The Role of High-Frequency Prices, Long Memory and Jumps for Value-at-Risk Predictions

Ana-Maria Fuertes[a,*], Jose Olmo[b]

[a] *Faculty of Finance, Cass Business School, City University London, U.K.*
[b] *Centro Universitario de la Defensa, Zaragoza, Spain*

May 2012

Abstract

This study investigates the practical importance of several VaR modeling and forecasting issues in the context of intraday stock returns. Value-at-Risk (VaR) predictions obtained from daily GARCH models extended with additional information such as the realized volatility and squared overnight returns, are confronted with those from ARFIMA realized volatility models. The out-of-sample evaluation is based on a novel difference-in-proportions test that exploits the frequency of individual VaR rejections and a block-bootstrap unconditional coverage test that is robust to estimation uncertainty and model risk. ARFIMA models produce better backtesting results than GARCH models but fare worse in terms of independence of the hits sequence. Encompassing tests further suggest that GARCH and ARFIMA models can be fruitfully combined to produce more competitive VaR measures. We find evidence that intraday jumps also have forecasting potential. The techniques are illustrated for a small portfolio of large-cap stocks.

Keywords: Encompassing; High-frequency data; Model uncertainty; Realized volatility; Risk Management

JEL Classification: C52; C53; G15.

*Corresponding author: e-mail: a.fuertes@city.ac.uk. T:+44 (0)20 7040 0186; F: +44 (0)20 7040 8881. We thank Marwan Izzeldin for help with processing the high frequency data. We acknowledge the valuable comments from Jerry Coakley, Carlos da Costa, Jean-Edouard Collard, Elena Kalotychou, Matthew Osborne and participants at the 7th Oxmetrics Conference at Cass Business School, London, in particular, Sir David Hendry and Sébastien Laurent, the 3rd CSDA Conference on Computational and Financial Econometrics in Cyprus, the 28th GdRE Annual International Simposium on Money, Banking and Finance at the University of Reading, and seminar at Cass Business School. Jose Olmo thanks the Spanish Government through project MICINN ECO2011-22650 for financial support.

1 Introduction

Current commercial banks routinely compute and disclose their daily Value-at-Risk (VaR) forecasts of the expected maximum loss over a target horizon (e.g. 1-day, 1-week) at a given confidence level (e.g. 95%, 99%). Despite strong criticisms over its mathematical properties, VaR has become the standard measure of market risk since Basel II.[1] Different VaR approaches are available but a common thread underlying most of them is their reliance on the assumption that returns belong to a location-scale family which implies that VaR is a linear function of the volatility. Relatively simple conditional volatility models in the GARCH class alongside Gaussian or Student-t quantiles remain widely used by banks for daily VaR prediction; e.g. see the recent RiskMetrics methodology (Zumbach, 2007). By assuming that the returns distribution belongs to the location-scale family, there is a direct mapping between volatility forecasts and VaR predictions.

The last decade has witnessed growing theoretical and empirical interest in model-free measures of volatility based on intraday prices. Special efforts have been devoted to try to improve the forecasts from GARCH models based on daily returns by exploiting intraday information. The main rationale for these efforts is that the squared return is an extremely noisy (albeit unbiased) estimator of *ex post* volatility. Several studies show that augmenting the daily GARCH model with the so-called *realized variance* (RV), or the sum of intraday squared returns, affords volatility forecast improvements in a statistical sense (MSE or Mincer-Zarnowitz criteria); see Martens (2001), Blair et al. (2001), Engle (2002) and Koopman et al. (2005) and Fuertes et al. (2009), inter alios.[2] Galbraith and Kisinbay (2002) illustrate that 1-day-ahead forecasts from AR models fitted to RV outperform those from GARCH in a MSE sense. Another interesting contribution is Gallo's (2001) analysis of the overnight news content for daily volatility prediction in the context of 20 large-cap NYSE stocks. By augmenting GARCH models with the squared overnight returns he demonstrates that the after-trading-hours 'surprise' has some conditional volatility forecasting potential according to the MAE criteria but rather less

[1] A critical overview of the Value-at-Risk approach and the Basel II Capital Accord is provided by Sollis (2009) together with examples illustrating the need to develop improved estimation techniques and backtesting procedures.

[2] Several other non-parametric volatility measures based on intraday data have been developed in the theoretical literature, partly, in an attempt to mitigate the bias introduced by market microstructure frictions (bid-ask bounce, screen fighting, price discreteness and irregular trading). Instances are the *realized power variation* that sums powers of the absolute intraday returns, *realized range* or the sum of intraday high-to-low price differences and realized *kernel-based* variance estimators. We direct the reader to McAleer and Medeiros (2008) and Andersen et al. (2009) for comprehensive reviews.

favourable evidence emerges from the RMSE.

There is a recent stream of research on high-frequency volatility modelling for daily VaR prediction. For two assets, DM/$ and Yen/$, Andersen et al. (2003) show that accurate VaRs can be obtained from a long-memory vector autoregression for RV coupled with the assumption of Gaussian standardized returns. Brownlees and Gallo (2010) document for various individual NYSE stocks that multiplicative error models (MEM) for realized volatility, realized bipower variation, two scale realized volatility and realized kernel produce VaR forecasts with better coverage properties than the daily return-based GARCH. In Clements et al. (2008) the information content of intraday FX quotes is exploited through several approaches that include MIxed DAta Sampling (MIDAS) and Heterogeneous Autoregressive (HAR) models, coupled with different methods to compute quantile forecasts; simple AR(5) models for RV coupled with Gaussian quantiles are shown to yield quite competitive VaR predictions. Giot and Laurent (2004) document that VaRs obtained from skewed Student t APARCH models are as adequate as those from ARFIMAX models fitted to daily realized variance. In a multi-period VaR forecasting framework, Louzis et al. (2011) run a horserace among daily range, realized range, realized variance, realized bipower variation, two scale realized variance and implied volatility. The jump-robust realized bipower variation fares quite well in terms of efficient capital allocation. In a similar vein, Shao et al. (2009) provide evidence in favor of the realized range compared to the realized volatility for daily VaR forecasting.

This paper contributes to the literature by shedding light on practical issues regarding how to improve the adequacy of daily VaR predictions in the context of a 7-year sample of intraday prices for a cross-section of 14 NYSE/Nasdaq stocks. As noted by Campbell et al. (2001) and Chen et al. (2012), inter alios, many investors are not fully diversified and maintain large holdings of a few individual stocks; hence, the modeling and forecasting of individual stock (as opposed to market index) volatility is relevant. For this purpose, we consider three distinct risk modeling approaches: i) The standard GARCH model based on daily returns and augmented GARCH versions that exploit the overnight returns or intraday-based realized volatilities, ii) Stochastic models given by ARMA and ARFIMA specifications fitted to logarithmic realized volatilities, iii) A novel naïve equal-weight combination of standard daily-based GARCH and intraday-based ARFIMA forecasts. As intraday measures of volatility, we employ the realized variance and realized bipower variation. The latter was proposed as an alternative to the former that excludes rare large jumps (i.e. extreme outliers) in the log price process. To our best knowledge, this is the first analysis that attempts (albeit indirectly) to decompose the degree of VaR backtesting success into the contributions from modeling the continuous component of log prices and rare extreme jumps. The quantile of the innovation distribution is estimated primarily from the standard Gaussian density but the Student-t density (with d.f. parameter estimated from the standardized returns) is also considered as a robustness check. VaR adequacy is defined both in terms of correct unconditional coverage and independence of the hits sequence. The present analysis departs from the extant literature in adopting a novel robustified version of Kupiec's unconditional backtesting approach, proposed in Escanciano and Olmo (2011), that is robust to estimation uncertainty and model misspecification. We propose as tool to assess relative VaR adequacy a 'panel' difference-in-proportions (DIP) test that is able to exploit the backtesting rejection frequencies obtained over a cross-section of time-series returns (i.e., pertaining to different assets). Last but not least, this is the first study to empirically demonstrate through encompassing tests that the GARCH (daily return based) and ARFIMA (intraday return based) forecasts contain distinct information.

We find that accounting for the slowly decaying empirical autocorrelations of realized volatility through long-memory specifications is not crucial since ARMA models perform as well as ARFIMA in terms of VaR adequacy. Intraday price variation can be useful for daily VaR prediction if appropriately exploited: augmenting the standard GARCH model with realized volatilities does not improve VaR adequacy but a rather effective approach is to combine the standard daily GARCH forecasts and intraday-based AR(FI)MA forecasts. This conclusion stems from robust unconditional coverage and independence tests on the out-of-sample sequence of hits for each of the stocks and, as a whole for the entire cross-section, using the DIP test. Accounting for rare large jumps matters for VaR forecasting but it can be accomplished either through an appropriate choice of realized volatility (e.g. RV that subsumes the jump risks) or through the choice of a fat-tailed density (e.g. Student t) for the quantile computation.

The rest of the paper is organized as follows. Section 2 presents the risk management framework. Section 3 discusses the empirical results, and a final Section 4 concludes.

2 VaR Prediction and Backtesting

Our VaR modeling approach builds upon the contributions of Clements et al. (2008) and Brownlees and Gallo (2010). Let r_t be the daily return at time t. The log return is assumed to follow a pure multiplicative process $r_t = \sqrt{\sigma_t^2}\varepsilon_t$ with $\varepsilon_t \sim F_\varepsilon(\cdot)$, where σ_t^2 is either a GARCH-type conditional variance of the daily return, a realized volatility conditional expectation, or a combination of both; the standardized return ε_t is an *iid* unit variance random variable with probability distribution F_ε. The VaR of r_t is essentially an α-percentage quantile of the conditional distribution of financial returns given the agent's information set Ω_{t-1}. Thus the predicted 1-day-ahead VaR, a measure of the maximum 1-day-ahead loss, is computed as

$$\widehat{VaR}_{t+1,\alpha} \equiv \sqrt{\hat{\sigma}^2_{t+1}(\widehat{\theta}_t)}\hat{F}_\varepsilon^{-1}(\alpha), \tag{1}$$

where $\widehat{\theta}_t$ is a consistent estimator of the parameters required to obtain $\hat{\sigma}^2_{t+1}$, and $\hat{F}_\varepsilon^{-1}(\alpha)$ is an α-quantile estimate.[3] Expression (1) reveals that the adequacy of VaR predictions hinges on two factors: the model chosen to generate the volatility forecasts, and the assumption made for the $\alpha-$quantile computation. Since our main goal is to compare volatility forecasts, for most of the analysis $F_\varepsilon(\cdot)$ is fixed at the standard Gaussian density but we also consider, as a robustness check, the unit-variance Student-t density with degrees of freedom parameter estimated by ML from the standardized returns. The models entertained to obtain $\hat{\sigma}^2_{t+1}(\widehat{\theta}_t)$ are presented next.

2.1 GARCH and AR(FI)MA Models for Volatility Forecasting

The augmented-GARCH class of models can be formalized as

$$r_t = \sqrt{h_t}\varepsilon_t,\ \varepsilon_t \sim iid(0,1) \tag{2a}$$

$$h_t = \omega + \sum_{i=1}^{r}\alpha_i r^2_{t-i} + \sum_{j=1}^{s}\beta_j h_{t-j} + \lambda z_{t-1} \tag{2b}$$

where r_t are daily returns, z_{t-1} is an intraday-based volatility predictor, and the lag orders (r, s) selection criteria is the removal of return volatility clustering according to the ARCH LM test. With $\lambda = 0$, equation (2b) becomes the standard GARCH. The candidates considered for z_{t-1} are the realized variance (RV), realized

[3] Our main focus is the 1-day-ahead prediction of downside tail risk (left quantile), that is, the VaR level for long traders who incur losses when stock prices fall. The 1-day-ahead predictions can be projected several days ahead using any of the existing approaches in the literature (see e.g., Kaplanski and Levy, 2010; Louzis et al., 2011).

bipower variation (RBP), or squared overnight returns. Model estimation is either by QML by assuming Gaussian errors or by ML on the basis of a Student t density.

