Optimally Harnessing Inter-day and Intra-day Information for Daily Value-at-Risk Prediction

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*

First version: June 2010; This version: April 2012

Abstract

We make use of quantile regression theory to obtain a combination of individual potentially-biased VaR forecasts that is optimal because it meets by construction ex post the correct out-of-sample conditional coverage criterion. This enables a Wald-type conditional quantile forecast encompassing test for any finite set of competing (semi/non)parametric models which can be nested. Two attractive properties of this backtesting approach are robustness to model risk and estimation uncertainty. We deploy the techniques to confront inter-day and high frequency intra-day VaR models for equity, FOREX, fixed income and commodity trading desks. Forecast combination of both types of models is especially warranted for more extreme-tail risks. Overall our empirical analysis supports the use of high frequency 5-minute price information for daily risk management.

Keywords: Quantile regression; Optimal forecast combination; Encompassing; Conditional coverage; High-frequency data; Realized variance.

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. The authors thank the editor, Esther Ruiz, and two anonymous referees for very useful comments. Wei Liu provided assistance with processing the high frequency data. We also acknowledge valuable suggestions from Katja Ahoniemi, Dick van Dijk, Tony Garratt, Kavita Sirichand, Ron Smith, William Pouliot, participants at the 2009 Oxmetrics Conference at Cass Business School in London, 2009 CSDA Conference on Computational and Financial Econometrics in Cyprus, 2011 GdRE International Simposium on Money, Banking and Finance at the University of Reading, 2011 MMF Conference at the University of Birmingham, 2011 CFE-ERCIM Conference at Birkbeck College London, 2012 CMS Conference at Imperial College London, and seminar at the Swiss Institute of Banking and Finance, University of St Gallen. Financial support from the Spanish Government through project MICINN ECO2011-22650 is gratefully acknowledged.

1 Introduction

Since the regulatory framework known as Basel II Accord (BCBS, 1996), the financial risk measure known as Value-at-Risk (VaR) has gained increasing popularity as tool for controling the risk of trading portfolios. Banks routinely compute VaR as a way to estimate the potential loss from adverse market moves. On a daily basis, as part of their risk management activities, banks undertake VaR calculations that are used for comparing risks across businesses and monitoring limits. The tail-risk measures thus obtained are reported to senior management and regulators, and they feed regulatory capital calculations. VaR is typically calculated using a 1-day time horizon and a 99% (or 95%) confidence level. This means the bank would expect to incur losses exceeding the VaR prediction one (or five) times in every 100 trading days, or about 2 to 3 (12 to 13) times a year. Basel II has promoted the internal model-based approach but banks have to convince the regulator that their VaR model is sound via so-called backtesting techniques. Significant progress has been made in the financial econometrics literature making available a number of VaR estimation methods.

Despite the fact that semi-parametric and non-parametric VaR techniques have gained a lot of ground, the parametric location-scale framework remains the most-widely used in the academic literature. In location-scale models, which require a distributional assumption for the standardized returns, the VaR forecast is a blend of conditional mean and variance forecasts. Essentially the VaR is made a linear function of the volatility. Although seemingly restrictive, this framework is quite popular because it is both tractable and flexible enough to accommodate diverse stylized facts of financial assets such as the asymmetric impact of good and bad news on volatility and fat-tailedness. Surprisingly, most of the risk management literature concerned with daily VaR prediction has neglected high frequency intra-day information despite the significant theoretical progress witnessed in the last decade regarding model-free "realized" volatility estimators constructed ex post from intra-day data (see, for instance, Barndorff-Nielsen et al., 2008). Empirically, special efforts have been devoted to improve the forecasts from GARCH models based on daily data. A good rationale for these efforts is that the squared close-to-close return is an extremely noisy estimator of ex post volatility.

Location-scale VaR models can easily make use of different information sets via the conditional volatility component. This paper is concerned with two specific types of information, *inter-day*, which is defined as daily-recorded closing, high and low prices, and *intra-day* that refers to high frequency 5-minute prices. Following existing studies, the inter-day data is exploited in a GARCH framework whereas the intra-day data is put to work in an AR(FI)MA realized volatility framework; see Giot and Laurent (2004), Clements et al. (2008), Angelidis and Degiannakis (2008), Shao et al. (2009), and Brownlees and Gallo (2010) inter alios. Going one way or the other is not a trivial issue, in fact, the literature has not reached yet any widely accepted conclusion. One attractive but unexplored avenue is combining the merits of both inter-day models and intra-day models into a single VaR forecast.

Despite the acclaimed success of *forecast combination* in many contexts, this tool has been barely utilized for tail risk (e.g., VaR) prediction where a plethora of modeling approaches are available which can be bewildering for risk managers. Giacomini and Komunjer (2005) represents the first attempt to combine quantile forecasts via a GMM estimation approach. They propose a conditional quantile forecast encompassing (CQFE) test in an environment with non-vanishing estimation uncertainty which, under suitable assumptions on the conditioning set, amounts to testing for correct out-of-sample VaR specification. Such CQFE test serves as a VaR predictive ability test that is naturally immune to model risk and estimation uncertainty and hence, it is superior to the usual statistical backtesting that focuses on *correct unconditional coverage* and *serial independence* of the hits sequence.[1] The implementation of Giacomini and Komunjer's (2005) approach is, however, not straightforward because it requires the numerical estimation of parameters inside discontinuous moment conditions.

The contributions of the present paper are both theoretical and empirical. Regarding the former,

[1] The Basel II traffic light regulatory backtesting also suffers from model risk and estimation uncertainty. One instance is the Daily Capital Charges or Market Risk Capital (MRC) measure which is set at the supremum of the last trading day VaR and the average VaR over the past 60 trading days multiplied by a violation penalty factor $(3 + k)$ from the Basel Accord Penalty Zone. MRC serves as a conservative estimate of capital required to cover daily market risk that corrects for past under-estimation of risk levels by the financial institution. We refer the reader to McAleer and da Veiga (2008) and Chen et al. (2011) for further discussion and applications of the regulatory-based VaR evaluation framework.

the first contribution is to enhance the literature by proposing an optimal conditional VaR forecast combination method that builds on quantile regression (QR) theory. The method is optimal because the combining weights are obtained by minimizing conditionally the 'tick' loss function that is implicit in quantile forecasting problems. The optimality of the method, in turn, implies that the resulting VaR forecast satisfies ex post out-of-sample the *correct conditional coverage* criterion. Our framework is inspired by Engle and Manganelli's (2004) CAViaR approach but differs from it in that the regressors are VaR forecasts based on different location-scale models instead of lagged values of the unobserved quantile of interest. In a sense, our method allows the risk manager to exploit different information sets for setting trading-loss limits more easily than the CAViaR approach. The second theoretical contribution is to deploy a tractable Wald-type conditional quantile forecast encompassing (CQFE) test that also builds on quantile regression theory and compares the distance between the optimal combined forecast and each of the individual VaR forecasts. The null hypothesis that one of the individual VaR forecasts is undistinguishable from the optimal combined forecast implies also that it encompasses the rest of competing VaR forecasts in the combination set and that it meets, by construction, out-of-sample ex post the correct conditional coverage criterion.

