *7*	5.09 *3*	1.08 *4*	0.57 *4*	9.15 *14*	5.19 *7*	1.39 *12*	0.62 *4*		5.19 *6*	1.18 *10*	0.57 *7*
GJR	9.92 *2*	5.34 *5*	1.75 *7*	1.13 *11*	10.33 *9*	5.04 *2*	1.13 *6*	0.46 *1*	9.51 *5*	5.19 *7*	1.44 *13*	0.67 *8*		5.29 *10*	1.28 *13*	**0.51** ***1***
ARMA-RV	9.66 *7*	5.14 *2*	1.80 *10*	1.13 *11*	9.82 *5*	4.88 *4*	0.82 *8*	0.41 *6*	9.30 *10*	4.93 *4*	1.23 *8*	0.62 *4*		4.78 *7*	**0.98** ***1***	0.41 *9*
ARFIMA-RV	9.82 *5*	5.55 *10*	**1.59** ***1***	0.98 *4*	9.82 *5*	4.88 *4*	0.82 *8*	0.31 *12*	9.40 *8*	5.40 *13*	1.18 *4*	0.72 *11*		5.34 *11*	0.92 *3*	0.46 *2*
HAR-RV	9.66 *7*	5.29 *4*	1.70 *4*	0.98 *4*	9.87 *2*	4.88 *4*	0.87 *5*	0.41 *6*	9.25 *11*	5.04 *3*	1.28 *10*	0.62 *4*		5.09 *3*	0.92 *3*	0.46 *2*
HAR-RPV	9.66 *7*	5.40 *6*	1.70 *4*	1.03 *7*	9.87 *2*	4.88 *4*	0.72 *10*	0.41 *6*	9.25 *11*	**4.98** ***1***	1.13 *3*	0.62 *4*		**5.04** ***1***	0.92 *3*	0.46 *2*
HAR-RV-G	9.66 *7*	5.50 *9*	1.64 *2*	0.92 *3*	9.87 *2*	4.83 *8*	0.92 *2*	0.41 *6*	9.25 *11*	**4.98** ***1***	1.23 *8*	0.67 *8*		5.14 *5*	0.82 *9*	0.46 *2*
AHAR-RPV-G	**9.97** ***1***	5.40 *6*	1.75 *7*	1.03 *7*	**9.92** ***1***	**4.98** ***1***	**0.98** ***1***	**0.46** ***1***	9.56 *4*	5.19 *7*	1.18 *4*	0.67 *8*		5.40 *13*	1.23 *12*	0.57 *7*
GARCH-RV	9.76 *6*	**5.04** ***1***	1.64 *2*	**0.72** ***1***	9.35 *10*	4.06 *13*	0.51 *12*	0.26 *13*	9.46 *7*	4.88 *5*	0.87 *2*	**0.57** ***1***		4.93 *2*	0.92 *3*	0.36 *11*
GARCH-RPV	9.92 *2*	5.19 *3*	1.70 *4*	**0.72** ***1***	9.76 *7*	4.42 *10*	0.51 *12*	0.36 *11*	9.61 *3*	5.14 *6*	**1.03** ***1***	**0.57** ***1***		5.09 *3*	**0.98** ***1***	0.26 *13*

This table presents the observed failure rates for the alternative VaR models for a 10%, 5%, 1% and 0.5% VaR coverage level. The relative performance ranking of each VaR model is shown in italics. The bold faced fonts denote the model ranking first. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 4 Unconditional coverage test (p-values)

	Normal				Skewed Student				FHS				EVT			
	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%
EWMA	**0.03** *14*	**0.01** *14*	**0.00** *13*	**0.00** *11*	**0.00** *14*	**0.03** *14*	0.23 *10*	0.69 *4*	0.79 *1*	0.63 *11*	0.43 *4*	0.20 *11*		0.58 *9*	0.57 *8*	0.20 *12*
GARCH	0.48 *13*	0.16 *12*	**0.00** *14*	**0.01** *4*	0.25 *12*	0.49 *9*	0.74 *2*	0.81 *1*	0.38 *8*	0.70 *7*	0.23 *10*	0.12 *13*		0.44 *12*	0.43 *9*	0.81 *2*
FIGARCH	0.58 *11*	0.16 *12*	**0.00** *12*	**0.00** *10*	0.22 *13*	0.23 *10*	0.57 *5*	0.57 *6*	0.67 *2*	0.63 *11*	0.43 *4*	0.69 *1*		0.63 *8*	0.73 *7*	0.57 *9*
EGARCH	0.86 *4*	0.23 *11*	**0.00** *7*	**0.00** *7*	0.32 *11*	0.14 *12*	**0.04** *12*	0.12 *13*	0.44 *6*	0.37 *14*	**0.00** *14*	0.12 *13*		0.16 *14*	**0.04** *14*	0.12 *13*
APARCH	0.56 *12*	0.37 *8*	**0.00** *10*	**0.00** *11*	0.73 *7*	0.86 *3*	0.73 *4*	0.69 *4*	0.20 *14*	0.70 *7*	0.10 *12*	0.48 *4*		0.70 *6*	0.43 *9*	0.69 *7*
GJR	0.90 *2*	0.49 *5*	**0.00** *7*	**0.00** *11*	0.63 *9*	0.94 *2*	0.57 *5*	0.81 *1*	0.46 *5*	0.70 *7*	0.07 *13*	0.32 *8*		0.56 *10*	0.23 *13*	0.93 *1*
ARMA-RV	0.62 *7*	0.78 *2*	**0.00** *10*	**0.00** *11*	0.79 *5*	0.81 *4*	0.42 *8*	0.57 *6*	0.30 *10*	0.89 *4*	0.32 *8*	0.48 *4*		0.65 *7*	0.92 *1*	0.57 *9*
ARFIMA-RV	0.79 *5*	0.27 *10*	**0.02** *1*	**0.01** *4*	0.79 *5*	0.81 *4*	0.42 *8*	0.20 *12*	0.38 *8*	0.43 *13*	0.43 *4*	0.20 *11*		0.49 *11*	0.74 *3*	0.81 *2*
HAR-RV	0.62 *7*	0.56 *4*	**0.01** *4*	**0.01** *4*	0.84 *2*	0.81 *4*	0.57 *7*	0.57 *6*	0.26 *11*	0.94 *3*	0.23 *10*	0.48 *4*		0.86 *3*	0.74 *3*	0.81 *2*
HAR-RPV	0.62 *7*	0.43 *6*	**0.01** *4*	**0.00** *7*	0.84 *2*	0.81 *4*	0.19 *11*	0.57 *6*	0.26 *11*	0.98 *1*	0.57 *2*	0.48 *4*		0.94 *1*	0.74 *3*	0.81 *2*
HAR-RV-G	0.62 *7*	0.32 *9*	**0.01** *2*	**0.02** *3*	0.84 *2*	0.73 *8*	0.74 *2*	0.57 *6*	0.26 *11*	0.98 *1*	0.32 *8*	0.32 *8*		0.78 *5*	0.42 *11*	0.81 *2*
AHAR-RPV-G	0.96 *1*	0.43 *6*	**0.00** *7*	**0.00** *7*	0.90 *1*	0.98 *1*	0.92 *1*	0.81 *1*	0.51 *4*	0.70 *7*	0.43 *4*	0.32 *8*		0.43 *13*	0.32 *12*	0.69 *7*
GARCH-RV	0.73 *6*	0.94 *1*	**0.01** *2*	0.20 *1*	0.34 *10*	**0.05** *13*	**0.02** *13*	0.09 *14*	0.42 *7*	0.81 *5*	0.57 *3*	0.69 *1*		0.89 *2*	0.74 *3*	0.36 *11*
GARCH-RPV	0.90 *2*	0.70 *3*	**0.01** *4*	0.20 *1*	0.73 *7*	0.23 *11*	**0.02** *13*	0.36 *11*	0.56 *3*	0.78 *6*	0.90 *1*	0.69 *1*		0.86 *3*	0.92 *1*	0.09 *14*