The *realized variance* is defined as the sum of squared returns over M intraday (length δ) intervals

$$RV_t \equiv \sum_{j=1}^{M} r_{t,j}^2. \tag{3}$$

where $r_{t,j} \equiv log(P_{t,j}) - log(P_{t,j-1})$ denotes the jth intraday return on day t. This estimator converges in probability (as $M \to \infty$) to the quadratic variation process that characterizes the latent true variance, $QV_t \equiv \int_{t-1}^{t} \sigma^2(u)du + \sum_{t-1<j\leq t} k^2(j)$, where the first term is the *integrated variance* (IV_t) that reflects the continuous component of the log price process, and the second term is the discontinuous *jump* component (J_t). Barndorff-Nielsen and Shephard (2004; BN-S) define the *realized bipower variation* as

$$RBP_t \equiv \frac{\pi}{2} \sum_{j=2}^{M} |r_{t,j}|\,|r_{t,j-1}|\,. \tag{4}$$

We use the term "realized volatilities" to refer to both RV and RBP hereafter.[4]

The ARFIMA modeling framework has been successfully employed in the literature to capture the stylized slow (less than exponentially) decay in autocorrelations of daily realized volatilities. We adopt it but, instead of searching for the "best" long-memory specification, we focus on the homoskedastic ARFIMA$(1, d, 0)$ model.[5] The conditional variance of the daily return process $r_t = \sqrt{\sigma_t^2}\varepsilon_t$ is consistently modeled via the ARFIMA model

$$(1 - \phi L)(1 - L)^d(s_t - \omega) = e_t,\ e_t|\Omega_{t-1} \sim iid(0, \sigma_e^2), \tag{5}$$

which has been shown to be a very good competitor to alternative time series methods of forecasting realized volatility (e.g. Andersen et al., 2003; Pong et al., 2004; Koopman et al., 2005); s_t is the daily RV or RBP sequence, as defined in (3) and (4), in logarithms; ω is the unconditional mean of s_t, and L is the lag operator ($Ls_t = s_{t-1}$). A well-known property of logarithmic realized volatilities is that they are effectively Gaussian; hence, estimation of the parameters in (5) including d is conducted by exact ML under normal innovations.

[4] The high-frequency volatility literature has grown considerably over the recent years. On the one hand, alternative measures have been developed for exploiting intraday information such as the two scale realized variance of Zhang et al. (2005) or the realized range of Christensen and Podolskij (2007) and Martens and van Dijk (2007). On the other hand, there is a literature that proposes different approaches to account for the overnight non-trading hours period (e.g., Hansen and Lunde, 2005; Ahoniemi et al., 2012).

[5] A small literature adds a refinement to capture the volatility of volatility, e.g. Ishida and Watanabe (2009) adopt an ARFIMA-GARCH for the Nikkei 225 index, and Corsi et al. (2008) introduce the HAR-GARCH model for S&P500 index futures.

We also consider an ARMA(2,1) specification for the log realized measures following Pong et al. (2004) who show for the £/\$, Yen/\$ and DM/\$ rates that low order ARFIMA$(1,d,0)$ and ARMA(2,1) models of $\log\sqrt{RV}$ produce forecasts of similar statistical (MSE and Mincer-Zarnowitz R^2) accuracy.[6] The AR(FI)MA volatility predictions, $\hat{RV}_{t+1|t}$, are obtained through the bias-corrected mapping $\widehat{RV}_{t+1|t} = \exp(\widehat{\log RV}_{t+1|t} + \frac{1}{2}\hat{\sigma}^2_{e,t})$ where $\hat{\sigma}^2_{e,t}$ is the estimated variance. The connection between equation (5) and the conditional volatility σ_t is made via the two-step estimation approach put forward by Giot and Laurent (2004) which, effectively, amounts to setting $\sigma^2_{t|t-1} = \sigma^2 RV_{t|t-1}$, i.e. the conditional variance of the daily return process is conceptualized as a fraction of the realized volatility, and σ^2 is a scaling factor that ensures a unit variance for the innovation ε_t.

2.2 Forecast Combination

The benefits of combining forecasts from a number of preferably distinct methods have been repeatedly demonstrated; e.g. see Clements and Hendry (2004) for a review. Timmermann (2006) provides a threefold rationale for why combined forecasts work well in practice: they exploit jointly the information contained in each individual forecast; they are less sensitive to possible misspecification of individual forecasting models; and they average across differences in the way individual forecasts are bedevilled by structural breaks. In this paper, the interest is in combining conditional variance forecasts with the aim of improving the accuracy of VaRs. Since our VaR predictions are obtained from a pure scale model, there is an immediate relationship between volatility forecast combination and VaR forecast combination, given by equation (1). In other words, the volatility forecast combination can be equivalently cast as a quantile (VaR) forecast combination.[7] Since forecast combining is particularly beneficial when the methods that are mixed differ substantially we focus on the two broad classes here considered, GARCH and AR(FI)MA, to obtain conditional variance forecasts as[8]

$$\hat{v}_{t+1} = w\hat{h}_{t+1} + (1-w)\hat{s}_{t+1}, \tag{6}$$

[6] An ARMA(2,1) process can be conceptualized as the aggregation of two AR(1) processes. Using spectral density analysis, Gallant et al. (1999) show that the sum of two AR(1) processes is able to capture much of the persistence in asset price volatility.

[7] Alternatively, Giacomini and Komunjer (2005) and Fuertes and Olmo (2012) put forward GMM-based and quantile regression-based methods to optimally combine quantile forecasts obtained from any approach which can be non-parametric (simulation), semi-parametric (CAViaR) or parametric (beyond pure scale), including nested VaR models which may be individually misspecified. The resulting combined VaR is optimal because it meets by construction ex post the correct out-of-sample VaR specification condition.

[8] The approach of combining conditional volatility forecasts and mapping them onto a conditional VaR prediction via (1) builds on the implicit assumption that the shape of the demeaned returns standardized by $\hat{h}_{t+1}$ and $\hat{s}_{t+1}$ is approximately identical.

where $0 < w < 1$ is a deterministic weight; we adopt $w = 0.5$. Equally-weighted forecast combinations occupy a special place in the literature having stood out as quite effective; for a recent survey and application see, respectively, Timmermann (2006) and Patton and Sheppard (2009). One motivation for GARCH and AR(FI)MA model averaging in the present context is that it offers a simple but novel way of incorporating intraday price variation into daily VaRs. Prior to this exercise, various encompassing regressions and Wald tests are utilized to provide formal empirical evidence that justifies this model combination.

2.3 Robust Daily VaR Backtesting

Theoretically, a correctly specified α-th conditional VaR model of an asset or portfolio returns r_t is defined as

$$P(r_t \leq VaR_{t,\alpha} \mid \Omega_{t-1}) = \alpha, \text{ almost surely (a.s.)}, \alpha \in (0,1), \ \forall t \in \mathbb{Z}, \tag{7}$$

a conditional moment restriction that has been used extensively in the VaR literature; see, for instance, Escanciano and Olmo (2011) and references therein. At an empirical level, given a target or nominal probability level α, the VaR model is considered to be adequate *iff* the out-of-sample *hits* or *exceedances* sequence associated with the VaR forecasts, defined as $I_{t+1,\alpha}(\theta_0) \equiv 1(r_{t+1} \leq VaR_{t+1,\alpha})$ for $t = R, ..., T-1$, exhibits both correct unconditional coverage and serial independence. This condition reads as follows

$$\{I_{t+1,\alpha}(\theta_0)\} \text{ is } iid \text{ Bernoulli}(\alpha) \text{ for some } \theta_0 \in \Theta, \ t = R, ..., T-1 \tag{8}$$

where Bernoulli(α) stands for a Bernoulli random variable with parameter α; this is the implicit "loss function" for out-of-sample evaluation of VaR forecasts, leading to the so-called *unconditional coverage* backtesting (H_{0u} : $E[I_{t+1,\alpha}(\theta_0)] = \alpha$) and *independence* backtesting (H_{0i} : $\{I_{t+1,\alpha}(\theta_0)\}_{t=R}^{T-1}$ is *iid*). In practice, the knowledge of the VaR model parameters is rare. Thus we need to replace θ_0 by a consistent estimator, denoted $\widehat{\theta}_t$, yielding the estimated out-of-sample hits sequence $I_{t+1,\alpha}(\widehat{\theta}_t) \equiv 1(r_{t+1} \leq \widehat{VaR}_{t+1,\alpha})$, for $t = R, ..., T-1$.

The pioneering Kupiec's (1995) test to test for correct unconditional coverage assumes $\{I_{t+1,\alpha}(\theta_0)\} \sim iid$ and is based on the standardized sample mean

$$S_P \equiv S_P(\widehat{\theta}_P) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} (I_{t+1,\alpha}(\widehat{\theta}_t) - \alpha), \tag{9}$$

where $\{\widehat{\theta}_t\}_{t=R}^{T-1}$ are the volatility model parameter estimates obtained iteratively as the information set Ω_{t-1} changes and $P = T - R$ is the length of the out-of-sample period. Inferences from (9) typically rely on the critical values of the asymptotic $N(0, \alpha(1-\alpha))$ distribution.

Condition (7) is sufficient but not necessary for the correct unconditional coverage and the independence of the out-of-sample hits sequence. There is a large class of VaR models which are misspecified in the sense that they do not satisfy (7) but nevertheless they yield an *iid* sequence of out-of-sample hits with the correct unconditional coverage probability α, that is, condition (8) is met; this mismatch is known as *model misspecification* (or model risk). Escanciano and Olmo (2011) derive the correct asymptotic distribution of Kupiec's test in the presence of model risk and estimation uncertainty; the extra terms that arise are too cumbersome to compute in practice. A block-bootstrap inference approach is suggested as a feasible and effective alternative.

The block bootstrap is an extension of the nonparametric *iid* bootstrap for serially dependent time series where the resampling refers to data blocks instead of individual data points. The aim is to construct artificial (i.e. bootstrap) time series that mimic the dependence structure observed in the original sample. The bootstrap algorithm is described next. Start by defining a blocks partition of the overall daily returns sample, $T = bl$, where b is the block size and l the total number of blocks, $\{B_1, ..., B_l, \}$, with $B_1 = \{r_1, \ldots, r_b\}$ and so forth. For each bootstrap iteration $j = 1, \ldots, B$ conduct the steps:

1. Generate a block-bootstrap returns sample $r^*_{1,j}, \ldots, r^*_{T,j}$, with the same size as the original sample, $T = R + P$, by concatenating the blocks $B^*_{1,j}, ..., B^*_{l,j}$ randomly drawn with replacement from $\{B_1, ..., B_l\}$.

2. Obtain an out-of-sample hits sequence $\{I^*_{R+k,j,\alpha}\}_{k=1}^{P}$ as follows:

   (a) Construct a sequence of R-length rolling samples $\{r^*_{t,j}\}_{t=k}^{R+k-1}$ for $k = 1, \ldots, P$.

   (b) Obtain the volatility model parameters, $\widehat{\theta}^*_{R+k-1,j}$, for each sequential sample $k = 1, \ldots, P$.

   (c) Compute the sequence of out-of-sample 1-day-ahead VaR forecasts $\{\widehat{VaR}_{R+k,j,\alpha}\}_{k=1}^{P}$ from which the hits can be obtained as $I^*_{R+k,j,\alpha}(\widehat{\theta}^*_{R+k-1,j}) = 1(r^*_{R+k,j} \leq \widehat{VaR}_{R+k,j,\alpha})$ for $k = 1, \ldots, P$.

3. Compute a block-bootstrap version of (9) denoted $S^{bb}_{P,j}(\widehat{\theta}^*_{P,j}) \equiv S_P(B^*_{1,j}, ..., B^*_{l,j}; \widehat{\theta}^*_{P,j})$ as

$$S^{bb}_{P,j}(\widehat{\theta}^*_{P,j}) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(I^*_{t+1,j,\alpha}(\widehat{\theta}^*_{t,j}) - \bar{I}_\alpha(\widehat{\theta}_P) \right) \tag{10}$$

where $\bar{I}_\alpha(\widehat{\theta}_P) = \frac{1}{P}\sum_{t=R}^{T-1} I_{t+1,\alpha}(\widehat{\theta}_t)$ is the average number of out-of-sample exceedances associated with $\widehat{\theta}_t$, the rolling parameter estimates from the actual returns sample.

From the centered statistics $\{S^{bb}_{P,j}(\widehat{\theta}^*_{P,j})\}^B_{j=1}$ one can compute the empirical p-value of Kupiec's test as[9]

$$\widehat{p}^{bb}_P = \frac{1}{B} \sum_{j=1}^{B} 1(|S^{bb}_{P,j}(\widehat{\theta}^*_{P,j})| > |S_P(\widehat{\theta}_P)|). \tag{11}$$

Escanciano and Olmo (2011) show that, under certain regularity conditions, a small ratio of out-of-sample to in-sample observations ($P/R < 0.5$) is a sufficient condition for estimation risk to become harmless and therefore, step 1 of the above algorithm can be simplified to random (block) draws from the hits sequence rather than from the returns sequence, and step 2 is redundant. We employ $B = 500$ iterations which is shown in Escanciano and Olmo (2011) to deliver a correctly-sized test with good power properties. Our choice of block size b is based on Politis et al. (2009) optimal data-driven algorithm.[10]

Financial regulation backtesting mainly focuses on the unconditional coverage property somehow understating the relevance of the *iid* condition (H_{0i}). However, it is possible to find a VaR approach yielding exceptions (i.e. larger losses than the maximum expected one) which, although adequate in number, happen to adversely occur over consecutive days; such VaR approach would imply greater stress for the corresponding trading desk (or bank) than a similar VaR that reports randomly scattered exceedances. Since for a Bernoulli random variable serial independence is equivalent to serial uncorrelation, it is natural to employ the test statistic

$$\xi_{P,k} \equiv \frac{1}{\sqrt{P-k}} \sum_{t=R+k}^{T-1} (I_{t+1,\alpha}(\widehat{\theta}_t) - E[I_{t+1,\alpha}(\widehat{\theta}_t)])(I_{t-k+1,\alpha}(\widehat{\theta}_{t-k}) - E[I_{t-k+1,\alpha}(\widehat{\theta}_{t-k})]), \quad k \geq 1 \tag{12}$$

in which the expectations are estimated by the average number of the corresponding out-of-sample hits. The test statistic $\xi_{P,k}$ is asymptotically Gaussian with zero mean and variance $\alpha^2(1-\alpha)^2$. As shown in Escanciano and Olmo (2011), the latter is not bedevilled by model risk, only by estimation uncertainty which is nevertheless negligible for small P/R ratios. In our empirical analysis we deploy $\xi_{P,1} \equiv \xi_P$.