At an empirical level, our contribution is to propose CQFE inference as a novel way of confronting *inter-day* and *intra-day* models for downside tail risk prediction. To the best of our knowledge, we are the first to develop an empirical exercise of this type. As highlighted by Granger (1989), the situation where forecasts obtained from different information sets are combined is a particularly interesting and potentially fruitful one because forecast combination lends itself as an effective (i.e. bias-correcting) information pooling device. Since VaR measurement is quite sensitive to the scant data points that fall in the sample distribution tails, forecast combination as a way of expanding the information set might be particularly worthwhile. For illustration, we combine the daily forecasts obtained from two distinct VaR models within the location-scale family: an inter-day Student t GJR-GARCH (GJR) model that exploits the daily close-to-close return and high-low range and an intra-day ARFIMA realized volatility model based on higher frequency 5-minute prices. Our goal is to formally assess whether GJR-based or ARFIMA-based forecasts subsume entirely the information content in the other for one-day-ahead 1% and 5% tail risk prediction. The techniques are illustrated in a univariate approach for individual equity, FOREX, fixed income and commodity trading desks.

As noted earlier, although a few studies have somehow confronted inter-day and intra-day VaR predictions, the evidence thus far is not entirely conclusive. For instance, in the context of data up to year 2001 on the CAC40 index, S&P500 futures and the YEN/USD and DEM/USD currencies, Giot and Laurent (2004) conclude that ARFIMA realized volatility-based VaR forecasts are unable to beat those obtained from the skewed Student (skt) APARCH model. Angelidis and Degiannakis (2008) illustrate for 1995-2003 data on the CAC40, DAX30 and FTSE100 indexes that the Gaussian GJR model produces as accurate VaRs as an ARFIMA realized volatility specification. On the other hand, observing the Shanghai Composite and Shenzhen Component indices over the period 2005-2007, Shao et al. (2009) conclude that the ARFIMA realized volatility model outperforms the skt-APARCH model. This mixed evidence may "hint" that neither inter-day nor intra-day based VaR forecasts fully encompass each other. Information pooling via forecast combination may prove useful for bank managers to control the discretion of their traders by setting optimal VaR limits on their portfolios.

Our novel quantile regression-based framework confirms that, generally, for the 1% tail there is no evidence of encompassing which calls for optimal combination of the intra-day and inter-day models. Overall the equal-weights combined VaR forecast, a natural benchmark, appears suboptimal out-of-sample in a *correct conditional coverage* sense. For the 5% tail risk, by contrast, ARFIMA forecasts encompass GJR forecasts in various contexts and hence, the intra-day conditional VaR model appears correctly specified out-of-sample. Thus our paper formally demonstrates through robust CQFE inference that intra-day information is worthwhile to measure and control the daily risk of trading portfolios. The rest of the paper unfolds as follows. Section 2 motivates our VaR modeling and forecasting choices. Section 3 proposes a novel conditional quantile forecast encompassing test in a quantile regression framework and Section 4 deploys it to analyze equity, FOREX, fixed income and commodity VaR forecasts. Section 5 concludes.

2 VaR Modeling and Forecasting

Despite the fact that VaR has unappealing properties (i.e., failure to meet subadditivity which makes it a non-coherent risk measure)[2] it is now established as the standard tool to measure and control the risk of trading portfolios. Commercial banks routinely calculate the 1-day-ahead VaR of their 'trading book' in order to quantify the mark-to-market loss that will not be exceeded during the day with a 99% (or 95%) probability. Formally, the task is forecasting the α-quantile of the conditional distribution of the return process, $\alpha = \{0.01, 0.05\}$, given the sigma-algebra $\Im_t$ generated by information up to time t, i.e. $P(r_{t+1} \leq VaR_{t+1,\alpha}|\Im_t) = \alpha$, where $VaR_{t+1,\alpha} \equiv F_t^{-1}(\alpha)$ is the inverse of the conditional return distribution function.[3] The literature has proposed various VaR approaches which, depending on the assumptions made on the returns distribution and model dynamics, can be grouped as non-parametric (e.g., historical simulation), semi-parametric (e.g., CAViaR) and parametric (e.g., location-scale).

A popular VaR approach assumes that the returns density belongs to a location-scale family

$$r_{t+1} = \mu_{t+1} + \sigma_{t+1}\varepsilon_{t+1} \tag{1}$$

where location μ_{t+1} and scale σ_{t+1} are $\Im_t$-measurable functions, and ε_{t+1} is an i.i.d. innovation with zero-location unit-scale probability density F_ε (see e.g., Kuester et al., 2006; Chernozhukov and Umantsev, 2001). By assuming independence between σ_{t+1} and ε_{t+1}, the conditional VaR process at nominal coverage probability α can be expressed as follows

$$VaR_{t+1,\alpha} = \mu_{t+1} + \sigma_{t+1}F_\varepsilon^{-1}(\alpha) \tag{2}$$

with $F_\varepsilon^{-1}(\alpha)$ the unconditional α-quantile of ε_{t+1}. This popular framework allows wide flexibility for choosing the specific form of μ_{t+1} and σ_{t+1}, and can accommodate many candidate probability distri-

[2]Subadditivity is a property of a mathematical function $q(\cdot)$ by which the value of the function at point $A+B$ is less or equal than the sum of the function's values at points A and B, that is, $q(A+B) \leq q(A) + q(B)$.

[3]Implicitly, we are interested in long trading positions. For short trading, one would forecast instead the right-tail risk of the distribution, i.e. the quantile $F_t^{-1}(1-\alpha)$. Commercial banks are required to report VaR at confidence level 99% to regulators but most banks adopt the 95% level for internal backtesting. We consider both levels given that the best VaR model among various competitors can depend on the specific quantile of choice, and may also be horizon- and asset-specific (see, for instance, Guidolin and Timmermann, 2006).

butions F_ε. A notable example is the pioneering RiskMetrics methodology initiated by J.P.Morgan in 1993 where σ_{t+1} can be cast as a specific Gaussian IGARCH type process. Departures from the latter are often motivated by a desire to accommodate: i) asymmetry in the response of volatility to positive and negative returns (e.g., GJR-GARCH or EGARCH models), ii) regime-switching nonlinearity in μ_{t+1} and/or σ_{t+1} (e.g., MS-GARCH models), and iii) departures from Gaussianity, particularly, in the form of fat-tailedness (e.g., Student t density commonly adopted by academics and practitioners).[4]

The semi-parametric VaR framework initiated by Engle and Manganelli (2004), known as conditional autoregressive Value-at-Risk (CAViaR), consists of modeling directly the dynamics of the α-quantile process, $F_t^{-1}(\alpha)$, as a function of its own unobserved lagged values, $F_{t-j}^{-1}(\alpha)$ and past observed returns r_{t-j}. This approach requires assumptions about the quantile dynamics but not about the conditional return distribution. In a recent paper, Chen et al. (2011) propose range-based CAViaR model extensions which are threshold nonlinear according to the daily-recorded high and low prices.

Our paper borrows elements from both of the above frameworks. On the one hand, and without loss of generality, it adopts location-scale models to generate individual VaR forecasts that exploit inter-day and intra-day data, respectively. On the other hand, the optimal VaR forecast combination approach we put forward is linked to the CAViaR modeling approach in that we are deriving a quantile measure in a linear regression framework as a function of other quantile measures as regressors.

The location-scale framework (2) lends itself as a feasible way of incorporating high-frequency (e.g. 5-minute) price information into daily VaRs through the so-called realized volatility. In this vein, our paper builds upon the studies by Giot and Laurent (2004), Angelidis and Degiannakis (2008) and Martens et al. (2009) on stock indices, Clements et al. (2008) on exchange rates and Brownlees and Gallo (2010) on specific NYSE stocks. Like these (and other) VaR studies, we focus on VaR as a tool to gauge market risk exposure, that is, to set loss limits for individual trading desks and, accordingly,

[4] Giot and Laurent (2004) stress the importance of acknowledging asymmetry in the conditional variance specification together with departures from Gaussianity in the ex ante standardized returns distribution. Haas et al. (2004), Guidolin and Timmermann (2006) and, more recently, Sajjad et al. (2008) advocate the use of Markov-switching GARCH (MS-GARCH) models to obtain more accurate VaR predictions, particularly, for risks in the far tail (i.e. $\alpha = 0.01$).

the one-day-ahead forecast horizon is very short term (BCBS, 1996).