This table presents the p-values for the unconditional coverage test for the alternative VaR models for a 10%, 5%, 1% and 0.5% VaR coverage level. The relative performance ranking of each VaR model is shown in italics and is based on the respective p-values. The bold faced fonts denote rejection of the null hypothesis of correct unconditional coverage at a 5% significance level. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 5 Conditional Coverage test (p-values)

	Normal				Skewed Student				FHS				EVT			
	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%
EWMA	0.08 *14*	**0.04** *14*	**0.00** *13*	**0.00** *11*	**0.01** *14*	0.09 *13*	0.30 *11*	0.87 *4*	0.59 *4*	0.63 *11*	0.40 *9*	0.26 *11*		0.63 *10*	0.44 *11*	0.43 *12*
GARCH	0.66 *11*	0.36 *11*	**0.00** *14*	**0.03** *4*	0.42 *11*	0.46 *9*	0.80 *2*	0.93 *1*	0.68 *6*	0.88 *7*	0.35 *10*	0.26 *13*		0.50 *11*	0.56 *8*	0.93 *2*
FIGARCH	0.83 *8*	0.36 *11*	**0.00** *12*	**0.01** *10*	0.46 *10*	0.08 *14*	0.66 *6*	0.82 *6*	0.91 *11*	0.85 *11*	0.56 *4*	0.87 *1*		0.68 *9*	0.75 *6*	0.82 *9*
EGARCH	0.19 *13*	0.31 *13*	**0.01** *7*	**0.01** *7*	**0.04** *13*	0.32 *11*	0.08 *12*	0.26 *13*	0.07 *13*	0.48 *14*	**0.01** *14*	0.26 *13*		0.32 *14*	0.08 *14*	0.26 *13*
APARCH	0.74 *9*	0.67 *4*	**0.00** *10*	**0.00** *11*	0.93 *6*	0.89 *6*	0.75 *4*	0.87 *4*	0.42 *12*	0.88 *7*	0.18 *12*	0.72 *4*		0.50 *12*	0.56 *8*	0.87 *7*
GJR	0.71 *10*	0.60 *6*	**0.01** *7*	**0.00** *11*	0.88 *7*	0.62 *8*	0.66 *6*	0.93 *1*	0.48 *14*	0.92 *7*	0.13 *13*	0.56 *8*		0.41 *13*	0.35 *13*	0.95 *1*
ARMA-RV	0.88 *4*	0.83 *2*	**0.00** *10*	**0.00** *11*	0.95 *4*	0.92 *2*	0.63 *8*	0.82 *6*	0.57 *3*	0.93 *4*	0.45 *7*	0.72 *4*		0.88 *4*	0.82 *1*	0.82 *9*
ARFIMA-RV	0.95 *2*	0.55 *7*	**0.03** *1*	**0.03** *4*	0.95 *4*	0.96 *1*	0.63 *8*	0.43 *12*	0.66 *7*	0.72 *13*	0.56 *4*	0.39 *11*		0.77 *7*	0.80 *3*	0.93 *2*
HAR-RV	0.84 *5*	0.67 *5*	**0.01** *5*	**0.03** *4*	0.95 *1*	0.92 *2*	0.73 *5*	0.82 *6*	0.48 *8*	0.90 *3*	0.35 *10*	0.72 *4*		0.87 *5*	0.80 *3*	0.93 *2*
HAR-RPV	0.84 *5*	0.54 *8*	**0.01** *5*	**0.01** *7*	0.95 *1*	0.92 *2*	0.38 *10*	0.82 *6*	0.48 *9*	0.92 *1*	0.66 *3*	0.72 *4*		0.90 *3*	0.80 *3*	0.93 *2*
HAR-RV-G	0.84 *5*	0.42 *10*	**0.02** *3*	0.05 *3*	0.95 *1*	0.91 *5*	0.80 *2*	0.82 *6*	0.48 *5*	0.92 *1*	0.45 *7*	0.56 *8*		0.83 *6*	0.63 *7*	0.93 *2*
AHAR-RPV-G	0.37 *12*	0.54 *8*	**0.01** *7*	**0.01** *7*	0.40 *12*	0.64 *7*	0.82 *1*	0.93 *1*	0.23 *10*	0.78 *7*	0.56 *4*	0.56 *8*		0.70 *8*	0.45 *10*	0.87 *7*
GARCH-RV	0.93 *3*	0.90 *1*	**0.03** *2*	0.39 *1*	0.61 *9*	0.13 *12*	0.06 *13*	0.24 *14*	0.72 *1*	0.92 *5*	0.73 *2*	0.87 *1*		0.93 *2*	0.35 *12*	0.64 *11*
GARCH-RPV	0.97 *1*	0.78 *3*	**0.02** *4*	0.39 *1*	0.78 *8*	0.40 *10*	0.06 *13*	0.64 *11*	0.82 *2*	0.96 *6*	0.81 *1*	0.87 *1*		0.98 *1*	0.82 *1*	0.24 *14*