[9] For a detailed discussion on the asymptotic properties of the test as $B \to \infty$ (and as $P \to \infty$) see Escanciano and Olmo (2011).

[10] We employ Politis et al.'s (2009) *Matlab* routine available from Andrew Patton's website which we gratefully acknowledge. The optimal b ranges between 30 and 60 across our sample of stocks.

2.4 Comparing VaR Models

One would wish to assess the statistical significance of differences in VaR adequacy between two risk models of interest, e.g. GARCH- versus ARFIMA-based VaR, but most of the models entertained above are nonnested implying that traditional approaches, such as LR tests, cannot be used.[11] Our novel test to compare VaR models, described below, exploits a cross-section of time-series returns $\{r_{t,i}\}_{i=1}^{N}$ since its inputs are the unconditional coverage (or *iid*) test p-values obtained over a set of N assets or portfolios. Therefore it can be seen as a 'panel' assessment of the relative ability of the VaR measure at hand to satisfy the conditions stated in (8).

Let V_1 and V_2 denote two competing VaR models. The statistical measure that we propose is a difference of proportions, $\widehat{p}_{V_1} - \widehat{p}_{V_2}$ with $\widehat{p}_{V_1} = \frac{1}{N}\sum_{i=1}^{N} 1(p\text{-}val\ S_{P,i(V_1)}^{bb} < c)$ or $\widehat{p}_{V_1} = \frac{1}{N}\sum_{i=1}^{N} 1(p\text{-}val\ \xi_{P,i(V_1)} < c)$, where $p\text{-}val\ S_{P,i(V_1)}^{bb}$ and $p\text{-}val\ \xi_{P,i(V_1)}$ are, respectively, the p-value of the unconditional coverage test (11) and the p-value of the *iid* test (12) for model V_1 over the ith time series of daily returns; we adopt a conservative $c = 0.10$ significance level to map each cross-section of p-values into a cross-section of 0s and 1s which are subsequently averaged to obtain an overall adequacy measure or proportion $\widehat{p}_{V_1}$. The difference-in-proportions $\widehat{p}_{V_1} - \widehat{p}_{V_2}$ enables a formal panel test to compare VaR models. The null hypothesis $H_0 : p_{V_1} \leq p_{V_2}$ can be tested via

$$\Pi_{N,P} \equiv \sqrt{N}\frac{\widehat{p}_{V_1} - \widehat{p}_{V_2}}{\sqrt{\widehat{p}_{V_1}(1-\widehat{p}_{V_1}) + \widehat{p}_{V_2}(1-\widehat{p}_{V_2})}}, \qquad (13)$$

against $H_1 : p_{V_1} > p_{V_2}$ using the asymptotic $N(0,1)$ critical values; the subscripts N and P denote that the test statistic exploits a time series of P daily returns and out-of-sample VaR forecasts across N assets. This test rests on the assumption that the p-values of the unconditional coverage (or *iid*) test are independent across the two VaR models V_1 and V_2. Seeking to add robustness to the comparative analysis, we also deploy a bootstrap relative VaR adequacy test $\Pi_{N,P}^{b}$ that does not hinge on this assumption since it does not necessitate a closed-form expression for the variance of the difference-in-proportions $\widehat{p}_{V_1} - \widehat{p}_{V_2}$. In order to maintain the dependence structure between VaR models, each bootstrap sample contains N pairs of the form $1(p\text{-}val\ S_{P,i(V_1)}^{bb} < c), 1(p\text{-}$

[11] Christoffersen et al. (2001) propose an elegant nonnested VaR comparison test based on the Kullback-Leibler Information criterion. However, their approach requires choosing an appropriate set of instrumental variables and adequate estimation of an unconditional long-run variance, two challenging tasks. Further, their test is 'univariate' in that it compares two VaR models deployed on the returns of a single asset or portfolio. Similarly, Giacomini and Komunjer (2005) propose a conditional forecast encompassing test for pairs of VaR models using single stocks/portfolios.

val $S^{bb}_{P,i(V_2)} < c$) randomly drawn with replacement from the cross-section of $i = 1, ..., N$ assets. For each bootstrap sample $j = 1, ..., B$ we compute the centered test statistic $\Pi^b_{N,P} \equiv \sqrt{N}(\Delta\widehat{p}^* - \Delta\widehat{p})$ where $\Delta\widehat{p} \equiv \widehat{p}_{V_1} - \widehat{p}_{V_2}$, $\Delta\widehat{p}^* \equiv \widehat{p}^*_{V_1} - \widehat{p}^*_{V_2}$ with $\widehat{p}_{V_1}$ defined as above, $\widehat{p}^*_{V_1} = \frac{1}{N}\sum_{i=1}^{N} 1^*(p\text{-}val\ S^{bb}_{P,i(V_1)} < c)$ and $1^*(\cdot)$ denotes a bootstrap observation; likewise, we deploy a relative VaR adequacy test that focuses on the serial independence (of the hits sequence) using *p-val* $\xi_{P,i(V_1)}$ instead to construct $\widehat{p}_{V_1}$ and $\widehat{p}^*_{V_1}$.

3 Empirical Results

3.1 Data and Summary Statistics

The analysis is based on high-frequency transaction prices from *Tick Data* for 14 large-cap NYSE/Nasdaq pertaining to the financial, industrial, technology, telecommunication and miscelaneous retailer sectors.[12] The 7-year sample period 02/01/97 to 02/01/04 amounts to $T = 1761$ days.[13] In order to compute daily realized volatilities, the official trading interval [9:30am-4:00pm] is divided into $M = 78$ five-minute subintervals.[14] The price at the start of the jth intraday interval is computed as the average of the close and open prices of intervals $j-1$ and j, respectively. The jth intraday return (on day t) is defined as

$$r_{t,j} \equiv \left(\frac{\log(p^c_{t,j}) + \log(p^o_{t,j+1})}{2} - \frac{\log(p^c_{t,j-1}) + \log(p^o_{t,j})}{2}\right), j = 2, ..., M-1 \tag{14}$$

where $p^c_{t,j}$ ($p^o_{t,j}$) is the close (open) price of the jth interval; $r_{t,1} \equiv \left(\frac{\log(p^c_{t,1})+\log(p^o_{t,2})}{2} - \log(p^o_{t,1})\right)$ and $r_{t,M} \equiv \left(\log(p^c_{t,M}) - \frac{\log(p^c_{t,M-1})+\log(p^o_{t,M})}{2}\right)$ are the first and last intraday return. The closing price on day t, denoted $p^c_{t,M}$ or simply p^c_t, is defined as the last price observed before 4:00pm; the intraday closing price $p^c_{t,j}$ is similarly defined as the last seen tick before the jth 5-min mark. The observed opening price on day t, denoted $p^o_{t,1}$ or p^o_t, is the first price recorded after 9:30am; likewise for $p^o_{t,j}$ with reference to the 5-min mark j-1.

The aggregation of all intraday returns gives the daily return $r_t = \sum_{j=1}^{M} r_{t,j} = \log(\frac{p^c_{t,M}}{p^o_{t,1}}) = \log(\frac{p^c_t}{p^o_t})$. The inter-daily (logarithmic close-to-close) return can be decomposed as the sum of the overnight return (previous-

[12] The stocks were chosen to give wide market coverage in terms of market capitalization and sector representation: American Express (AXP), AT&T (ATT), Boeing (BA), Caterpillar (CAT), DELL, General Electric (GE), General Motors (GM), IBM, J.P. Morgan (JPM), KO (Coca-Cola), McDonald (MCD), Microsoft (MSFT), Procter & Gamble (PG) and WAL-MART (WMT)

[13] The G@RCH 4.2 module (Laurent and Peters, 2004) and ARFIMA package 1.04 (Doornik and Ooms, 2006) for *OxMetrics* 5 are used in modeling and forecasting. *Matlab* 6.5 is used for the VaR estimation and backtesting.

[14] The 5-min grid is the most widely adopted in the empirical literature because it is short enough for the daily volatility dynamics to be picked up with reasonable accuracy, and long enough for the adverse effects of market microstructure noise not to be excessive.

day close to open) and the daily return, i.e. $\log(\frac{p_t^c}{p_{t-1}^c}) = \log(\frac{p_t^o}{p_{t-1}^c}) + \log(\frac{p_t^c}{p_t^o})$. As in Liu and Maheu (2009) and Gallo (2001), the modeling object of interest is the daily return defined as open-to-close logarithmic price differences excluding the overnight (ON) return. The argument for this choice is twofold. First, this allows us to complement and extend Gallo's (2001) analysis, based exclusively on the RMSE and MAE criteria, by assessing whether the information content in the squared ON return, $r_{o,t}^2 \equiv (\log \frac{p_t^o}{p_{t-1}^c})^2$, can enhance in a GARCH framework the adequacy of daily VaR predictions. Second, a practical problem with adopting instead the inter-daily return as the object of interest is having to determine the weight that $r_{o,t}^2$ should deserve in the realized measures since the ON return is far more volatile than the intraday 5-min returns which would introduce extra noise. Hansen and Lunde (2005) propose an optimal weighting scheme for incorporating intraday and overnight information into daily volatility measures. Other studies concerned with the importance of overnight information in a volatility forecasting framework are Engle et al. (2006) and Ahoniemi and Lanne (2010).

Ljung-Box portmanteau tests confirm the well-known absence of serial correlation in daily stock returns and the presence of strong volatility clustering. Table 1 reports summary statistics for several daily unconditional volatility measures. Relative to their mean, the realized volatilities exhibit much smaller dispersion than the squared daily and overnight returns; RV is the least noisy and the squared overnight return the most noisy. The mean of RV is invariably higher than the mean of the jump-immune RBP measure. Both realized volatilities are markedly right-skewed and leptokurtic. In contrast, the (unreported, to preserve space) sample skewness and kurtosis of logRV and logRBP suggest that their distribution is approximately Gaussian. The skewness of logRV ranges between 0.0189 (stock PG) and 0.359 (stock MCD) and the kurtosis between 2.969 (stock DELL) and 4.004 (stock JPM); for logRBP the range is [0.0050, 0.359] and [2.958, 3.903], respectively. The Ljung-Box statistics indicate that volatility clustering is not a distinctive feature of the overnight returns although this may be due to their noisiness, i.e. the autocorrelation signal is difficult to pick up, rather than its true absence.

Prior to the ARFIMA modeling of the realized volatilities, we compute the long-memory parameter d using the Gaussian Semi-Parametric estimator (see Robinson and Henry, 1998). The estimates $\hat{d}^{GSP}$ for RV and RBP are significantly positive, generally below 0.4. The estimates $\hat{d}^{GSP}$ for logRV and logRBP (unreported, to preserve space) are closer to the stationarity boundary of 1/2; for instance, for AXP the estimate is 0.401 (RV) and 0.414 (RBP), and increases to 0.435 (logRV) and 0.426 (logRBP). Nevertheless, in both levels and logs none of the estimated long-memory parameters is significantly different from 1/2. The stationarity of realized volatilities in levels (and logs) is also borne out by the ADF test statistic. Thus our dataset confirms two stylized facts of daily realized volatilities: covariance stationarity and slow hyperbolic decay of autocorrelations.[15]

3.2 Out-of-Sample VaR Backtesting

The volatility models' parameters are updated over rolling windows of length $R = 1261$ days.[16] This forecasting scheme facilitates 500 out-of-sample daily VaR predictions.[17] Tables 2 and 3 summarize the backtesting of daily VaR predictions when $\hat{F}_{\varepsilon}^{-1}(\alpha)$ is a Gaussian quantile at, respectively, the nominal level $\alpha = 5\%$ (often adopted by banks internally) and the mandatory 1% level to set minimal capital requirements.[18]

We start by examining the role of the overnight surprise. Both the 5% and 1% VaRs suggest that the simple approach of extending GARCH models with the squared previous-close-to-open overnight return is rather futile.[19] If anything, it adds noise to the VaR prediction by slightly increasing the number of VaR adequacy rejections regarding correct unconditional coverage (S_P^{bb} test) and the *iid* property (ξ_P^{asy} test). Another practical question of interest is whether augmenting the standard daily GARCH model with intraday-based realized volatilities enhances VaR adequacy. The results indicate that the GARCH-RV or GARCH-RBP models do not improve VaR adequacy relative to GARCH. Hence, augmenting the standard GARCH equation with realized volatilities is not an effective way of exploiting intraday data for assessing downside tail risk exposure.