Our ultimate goal is to test formally whether it is worthwhile to exploit jointly two information sets: i) *daily-recorded* close, high and low prices, and ii) higher frequency *intra-day* prices. The first information set is mapped into a daily return, defined as close-to-close price differences $r_t \equiv \log \frac{p_t^C}{p_{t-1}^C}$, and a daily volatility measure defined according to Parkinson's (1980) range estimator

$$HL_t = \frac{1}{4 \log 2} \left[\log \frac{p_t^H}{p_t^L} \right]^2 \tag{3}$$

where $t = 1, ..., T$ are sample days, $4 \log 2$ is a scaling factor, p_t^H and p_t^L are high and low prices, respectively. Parkinson (1980) demonstrates that scaled daily highlow range is not only an unbiased estimator of daily volatility but is 5 times more efficient than the squared daily close-to-close return.

The second higher frequency (5-minute) information set is mapped into a daily *realized variance* computed as the sum of the M intra-day squared returns corresponding to equal-length intervals from day open to close, $\widetilde{RV}_t = \sum_{j=1}^{M} r_{t,j}^2$. Under certain conditions, this estimator converges in probability (as $M \to \infty$) to the quadratic variation process $QV_t = \int_{t-1}^{t} \sigma^2(u)du + \sum\limits_{t-1<j\leq t} \kappa^2(j)$ which comprises a continuous part (integrated variance) and a discontinuous part (jump variation).[5] In markets with no (or thin) trading during the so-called overnight period, the realized variance can underestimate the latent QV_t. To mitigate this problem, we incorporate the overnight 'surprise' using Hansen and Lunde's (2005) optimal weighting method which fares well in comparisons with alternative methods (see Ahoniemi and Lanne, 2011). Thus our overnight-adjusted realized variance (RV_t) estimator is

$$RV_t = \tau_1^* r_{o,t}^2 + \tau_2^* \sum_{j=1}^{M} r_{t,j}^2 \tag{4}$$

where $r_{o,t}$ is the overnight close-to-open return; τ_1^* and τ_2^* are the weights that solve the optimization problem $\min\limits_{\tau_1,\tau_2} \text{var}(RV_t)$, s.t. $\tau_1\eta_1 + \tau_2\eta_2 = \eta_0$ where $\eta_1 \equiv E(r_{o,t}^2)$, $\eta_2 \equiv E(\widetilde{RV}_t)$ and $\eta_0 \equiv E(QV_t)$.

We consider two forms for σ_{t+1} in (2) which are broadly representative of much of the research in this area; both of them are coupled with an AR(1) formulation, $r_{t+1} = \rho_0 + \rho_1 r_t + z_{t+1}$, so that 'location'

[5]Alternatives include the *realized bipower variation* which converges in probability to the integrated variance (i.e. excluding the jump component), and various *kernel-based realized variance* estimators proposed by Barndorff-Nielsen et al. (2008) to mitigate market microstructure noise. See McAleer and Medeiros' (2008) survey and references therein.

is given by $\mu_{t+1} \equiv \rho_0 + \rho_1 r_t$. Our inter-day model belongs to the GARCH class that accommodates the asymmetry known as *leverage effect*: past negative returns have a larger impact on current volatility than positive returns of the same magnitude.[6] Various reasonable candidates are the GJR-GARCH (GJR), TGARCH, EGARCH and APARCH whose properties are amply discussed in Rodríguez and Ruiz (2012). Following Angelidis and Degiannakis (2008) and Corrado and Truong (2008), we opt for the GJR model of Glosten et al. (1993) enhanced with range information HL_t as defined in (3). This enhancement is motivated by the emphasis made in a number of recent papers on daily range as very effective for out-of-sample financial risk prediction (e.g., Chen et al., 2011; Brownlees and Gallo, 2010; Corrado and Truong, 2007; Brandt and Jones, 2006). In particular, Corrado and Truong (2007) provide empirical evidence that the range-augmented GJR model produces similarly accurate volatility forecasts as those stemming from high-quality implied volatility indexes and more accurate ones than the baseline GJR model. Because the distribution of asset returns is admittedly fat-tailed we work with standardized Student t innovations. Formally, our inter-day model is given by

$$h_t = \alpha_0 + \alpha_1 z_{t-1}^2 + \beta_1 h_{t-1} + \gamma_1 I_{t-1}^- z_{t-1}^2 + \delta HL_{t-1} \tag{5}$$

with $z_t = \sqrt{h_t}\varepsilon_t$ and ε_t is an i.i.d. Student $t(0,1,\upsilon)$ innovation where υ is the degrees of freedom parameter that reflects fat-tailedness; $I_t^- = 1$ if $z_t < 0$ and $I_t^- = 0$ otherwise; $\gamma_1 > 0$ implies that large negative returns increase the conditional volatility of returns more than past positive outcomes[7]. The one-day-ahead long trading VaR is thus obtained by plugging in (2) the conditional mean and volatility forecasts, $\hat{\mu}_{t+1}$, and $\sqrt{\hat{h}_{t+1}}$, together with the α-quantile of the Student $t(0,1,\upsilon)$ density where υ is estimated on the basis of the information set available at time t. Each trading day $t = R, ..., T-1$, the parameter vector $(\rho_0, \rho_1, \alpha_0, \alpha_1, \beta_1, \gamma_1, \delta, \upsilon)'$ is updated using the most recent $R-$length rolling estimation window in order to obtain a sequence of $n = T - R$ out-of-sample forecasts.

[6] The volatility-feedback hypothesis put forward by Campbell and Hentschel (1992) states a slightly different asymmetry where causality runs in the other direction: current positive shocks to volatility drive down future returns.

[7] In this augmented GJR model, the coefficient δ plays a role to ensure positivity and weak stationarity of the conditional volatility process. A sufficient condition for positivity is $\alpha_0 > 0$, β_1, $\gamma_1 \geq 0$, and $\alpha_1 z_{t-1}^2 + \delta HL_{t-1} \geq 0$ for all t. A sufficient condition for weak stationarity is $\gamma_1 < 2(1 - \alpha_1 - \beta_1 - \delta)$. The kurtosis expression is more involved than in the baseline GJR model due both to the kurtosis of the range HL_{t-1} and the interaction of HL_{t-1} and z_{t-1}^2.

Our second set of daily VaR predictions is constructed from 5-minute intra-daily prices through the following long-memory ARFIMAX$(0, d, 1)$ model for the essentially-Gaussian daily $\log RV_t$ series[8]

$$(1-L)^d(\log RV_t - \omega_0 - \omega_1 r_{t-1} - \omega_2 r_{t-1}^-) = (1+\theta_1 L)u_t \tag{6}$$

where u_t is i.i.d. $N(0, \sigma_u^2)$; $r_t^- = r_t$ if $r_t < 0$ and $r_t^- = 0$ otherwise. A parameter value $\omega_2 < 0$ is suggestive of the leverage effect. ARFIMAX models were initially developed by Granger and Joyeux (1980) who provide a detailed account of their theoretical properties. Giot and Laurent (2004) originally deploy (6) for VaR prediction and put forward the two-step approach that we adopt. In step one, the conditional variance obtained by exact ML estimation of (6) under normality is transformed into levels using the bias-corrected mapping $\widehat{RV}_{t|t-1} = \exp(\widehat{\log RV}_{t|t-1} + 0.5\hat{\sigma}_u^2), t = 1, ..., T$. In step two, the conditional mean and variance forecasts, $\widehat{\mu}_{t+1}$ and $\widehat{\sigma}_{t+1}^2$, are obtained by QML estimation of an AR(1) model for the daily returns assuming that the conditional heteroskedasticity is of ARFIMAX type, i.e. $r_{t+1} = \rho_0 + \rho_1 r_t + \sqrt{\sigma^2 \cdot \widehat{RV}_{t|t-1}} \cdot \varepsilon_t$ where $\varepsilon_t \sim$ i.i.d. Student $t(0, 1, \upsilon)$. Out-of-sample VaR forecasts are obtained by combining $\hat{\mu}_{t+1}$, $\hat{\sigma}_{t+1}$, $t = R, ..., T-1$, and a Student $t(0, 1, \upsilon)$ α-quantile with υ estimated sequentially from the standardized returns as in the inter-daily framework.