This table presents the p-values for the conditional coverage test for the alternative VaR models for a 10%, 5%, 1% and 0.5% VaR coverage level. The relative performance ranking of each VaR model is shown in italics and is based on the respective p-values. The bold faced fonts indicate rejection of the null hypothesis of correct conditional coverage at a 5% significance level. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 6 Dynamic quantile test (p-values)

	Normal				Skewed Student				FHS				EVT			
	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%
EWMA	0.36 *12*	0.13 *14*	**0.00** *12*	**0.00** *13*	0.10 *14*	0.29 *11*	**0.00** *13*	0.97 *9*	0.70 *9*	0.65 *4*	**0.00** *12*	0.60 *10*		0.42 *8*	**0.00** *13*	0.91 *12*
GARCH	0.59 *10*	0.13 *13*	**0.00** *14*	**0.00** *8*	0.45 *11*	0.13 *14*	0.36 *12*	1.00 *3*	0.76 *8*	0.32 *14*	**0.02** *8*	**0.00** *13*		0.06 *14*	**0.02** *10*	1.00 *1*
FIGARCH	0.73 *7*	0.63 *3*	**0.00** *13*	**0.00** *14*	0.48 *10*	0.21 *13*	0.40 *10*	0.87 *12*	0.80 *7*	0.60 *6*	**0.02** *10*	**0.02** *12*		0.58 *3*	**0.01** *12*	0.89 *12*
EGARCH	0.33 *13*	0.30 *9*	**0.00** *11*	**0.00** *9*	0.31 *13*	0.40 *10*	**0.00** *14*	**0.02** *14*	0.31 *13*	0.61 *5*	**0.00** *14*	**0.00** *14*		0.35 *13*	**0.00** *14*	**0.00** *14*
APARCH	0.74 *6*	0.31 *8*	**0.00** *10*	**0.00** *12*	0.81 *3*	0.28 *12*	0.40 *11*	**0.02** *13*	0.41 *12*	0.53 *8*	**0.02** *9*	0.92 *3*		0.45 *6*	**0.02** *11*	**0.02** *13*
GJR	0.31 *14*	0.15 *12*	**0.00** *9*	**0.00** *10*	0.44 *12*	0.55 *5*	0.63 *9*	1.00 *4*	0.14 *14*	0.50 *9*	**0.00** *13*	0.91 *5*		0.56 *4*	**0.04** *9*	1.00 *2*
ARMA-RV	0.92 *1*	0.69 *2*	**0.00** *8*	**0.00** *10*	0.87 *2*	0.55 *3*	0.96 *4*	1.00 *2*	0.92 *4*	0.66 *3*	0.55 *6*	0.92 *4*		0.44 *7*	0.47 *7*	0.96 *7*
ARFIMA-RV	0.80 *5*	0.46 *5*	**0.01** *5*	**0.00** *6*	0.73 *8*	0.41 *9*	0.89 *5*	0.98 *8*	0.86 *6*	0.37 *13*	**0.02** *11*	0.46 *11*		0.38 *12*	0.35 *8*	0.97 *6*
HAR-RV	0.87 *3*	0.22 *10*	**0.00** *7*	**0.00** *6*	0.76 *4*	0.53 *7*	0.96 *3*	1.00 *6*	0.92 *1*	0.39 *12*	0.44 *7*	0.81 *7*		0.39 *10*	0.80 *4*	0.99 *5*
HAR-RPV	0.87 *4*	0.37 *7*	**0.00** *6*	**0.00** *5*	0.75 *6*	0.53 *8*	0.87 *6*	1.00 *5*	0.92 *2*	0.44 *11*	0.56 *5*	0.82 *6*		0.39 *11*	0.80 *3*	0.99 *4*
HAR-RV-G	0.90 *2*	0.19 *11*	**0.01** *4*	0.21 *3*	0.69 *9*	0.55 *4*	0.99 *1*	0.99 *7*	0.92 *3*	0.45 *10*	0.79 *3*	0.75 *9*		0.40 *9*	0.86 *2*	0.99 *3*
AHAR-RPV-G	0.70 *8*	0.41 *6*	**0.03** *2*	**0.00** *4*	0.74 *7*	0.57 *2*	0.98 *2*	1.00 *1*	0.64 *11*	0.59 *7*	0.79 *4*	0.79 *8*		0.52 *5*	0.74 *5*	0.94 *9*
GARCH-RV	0.63 *9*	0.95 *1*	0.06 *1*	0.67 *2*	0.89 *1*	0.54 *6*	0.73 *7*	0.92 *11*	0.87 *5*	0.78 *2*	0.96 *1*	0.99 *1*		0.90 *1*	0.52 *6*	0.95 *8*
GARCH-RPV	0.57 *11*	0.58 *4*	**0.03** *3*	0.83 *1*	0.75 *5*	0.90 *1*	0.72 *8*	0.95 *10*	0.65 *10*	0.78 *1*	0.79 *2*	0.99 *2*		0.83 *2*	0.90 *1*	0.91 *11*

This table presents the p-values for the dynamic quantile test for the alternative VaR models for a 10%, 5%, 1% and 0.5% VaR coverage level. The relative performance ranking of each VaR model is shown in italics and is based on the respective p-values. The bold faced fonts denote rejection of the null hypothesis of correct VaR estimates at a 5% significance level. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 7 Average Quadratic Loss Function (AQLF)