[15]For all stocks, the unconditional distribution of daily stock returns is fat-tailed with mild skewness. Daily returns scaled by *ex post* RV are far closer to Gaussian than GARCH-scaled returns, consistent with the literature (e.g. Andersen et al., 2003). The average contribution of rare large jumps to the realized variance is 9.8% over trading hours.

[16]The GARCH equation (2b) for CAT, JPM, KO and MCD has lags $r = 2$ and $s = 1$ whereas for all other stocks a GARCH(1,1) sufficed to absorb the autocorrelation in squared daily returns.

[17]Several volatility forecast competitions have been based on fixed model parameters over the out-of-sample period (e.g. Ghysels et al., 2006; Giot and Laurent, 2004; Andersen et al., 2003). However, as illustrated empirically in Clements et al. (2008) and theoretically argued in Eklund et al. (2009), a rolling-window scheme facilitates some 'shield' against abrupt changes in the dynamics of the volatility process during the out-of-sample period.

[18]The backtesting procedure enforced by Basel II for market risk VaR boils down to assessing out-of-sample whether the observed frequency with which daily returns fall below the daily VaR ("exceedances") exceeds the nominal coverage level; the observed daily losses can exceed the 99% VaR reported by the institution no more often than once every one hundred days. The capital charge for market risk for banks using internal models is set at the maximum of the previous day's VaR and three times (plus a penalty) the previous 60-day average of the daily VaR. The penalty component seeks to reflect too frequent exceedances.

[19]In a recent paper, Ahoniemi et al. (2012) compare different modeling approaches to incorporate the overnight period into daily VaR predictions. For the large cap S&P 500 index, the evidence supports a bivariate framework where separate efforts are devoted to model the daytime and overnight return processes, and which can also account for the intraday-overnight return covariance.

Regarding the comparison among the two realized volatility measures, we observe that the empirical coverage rates for GARCH-RBP and GARCH-RV are very close; on average across stocks the actual 5% VaR coverage is 3.600% for GARCH-RBP and 3.743% for GARCH-RV. The unconditional coverage and *iid* backtesting outcome is also very similar for both models. However, it would be too hasty thus to conclude that rare large jumps play no role in VaR prediction since the augmented GARCH models miss the autocorrelation dynamics of realized volatilities.[20] In fact, a somewhat different picture emerges when facing the choice between ARFIMA(logRV) and ARFIMA(logRBP): the number of stocks where $\bar{I}_\alpha(\%) > \alpha$ (risk underprediction) with $\alpha = 5\%$ is 7 for ARFIMA(logRV) and increases to 11 for ARFIMA(logRBP); the average empirical coverage probability across stocks is 5.2% for ARFIMA(logRV) and 5.9% for ARFIMA(logRBP). Moreover, the ARFIMA(logRV) forecasts appear to outperform the corresponding logRBP forecasts in two senses: the resulting 5% and 1% VaRs pass more often the unconditional coverage and *iid* backtesting. Figure 1 (bottom panel) depicts this contrast.

Confronting next the GARCH and ARFIMA frameworks, as illustrated in Figure 1 (top panel), VaR predictions based on GARCH models tend to appear more conservative (downside tail risk appears overstated) than those associated with ARFIMA models fitted to realized volatilities. This finding together with the fact that the GARCH framework remains widely used in the financial industry (e.g. J.P.Morgan Riskmetrics can be cast as a Gaussian IGARCH) squares well with the evidence presented in Pérignon et al. (2006) for several commercial banks suggesting a tendency to report inflated VaRs.[21] The long-memory models of realized volatilities tend to outperform the GARCH models in terms of correct unconditional coverage backtesting; e.g. the VaR based on ARFIMA(logRV) forecasts is rejected as inadequate in one case only (stock GM) whereas the GARCH models tend to produce too few exceedances ($\bar{I}_{t+1,\alpha} < \alpha$). Hence, modeling the dynamics of realized volatility is quite effective to achieve correct VaR coverage but the use of a long-memory specification does not seem crucial since the ARMA(2,1) forecasts yield very similar backtesting results. Therefore the genuinely advantageous feature

[20] In an earlier version of the paper, we also explored the information role of jumps by incorporating the jump variation measure $\hat{J}_t \equiv max\{0, (RV_t - RBP_t)\}$ lagged one day as regressor in the GARCH equation, and Andersen et al.'s (2007) shrinkage refinement of this jump measure. The resulting VaR backtesting results fail to improve also upon those from the standard GARCH.

[21] Pérignon et al. (2006) rationalize their evidence on over-conservatism in the banks' overall market risk VaRs using two different arguments which do not relate to the volatility modeling framework. One is that banks are deliberately cautious because they do not want to taint their reputation by reporting too many exceedances. Another is that, by only taking partial account of diversification across portfolios (and risk classes) some of the offsetting effects are lost, resulting in inflated VaRs.

of the AR(FI)MA framework is the effective incorporation of intraday information by enabling the realized volatility forecasts to quickly adapt to changes in the underlying latent volatility process. However, a finding in favour of the GARCH-based volatility forecasts is that they tend to produce VaRs which satisfy more often the *iid* backtesting criteria (ξ_P^{asy}). Hence, in terms of coverage rates the most competitive VaRs come from the ARFIMA(logRV) forecasts which outperform the (augmented) GARCH forecasts. But the GARCH-based VaR framework is better able to filter out the serial dependence in the hits sequence which may indirectly suggest that it is more reactive to actual P&L shocks. These findings provide *prima facie* evidence that GARCH and AR(FI)MA forecasts exhibit complementary 'skills' from the point of view of VaR adequacy which points to the potential usefulness of forecast combining. This issue is formally examined in Section 3.3.

3.3 Exploiting Intraday Returns Through Forecast Combination

We run encompassing tests formally to corroborate that there is distinct information in the GARCH and ARFIMA volatility forecasts which can be usefully combined. A typical approach adopted in the literature is to run a regression of the observed data on the competing forecasts, in our context this is

$$\text{ENC1:}\quad \tilde{\sigma}_t^2 = \varphi_0 + \varphi_1 \hat{h}_t + \varphi_2 \hat{s}_t + e_{1,t}, \tag{15}$$

where $\hat{s}_t$ is the forecasted daily variance conditional on information up to day $t-1$ using the ARFIMA(logRV) model, $\hat{h}_t$ is the forecasted variance using the daily GARCH model and $\tilde{\sigma}_t^2$ is the 'actual' or realized daily variance proxied by the sum of intraday 5-min squared returns. The practice of forecast combination implicitly acknowledges the possibility of model misspecification. We seek to combine the 'best' models considered within the GARCH and AR(FI)MA classes. Following the parsimony principle, the standard GARCH is chosen because it was not outperformed by any of the augmented GARCH models. ARMA and ARFIMA forecast performance proved very similar but on the basis of the $\hat{d}^{GSP}$ estimates discussed in Section 3.1 we opted for the latter. Within the ARFIMA class, the forecasting properties of ARFIMA(logRV) proved somewhat superior to those of ARFIMA(logRBP) as noted earlier. Forecast $\hat{h}_t$ encompasses forecast $\hat{s}_t$ when the parameter restriction (φ_0 φ_1 φ_2) = (0 1 0) holds. Conversely, if forecast $\widehat{s_t}$ encompasses forecast $\hat{h}_t$ we have (φ_0 φ_1 φ_2) = (0 0 1). A potential problem with the above encompassing regression (hereafter, ENC1) is the multicollinearity arising from the high correlation between the two sets of forecasts. In our sample, the correlation between standard GARCH and ARFIMA(logRV) forecasts ranges across stocks from a low of 55.13 (stock CAT) to a high of 91.69 (stock GM) with mean and median equal to 78.58 and 79.92, respectively. As an *ad hoc* solution to this problem, following Timmermann (2006) we implement a more general test of the hypothesis that $\widehat{s}_t$ encompasses $\hat{h}_t$ by fitting

$$\text{ENC2:}\quad \tilde{\sigma}_t^2 - \hat{s}_t = \gamma_{0,1} + \gamma_1 \hat{h}_t + e_{2,t} \tag{16}$$

and testing that $\gamma_1 = 0$; likewise, to investigate whether $\hat{h}_t$ encompasses $\widehat{s}_t$ we test for $\gamma_2 = 0$ in $\tilde{\sigma}_t^2 - \hat{h}_t = \gamma_{0,2} + \gamma_2 \hat{s}_t + e_t$. To make our inferences more robust, we also deploy the encompassing test suggested by Fair and Shiller (1990) based on the first-difference regression

$$\text{ENC3:}\quad \Delta\tilde{\sigma}_t^2 = \eta_0 + \eta_1(\hat{h}_t - \tilde{\sigma}_{t-1}^2) + \eta_2(\hat{s}_t - \tilde{\sigma}_{t-1}^2) + e_{3,t} \tag{17}$$

that relates the actual changes to the predicted changes from the two competing models. On this basis, we conduct a Wald test for the restriction $(\eta_1,\ \eta_2) = (1,\ 0)$ pertaining to the hypothesis that $\hat{s}_t$ contains no information relevant to predict $\tilde{\sigma}_t^2$ not already contained in the constant term and in $\hat{h}_t$; conversely, the restriction $(\eta_1,\ \eta_2) = (0,\ 1)$ is tested to falsify the hypothesis that $\hat{h}_t$ contains no information relevant to predict $\tilde{\sigma}_t^2$ not already contained in the constant term and in $\hat{s}_t$. The main motivation for ENC3 is that the regressand is less persistent than that in ENC1. Table 4 sets out the OLS coefficient estimates and p-values of Wald type tests for the above restrictions. All inferences are based on the Newey-West h.a.c. covariance matrix. Although the constraint $\varphi_1 + \varphi_2 = 1$ is not imposed in (15) the $\hat{\varphi}_1$ and $\hat{\varphi}_2$ estimates often sum quite reasonably close to one. Overall there is evidence that none of the two forecasts clearly dominates the other.

Thus motivated we compute VaRs based on the combination of GARCH and ARFIMA(logRV) forecasts through a model averaging approach corresponding to $w = 0.5$ in (6). The results in Table 5 are rather encouraging despite the naïve equal-weighting nature of our approach. VaR adequacy is virtually supported for all stocks with both the unconditional coverage S_P^{bb} and independence ξ_P^{asy} tests. One key message from these findings is that the informational value of realized volatility-based ARFIMA forecasts becomes more apparent when it is combined with daily return-based GARCH forecasts. Augmenting GARCH with realized volatilities did not materialize in VaR adequacy improvement. However, combining forecasts from GARCH models fitted to daily returns and AR(FI)MA models for logRV yields more adequate VaR predictions for several stocks than any of them individually in terms of *both* correct coverage and independence of the hits sequence.

3.4 Testing for Relative VaR Adequacy

The comparison of risk models thus far has relied on observations about the backtesting results for the individual sampled stocks. But how does one *statistically* decide between two VaR models V_1 and V_2? The statistic $\Pi_{N,P}$ outlined in Section 3.4 is useful should a risk manager want to perform a pairwise (nonnested) VaR comparison testing across a set of assets/portfolios or classes of risk. It provides a formal way to gauge 'relative VaR-adequacy' or model ranking in terms of two desirable properties for the hits sequence: correct unconditional coverage and serial independence. In order to gather evidence that is robust to cross-section dependence (across forecasting models) and sample size ($N = 14$ stocks in our application), the discussion focuses on the bootstrap $\Pi^b_{N,P}$ test with $B = 500$ iterations. Table 6 sets out the results.