3 Combining and Backtesting Competing VaR Forecasts

The benefits of combining forecasts from a number of preferably distinct methods have been repeatedly demonstrated in diverse areas; see Clements and Hendry (2004) and Clemen (1989). Timmermann (2006) provides a threefold rationale for why combined forecasts work well: they are less influenced by possible misspecification of individual models; they average across differences in the way individual forecasts are affected by structural breaks; and they exploit jointly the information contained in each individual forecast. Information pooling can be especially fruitful for VaR because tail risk measurement is very sensitive to the scant observations in the sample distribution tails.

[8]Other specifications that have been shown in the empirical finance literature to approximate well the long memory properties of the realized volatility, namely, the hyperbolic decay of its autocorrelation function, are Corsi's (2004) Heterogeneous AutoRegressive (HAR) model and Ghysels et al.'s (2004) MIxed DAta Sampling (MIDAS) model. Both of these specifications combine information sampled at different frequencies.

3.1 Optimal Out-of-Sample VaR Forecast Combination

This section proposes a novel estimation technique for constructing optimal combinations of a finite set of (potentially biased) out-of-sample VaR forecasts. For simplicity in the exposition and given that our empirical application compares VaRs obtained from inter-day and intra-day models, without loss of generality, we focus the theoretical discussion on bivariate combinations alongside a constant.

In what follows the subscript t for an $\alpha-$quantile forecast $\widehat{q}_t$ or density function $f_t(\cdot)$ indicates conditioning on the information set available at time t, denoted $\Im_t$. Consider $\widehat{q}_t \equiv (1, \widehat{q}_{1,t}, \widehat{q}_{2,t})', t = R, ..., T-1$, a sequence of vectors containing two competing out-of-sample $VaR_{t+1,\alpha}$ forecasts, such that $corr(\widehat{q}_{1,t}, \widehat{q}_{2,t}) \neq 1$. Let $\widehat{q}_{c,t} \equiv \lambda' \cdot \widehat{q}_t$ with $\lambda \equiv (\lambda_0, \lambda_1, \lambda_2)'$ denote a quantile forecast combination that depends on the combination weights (λ_1, λ_2) corresponding to the individual forecasts $(\widehat{q}_{1,t}, \widehat{q}_{2,t})$, and on an intercept term λ_0 whose role is to correct for potential biases in the individual forecasts induced by model misspecification. Nonlinear combinations of quantile forecasts can be entertained in our framework. As in Giacomini and Komunjer (2005), we impose a linear structure $\widehat{q}_{c,t} \equiv \lambda_0 + \lambda_1 \widehat{q}_{1,t} + \lambda_2 \widehat{q}_{2,t}$, but the combining framework is quite flexible in that the weights λ_1 and λ_2 are not constrained to lie in $(0, 1)$ nor to sum up to one.

Following standard practice, we assume that the decision-maker or forecast user adopts a loss function that only depends on the forecast error. Let $\mathbb{L}$ denote an arbitrary loss function that maps forecast errors into losses $\mathbb{L} : \mathbb{R} \to \mathbb{R}^+$. A combined forecast is generally said to be conditionally *optimal* if it minimizes the expected loss $\mathbb{L}$ conditional on a given information set. The interest in optimally combining conditional quantile forecasts leads us naturally to adopt the asymmetric (piecewise linear) loss function $\varphi_\alpha(e_{t+1}) \equiv (\alpha - 1(e_{t+1} < 0))e_{t+1}$ where the error is defined as $e_{t+1} \equiv r_{t+1} - \widehat{q}_{c,t}$ and $1(\cdot)$ is the indicator function; this so-called 'tick' or 'check' loss function penalizes (for the left tail) negative errors or exceedances $r_{t+1} < \widehat{q}_{c,t}$ more heavily with weight $(1-\alpha)$ than positive errors with weight α. The optimal forecast combination problem can be stated as

$$(\lambda_0^*, \lambda_1^*, \lambda_2^*) \equiv \underset{(\lambda_0, \lambda_1, \lambda_2) \in \Lambda}{\arg\min} \; E[\varphi_\alpha(r_{t+1} - \widehat{q}_{c,t}) | \Im_t] \quad (7)$$

with $\Lambda \subset \mathbb{R}^3$ a compact set. The above expected loss is *conditional* on the information set $\Im_t$, and so it fundamentally differs from the unconditional framework that underlies extant tests of equal predictive ability; see Diebold and Mariano (1995), Corradi and Swanson (2002) or Clements et al. (2008).[9] This aspect will have implications for the interpretation of our encompassing test in Section 3.2.

The first-order condition corresponding to (7) is given by

$$E[g(\lambda^*; r_{t+1}, \widehat{q}_t)|\Im_t] = 0, \tag{8}$$

with $g(\lambda^*; r_{t+1}, \widehat{q}_t) = (\alpha - 1(r_{t+1} < \lambda_0^* + \lambda_1^* \widehat{q}_{1,t} + \lambda_2^* \widehat{q}_{2,t}))\, \widehat{q}_t$. Next we make the following assumption:

Assumption A.1 (Information set). The conditional density function of r_{t+1} defined as $f_t(r_{t+1}) \equiv f(r_{t+1}|\Im_t)$ satisfies that $f_t(r_{t+1}) = f(r_{t+1}|\widehat{q}_t)$ for all $y \in \mathbb{R}$ and $t = R, \ldots, T-1$.

This assumption amounts to saying that the information contained in $\Im_t$ can be encapsulated in the vector of forecasts $\widehat{q}_t$.[10] Under assumption A.1, the above first-order condition becomes

$$E[\alpha - 1(r_{t+1} < \widehat{q}_{c,t})|\widehat{q}_t] = 0. \tag{9}$$

We can recast the optimal weight estimation problem in terms of the quantile regression (QR) model

$$r_{t+1} = \lambda_0 + \lambda_1 \widehat{q}_{1,t} + \lambda_2 \widehat{q}_{2,t} + \epsilon_{t+1} \tag{10}$$

where ϵ_{t+1} is an *iid* error process with finite variance and $F_\epsilon^{-1}(\alpha) = 0$. The essence of equations (9)-(10) is to state that the optimal α-quantile combination is the linear combination of quantiles $\widehat{q}_{1,t}$ and $\widehat{q}_{2,t}$, given by the weights $(\lambda_0, \lambda_1, \lambda_2)$, that minimizes the conditional expected 'tick' loss over Λ.

This leads to a very tractable weight estimation procedure that builds on QR theory. In particular, we rely on Koenker and Xiao's (2006) theoretical setup in the context of conditional autoregressions

[9] Moreover, such tests are not valid for nested models or when the VaRs are obtained semi- or non-parametrically.