	Normal				Skewed Student				FHS				EVT			
	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%
EWMA	21.60 *13*	10.81 *13*	2.85 *11*	1.69 *11*	22.85 *14*	10.36 *14*	1.83 *13*	0.89 *13*	17.92 *7*	8.41 *7*	1.65 *5*	1.03 *10*		**7.38** ***1***	1.55 *9*	0.58 *12*
GARCH	21.60 *14*	10.95 *14*	3.21 *14*	1.61 *9*	22.06 *13*	9.79 *12*	1.42 *8*	0.74 *10*	19.65 *11*	9.33 *13*	1.90 *11*	1.14 *13*		8.15 *8*	1.75 *10*	0.75 *10*
FIGARCH	20.74 *12*	10.45 *12*	2.77 *10*	1.58 *8*	21.69 *12*	9.68 *11*	1.54 *10*	0.70 *8*	19.94 *12*	9.24 *11*	1.62 *4*	0.87 *4*		8.94 *12*	1.51 *8*	0.67 *5*
EGARCH	19.96 *10*	10.10 *9*	2.72 *9*	1.62 *10*	20.86 *11*	10.17 *13*	2.24 *14*	1.20 *14*	21.79 *14*	9.47 *14*	2.56 *14*	1.19 *14*		9.43 *14*	2.24 *14*	1.17 *14*
APARCH	19.69 *9*	10.21 *10*	2.89 *12*	1.78 *12*	19.41 *9*	8.97 *9*	1.65 *11*	0.87 *12*	19.00 *10*	8.77 *10*	2.10 *12*	0.96 *7*		8.72 *10*	1.81 *12*	0.88 *12*
GJR	20.52 *11*	10.43 *11*	2.98 *13*	1.85 *13*	20.82 *10*	9.35 *10*	1.74 *12*	0.79 *11*	20.09 *13*	9.33 *12*	2.21 *13*	1.10 *12*		9.20 *13*	2.03 *13*	0.89 *13*
ARMA-RV	**17.32** ***1***	8.15 *2*	2.37 *4*	1.85 *13*	17.62 *3*	7.54 *3*	1.16 *4*	0.60 *4*	**16.43** ***1***	7.69 *2*	1.68 *8*	0.87 *3*		7.48 *2*	1.31 *3*	0.63 *4*
ARFIMA-RV	18.95 *8*	9.20 *7*	2.30 *3*	1.41 *4*	19.16 *8*	8.03 *7*	1.24 *5*	0.54 *3*	18.09 *9*	8.47 *9*	1.66 *6*	1.06 *11*		8.46 *9*	1.37 *5*	0.71 *9*
HAR-RV	18.09 *5*	8.71 *4*	2.46 *7*	1.41 *4*	18.53 *6*	7.88 *6*	1.31 *6*	0.64 *6*	17.22 *5*	8.08 *6*	1.77 *10*	0.95 *5*		8.13 *6*	1.38 *6*	0.71 *7*
HAR-RPV	18.05 *4*	8.81 *5*	2.45 *6*	1.49 *6*	18.49 *5*	7.88 *5*	1.12 *3*	0.64 *7*	17.21 *4*	8.04 *5*	1.61 *3*	0.95 *6*		8.08 *5*	1.38 *7*	0.71 *8*
HAR-RV-G	17.95 *2*	8.84 *6*	2.38 *5*	1.37 *3*	18.37 *4*	7.78 *4*	1.36 *7*	0.64 *5*	17.01 *3*	8.00 *4*	1.71 *9*	0.98 *8*		8.13 *7*	1.26 *2*	0.70 *6*
AHAR-RPV-G	18.76 *6*	9.29 *8*	2.62 *8*	1.56 *7*	18.79 *7*	8.39 *8*	1.49 *9*	0.73 *9*	18.09 *8*	8.43 *8*	1.68 *7*	1.00 *9*		8.82 *11*	1.77 *11*	0.85 *11*
GARCH-RV	18.02 *3*	**8.15** ***1***	**2.15** ***1***	**1.02** ***1***	**15.59** ***1***	**5.98** ***1***	**0.75** ***1***	**0.42** ***1***	16.94 *2*	**7.42** ***1***	**1.19** ***1***	**0.80** ***1***		7.49 *3*	**1.26** ***1***	0.56 *2*
GARCH-RPV	18.81 *7*	8.61 *3*	2.24 *2*	1.04 *2*	17.22 *2*	6.74 *2*	0.79 *2*	0.53 *2*	17.66 *6*	7.85 *3*	1.36 *2*	0.82 *2*		7.86 *4*	1.33 *4*	**0.46** ***1***

This table presents the Average QLF for the alternative VaR models for a 10%, 5%, 1% and 0.5% VaR coverage level. The relative performance ranking of each VaR model is shown in italics. The bold faced fonts denote the model ranking first. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 8 Quadratic Loss Function relative to the number of exceptions