The comparison testing results, in general, square well with our earlier observations. Regarding the unconditional coverage property, for instance, the ARMA(logRV) forecasts are preferred to the GARCH forecasts as suggested by the p-value=0.058 (95% VaR) and p-value=0.084 (99% VaR) in the first row of Table 6. Predominantly for the 95% VaRs, the vast majority of the significant pairwise statistics are located in the top-right area of the table; this outcome formally corroborates our initial observation that forecasts from the (augmented) GARCH family of models tend to produce inferior VaR adequacy relative to the AR(FI)MA forecasts in terms of unconditional coverage backtesting. Table 7 pertains to the comparison based on the *iid* property; it shows most of the significant cases in the bottom-left area suggesting that forecasts from the (augmented) GARCH family of models tends to yield superior VaR adequacy than the AR(FI)MA forecasts in terms of independence of the hits sequence. In both tables and particularly so with regard to the unconditional coverage criteria (Table 6), there is a striking contrast between the large number of rejections reported in the last column and the invariably insignificant test statistics in the bottom row; this pattern is a reflection of the overall superior adequacy of the VaR model based on combined GARCH-ARFIMA forecasts. The less reliable asymptotic p-values of the pairwise comparison tests are quantitatively (and in some cases, qualitatively) different but the above picture can be seen broadly to remain, particularly, for the 95% VaRs.

3.5 Predicting VaR with Fat-Tailed Densities

As a robustness check, we now re-conduct the 'horse race' relying on the Student-t family for the quantile $\hat{F}_{\varepsilon}^{-1}(\alpha)$ computation. This choice instead of the skewed counterpart densities obeys the non-rejection of the symmetry null for the standardized returns distribution on the basis of Delgado and Escanciano's (2007) nonparametric conditional test. To illustrate this finding graphically, we plot in Figure 2 for four stocks the kernel smoothed finite-sample density of $r_t/\widehat{h}_t^{1/2}$ and $r_t/\widehat{s}_t^{1/2}$ corresponding to the first estimation window, t=1,...,1261 days, where $\widehat{h}_t$ and $\widehat{s}_t$ are the in-sample GARCH and ARFIMA(logRV) volatility forecasts, alongside the N(0,1) and standardized Student-t density with d.f. parameter estimated by ML. The plots reveal negligible asymmetries.[22]

A controversial empirical question is whether Student t quantiles add accuracy to VaRs relative to Gaussian quantiles.[23] In order to address this question it is key to confront again the results from the two realized volatility measures, RV and RBP. The answer from our analysis, summarised in Figure 1 (bottom panel), is: *yes* and *no*. Mostly for the 99% VaRs, the unconditional coverage properties associated with ARFIMA(logRBP) forecasts together with Student-t quantiles show improvements relative to the Gaussian framework (likewise, for the unreported GARCH-based VaRs). However, this improvement is virtually absent in the ARFIMA(logRV)-based VaRs and this may relate to the fact that the RV measure fully incorporates the intraday jump contribution. Indirectly, our empirical analysis provides an answer to the question: how do rare but large jumps manifest themselves in daily VaR predictions if they are ignored? The estimated d.f. from the standardized returns are almost invariably smaller for the ARFIMA(logRBP), 8 on average and ranging between 5 and 11, than for the ARFIMA(logRV) forecasts, 12 on average and ranging between 7 and 22. This result is in line with the fact that extreme occassional jumps are fully accounted for in the logRV measure and so the standardization of returns based on ARFIMA(logRV) forecasts brings them closer to Gaussianity. Relatedly, the differences previously

[22] A complete set of density plots and test results are available from the authors upon request. The VaR measures in this section are obtained from (1) using as d.f. parameter for the Student t quantile computation the estimated parameter using at each point in time the available sample at the time, day t, the forecast is made. Thus the forecast, $\widehat{VaR}_{t+1,\alpha}$, is strictly out of sense.

[23] Andersen et al. (2003) conclude that accurate daily VaRs for DM/$ and Yen/$ returns can be obtained from long-memory AR models for realized volatility alongside Gaussian quantiles. Clements et al. (2008) document that simple models such as AR(5) fitted to $\sqrt{RV}$ together with Gaussian quantiles yield good VaRs for currencies. In contrast, Giot and Laurent (2004) strongly advocate the use of Skewed Student t quantiles for VaR prediction in the context of the CAC40, S&P500 and two currencies.

observed in terms of VaR backtesting between the forecasts from ARFIMA fitted to logRBP and logRV (i.e. the superiority of the latter over the former) coupled with Gaussian quantiles are virtually absent now, possibly because the task of accounting for jumps is also given to the 'freely estimated' fat tails of the Student-t density. Table 5 also bears this out by showing that the VaR backtesting results of combined GARCH-ARFIMA(logRV) forecasts alongside Gaussian quantiles remain virtually unchanged by using Student-t quantiles.

The upshot is that, by adopting the standard Gaussian density for the quantile estimation, $\hat{F}_{\varepsilon}^{-1}(\alpha)$, a larger role is left to the volatility forecasts in capturing rare but large jumps for accurate VaR prediction. Put differently, the use of a Student-t density with freely estimated d.f. parameter from the standardized returns for the quantile computation inexorably obscures the link (relatively to the Gaussian case) between the importance of incorporating the intraday jumps in daily volatility measurement and VaR adequacy. Therefore it appears that from the lens of VaR backtesting one can choose either to pay more "attention" into the volatility measurement (e.g. choosing an appropriate realized measure such that forecasts based on it delivers near Gaussian standardized returns) or to the quantile computation using non-Gaussian distributions.

4 Conclusions

This paper examines in a robust backtesting framework the practical importance of several issues with a view to improve the adequacy of daily VaR predictions. Two novel aspects of our VaR validation framework are that it deploys a block-bootstrap version of Kupiec's unconditional coverage test which is robust to estimation uncertainty and model misspecification, and that it proposes a panel test for differences in VaR adequacy between models. The 'horse race' includes two distinct classes of volatility models: GARCH specifications based exclusively on daily returns and extensions thereof with squared overnight returns or intraday-based realized volatilities, and AR(FI)MA specifications fitted directly to the realized volatilities in order to capture their slowly-decaying autocorrelation dynamics.

Our findings suggest that GARCH augmentation with lagged realized volatility does not enhance VaR adequacy, vis-à-vis the standard daily return-based GARCH, possibly because in this framework the autocorrelation dynamics of realized volatility is not explicitly modeled. ARMA or ARFIMA models outperform GARCH in terms of VaR coverage backtesting but, regarding the serial independence of the hits sequence, GARCH forecasts lead to superior backtesting results than AR(FI)MA. This mixed picture prompts the thought that forecast combining may be fruitful. Combination of forecasts from both classes of models, GARCH and AR(FI)MA, is further motivated formally through various encompassing tests. To the best of our knowledge, this is the first study to highlight the merits of forecast combination from standard GARCH fitted to daily returns and AR(FI)MA fitted to logarithmic realized variance as a way of subsuming intraday information into VaRs. A naïve model averaging approach produced rather satisfactory VaR backtesting, in terms of both coverage and independence. There is evidence that daily realized variance forecasts together with the assumption of Gaussian standardized returns are as effective, from the viewpoint of VaR adequacy, as volatility forecasts from models that neglect rare large intraday jumps but are coupled with quantiles from appropriate fat-tailed Student t densities. In future research it might be fruitful to dig deeper into forecast combining issues in a VaR context.

References

[1] Ahoniemi, K., Fuertes, A.-M., Olmo, J. (2012). Overnight News and Daily Equity Trading Risk Limits. SSRN Working Paper http://ssrn.com/abstract=2065017.

[2] Ahoniemi, K., Lanne, M., (2010). Realized Volatility and Overnight Returns. Bank of Finland Research Discussion Paper 19.

[3] Andersen, T., Bollerslev, T., Diebold, F.X, & Labys, P. (2003). Modelling and Forecasting Realized Volatility. *Econometrica, 71,* 579-625.

[4] Andersen, T., Bollerslev, T., & Diebold, F.X. (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics & Statistics, 89*, 701-720.

[5] Andersen, T., Bollerslev, T., & Diebold, F.X. (2009). Parametric and Non-parametric Volatility Measurement. In *Handbook of Financial Econometrics* (eds Y. Ait-Sahalia and L. Hansen), Handbooks in Finance, volume 1, chapter 2. Forthcoming.

[6] Barndorff -Nielsen, O. E., & Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps . *Journal of Financial Econometrics, 2*, 1-48.

[7] Blair, J.B., Poon, SH., & Taylor, S.J. (2001). Forecasting S&P100 volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns'. *Journal of Econometrics* 105, 5-26.

[8] Brownlees, C. T. and G. M. Gallo (2010). Comparison of Volatility Measures: A Risk Management Perspective. *Journal of Financial Econometrics* 8, 29-56.

[9] Campbell, J. Y., Lettau, M., Malkiel, B.G., & Xu, Y. (2001). Have Individual Stocks become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *Journal of Finance* 56, 1-43.

[10] Chen, C. H., Yu, M. C., & Zivot, E. (2012). Predicting Stock Volatility using After-Hours Information: Evidence from the NASDAQ Actively Traded Stocks. *International Journal of Forecasting* 28, 366-383.

[11] Christensen, K., and M. Podolskij (2007). Realized range-based estimation of integrated variance. *Journal of Econometrics*, 141, 323-349.

[12] Christoffersen, P., Hahn, J., & A. Inoue (2001) Testing and Comparing Value-at-Risk Measures. *Journal of Empirical Finance 8*, 325-352.

[13] Clements, M. P., Galvão, A.B., & J.H. Kim (2008). Quantile Forecasts of Daily Exchange Returns from Forecasts of Realized Volatility. *Journal of Empirical Finance, 15*, 729-750.

[14] Clements, M.P., & Hendry, D. F. (2004). *A Companion to Economic Forecasting*. Blackwell.

[15] Corsi, F., Mittnik, S., Pigorsch, C, & Pigorsch, U. (2008). The Volatility of Realized Volatility, *Econometric Reviews*, *27,* 46-78.

[16] Delgado, M.A., & Escanciano, J.C. (2007). Nonparametric Tests for Conditional Symmetry in Dynamic Models. *Journal of Econometrics* 141, 652-682.

[17] Doornik, J. A., & Ooms, M. (2006). A Package for Estimating, Forecasting and Simulating Arfima Models: Arfima packeage 1.04 for Ox. Downloadable at www.doornik.com.

[18] Eklund, J., Kapetanios, S., & Price, S. (2009). Forecasting in the Presence of Recent Structural Change. Bank of England, mimeo.

[19] Engle, R.F. (2002). New Frontiers for ARCH Models. *Journal of Applied Econometrics, 17*, 425-446.

[20] Engle, F., Sokalska, M.E., & Chanda, A. (2006). Forecasting Intraday Volatility in the US Market. Stern School of Business, NYU, mimeo.

[21] Escanciano, J. C., & Olmo, J. (2011). Robust Backtesting Tests for VaR Models. *Journal of Financial Econometrics*, 9, 132-161.

[22] Fair, R. C., & Shiller, R. J. (1990). Comparing Information in Forecasts from Econometric Models. *American Economic Review*, 80, 375–389.

[23] 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.

[24] Fuertes, A.M., & Olmo, J. (2012). Optimally Harnessing Inter-day and Intra-day Information for Daily Value-at-Risk Prediction. *International Journal of Forecasting*, forthcoming.

[25] 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.

[26] Galbraith, J, & Kisinbay, T. (2002). Information Content of Volatility Forecasts at Medium-Term Horizons. *CIRANO Working Papers 2002s-21.*

[27] Gallo, G. (2001). Modelling the Impact of Overnight Surprises on Intra-daily Volatility. *Australian Economic Papers*, *40*, 567-580.

[28] Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies. *Journal of Econometrics, 131*, 59-95.

[29] Giacomini, R., & Komunjer, I. (2005). Evaluation and Combination of Conditional Quantile Forecasts. *Journal of Business & Economic Statistics 23*, 416-431.

[30] Giot, P., & Laurent, S. (2004). Modeling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models. *Journal of Empirical Finance 11*, 379-398.

[31] Hansen, P.R., & Lunde, A. (2005). A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics,* 3, 525–554.

[32] Hansen, P.R., & Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics*, *24*, 127-161.

[33] Ishida, I., & Watanabe, T. (2009). Modeling and Forecasting the Volatility of the Nikkei 225 Realized Volatility using the ARFIMA-GARCH Model, Global COE Hi-Stat Discussion Paper 032.

[34] Kaplanski, G., & Levy, H. (2010). The Two-Parameter Long-Horizon Value-at-Risk, *Frontiers in Finance & Economics 7*, 1-20.

[35] Koopman, S.J., Jungbacker, B., & Hol, E. (2005). Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realized and Implied Volatility Measurements. *Journal of Empirical Finance*, *12*, 445-475.

[36] Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives* 3, 73-84.

[37] Laurent, S., & Peters, J.-P. (2006). Estimating and Forecasting ARCH Models using G@ARCH 4.2. Timberlake Consultants Ltd.

[38] Liu, C., & Maheu, J. (2009). Forecasting Realized Volatility: A Bayesian Model-Averaging Approach. *Journal of Applied Econometrics 24*, 709-733.