[10] This type of assumption on the conditioning information set is standard in the forecasting time-series literature. More specifically, it has been adopted by Su and Xiao (2008) for parameter stability testing in the context of quantile regressions, and by Giacomini and Komunjer (2005) for quantile forecast evaluation.

for quantiles of stationary and $\beta-$mixing processes. Accordingly, the weight estimates are

$$\widehat{\lambda}_n^{*\prime} \equiv (\widehat{\lambda}_{0n}^*, \widehat{\lambda}_{1n}^*, \widehat{\lambda}_{2n}^*) \equiv \arg\min_{\lambda\in\Lambda} \frac{1}{n}\sum_{t=R}^{T-1} \varphi_\alpha(r_{t+1} - (\lambda_0 + \lambda_1 \widehat{q}_{1,t} + \lambda_2 \widehat{q}_{2,t})), \tag{11}$$

Equations (10) and (11) can be seen as an extension to a quantile context of the typical linear regression setup where the optimal combined forecast is the conditional mean, that is, the expected value of r_{t+1} given the individual mean forecasts; the loss function $\mathbb{L}(e_{t+1}) \equiv e_{t+1}^2$ is quadratic and hence, the minimization criterion is the mean square error (MSE) instead of (11). Under suitable regularity conditions, Koenker and Xiao (2006; Theorem 3.1) establish the statistical consistency and asymptotic distribution of QR parameter estimates. Accordingly, in the present setting we have

$$\sqrt{n}(\widehat{\lambda}_n^* - \lambda^*) \xrightarrow{d} N(0, \Sigma_{QR}), \tag{12}$$

with $\Sigma_{QR} \equiv \alpha(1-\alpha)\Omega_1^{-1}\Omega_0\Omega_1^{-1}$ where $\Omega_0 \equiv E\left[\widehat{q}_t \widehat{q}_t'\right]$, $\Omega_1 \equiv E\left[f_t(\lambda^{*\prime}\widehat{q}_t)\widehat{q}_t\widehat{q}_t'\right]$ and $f_t(\lambda^{*\prime}\widehat{q}_t)$ is the conditional density function of r_{t+1} evaluated at the optimal conditional quantile combination. Timmermann (2006) discusses GMM estimation methods for optimal forecast combinations based on loss functions of general form. Focusing also on the out-of-sample 'tick' loss function, Giacomini and Komunjer (2005) make use of GMM theory to obtain the weight estimates for optimal conditional quantile forecast combination. Their weights estimator $\widetilde{\lambda}_n^*$ is defined as

$$(\widetilde{\lambda}_{0n}^*, \widetilde{\lambda}_{1n}^*, \widetilde{\lambda}_{2n}^*) \equiv \min_{\lambda\in\Lambda} g_n(\lambda; r_{t+1}, \widehat{q}_t)' \widehat{S}_n^{-1} g_n(\lambda; r_{t+1}, \widehat{q}_t) \tag{13}$$

with $g_n(\lambda; r_{t+1}, \widehat{q}_t) \equiv \frac{1}{n}\sum_{t=R}^{T-1} g(\lambda; r_{t+1}, \widehat{q}_t)$, and $\widehat{S}_n$ the empirical version of $S \equiv E[g(\lambda; r_{t+1}, \widehat{q}_t) g(\lambda; r_{t+1}, \widehat{q}_t)']$.

Under assumption A.1, it can be shown that our QR estimator $\widehat{\lambda}_n^*$ and Giacomini and Komunjer's (2005) GMM estimator $\widetilde{\lambda}_n^*$ are asymptotically equivalent. The limiting properties of the latter are

$$\sqrt{n}(\widetilde{\lambda}_n^* - \lambda^*) \xrightarrow{d} N(0, \Sigma_{GMM}), \tag{14}$$

where $\widetilde{\lambda}_n^* \equiv (\widetilde{\lambda}_{0n}^*, \widetilde{\lambda}_{1n}^*, \widetilde{\lambda}_{2n}^*)'$ is the solution of (13) and $\Sigma_{GMM} \equiv (\Omega_1 S^{-1} \Omega_1)^{-1}$ with Ω_1 as defined above. By applying the Law of Iterated Expectations conditional on $\widehat{q}_t$, we observe that $S = \alpha(1-\alpha)\Omega_0$ and hence, both QR and GMM estimation methods are equivalent in terms of asymptotic efficiency.

The main difference between them lies in their finite sample properties and practical implementation. Regarding the latter, the QR-based optimal forecast combination we propose builds on standard kernel density estimation methods. By contrast, GMM is technically more challenging: the search for the minimum is computationally more intensive and depends heavily on the initial conditions provided; the method also requires several iterations to achieve an efficient estimator. Moreover, the specific approach put forward by Giacomini and Komunjer (2005) hinges on the choice of a parameter τ that suitably smooths the sample moment function $g_n(\cdot)$ in order to estimate the matrix Ω_1.

The weight estimates that define the optimal forecast combination are a combination of in-sample information $\Im_t$, used to obtain the individual forecasts $\widehat{q}_{1t}$ and $\widehat{q}_{2t}$, and out-of-sample information $\{r_{t+1}\}_{t=P}^{T-1}$. Therefore the resulting optimal combined VaR forecast is not a fair competitor to the individual VaR forecasts. However, it provides the basis for a conditional quantile forecast encompassing (CQFE) test that serves for backtesting analysis because it can detect which of two (or more) competing VaR models meets ex post the correct conditional coverage criterion out-of-sample.[11]

3.2 A Test for Correct Out-of-Sample VaR Specification

A correctly specified α-th conditional VaR model of an asset or portfolio returns r_t satisfies that

$$P(r_{t+1} \leq VaR_{t+1,\alpha} \mid \Im_t) = \alpha, \text{ almost surely (a.s.)}, \ \alpha \in (0,1), \ \forall t \in \mathbb{Z}. \tag{15}$$

This conditional moment restriction has been used in various theoretical papers (Christoffersen et al., 2001; Engle and Manganelli, 2004; Gourieroux and Jasiak, 2006; Koenker and Xiao, 2006).

However, condition (15) is not what is usually backtested in practice. Most backtesting approaches focus instead on assessing some of the implications of criterion (15) rather than the criterion itself. More specifically, a VaR model is typically considered adequate *iff* the out-of-sample hits or exceedances associated with the VaR forecasts, defined as $I_{t+1,\alpha} \equiv 1(r_{t+1} \leq VaR_{t+1,\alpha})$, exhibit both correct unconditional coverage, $E[I_{t+1,\alpha}] = \alpha$, and serial independence. Condition (15) is sufficient

[11] The optimal weight estimation approach (11) exploits the *entire* out-of-sample period and hence, we obtain a single weights vector as opposed to time-varying weights. Nevertheless, our optimal weight combination framework would be deployed in real time over out-of-sample windows that are rolled forward day by day so the weights evolve accordingly.

(but not necessary) for the *correct unconditional coverage* and *independence* properties to be fulfilled. There is a large class of VaR models which are misspecified in the sense that they do not satisfy (15) but nevertheless they yield an i.i.d. sequence of out-of-sample hits with correct unconditional coverage; this mismatch is known as model misspecification or model risk in the VaR literature. Escanciano and Olmo (2011) circumvent the bias induced by the latter in the context, for instance, of Kupiec's (2005) unconditional coverage test statistic by deriving the correct asymptotic variance in the presence of model risk (and estimation uncertainty); a block-bootstrap is suggested as a feasible alternative in cases when the adjusted variance is too cumbersome to estimate in practice.