	Normal				Skewed Student				FHS				EVT			
	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%
EWMA	1.88 *6*	1.72 *9*	1.46 *9*	1.50 *8*	1.91 *8*	1.69 *9*	1.42 *3*	1.57 *8*	1.83 *3*	1.60 *5*	1.40 *7*	1.44 *3*		1.56 *3*	1.37 *4*	1.89 *12*
GARCH	2.06 *13*	1.92 *13*	1.49 *10*	1.65 *14*	2.04 *14*	1.83 *13*	1.54 *13*	1.59 *10*	2.08 *13*	1.80 *14*	1.48 *12*	1.49 *7*		1.76 *14*	1.48 *8*	1.62 *10*
FIGARCH	2.00 *11*	1.83 *11*	1.46 *8*	1.47 *6*	2.00 *12*	1.73 *10*	**1.37** *1*	1.70 *12*	2.05 *10*	1.76 *12*	1.37 *4*	1.54 *11*		1.71 *12*	1.40 *5*	1.64 *11*
EGARCH	1.97 *10*	1.80 *10*	1.56 *12*	1.58 *11*	1.95 *9*	1.77 *12*	1.50 *7*	1.55 *4*	2.06 *11*	1.74 *11*	1.47 *11*	1.55 *12*		1.65 *10*	1.50 *11*	1.52 *3*
APARCH	2.05 *12*	1.87 *12*	1.60 *13*	1.58 *10*	1.99 *11*	1.76 *11*	1.53 *10*	1.53 *3*	2.07 *12*	1.69 *10*	1.51 *13*	1.55 *13*		1.68 *11*	1.53 *12*	1.56 *8*
GJR	2.07 *14*	1.95 *14*	1.71 *14*	1.63 *12*	2.02 *13*	1.86 *14*	1.54 *11*	1.70 *13*	2.11 *14*	1.80 *13*	1.53 *14*	1.65 *14*		1.74 *13*	1.58 *14*	1.72 *12*
ARMA	**1.79** *1*	**1.59** *1*	1.32 *2*	1.63 *12*	1.80 *3*	1.54 *3*	1.41 *2*	1.47 *2*	**1.76** *1*	1.56 *3*	1.36 *3*	**1.41** *1*		1.57 *4*	**1.34** *1*	1.53 *4*
ARFIMA	1.93 *9*	1.66 *6*	1.44 *4*	1.45 *3*	1.95 *10*	1.65 *7*	1.51 *8*	1.74 *14*	1.92 *9*	1.57 *4*	1.40 *8*	1.47 *6*		1.58 *6*	1.48 *7*	1.54 *7*
HAR-RV	1.87 *5*	1.65 *5*	1.45 *6*	1.45 *3*	1.88 *6*	1.61 *6*	1.50 *6*	1.56 *6*	1.86 *7*	1.61 *6*	1.38 *5*	1.54 *9*		1.60 *7*	1.49 *9*	1.53 *5*
HAR-RPV	1.87 *4*	1.63 *4*	1.45 *5*	1.45 *5*	1.87 *5*	1.61 *5*	1.55 *14*	1.56 *7*	1.85 *6*	1.61 *8*	1.42 *10*	1.54 *10*		1.60 *8*	1.49 *10*	1.54 *6*
HAR-RV-G	1.86 *3*	1.61 *2*	1.45 *7*	1.48 *7*	1.86 *4*	1.61 *4*	1.47 *5*	1.55 *5*	1.83 *5*	1.61 *7*	1.39 *6*	1.47 *5*		1.58 *5*	1.54 *13*	1.52 *2*
AHAR-RPV-G	1.88 *7*	1.72 *8*	1.50 *11*	1.52 *9*	1.89 *7*	1.68 *8*	1.52 *9*	1.58 *9*	1.89 *8*	1.62 *9*	1.42 *9*	1.50 *8*		1.63 *9*	1.43 *6*	**1.50** *1*
GARCH-RV	1.85 *2*	1.62 *3*	**1.31** *1*	**1.42** *1*	**1.67** *1*	**1.47** *1*	1.46 *4*	1.62 *11*	1.79 *2*	**1.52** *1*	1.36 *2*	1.42 *2*		**1.52** *1*	**1.36** *2*	1.56 *9*
GARCH-RPV	1.90 *8*	1.66 *7*	1.32 *3*	1.45 *2*	1.76 *2*	1.53 *2*	1.54 *12*	**1.46** *1*	1.83 *4*	1.53 *2*	**1.32** *1*	1.44 *4*		1.55 *2*	1.36 *3*	1.80 *13*

This table presents the QLF relative to the number of exceptions for the alternative VaR models for a 10%, 5%, 1% and 0.5% VaR coverage level. The relative performance ranking of each VaR model is shown in italics. The bold faced fonts denote the model ranking first. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 9 Mean Relative Scaled Bias (MRSB)

	Normal				Skewed Student				FHS				EVT			
	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%	10%	5%	1%	0.5%
EWMA	0.239 *13*	0.253 *14*	0.235 *10*	0.238 *9*	0.236 *14*	0.248 *13*	0.226 *10*	0.217 *9*	0.241 *14*	0.241 *13*	0.240 *12*	0.227 *8*		0.245 *12*	0.255 *14*	0.226 *12*
GARCH	0.227 *12*	0.228 *11*	0.230 *9*	0.290 *12*	0.233 *13*	0.238 *12*	0.230 *11*	0.246 *11*	0.235 *13*	0.234 *10*	0.256 *14*	0.208 *6*		0.245 *11*	0.248 *12*	0.265 *13*
FIGARCH	0.218 *10*	0.242 *13*	0.259 *13*	0.295 *13*	0.210 *7*	0.249 *14*	0.243 *12*	0.263 *12*	0.226 *11*	0.251 *14*	0.232 *10*	0.235 *9*		0.255 *14*	0.249 *13*	0.270 *14*
EGARCH	0.246 *14*	0.227 *10*	0.240 *11*	0.233 *8*	0.227 *11*	0.237 *11*	0.295 *14*	0.275 *13*	0.227 *12*	0.240 *12*	0.250 *13*	0.263 *14*		0.238 *10*	0.236 *11*	0.261 *12*
APARCH	0.205 *5*	0.213 *8*	0.250 *12*	0.321 *14*	0.198 *1*	0.222 *10*	0.226 *9*	0.234 *10*	0.194 *1*	0.240 *11*	0.209 *7*	0.202 *5*		0.233 *9*	0.215 *8*	0.208 *7*
GJR	0.222 *11*	0.232 *12*	0.264 *14*	0.256 *10*	0.230 *12*	0.221 *9*	0.263 *13*	0.276 *14*	0.224 *10*	0.225 *9*	0.233 *11*	0.199 *4*		0.249 *13*	0.220 *9*	0.253 *11*
ARMA	0.208 *6*	0.209 *6*	0.185 *3*	0.256 *10*	0.208 *6*	0.204 *7*	0.204 *6*	0.208 *7*	0.199 *3*	0.195 *6*	0.213 *9*	0.170 *3*		0.187 *3*	0.200 *6*	0.204 *5*
ARFIMA	**0.198** ***1***	0.204 *5*	0.214 *8*	**0.130** ***1***	**0.200** ***2***	0.200 *3*	0.194 *5*	0.203 *6*	**0.198** ***2***	0.213 *8*	0.191 *3*	0.218 *7*		0.206 *8*	0.188 *3*	0.194 *3*
HAR-RV	0.199 *2*	0.195 *4*	0.206 *7*	**0.130** ***1***	0.201 *4*	0.202 *5*	0.189 *4*	0.199 *5*	0.202 *4*	0.191 *3*	0.202 *5*	0.259 *13*		0.191 *6*	0.199 *5*	0.205 *6*
HAR-RPV	0.202 *3*	0.193 *3*	0.205 *6*	0.176 *6*	0.200 *3*	0.201 *4*	0.206 *7*	0.188 *3*	0.203 *5*	**0.189** ***1***	0.201 *4*	0.257 *12*		0.187 *4*	0.200 *7*	0.234 *9*
HAR-RV-G	0.204 *4*	0.212 *7*	0.199 *5*	0.165 *5*	0.204 *5*	0.203 *6*	0.187 *3*	0.195 *4*	0.210 *7*	0.191 *4*	0.211 *8*	0.243 *11*		0.190 *5*	0.195 *4*	0.200 *4*
AHAR-RPV-G	0.208 *7*	0.217 *9*	0.198 *4*	0.201 *7*	0.211 *8*	0.207 *8*	0.209 *8*	0.214 *8*	0.205 *6*	0.208 *7*	0.209 *6*	0.237 *10*		0.206 *7*	0.236 *10*	0.240 *10*
GARCH-RV	0.213 *9*	**0.184** ***1***	0.162 *2*	0.163 *4*	0.224 *10*	**0.180** ***1***	0.170 *2*	**0.121** ***1***	0.215 *8*	0.189 *2*	**0.173** ***1***	**0.140** ***1***		0.185 *2*	**0.176** ***1***	0.122 *2*
GARCH-RPV	0.210 *8*	0.190 *2*	**0.151** ***1***	0.144 *3*	0.219 *9*	0.187 *2*	**0.160** ***1***	0.162 *2*	0.222 *9*	0.192 *5*	0.180 *2*	0.141 *2*		**0.182** ***1***	0.182 *2*	**0.119** ***1***