[39] Louzis, D.P., Xanthopoulos, S.S., and A.P. Refenes (2011). The Role of High Frequency Intra-Daily Data, Daily Range and Implied Volatility in Multi-Period Value-at-Risk Forecasting. SSRN Working Paper http://ssrn.com/abstract=1970341

[40] Martens, M. (2001). Forecasting Daily Exchange Return Volatility Using Intraday Returns. *Journal of International Money & Finance*, *20*, 1-23.

[41] Martens, M., and van Dijk, D. (2007). Measuring volatility with realized range. *Journal of Econometrics,* 138, 181-207.

[42] McAleer, M. and Medeiros, M. C. (2008). Realized Volatility: A Review. *Econometric Reviews*, 27, 10-45.

[43] Patton, A., & Sheppard, K. (2009). Optimal Combinations of Realized Volatility Estimators. *International Journal of Forecasting*, *25*, 218-238.

[44] Pérignon, C., Deng, Z.Y, & Wang, Z.J. (2006). Do Banks Overstate their Value-at-Risk?, Simon Fraser University CIBC Working Paper 7.

[45] Politis, D.N., White, H., & Patton, A. (2009). Correction: Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews*, 28, 373-375.

[46] Pong., S., M. Shackleton, S. Taylor, & Xu, X. (2004). Forecasting Sterling/Dollar Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models. *Journal of Banking & Finance*, *28*, 2541-2563.

[47] Robinson, P.M., & Henry, M. (1998). Long and Short Memory Conditional Heteroskedasticity in Estimating the Memory Parameter of Levels. STICERD Econometrics EM/98/357, London School of Economics.

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

[49] Sollis, R. (2009). VaR: A Critical Overview. *Journal of Financial Regulation & Compliance, 17*, 398-414.

[50] Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. G. Timmermann (Eds.), Handbook of Economic Forecasting. Amsterdam: North Holland Press.

[51] Zhang, L., Mykland, P.A., Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100, 1394-1411.

[52] Zumbach, G. (2007). A Gentle Introduction to the RM2006 Methodology. Technical document, RiskMetrics.

Table 1. Summary statistics for model-free measures of volatility.

	ATT	**AXP**	**BA**	**CAT**	**DELL**	**GE**	**GM**	**IBM**	**JPM**	**KO**	**MCD**	**MSFT**	**PG**	**WMT**
							Squared daily returns (r_t^2)							
Mean	5.150	4.683	3.940	3.991	8.965	3.481	3.567	3.714	5.535	2.740	3.189	4.765	2.690	4.021
StDev	10.141	9.289	8.108	7.279	20.529	7.206	6.251	7.890	19.972	6.093	7.610	8.328	6.597	8.522
Skewness	5.434	5.820	5.568	4.352	9.916	7.059	4.100	8.178	20.758	8.601	9.084	5.035	8.003	6.898
Kurtosis	46.480	56.130	47.390	29.840	160.129	82.450	27.220	113.350	586.580	124.030	121.660	44.530	92.770	87.457
Q(10)	196.91*	242.16*	98.246*	90.423*	121.87*	116.61*	271.85*	81.862*	166.53*	128.27*	67.183*	142.29*	265.57*	127.67*
							Squared overnight returns ($r_{o,t}^2$)							
Mean	7.372	1.516	1.815	1.318	3.731	1.163	1.289	2.079	2.319	0.917	1.177	1.893	1.594	1.220
StDev	221.644	9.455	13.876	4.748	14.101	4.047	7.519	15.257	9.764	2.952	5.039	9.237	39.8241	3.476
Skewness	41.759	31.221	24.377	16.359	13.361	14.599	23.177	23.511	14.839	13.675	25.116	17.324	41.668	8.717
Kurtosis	1748.8	1142.9	660.37	396.39	275.33	311.74	632.078	680.00	304.34	295.54	821.77	401.45	1743.30	112.188
Q(10)	0.013	28.747*	8.661	5.809	28.215*	213.87*	4.817	2.169	52.445*	26.728*	1.726	6.193	0.021	107.71*
							Realized variance ($\mathbf{RV}_t$)							
Mean	4.506	4.673	4.078	3.782	8.170	3.648	3.021	3.572	5.622	2.836	3.550	4.441	2.905	4.140
StDev	4.356	5.065	3.950	3.176	7.689	3.776	2.950	3.663	7.717	2.417	3.337	3.892	3.080	4.298
Skewness	4.190	4.287	6.773	3.054	3.570	4.629	4.145	6.442	8.931	3.206	4.511	3.486	5.152	6.340
Kurtosis	31.610	32.033	105.245	17.800	25.870	38.220	33.070	92.930	136.670	19.070	35.830	24.562	49.260	90.930
ADF(k)	-4.73(18)*	-5.20(17)*	-8.63(7)*	-4.68(18)*	-3.75(24)*	-5.48(20)*	-6.70(7)*	-5.04(18)*	-5.31(20)*	-4.85(21)*	-5.68(15)*	-5.60(16)*	-4.49(21)*	-5.72(13)*
d^{GSP}	0.364	0.401	0.367	0.361	0.446	0.393	0.350	0.359	0.461	0.408	0.320	0.416	0.412	0.355
							Realized bipower variation ($\mathbf{RBP}_t$)							
Mean	4.090	4.292	3.702	3.401	7.652	3.416	2.734	3.340	5.175	2.598	3.252	4.146	2.676	3.724
StDev	4.252	4.732	3.775	3.048	7.480	3.710	2.783	3.210	7.024	2.369	3.263	3.753	2.879	3.888
Skewness	4.506	3.931	7.403	3.529	3.656	4.945	3.818	3.541	7.799	3.549	4.798	3.658	5.253	4.961
Kurtosis	34.320	25.598	124.150	23.793	25.351	42.870	25.731	23.868	101.450	22.800	40.060	28.230	51.281	51.810
ADF(k)	-4.79(18)*	-6.46(11)*	-8.67(7)*	-4.71(19)*	-3.82(24)*	-5.55(24)*	-6.69(7)*	-7.00(7)*	-4.98(21)*	-5.00(21)*	-6.33(12)*	-6.96(6)*	-4.42(21)*	-5.01(17)*
d^{GSP}	0.352	0.414	0.368	0.339	0.422	0.397	0.364	0.393	0.478	0.402	0.309	0.414	0.401	0.366

The sample period is January 2, 1997 to December, 31 2003 (1761 days). The daily RV and RBP series are based on 5-min prices. Q(10) is the Ljung-Box statistic for the null of no autocorrelation in squared returns up to 10 days. ADF(k) is the augmented Dickey-Fuller statistic for the unit root null with maximum lag k selected using AIC. d^{GSP} is the Gaussian semi-parametric estimator of the long memory parameter, as discussed in Robinson and Henry (1998), with standard error 0.0169. *, ** and *** denote significant at the 1%, 5% and 10% levels, respectively.

Table 2. Unconditional coverage and serial independence backtesting of Gaussian quantile-based 95% VaRs.

	ATT	**AXP**	**BA**	**CAT**	**DELL**	**GE**	**GM**	**IBM**	**JPM**	**KO**	**MCD**	**MSFT**	**PG**	**WMT**
GARCH														
$\bar{I}_\alpha(\%)$	5.000	4.200	3.800	2.800	3.800	4.000	4.400	4.200	4.200	3.400	4.400	3.200	3.800	4.600
p-val S_P^{bb}	0.924	0.328	**0.094**	**0.000**	0.126	0.212	0.400	0.356	0.358	**0.056**	0.464	**0.020**	**0.096**	0.436
p-val ξ_P^{asy}	0.100	0.293	0.518	0.711	0.794	0.852	0.361	0.293	0.913	0.180	0.977	0.646	0.495	0.156
GARCH augmented with $\mathbf{RV}_{t-1}$														
$\bar{I}_\alpha(\%)$	5.600	3.600	3.000	3.200	2.000	3.400	4.600	3.600	4.600	3.400	4.800	3.200	3.600	3.800
p-val S_P^{bb}	0.548	**0.098**	**0.012**	**0.010**	**0.000**	**0.018**	0.728	0.116	0.810	**0.028**	0.756	**0.032**	**0.038**	**0.086**
p-val ξ_P^{asy}	0.178	0.541	0.671	0.646	0.850	0.585	0.318	0.203	0.376	0.585	0.425	0.646	0.541	0.229
GARCH augmented with $\mathbf{RBP}_{t-1}$														
$\bar{I}_\alpha(\%)$	5.400	3.400	3.000	2.800	2.600	3.200	4.000	5.000	4.000	2.800	4.800	3.400	3.800	3.600
p-val S_P^{bb}	0.772	**0.032**	**0.004**	**0.000**	**0.004**	**0.010**	0.384	0.926	0.274	**0.000**	0.776	**0.058**	0.116	**0.054**
p-val $\xi_{P,1}^{asy}$	**0.017**	0.585	0.671	0.567	0.533	0.629	0.361	0.100	0.259	0.711	0.425	0.692	0.495	0.203
GARCH augmented with overnight $\mathbf{r}_{o,t}^2$														
$\bar{I}_\alpha(\%)$	5.400	4.200	3.600	3.400	4.200	3.400	5.200	3.200	4.400	3.600	3.800	2.000	3.200	4.400
p-val S_P^{bb}	0.780	0.356	**0.050**	**0.036**	0.322	**0.020**	0.722	**0.006**	0.514	0.116	0.160	**0.000**	**0.010**	0.466
p-val ξ_P^{asy}	0.611	0.293	0.541	0.585	0.293	0.585	0.738	0.629	**0.056**	0.203	0.229	0.451	0.629	**0.056**
ARMA$(\log \mathbf{RV}_t)$														
$\bar{I}_\alpha(\%)$	6.200	5.000	5.400	5.200	5.600	4.400	7.200	5.200	5.600	4.400	4.800	5.400	3.800	4.200
p-val S_P^{bb}	0.186	0.920	0.876	0.808	0.574	0.438	**0.040**	0.788	0.530	0.490	0.800	0.804	**0.080**	0.276
p-val ξ_P^{asy}	**0.023**	0.812	0.715	0.511	0.741	0.977	0.704	0.143	0.178	**0.056**	0.425	0.147	0.495	**0.003**
ARMA$(\log \mathbf{RBP}_t)$														
$\bar{I}_\alpha(\%)$	8.000	5.400	6.000	5.400	5.200	4.600	8.000	5.600	6.600	4.600	5.000	6.600	4.600	5.400
p-val S_P^{bb}	**0.056**	0.792	0.230	0.750	0.756	0.726	**0.010**	0.496	0.170	0.750	0.932	0.140	0.772	0.772
p-val ξ_P^{asy}	**0.000**	0.611	0.822	0.576	**0.082**	0.955	0.846	**0.046**	**0.087**	**0.005**	0.481	**0.087**	0.318	**0.000**
ARFIMA$(\log \mathbf{RV}_t)$														
$\bar{I}_\alpha(\%)$	6.000	5.000	5.400	5.000	5.600	3.800	7.200	5.400	5.800	4.400	5.000	5.400	4.000	4.800
p-val S_P^{bb}	0.324	0.942	0.602	0.904	0.536	0.142	**0.046**	0.768	0.416	0.512	0.934	0.770	0.172	0.738
p-val $\xi_{P,1}^{asy}$	**0.001**	0.812	0.715	0.452	0.794	0.794	0.704	0.111	**0.029**	**0.056**	0.481	0.147	0.450	**0.010**
ARFIMA$(\log \mathbf{RBP}_t)$														
$\bar{I}_\alpha(\%)$	8.400	5.400	6.000	5.400	5.200	4.800	8.000	5.600	6.400	5.000	5.200	6.600	4.400	5.400
p-val S_P^{bb}	**0.040**	0.796	0.260	0.764	0.776	0.744	**0.006**	0.498	0.222	0.948	0.782	0.132	0.422	0.732
p-val ξ_P^{asy}	**0.000**	0.611	0.822	0.576	**0.082**	0.884	0.846	**0.046**	**0.066**	**0.010**	0.543	**0.087**	0.361	**0.017**

The table reports the percentage of exceedances $\bar{I}_\alpha(\%)$ and p-values of the unconditional coverage (block bootstrap) test S_P^{bb} and of the serial independence $\xi_{P,1}^{asy}$ asymptotic test. $B = 500$ bootstrap replications.

Table 3. Unconditional coverage and serial independence backtesting of Gaussian quantile-based 99% VaRs.