In what follows, we propose instead backtesting directly criterion (15) via a novel CQFE inference approach; this represents a novel test for correct out-of-sample specification ex post. It builds upon the first-order condition (8) and assumption A.1 that together imply that the optimal forecast combination $\widehat{q}_{c,t} \equiv \lambda_0^* + \lambda_1^*\widehat{q}_{1,t} + \lambda_2^*\widehat{q}_{2,t}$ satisfies property (15) by construction. Intuitively, the latter is possible because, in contrast to $\widehat{q}_{1,t}$ and $\widehat{q}_{2,t}$, the optimally combined forecast $\hat{q}_{c,t}$ is obtained by exploiting both in-sample and out-of-sample information. Thus our test is similar in spirit to Giacomini and Komunjer's (2005) CQFE inference approach built around the following definition:

Definition 1 (Encompassing). Let $\widehat{q}_{1,t}$ and $\widehat{q}_{2,t}$ be two alternative forecasts of the $VaR_{t+1,\alpha}$ quantile of interest. Forecast $\widehat{q}_{1,t}$ encompasses forecast $\widehat{q}_{2,t}$ conditionally at time $t = R, ..., T-1$ *iff*

$$E[\varphi_\alpha(r_{t+1} - \widehat{q}_{1,t})|\Im_t] = E[\varphi_\alpha(r_{t+1} - \widehat{q}_{c,t})|\Im_t]. \tag{16}$$

The above definition departs from the classical encompassing literature in that it is based on conditional expected losses as opposed to unconditional ones. The main hypotheses of interest, $H_{10} : \lambda^* = (0, 1, 0)'$ against $H_{1a} : \lambda^* \neq (0, 1, 0)'$, and $H_{20} : \lambda^* = (0, 0, 1)'$ against $H_{2a} : \lambda^* \neq (0, 0, 1)'$, correspond to testing, respectively, whether $\widehat{q}_{1,t}$ encompasses $\widehat{q}_{2,t}$, and whether $\widehat{q}_{2,t}$ encompasses $\widehat{q}_{1,t}$, conditionally.

We propose the following Wald test statistics in a QR estimation framework:

$$ENC_1 = n(\widehat{\lambda}'^*_n - (0, 1, 0))\widehat{\Sigma}_n^{-1}(\widehat{\lambda}_n^* - (0, 1, 0)'), \tag{17}$$

and

$$ENC_2 = n(\widehat{\lambda}'^*_n - (0,0,1))\widehat{\Sigma}_n^{-1}(\widehat{\lambda}^*_n - (0,0,1)'), \tag{18}$$

where $\widehat{\Sigma}_n \equiv \alpha(1-\alpha)\widehat{\Omega}_{1,n}^{-1}\widehat{\Omega}_{0,n}\widehat{\Omega}_{1,n}^{-1}$ is the estimator of the covariance matrix Σ_{QR} in (12) with $\widehat{\Omega}_{0,n} \equiv \frac{1}{n}\sum_{t=R}^{T-1}\widehat{q}_t\widehat{q}_t'$, $\widehat{\Omega}_{1,n} \equiv \frac{1}{2nh_n}\sum_{t=R}^{T-1}1(|r_{t+1} - \widehat{q}_{c,t}| \leq h_n)\widehat{q}_t\widehat{q}_t'$, and $h_n = \nu \cdot n^{-1/3}$ with $\nu > 0$ is a bandwidth parameter satisfying that $h_n \to 0$ and $nh_n^2 \to \infty$ as $n \to \infty$. This kernel-type matrix estimator builds upon the method proposed by Powell (1991) and applied by Angrist et al. (2006) inter alios; see also Koenker (2005). It follows that $ENC_i \xrightarrow{d} \chi_3^2$ under H_{i0} and $ENC_i \xrightarrow{d} \infty$ under H_{ia}, $i = 1, 2$, as $n \to \infty$. The finite-sample performance of these Wald tests mainly depends on the estimates of Σ_{QR}; this inference methodology is standard in the QR literature and hence, we do not pursue further a more in-depth study of finite-sample size and power properties (see Koenker, 2005).

Akin to Giacomini and Komunjer's (2005) CQFE test, our encompassing test for correct ex post out-of-sample specification of the conditional VaR measure lets the out-of-sample size n go to infinity, while the in-sample size R remains finite. The use of fixed or rolling (but not recursive) schemes to obtain the VaR predictions $\widehat{q}_{1,t}$ and $\widehat{q}_{2,t}$ implies that estimation risk does not vanish asymptotically as $n \to \infty$; thus the test naturally controls for the effect of parameter uncertainty. The latter implies that general VaR modeling approaches are accommodated, e.g. parametric GARCH-based and semi-parametric CAViaR, and that the encompassing test is also valid in the context of nested models.

We conclude this section with comments on the CQFE test interpretation. Inference on the constant allows us to ascertain whether the individual forecasts are both unbiased ($H_{00} : \lambda_0^* = 0$) or there is bias at least in one of them ($H_{0a} : \lambda_0^* \neq 0$). If only one of the two encompassing hypotheses is rejected, say, H_{10}, the implication is that $\hat{q}_{2,t}$ encompasses $\hat{q}_{1,t}$ and, in turn, the VaR_2 model used to obtain $\widehat{q}_{2,t}$ represents itself the optimal "combination" that meets the correct out-of-sample conditional VaR specification criterion. When both hypotheses, H_{10} and H_{20}, are rejected it is concluded that neither competing model, VaR_1 or VaR_2, is correctly conditionally specified out-of-sample and hence, combination is beneficial. However, if the estimation results further indicate, say, that λ_1 is insignificantly different from zero but λ_0 and λ_2 are both significant, one can conclude that the *bias-corrected* VaR_2

model has correct conditional coverage out-of-sample. If none of the tests is rejected, the backtesting is not informative enough (low signal-noise ratio) and so one could choose either model. Lastly, we should stress that the dependence between individual VaR forecasts, which plays an important role in the reliability of the ENC_i test results, is captured by the covariance matrix $\widehat{\Sigma}_n$.

4 Empirical Application

4.1 Data and Preliminary Statistics

We apply the outlined quantile combination methods and encompassing tests in a univariate fashion to the problem of setting VaR limits for equity, FOREX, fixed income and commodity trading desks.[12] We have 6 datasets comprising open, close, high and low prices sampled every day and 5-minute sampled prices pertaining to a recent 14-year period ending 31/05/2011 that is currently available. This allows us to make the forecasting task more challenging by examining how the risk models perform during the recent global financial crisis (GFC) period. In particular, we choose 01/09/2008 as start date for the holdout (or evaluation) period.[13] The intra-day sampling frequency is 5 minutes.[14]

Three equity indices are chosen to cover different segments of the stock market: the S&P 500 index, by far the most common benchmark for funds (mutual funds, ETFs, pension funds) that identify themselves as large cap; the Russell 2000 index which is the typical benchmark for funds that categorize themselves as small cap; and the Nasdaq Composite which serves as indicator of the

[12]For the purpose of reporting the overall trading VaR of the financial institution one can either assume a known vector of weights and model the time series of portfolio returns (univariate approach) or model the joint dynamics of the assets contained in the overall trading portfolio (multivariate approach). The recommendations arising from the literature in this regard lack consensus as yet. Whereas some papers endorse the univariate modeling approach (e.g., Brooks and Persand, 2003; Bauwens et al., 2006; Christoffersen, 2009), others produce mixed evidence (McAleer and da Veiga, 2008) or strongly support multivariate modeling (Santos et al., 2009). A distinctive aspect of the latter study is that it deals with relatively large portfolios (of up to 81 assets) and allows for dynamic conditional correlation structures.