This table presents the Mean Relative Scaled Bias for the alternative VaR models for a 10%, 5%, 1% and 0.5% VaR coverage level. The relative performance ranking of each VaR model is shown in italics. The bold faced fonts denote the model ranking first. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 10 Regulatory Loss Function - Capital requirements

	Normal					Skewed Student					FHS					EVT				
	Green	Yellow	Red	Mean	St.dev	Green	Yellow	Red	Mean	St.dev	Green	Yellow	Red	Mean	St.dev	Green	Yellow	Red	Mean	St.dev
EWMA	**48.3**	**36.9**	**14.9**	**306**	**195**	**68.3**	**31.7**	**0.0**	**312**	**194**	**77.6**	**22.4**	**0.0**	**304**	**181**	**86.4**	**13.6**	**0.0**	**308**	**183**
GARCH	**63.1**	**9.8**	**27.1**	**307**	**178**	75.2	24.8	0.0	318	184	**71.6**	**23.3**	**5.1**	**314**	**189**	**71.6**	**23.3**	**5.1**	**319**	**192**
FIGARCH	**53.8**	**27.3**	**18.9**	**309**	**180**	72.0	28.0	0.0	319	179	**72.4**	**27.6**	**0.0**	**314**	**183**	**72.4**	**27.6**	**0.0**	**318**	**184**
EGARCH	**58.3**	**28.3**	**13.4**	**311**	**179**	**59.7**	**35.2**	**5.1**	**316**	**180**	**50.8**	**42.3**	**6.9**	**312**	**189**	**56.7**	**40.1**	**3.2**	**301**	**163**
APARCH	**63.1**	**13.2**	**23.6**	**304**	**179**	70.4	29.6	0.0	318	180	**63.7**	**34.4**	**1.9**	**303**	**183**	**65.8**	**34.2**	**0.0**	**303**	**178**
GJR	**64.7**	**19.3**	**16.0**	**300**	**174**	**74.9**	**24.9**	**0.2**	**318**	**178**	**71.5**	**21.0**	**7.5**	**306**	**196**	**71.5**	**27.1**	**1.5**	**303**	**182**
ARMA	**63.1**	**22.1**	**14.7**	**305**	**179**	83.2	16.8	0.0	307	160	**64.2**	**33.7**	**2.1**	**312**	**179**	74.9	25.1	0.0	308	175
ARFIMA	**63.8**	**23.1**	**13.1**	**298**	**163**	81.8	18.2	0.0	306	146	**72.9**	**22.0**	**5.1**	**302**	**164**	78.2	21.8	0.0	300	156
HAR-RV	**63.2**	**24.2**	**12.6**	**300**	**168**	77.0	23.0	0.0	313	163	**69.6**	**25.1**	**5.2**	**313**	**185**	77.6	22.4	0.0	306	169
HAR-RPV	**63.2**	**24.2**	**12.6**	**300**	**168**	82.4	17.6	0.0	314	156	**74.6**	**21.0**	**4.3**	**308**	**176**	77.6	22.4	0.0	306	169
HAR-RV-G	**63.1**	**31.6**	**5.2**	**298**	**165**	68.9	31.1	0.0	319	175	66.9	33.1	0.0	312	182	82.3	17.7	0.0	300	164
AHAR-RPV-G	**62.5**	**29.0**	**8.5**	**299**	**169**	71.5	28.5	0.0	312	164	75.1	24.9	0.0	306	174	68.1	31.9	0.0	304	173
GARCH-RV	**63.4**	**30.2**	**6.4**	**298**	**163**	**100.0**	**0.0**	**0.0**	**304**	**150**	80.1	19.9	0.0	295	159	77.7	22.3	0.0	296	157
GARCH-RPV	**59.7**	**34.0**	**6.4**	**295**	**153**	**97.0**	**3.0**	**0.0**	**298**	**141**	76.7	23.3	0.0	295	154	70.5	29.5	0.0	299	159

This table summarizes the capital requirements imposed by the alternative VaR models. For each model the table presents the percentage of days during the out of sample forecasting period that the model is placed in the green, yellow and red zone according to the Basel traffic light system, the average daily capital requirements and its standard deviation. The bold faced fonts denote that the model has been placed in the red zone or it has failed the statistical tests ((un)conditional coverage test and dynamic quantile test), or both of them. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 11 EPA and SPA tests for the Market Risk Capital (MRC) loss function

	% days	EPA-average (p-values)	Median Sign-test (p-values)	SPA (p-values)
GARCH-skst	72.94	**0.00**	**0.00**	**0.00**
FIGARCH-skst	74.76	**0.00**	**0.00**	**0.00**
APARCH-skst	67.69	**0.00**	**0.00**	**0.00**
ARMA-skst	70.58	0.05	**0.00**	**0.00**
ARMA-EVT	79.78	**0.00**	**0.00**	**0.00**
ARFIMA-skst	67.45	0.14	**0.00**	**0.00**
ARFIMA-EVT	58.37	0.94	**0.00**	**0.00**
HAR-RV-skst	73.11	**0.00**	**0.00**	**0.00**
HAR-RV-EVT	78.36	**0.03**	**0.00**	**0.00**
HAR-RPV-skst	71.23	**0.00**	**0.00**	**0.00**
HAR-RPV-EVT	78.13	**0.04**	**0.00**	**0.00**
HAR-RV-G-skst	80.96	**0.00**	**0.00**	**0.00**
HAR-RV-G-FHS	75.83	**0.00**	**0.00**	**0.00**
HAR-RV-G-EVT	66.51	0.99	**0.00**	**0.00**
AHAR-RPV-G-skst	69.87	**0.00**	**0.00**	**0.00**
AHAR-RPV-G-FHS	77.30	**0.00**	**0.00**	**0.00**
AHAR-RPV-G-EVT	-	-	-	**0.00**
GARCH-RV-FHS	56.78	1.00	**0.00**	0.63
GARCH-RV-EVT	61.50	1.00	**0.00**	**0.00**
GARCH-RPV-FHS	60.44	1.00	**0.00**	0.45
GARCH-RPV-EVT	64.68	0.99	**0.00**	**0.00**

The first column depicts the volatility models and the VaR method used. The bold faced p-values denote rejection of the null hypothesis at a 5% significance level.