	ATT	AXP	BA	CAT	DELL	GE	GM	IBM	JPM	KO	MCD	MSFT	PG	WMT
GARCH														
$\bar{I}_\alpha(\%)$	2.200	1.000	1.200	0.400	0.500	0.600	0.800	1.000	0.600	1.200	0.800	0.400	1.000	1.000
p-val S_P^{bb}	**0.028**	0.814	0.614	**0.030**	**0.048**	0.396	0.410	0.816	0.320	0.496	0.552	**0.008**	0.816	0.788
p-val $\xi_{P,1}^{asy}$	0.273	0.821	0.744	0.885	0.993	0.935	0.885	0.821	0.821	0.744	0.657	0.971	0.821	0.821
GARCH augmented with $\mathbf{RV}_{t-1}$														
$\bar{I}_\alpha(\%)$	2.400	0.600	1.000	0.400	0.200	0.600	0.600	0.800	1.000	0.800	0.800	0.200	1.400	1.000
p-val S_P^{bb}	**0.040**	0.369	0.828	**0.014**	**0.000**	0.366	0.366	0.468	0.828	0.426	0.604	**0.000**	0.550	0.824
p-val ξ_P^{asy}	0.192	0.935	0.821	0.971	0.993	0.935	0.935	0.885	0.821	0.885	**0.000**	0.993	0.657	0.821
GARCH augmented with $\mathbf{RBP}_{t-1}$														
$\bar{I}_\alpha(\%)$	2.400	0.600	1.000	0.400	0.200	0.600	0.800	0.800	1.000	0.600	1.000	0.400	1.200	1.000
p-val S_P^{bb}	**0.024**	0.369	0.816	**0.014**	**0.000**	0.358	0.408	0.426	0.828	0.324	0.812	**0.008**	0.500	0.846
p-val ξ_P^{asy}	0.192	0.935	0.821	0.971	0.993	0.935	0.885	0.885	0.821	0.935	0.821	0.971	0.744	0.821
GARCH augmented with overnight $\mathbf{r}_{o,t}^2$														
$\bar{I}_\alpha(\%)$	2.200	1.000	1.400	0.400	0.600	0.200	0.600	1.000	1.600	1.200	0.400	0.600	0.800	1.000
p-val S_P^{bb}	**0.020**	0.804	0.562	**0.008**	0.344	**0.000**	0.366	0.814	0.158	0.608	**0.080**	0.478	0.444	0.784
p-val ξ_P^{asy}	0.273	0.821	0.657	0.971	0.935	0.993	0.935	0.821	0.562	0.744	0.463	**0.000**	0.885	0.821
ARMA$(\log \mathbf{RV}_t)$														
$\bar{I}_\alpha(\%)$	2.800	1.400	1.000	1.200	0.600	0.600	1.600	1.800	1.200	0.800	2.000	0.800	1.200	1.000
p-val S_P^{bb}	**0.008**	0.560	0.824	0.550	0.372	0.374	0.256	0.138	0.582	0.396	**0.076**	0.446	0.494	0.836
p-val ξ_P^{asy}	**0.076**	0.657	0.821	0.744	0.935	0.935	0.562	**0.000**	0.744	0.885	**0.000**	0.885	0.744	**0.000**
ARMA$(\log \mathbf{RBP}_t)$														
$\bar{I}_\alpha(\%)$	3.400	2.000	1.000	2.000	1.400	0.800	2.200	1.800	2.000	1.000	2.400	1.200	1.200	1.200
p-val S_P^{bb}	**0.000**	**0.088**	0.794	**0.058**	0.612	0.460	**0.056**	0.114	**0.090**	0.818	**0.022**	0.488	0.518	0.578
p-val ξ_P^{asy}	**0.057**	0.365	0.821	**0.002**	0.657	0.885	**0.073**	**0.000**	0.365	0.821	**0.001**	0.744	0.744	**0.000**
ARFIMA$(\log \mathbf{RV}_t)$														
$\bar{I}_\alpha(\%)$	2.800	1.400	1.000	1.200	1.400	0.600	1.800	1.600	1.200	0.800	2.000	0.800	1.200	1.200
p-val S_P^{bb}	**0.010**	0.546	0.786	0.536	0.352	0.346	0.136	0.216	0.530	0.472	**0.064**	0.454	0.520	0.554
p-val ξ_P^{asy}	**0.076**	0.657	0.821	0.657	0.935	0.935	0.463	**0.000**	0.744	0.885	**0.000**	0.885	0.744	**0.000**
ARFIMA$(\log \mathbf{RBP}_t)$														
$\bar{I}_\alpha(\%)$	3.400	2.000	1.000	2.200	1.400	0.600	2.200	1.800	2.000	0.800	2.400	1.200	1.200	1.200
p-val S_P^{bb}	**0.004**	**0.052**	0.790	**0.036**	0.586	0.390	**0.064**	0.142	**0.088**	0.408	**0.034**	0.522	0.494	0.544
p-val ξ_P^{asy}	**0.057**	0.365	0.821	**0.001**	0.657	0.935	**0.073**	**0.000**	0.463	0.885	**0.001**	0.744	0.744	**0.000**

See note to Table 2.

Table 4. Encompassing regressions and tests for daily GARCH and ARFIMA(logRV) forecasts.

	Regression coefficients										Wald tests					
	ENC1			ENC2				ENC3			$H_o: \widehat{s}_t$ encompasses $\widehat{h}_t$			$H_o: \widehat{h}_t$ encompasses $\widehat{s}_t$		
Stocks	$\hat{\varphi}_0$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\gamma}_{0,1}$	$\hat{\gamma}_1$	$\hat{\gamma}_{0,2}$	$\hat{\gamma}_2$	$\hat{\eta}_0$	$\hat{\eta}_1$	$\hat{\eta}_2$	**ENC1**	**ENC2**	**ENC3**	**ENC1**	**ENC2**	**ENC3**
ATT	-0.117	0.422	0.878	-1.251	0.354	-2.091	0.325	0.003	0.391	0.325	**0.003**	**0.039**	**0.088**	**0.000**	**0.033**	**0.000**
AXP	-0.139	-0.304	1.515	-0.149	0.136	-2.176	0.534	0.631	-0.357	1.262	**0.022**	**0.063**	**0.000**	**0.000**	**0.012**	**0.000**
BA	-0.793	0.210	1.093	-0.735	0.288	-1.899	0.583	0.258	0.160	0.676	**0.002**	**0.001**	**0.014**	**0.000**	**0.000**	**0.000**
CAT	-0.478	0.014	1.234	-0.421	0.166	-3.005	0.770	0.288	-0.092	0.953	0.104	0.114	0.101	**0.000**	**0.000**	**0.000**
DELL	-0.344	0.038	1.061	-0.320	0.080	-1.410	0.052	-0.034	0.116	0.714	0.569	0.186	0.301	**0.000**	0.739	**0.000**
GE	-0.710	0.248	1.067	-0.731	0.315	-1.731	0.548	-0.281	0.226	0.438	**0.001**	**0.003**	**0.000**	**0.000**	**0.000**	**0.000**
GM	-0.625	0.431	0.788	-0.648	0.281	-0.894	0.114	-0.216	0.554	0.246	**0.002**	**0.005**	**0.007**	**0.000**	0.421	**0.002**
IBM	-0.078	-0.160	1.330	-0.285	0.151	-1.730	0.459	0.398	-0.253	1.186	**0.018**	**0.047**	**0.068**	**0.000**	**0.001**	**0.000**
JPM	-0.808	0.571	0.587	-1.839	0.414	-0.058	-0.129	-0.140	0.449	0.193	0.170	**0.041**	**0.008**	**0.073**	0.711	**0.029**
KO	-0.572	-0.103	1.261	-0.513	0.317	-1.134	0.567	0.027	0.263	0.824	0.211	**0.070**	**0.000**	**0.000**	**0.018**	**0.000**
MCD	-0.124	-0.105	1.332	0.321	0.096	-1.580	0.672	0.536	-0.129	0.974	**0.000**	0.193	0.523	**0.000**	**0.003**	**0.000**
MSFT	-0.404	-0.063	1.280	-0.287	0.122	-1.808	0.273	0.204	-0.035	0.640	**0.042**	0.275	0.142	**0.000**	0.240	**0.000**
PG	-0.248	0.115	1.092	-0.238	0.185	-0.476	0.185	-0.032	0.240	0.316	0.225	**0.054**	**0.004**	**0.000**	0.335	**0.000**
WMT	-0.208	-0.317	1.543	-0.330	0.212	-1.334	0.526	0.089	0.106	0.373	**0.020**	0.206	**0.001**	**0.000**	**0.015**	**0.000**

The table reports OLS coefficient estimates and Wald test p-values for the forecast encompassing regressions ENC1, ENC2 and ENC3 in (15), (16) and (17), respectively; $\hat{h}_t$ and $\widehat{s}_t$ are the one-day-ahead volatility forecasts using, respectively, the GARCH model and the ARFIMA(logRV) model. p-values in bold denote 'no encompassing'. Newey-West h.a.c. standard errors are used.

Table 5. Unconditional coverage and serial independence backtesting of VaRs based on GARCH and ARFIMA(logRV) combined forecasts.

	ATT	**AXP**	**BA**	**CAT**	**DELL**	**GE**	**GM**	**IBM**	**JPM**	**KO**	**MCD**	**MSFT**	**PG**	**WMT**
Gaussian quantile-based 95% VaR														
$\bar{I}_\alpha(\%)$	6.600	4.400	3.800	4.400	4.000	4.200	6.600	4.400	5.200	3.800	5.000	4.000	3.600	4.400
p-val S_P^{bb}	0.208	0.484	0.760	0.332	0.232	0.300	0.156	0.416	0.884	0.212	0.892	0.272	0.140	0.480
p-val ξ_P^{asy}	**0.087**	0.332	0.767	0.332	0.259	0.405	0.864	0.332	0.182	0.229	0.812	0.852	0.541	0.156
Gaussian quantile-based 99% VaR														
$\bar{I}_\alpha(\%)$	2.600	1.000	0.800	0.800	0.600	0.600	1.400	1.200	1.200	0.800	1.800	1.000	1.000	1.000
p-val S_P^{bb}	**0.010**	0.828	0.448	0.416	0.388	0.382	0.590	0.518	0.488	0.446	0.138	0.832	0.822	0.824
p-val ξ_P^{asy}	0.126	0.821	0.885	0.885	0.935	0.935	0.657	0.744	0.744	0.885	0.178	0.821	0.821	0.821
Student t quantile-based 95% VaR														
d.f.	9	12	9	10	10	15	9	13	14	13	13	14	18	11
$\bar{I}_\alpha(\%)$	6.200	4.400	3.800	4.200	4.000	4.600	5.800	4.400	5.600	4.200	5.000	4.400	4.600	4.400
p-val S_P^{bb}	0.314	0.526	0.132	0.304	0.212	0.746	0.374	0.466	0.570	0.330	0.904	0.482	0.774	0.428
p-val ξ_P^{asy}	**0.051**	0.332	0.668	0.293	0.259	0.955	0.767	0.332	0.178	0.146	0.812	0.332	0.318	**0.096**
Student t quantile-based 99% VaR														
d.f.	9	12	9	10	10	15	9	13	14	13	13	14	18	11
$\bar{I}_\alpha(\%)$	2.000	0.600	0.800	0.600	0.500	0.600	0.800	1.000	1.200	0.800	1.400	0.800	1.000	0.600
p-val S_P^{bb}	**0.056**	0.412	0.494	0.312	0.296	0.448	0.440	0.794	0.518	0.434	0.534	0.448	0.842	0.400
p-val ξ_P^{asy}	0.365	0.935	0.885	0.993	0.993	0.971	0.885	0.821	0.821	0.885	0.657	0.885	0.821	0.935

The table reports the percentage of exceedances $\bar{I}_\alpha(\%)$ and p-values of the unconditional coverage (block bootstrap) test S_P^{bb} and of the serial independence $\xi_{P,1}$ test. *d.f.* is the Student-t degrees of freedom parameter estimated from the returns standardized by the combined forecasts.

Table 6. Panel tests of relative VaR adequacy: unconditional coverage property.