[13]September 2008 is an important landmark of the late 2000s GFC because several important events occurred during this month, for instance: i) Fannie Mae and Freddie Mac, two U.S. government sponsored enterprises, owned or guaranteed nearly $5 trillion in mortgage obligations at the time they were placed into conservatorship by the U.S. government in September 7, ii) Lehman Brothers filed for Chapter 11 bankruptcy protection on September 15.

[14]The data source is Disk Trading http://www.is99.com/disktrading/. The 5-minute frequency is the most typical in the literature because it appears short enough for the daily volatility dynamics to be picked up with reasonable accuracy (small estimation error) and long enough for the adverse effects of market microstructure noise not to be excessive.

performance of technology and growth stocks. The fact that the latter index considers non-U.S. companies differentiates it further from the former two indices. The observations are available from 9:30-16:00 Eastern time (EST) which amounts to $M = 78$ five-minute intra-day intervals. Thus the closing price on day t is defined as the last transaction price observed before 16:00, and the intra-day 'closing' price $p^c_{t,j}$ is similarly defined as the last seen tick before the jth 5-minute mark. The observed opening price on day t is the first transaction price recorded after 9:30. The three equity samples span the period from 12/11/1997 to 31/05/2011 ($T = 3404$ trading days).

Fourth, we consider the US Dollar Index (USDX) which is a weighted geometric mean of the dollar's value vis-à-vis Euro (57.6% weight), Japanese Yen (13.6%), Pound sterling (11.9%), Canadian dollar (9.1%), Swedish krona (4.2%), and Swiss franc (3.6%). The underlying spot rates are averages of bid-ask quotes. Like for the above equity scenarios, we consider a portfolio consisting of a long position in the index which can be easily achieved through ETFs. The FOREX market operates round-the-clock (24 hr) and so we have $M = 288$ 5-minute intervals for each of $T = 3361$ days spanning the period from 16/04/1998 to 31/05/2011. As is standard practice, we take 21:00 GMT as the 'closing' time each day and remove all returns from Friday 21:05 GMT to Sunday 21:00 GMT, and the very slow trading days around Christmas day (24th-26th December), New Year (31st December, 1st-2nd January), Good Friday, Easter Monday, Memorial day, Labour day and Thanksgiving day.

In order to include bond (fixed income) trading desks in our tail-risk analysis, we gather prices on 10-year US Treasury Notes futures contracts from 12/06/1997 to 31/05/2011. This choice obeys the fact that the 10Y T-Note has become the most frequently quoted security when discussing the performance of the U.S. government bond market and is widely thought to convey the market's take on long term macroeconomic expectations. The prices are for contracts traded on the open outcry Chicago Board of Trade (CBOT) from 8:20-15:00 EST; thus $M = 76$ intraday 5-minute intervals.[15]

[15] The Chicago Mercantile Exchange (CME) acquired the CBOT in 2007. Although there is an electronic trading system taking place nearly round-the-clock, we focus on the main open outcry CME trading hours like Giot and Laurent (2004) and Žikěs (2009), among others, who use price quotes for S&P500 futures contracts and WTI Crude Oil futures contracts traded, respectively on the CME from 9:30-16:00 EST and the NYMEX from 9:00-15:00 EST, thus ignoring the electronic trading hours. Flemming (1997) shows that the U.S. Treasury market behaves more like U.S. equity markets,

Our final dataset consists of Gold futures prices over the period 03/12/1997 to 31/05/2011 ($T =$ 3395 days). Our choice relates to the role of Gold as a key component of global monetary reserves and also for currency hedging and trading (i.e. traditional safe haven). Our quotes pertain to the open outcry CBOT from 8:20-13:30 EST (thus $M = 58$) and, like for the T-Notes futures, refer to the nearest-to-maturity contract until the next contract becomes more active in terms of trading volume.

Figure 1 plots the three daily volatility series, squared returns (r_t^2), range volatility (HL_t) and realized volatility (RV_t), as defined in Section 2. The graphs reflect clear episodes of heightened turbulence in financial markets, particularly, in the final (initial) months of 2008 (2009) following the effective nationalization of Fannie Mae and Freddie Mac on 07/09/2008 (owning about 1/2 of the $12 trillion US mortgage market) which sent waves of panic to home mortgage lenders and Wall Street, and the filling for bankruptcy of Lehman Brothers one week afterwards. The big spike in the graph for Gold in October 1999 mirrors an agreement by fifteen European central banks to limit sales which pushed the price to a high of $338 per ounce. Other turmoil episodes borne out by the graphs are the dotcom bubble (particularly, reflected in Nasdaq) which burst on March 2000 and was deflating at full speed by 2001. Market jitters were felt, particularly, in the S&P 500 and Russell 2000 indices, following an announcement on April 2010 that the Securities and Exchange Commission sues Goldman Sachs for failing to disclose important information on one of their mortgage-backed CDOs in 2007 to the benefit of John Paulson's hedge fund. Descriptive statistics for the daily returns, squared returns, $\log HL_t$ and $\log RV_t$, set out in Table 1, confirm various stylized facts.[16] The volatility autocorrelation function shows a very slow (hyperbolic) declining rate, particularly, in the case of $\log RV_t$. Approximately Gaussian properties are observed for both $\log HL_t$ and $\log RV_t$. In line with the theoretical discussion in Parkinson (1980), the scaled range is a rather efficient volatility measure whereas the squared return

with limited trading hours, than like the round the-clock FOREX market. Not only price volatility and trading volume are highly concentrated during 8:30-15:00 EST but also over this period the U-shaped patterns of price volatility, trading volume, and the bid-ask spread resemble those found in equity markets (but not in the FOREX market).

[16]Hansen and Lunde's (2005) approach deployed on our data (except for USDX) to mitigate the overnight bias produced weights that place disproportionately more importance on $\widetilde{RV}_t$ than on the overnight return; e.g, $\tau_1^* = 0.181$ and $\tau_2^* = 1.022$ for S&P 500 and $\tau_1^* = 0.322$ and $\tau_2^* = 1.357$ for Gold in equation (4). See also Ahoniemi and Lanne (2011).

is very noisy. Returns are mildly skewed but highly kurtosed, and the positive skewness for Gold is not an unusual phenomenon in commodity futures (e.g., see Rallis et al., 2012).

The estimation results and diagnostics for the models are set out in Table 2.[17] For the USDX and the 10Y T-Note volatility, the GJR model parameter γ_1 suggests that there is no leverage effect in line with previous studies (e.g., Giot and Laurent, 2004; Andersen et al., 2011; Rodriguez and Ruiz, 2012). For equities, the significantly positive γ_1 confirms the stylized fact that bad news lead to higher subsequent volatility than good news. In the case of Gold, the significantly negative coefficient γ_1 reflects the so called 'inverse leverage effect' which stems from the distinctive influential forces on commodity futures prices such as, for instance, supply (production) and demand imbalances, and the interaction of hedgers and speculators. In fact, it is not uncommon to witness periods where both equity and commodity markets experience extreme volatility but with prices falling in the former and rising in the latter. The parameter ω_2 in the ARFIMA equation mirrors the leverage findings from the GJR model. The long-memory parameter d confirms the weak stationarity but slow autocorrelation decay of $\log RV_t$; it is relatively high for equities but below 1/2 according to statistical tests.

Both frameworks suggest that there is considerable fat-tailedness in standardized returns (except for the Russell index) as borne out by the dof parameter υ which justifies the Student t density for the innovation process. The extremely high leptokurtosis (small υ) observed for Gold may be linked to the occasional but very aggressive price swings that are characteristic of energy and metals (Füss et al., 2010; Rallis et al., 2012). The Ljung-Box test applied to the standardized residuals in levels and squares suggests that the models are reasonably well specified to capture the dynamics of returns.