Table 12 EPA and SPA tests for the Firm Loss Function (FLF)

	Average FLF	% days	EPA-average (p-values)	Median Sign-test (p-values)	SPA (p-values)
GARCH-skst	4.47	66.03	**0.00**	**0.00**	**0.00**
FIGARCH-skst	4.48	66.29	**0.00**	**0.00**	**0.00**
APARCH-skst	4.45	68.55	**0.00**	**0.00**	**0.00**
ARMA-skst	4.44	72.15	**0.00**	**0.00**	**0.00**
ARMA-EVT	4.40	64.39	**0.00**	**0.00**	**0.00**
ARFIMA-skst	4.43	73.84	**0.00**	**0.00**	**0.00**
ARFIMA-EVT	4.34	63.16	**0.00**	**0.00**	**0.00**
HAR-RV-skst	4.45	73.84	**0.00**	**0.00**	**0.00**
HAR-RV-EVT	4.38	66.34	**0.00**	**0.00**	**0.00**
HAR-RPV-skst	4.55	78.52	**0.00**	**0.00**	**0.00**
HAR-RPV-EVT	4.38	65.88	**0.00**	**0.00**	**0.00**
HAR-RV-G-skst	4.43	74.72	**0.00**	**0.00**	**0.00**
HAR-RV-G-FHS	4.28	62.49	0.17	**0.00**	**0.00**
HAR-RV-G-EVT	4.38	67.06	**0.00**	**0.00**	**0.00**
AHAR-RPV-G-skst	4.39	75.13	**0.00**	**0.00**	**0.00**
AHAR-RPV-G-FHS	4.33	79.09	**0.00**	**0.00**	**0.00**
AHAR-RPV-G-EVT	4.26	-	-	-	**0.02**
GARCH-RV-FHS	4.23	60.02	0.80	**0.00**	**0.00**
GARCH-RV-EVT	4.23	60.43	0.77	**0.00**	**0.00**
GARCH-RPV-FHS	4.19	58.63	0.97	**0.00**	0.24
GARCH-RPV-EVT	4.19	58.89	0.97	**0.00**	0.76

The first column depicts the volatility models and the VaR method used. The bold faced p-values denote rejection of the null hypothesis at a 5% significance level.

Table 13 1% VaR forecasting during the latest financial crisis period (07.01.2007 - 09.30.2009)

	Skewed Student						FHS						EVT					
	FR	AQLF	UC	CC	DQ	%Red	FR	AQLF	UC	CC	DQ	%Red	FR	AQLF	UC	CC	DQ	%Red
EWMA	1.60_{6}	2.14_{13}	0.19_{6}	0.36_{6}	$\mathbf{0.01}_{10}$	0.0	1.6_{3}	1.66_{5}	0.19_{3}	0.19_{4}	$\mathbf{0.03}_{4}$	0.0	1.24_{2}	1.56_{9}	0.58_{2}	0.78_{2}	$\mathbf{0.03}_{9}$	0.0
GARCH	1.78_{9}	1.42_{8}	0.10_{9}	0.21_{9}	$\mathbf{0.01}_{11}$	0.0	2.5_{11}	1.90_{11}	$\mathbf{0.00}_{11}$	$\mathbf{0.01}_{11}$	$\mathbf{0.00}_{12}$	15.3	2.49_{13}	1.76_{10}	$\mathbf{0.00}_{13}$	$\mathbf{0.01}_{13}$	$\mathbf{0.00}_{14}$	15.3
FIGARCH	1.60_{6}	1.54_{10}	0.19_{6}	0.36_{6}	$\mathbf{0.01}_{9}$	0.0	2_{5}	1.63_{4}	$\mathbf{0.04}_{5}$	0.11_{5}	$\mathbf{0.00}_{11}$	0.0	1.95_{10}	1.51_{8}	$\mathbf{0.04}_{10}$	0.11_{10}	$\mathbf{0.00}_{13}$	0.0
EGARCH	2.66_{14}	2.24_{14}	$\mathbf{0.00}_{14}$	$\mathbf{0.00}_{14}$	$\mathbf{0.00}_{14}$	15.3	3_{13}	2.57_{14}	$\mathbf{0.00}_{11}$	$\mathbf{0.00}_{13}$	$\mathbf{0.00}_{14}$	20.8	2.31_{12}	2.25_{14}	$\mathbf{0.01}_{12}$	$\mathbf{0.02}_{12}$	$\mathbf{0.00}_{12}$	9.8
APARCH	2.13_{12}	1.65_{11}	$\mathbf{0.02}_{12}$	$\mathbf{0.05}_{12}$	$\mathbf{0.00}_{13}$	0.0	2.7_{12}	2.11_{12}	$\mathbf{0.00}_{11}$	$\mathbf{0.00}_{12}$	$\mathbf{0.00}_{10}$	5.9	2.13_{11}	1.82_{12}	$\mathbf{0.02}_{11}$	$\mathbf{0.05}_{11}$	$\mathbf{0.00}_{10}$	0.0
GJR	2.31_{13}	1.74_{12}	$\mathbf{0.01}_{13}$	$\mathbf{0.02}_{13}$	$\mathbf{0.00}_{12}$	0.7	3_{13}	2.22_{13}	$\mathbf{0.00}_{11}$	$\mathbf{0.00}_{13}$	$\mathbf{0.00}_{13}$	22.7	2.66_{14}	2.04_{13}	$\mathbf{0.00}_{14}$	$\mathbf{0.00}_{14}$	$\mathbf{0.00}_{11}$	4.4
ARMA	1.42_{5}	1.16_{4}	0.35_{5}	0.57_{5}	0.79_{5}	0.0	2.3_{9}	1.69_{8}	$\mathbf{0.01}_{9}$	$\mathbf{0.02}_{9}$	$\mathbf{0.00}_{6}$	6.4	1.42_{4}	1.32_{3}	0.35_{4}	0.57_{4}	0.57_{4}	0.0
ARFIMA	1.24_{4}	1.24_{5}	0.58_{4}	0.78_{4}	0.85_{4}	0.0	2.1_{6}	1.66_{6}	$\mathbf{0.02}_{6}$	$\mathbf{0.05}_{6}$	$\mathbf{0.00}_{8}$	15.5	1.24_{2}	1.37_{5}	0.58_{2}	0.78_{2}	0.64_{3}	0.0
HAR-RV	1.60_{6}	1.31_{6}	0.19_{6}	0.36_{6}	0.49_{6}	0.0	2.3_{9}	1.77_{10}	$\mathbf{0.01}_{9}$	$\mathbf{0.02}_{9}$	$\mathbf{0.00}_{9}$	15.8	1.42_{4}	1.38_{6}	0.35_{4}	0.57_{4}	0.46_{7}	0.0
HAR-RPV	1.07_{1}	1.12_{3}	0.88_{1}	0.93_{1}	0.96_{3}	0.0	2.1_{6}	1.62_{3}	$\mathbf{0.02}_{6}$	$\mathbf{0.05}_{6}$	$\mathbf{0.00}_{7}$	13.0	1.42_{4}	1.39_{7}	0.35_{4}	0.57_{4}	0.47_{6}	0.0
HAR-RV-G	1.78_{9}	1.36_{7}	0.10_{9}	0.21_{9}	0.44_{7}	0.0	2.1_{6}	1.72_{9}	$\mathbf{0.02}_{6}$	$\mathbf{0.05}_{6}$	$\mathbf{0.03}_{5}$	0.0	1.07_{1}	1.27_{2}	0.88_{1}	0.93_{1}	0.95_{1}	0.0
AHAR-RPV-G	1.78_{9}	1.49_{9}	0.10_{9}	0.21_{9}	0.27_{8}	0.0	1.6_{3}	1.69_{7}	0.19_{3}	0.36_{3}	0.45_{3}	0.0	1.78_{9}	1.78_{11}	0.10_{9}	0.21_{9}	0.27_{8}	0.0
GARCH-RV	0.89_{2}	0.75_{1}	0.79_{2}	0.92_{2}	0.99_{2}	0.0	1.4_{1}	1.20_{1}	0.35_{1}	0.57_{1}	0.67_{1}	0.0	1.42_{4}	1.26_{1}	0.35_{4}	0.57_{4}	0.66_{2}	0.0
GARCH-RPV	0.89_{2}	0.79_{2}	0.79_{2}	0.92_{2}	0.99_{1}	0.0	1.4_{1}	1.36_{2}	0.35_{1}	0.57_{1}	0.66_{2}	0.0	1.60_{8}	1.33_{4}	0.19_{8}	0.36_{8}	0.53_{5}	0.0