Models $V_1 \backslash V_2$	$\alpha(\%)$	GARCH	GARCH-RV	GARCH-RBP	GARCH-r_o^2	ARMA	ARMA (logRV)	ARFIMA (logRBP)	ARFIMA (logRV)	COMB. (logRBP)
GARCH	5	–	0.995	0.886	0.620	**0.058**	**0.086**	**0.035**	**0.093**	**0.000**
	1	–	0.491	0.473	0.393	**0.084**	0.729	**0.099**	0.710	**0.020**
GARCH-RV	5	0.108	–	0.156	0.102	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	1	0.631	–	0.410	0.379	**0.097**	0.758	**0.095**	0.746	**0.024**
GARCH-RBP	5	0.131	0.936	–	0.187	**0.002**	**0.005**	**0.000**	**0.005**	**0.000**
	1	0.825	0.639	–	0.434	**0.097**	0.749	**0.099**	0.751	**0.036**
GARCH-r_o^2	5	0.364	0.956	0.744	–	**0.044**	**0.063**	**0.017**	**0.066**	**0.000**
	1	0.410	0.434	0.401	–	**0.033**	0.794	**0.031**	0.777	**0.033**
ARMA(logRV)	5	0.931	1.000	0.985	0.951	–	0.337	0.188	0.362	**0.033**
	1	0.785	0.806	0.782	0.884	–	0.973	0.631	0.978	**0.085**
ARMA(logRBP)	5	0.874	0.996	0.990	0.918	0.343	–	0.387	0.792	**0.036**
	1	0.153	0.187	0.160	0.120	**0.012**	–	**0.011**	0.108	**0.002**
ARFIMA(logRV)	5	0.942	1.000	1.000	0.979	0.740	0.740	–	0.729	**0.078**
	1	0.806	0.792	0.779	0.880	**0.000**	0.977	–	0.976	**0.093**
ARFIMA(logRBP)	5	0.868	0.997	0.988	0.919	0.347	**0.000**	**0.088**	–	**0.033**
	1	0.151	0.164	0.182	0.118	**0.010**	0.631	**0.017**	–	**0.009**
COMBINED	5	0.992	1.000	1.000	0.996	0.860	0.882	0.756	0.867	–
	1	0.935	0.934	0.945	0.932	0.732	0.993	0.746	0.996	–

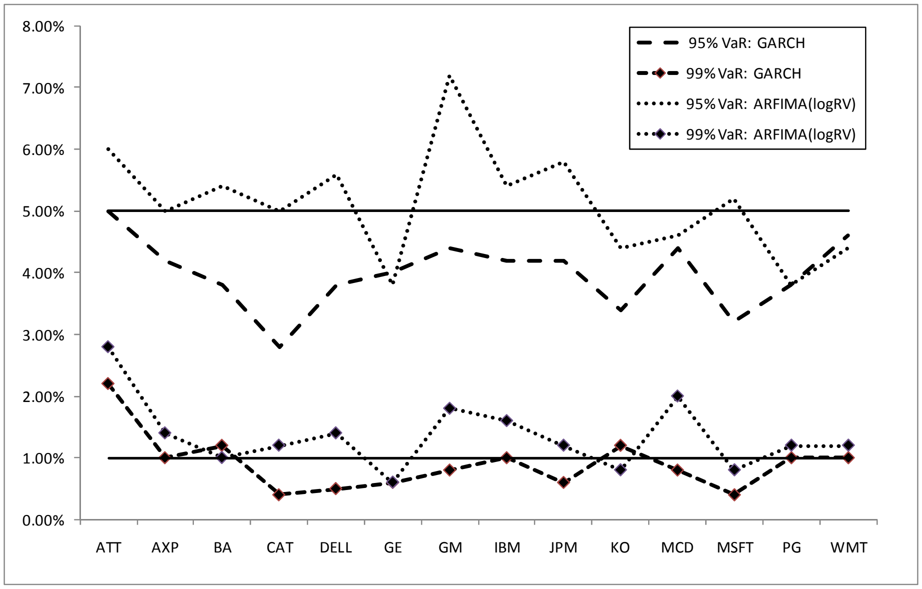
The table reports bootstrap p-values for the difference-in-proportions test $\Pi_{N,P}$ based on $B = 500$ replications; $N = 14$ stocks; $P = 500$ out-of-sample days. Bold denotes rejection at the 1%, 5% or 10% level. $H_0 : p_{V_1} \leq p_{V_2}$, where p_{V_1} is the proportion of violations of the *unconditional coverage* hypothesis for VaR model V_1. VaR predictions based on the forecasts from each of the volatility models together with Gaussian quantiles of the innovation's distribution. Small p-values (< 0.10) in bold indicate that model V_1 (first column) leads to significantly worse VaR coverage than model V_2 (first row).

Table 7. Panel tests of relative VaR adequacy: serial independence property.

Models $V_1 \backslash V_2$	$\alpha(\%)$	GARCH	GARCH-RV	GARCH-RBP	GARCH-r_o^2	ARMA	ARMA (logRV)	ARFIMA (logRBP)	ARFIMA (logRV)	COMB. (logRBP)
GARCH	5	–	0.500	0.736	0.885	0.942	1.000	0.974	1.000	0.857
	1	–	0.713	0.500	0.748	0.974	0.998	0.986	0.999	0.500
GARCH-RV	5	0.500	–	0.731	0.871	0.939	1.000	0.976	1.000	0.876
	1	**0.084**	–	**0.082**	0.334	0.931	0.994	0.949	0.994	**0.067**
GARCH-RBP	5	**0.070**	**0.071**	–	0.616	0.857	1.000	0.942	1.000	0.169
	1	0.500	0.745	–	0.737	0.974	1.000	0.977	0.998	0.500
GARCH-r_o^2	5	**0.040**	**0.039**	0.191	–	0.624	0.995	0.883	0.995	**0.094**
	1	**0.063**	0.371	0.175	–	0.893	0.983	0.901	0.973	**0.079**
ARMA(logRV)	5	**0.013**	**0.018**	**0.039**	0.199	–	0.980	0.747	0.977	**0.079**
	1	**0.006**	**0.041**	**0.007**	**0.073**	–	0.871	0.338	0.864	**0.009**
ARMA(logRBP)	5	**0.000**	**0.000**	**0.000**	**0.000**	**0.016**	–	0.112	0.616	**0.000**
	1	**0.000**	**0.006**	**0.000**	**0.018**	**0.064**	–	**0.051**	0.378	**0.000**
ARFIMA(logRV)	5	**0.005**	**0.004**	**0.030**	**0.084**	0.112	0.967	–	0.951	**0.005**
	1	**0.008**	**0.029**	**0.003**	**0.073**	0.627	0.864	–	0.876	**0.004**
ARFIMA(logRBP)	5	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	0.352	**0.017**	–	**0.000**
	1	**0.000**	**0.008**	**0.000**	**0.019**	**0.057**	0.188	**0.069**	–	**0.000**
COMBINED	5	**0.073**	**0.094**	0.726	0.881	0.937	1.000	0.984	0.999	–
	1	0.500	0.760	0.500	0.742	0.975	1.000	0.987	0.998	–

The table reports bootstrap p-values for the difference-in-proportions test $\Pi_{N,P}$ based on $B = 500$ replications; $N = 14$ stocks; $P = 500$ out-of-sample days. Bold denotes rejection at the 1%, 5% or 10% level. $H_0 : p_{V_1} \leq p_{V_2}$, where p_{V_1} is the proportion of violations of the *serial independence* hypothesis for VaR model V_1. VaR predictions based on the forecasts from each of the volatility models together with Gaussian quantiles of the innovation's distribution. Small p-values (< 0.10) in bold indicate that model V_1 (first column) leads to significantly worse VaR adequacy in terms of serial independence than model V_2 (first row).

Daily VaR exceedances (%)



Proportion of backtesting rejections

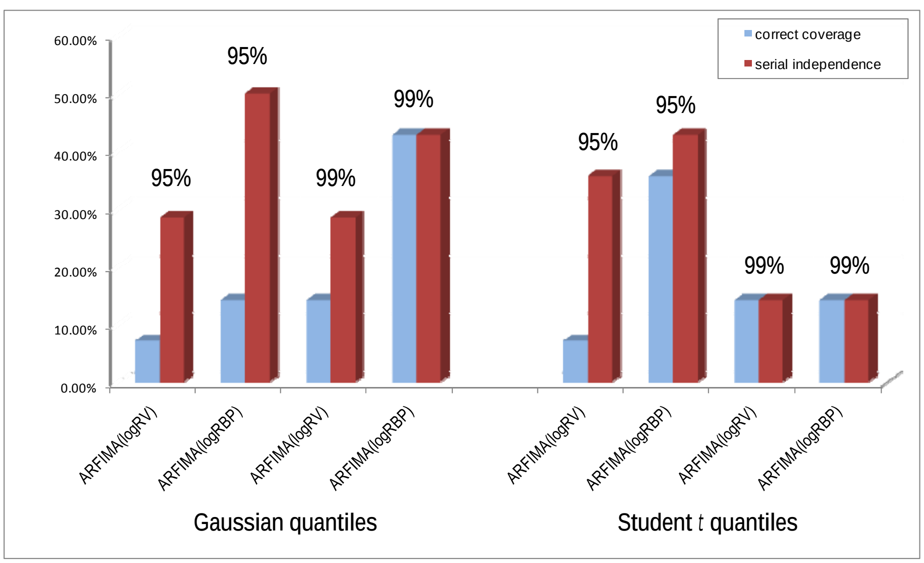


Figure 1. The top panel displays the percentage of daily VaR exceedances over a time-series of 500 days. The bottom panel depicts the proportion of VaR adequacy rejections across 14 stocks.

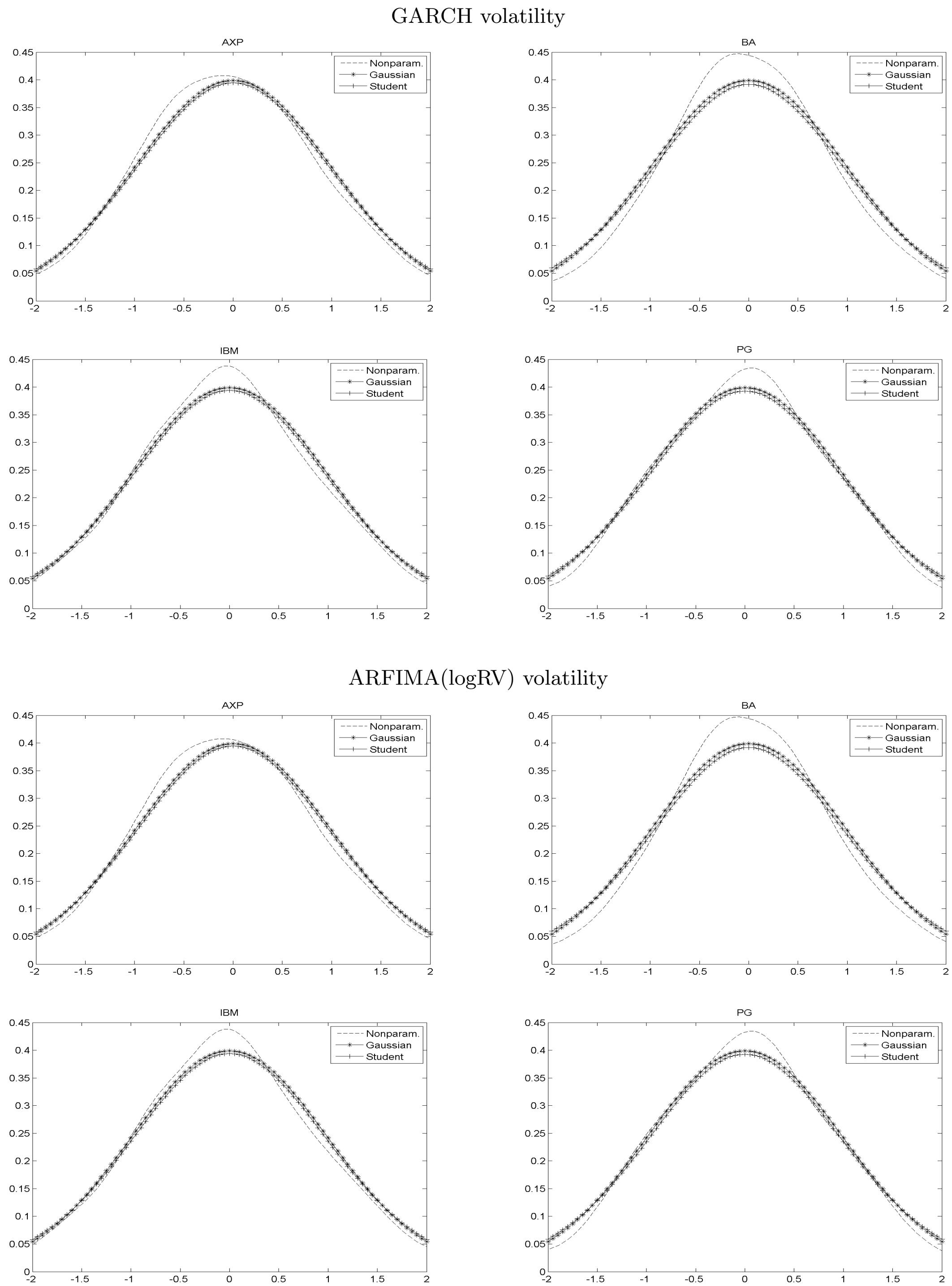


Figure 2. Density functions of returns standardized by GARCH and ARFIMA(logRV)forecasts: American Express (AXP; Financial), British Airways (BA; Industrial), IBM (Technology), Procter & Gamble (PG; Consumer Goods).