This table summarizes the failure rate (FR), the Average QLF, the p-values for the unconditional coverage (UC), the conditional coverage (CC) and dynamic quantile (DQ) test results and the percentage of days in red zone (%Red). The bold faced numbers for the UC, CC and DQ denote rejection of the null at a 5% significance level. The HAR-RV-G and the AHAR-RPV-G are the HAR-RV and the Asymmetric HAR-RPV models respectively, with a GARCH specification for their residuals.

Table 14 EPA and SPA tests for the MRC and FLF during the latest financial crisis period (07.01.2007 - 09.30.2009)

	MRC					FLF				
	Average	% days	EPA-average (p-values)	Median Sign-test (p-values)	SPA (p-values)	Average	% days	EPA-average (p-values)	Median Sign-test (p-values)	SPA (p-values)
ARMA-skst	484	25	1.00	1.00	**0.00**	6.66	45	0.50	0.98	**0.00**
ARMA-EVT	509	48	0.66	0.83	**0.00**	6.83	52	**0.01**	0.21	**0.00**
ARFIMA-skst	478	12	1.00	1.00	**0.00**	6.56	48	0.94	0.81	**0.00**
ARFIMA-EVT	487	25	1.00	1.00	**0.00**	6.61	51	0.79	0.32	**0.00**
HAR-RV-skst	496	29	1.00	1.00	**0.00**	6.60	47	0.88	0.92	**0.00**
HAR-RV-EVT	503	49	0.99	0.74	**0.00**	6.74	53	0.07	0.11	**0.00**
HAR-RPV-skst	488	24	1.00	1.00	**0.00**	6.77	55	**0.02**	**0.01**	**0.00**
HAR-RPV-EVT	503	50	0.98	0.58	**0.00**	6.74	53	0.07	0.10	**0.00**
HAR-RV-G-skst	488	53	0.07	0.08	**0.00**	6.60	50	0.89	0.48	**0.00**
HAR-RV-G-EVT	487	19	1.00	1.00	**0.00**	6.75	56	**0.04**	**0.00**	**0.00**
AHAR-RPV-G-skst	506	20	1.00	1.00	**0.00**	6.57	16	1.00	1.00	**0.00**
AHAR-RPV-G-FHS	510	66	0.54	**0.00**	**0.00**	6.75	62	**0.00**	**0.00**	**0.00**
AHAR-RPV-G-EVT	510	-	-	-	**0.00**	6.66	-	-	-	**0.00**
GARCH-RV-skst	469	8	1.00	1.00	**0.00**	6.93	60	**0.01**	**0.00**	**0.00**
GARCH-RV-FHS	469	12	1.00	1.00	**0.00**	6.27	35	1.00	1.00	**0.00**
GARCH-RV-EVT	467	8	1.00	1.00	**0.00**	6.23	34	1.00	1.00	**0.00**
GARCH-RPV-skst	451	0	1.00	1.00	0.55	6.51	49	0.94	0.68	**0.00**
GARCH-RPV-FHS	469	21	1.00	1.00	**0.00**	6.16	35	1.00	1.00	**0.00**
GARCH-RPV-EVT	478	18	1.00	1.00	**0.00**	6.06	32	1.00	1.00	0.72