[17]Model estimation and forecasting are carried out using the G@RCH 4.2 and ARFIMA 1.0 packages for *Oxmetrics*. The reported GJR model estimates suggest that the conditional volatility process satisfies the weak stationarity condition for all datasets. For equities, the parameter α_1 is negative but very small and positivity of the conditional variance is still guaranteed because the larger (in absolute value) positive δ coefficient of HL_{t-1} has an offsetting effect, namely, the condition $\widehat{\alpha}_1\widehat{z}_{t-1}^2 + \widehat{\delta}HL_{t-1} \geq 0$, where $\widehat{z}_t$ are residuals, is met for all t. Thus for equities the contribution of the lagged range HL_{t-1} is larger than that of the lagged squared return innovation z_t^2.

4.2 Encompassing test and optimal VaR combinations

We now discuss the empirical findings from the CQFE test deployed out-of-sample. The number of observations n ranges from 692 days (equity) to 716 days (FOREX). Table 3 summarizes the results. The top and bottom panels are for the 5% VaR and 1% VaR, respectively. Columns with the heading $fr_i\%$ report the failure rate or percentage of days when the daily loss exceeds the VaR prediction; $i = 1, 2, c$ denote the GJR model, ARFIMAX model and optimally combined model, respectively.

We start by discussing the results from our QR weight estimation framework. For the 5% VaR, it is observed in the case of Nasdaq and USDX that the ENC_1 test rejects the null hypothesis of conditional encompassing but ENC_2 does not; this implies that the ARFIMA forecasts encompass the GJR forecasts. Moreover, for the S&P 500 index the weight estimate on the GJR forecast is not significantly different from zero suggesting that overall in three (out of six) cases the optimal combination is the (bias-corrected) ARFIMA forecast. For Gold futures, the results are difficult to interpret. Whereas the ENC_i $(i = 1, 2)$ tests point toward the encompassing of the ARFIMA forecast by the GJR forecast, the optimal combination weights virtually suggest the opposite. This inconclusive evidence may arise because the encompassing inference is contaminated by high dependence in the individual forecasting errors which is reflected in the off-diagonal terms of the covariance matrix $\widehat{\Sigma}_{QR}$ (see Timmermann, 2006). Only two cases, Russell 2000 and 10Y T-Notes, visibly call for forecast combination since both ENC_i $(i = 1, 2)$ statistics reject the encompassing null hypothesis.

For the 1% VaR, there is more evidence calling for combination: in 4 out of 6 cases (S&P 500, Russell, T-Notes and Gold) both ENC_i tests plainly reject suggesting that neither forecast encompasses its competitor. With the remaining two datasets, the evidence is suggestive of encompassing. For the Nasdaq portfolio, the GJR-based 1% tail risk forecast is statistically not different from the optimal one, whereas for the USDX portfolio the ARFIMA 1% tail risk forecast emerges as optimal. Thus the overall findings differ for the two quantiles under consideration, 5% and 1%. In particular, the merit of combining inter-day and intra-day information becomes more apparent as one moves further inside the tail. This may be linked to the fact that reliably forecasting extreme quantiles is challenging because of the scant data points that fall in the corresponding sample distribution tails and hence, information pooling through forecast combination becomes more crucial.

The ENC_e test in Table 3 corresponds to the null hypothesis that the simple combination of VaR forecasts $\widehat{q}_{e,t} = \frac{1}{2}(\widehat{q}_{1,t} + \widehat{q}_{2,t})$ is optimal, that is, it meets the correct conditional coverage condition, i.e. $H_{e0} : \lambda^* = (0, 0.5, 0.5)'$ against $H_{ea} : \lambda^* \neq (0, 0.5, 0.5)'$.[18] Focusing on the 6 cases (2 for 5% VaR and 4 for 1% VaR) where no evidence of encompassing was found, the optimal combining weights are significantly different from $(0, 0.5, 0.5)'$ so the test rejects H_{e0}. Indirectly, this suggests that the benefits from freely-estimating the combining weights versus fixing them a priori at (0, 0.5, 0.5) outweigh the noise contamination arising from the high correlation between the VaR forecasts being combined.

The optimal QR weight estimates $\widehat{\lambda}^*_n$ in Table 3 deserve some attention. They are obtained using (11) and their standard errors are the square root of the diagonal entries of $\widehat{\Sigma}_n$ divided by $\sqrt{n}$. Following Koenker's (2005) recommendation, our bandwidth parameter is $h_n = \nu n^{-1/3}$ with $\nu = 1$; the results hardly vary for alternative ν.[19] Despite the fact that the first two columns $fr_i\%, i = 1, 2$ show often that both individual VaR models underestimate the risk exposure, the optimal VaR forecast combination can plausibly be convex or non-convex because the intercept plays a bias-correcting role.[20] We observe that for the 10Y T-Notes the ARFIMAX forecast has a large positive contribution ($\hat{\lambda}^*_{2n} > 1$) whereas, on the contrary, the GJR forecast has a negative weight ($\hat{\lambda}^*_{1n} < 0$). The former acts as a correction towards making the combined VaR forecast more conservative than the individual $VaR_{ARFIMAX}$ forecast; in fact, the size of the weight represents an over-correction which is compensated by the negative weight on the VaR_{GJR} forecast. Lastly, a noteworthy result is that the rejection rate of the optimally combined forecast ($fr_c\%$) is virtually identical to the nominal coverage; this is

[18]This test is motivated by the empirical forecasting literature which has shown in diverse scenarios, other than VaR, that the simple equal-weight combination is rather effective in out-of-sample evaluations (e.g. Patton and Sheppard, 2009) even though weight estimates from optimal combination methods often do not bear it out.

[19]We also considered values $\nu = \{0.5, 3\}$. Detailed results are available from the authors upon request.

[20]In the present context that combines individual GJR-based and ARFIMA-based quantile forecasts, denoted $\widehat{q}_{1t}$ and $\widehat{q}_{2t}$, the expression 'convexity' refers to a linear combination such that $\widehat{\lambda}^*_{1n} > 0$, $\widehat{\lambda}^*_{2n} > 0$ and $\widehat{\lambda}^*_{1n} + \widehat{\lambda}^*_{2n} = 1$.

to be expected because, by construction, the optimally combined forecast satisfies condition (15) ex post and, in turn, the *correct unconditional coverage* and *iid* properties are met.

For completeness, we also report in the right-hand-side of Table 3 the counterpart results from Giacomini and Komunjer's (2005) GMM weight estimation framework. We use $\tau = \{0.006, 0.01, 0.05\}$ as smoothing parameter values; the first two are adopted in Giacomini and Komunjer (2005) whereas the third one may conform better with our out-of-sample size n at about 700 days; the outcome is very close in all three cases and Table 3 pertains to $\tau = 0.01$. Two striking aspects of these results are: *i*) weight estimation noise as suggested by the standard errors is generally larger than in the QR framework, *ii*) overall the results are less informative. The encompassing of either individual 5% VaR forecast cannot be falsified for Nasdaq, USDX and 10Y T-Notes, i.e. neither H_{10} nor H_{20} are rejected, suggesting that both the GJR-based and ARFIMAX-based tail risk models are correctly specified out-of-sample. There are another five similar inconclusive cases for the 1% VaR. At the same time, there is virtually no statistical evidence against the optimality of the equal-weight combination. Thus the CQFE test is not helpful in these eight cases since it cannot discriminate in terms of correct out-of-sample VaR specification, criterion (15), between the GJR and ARFIMAX forecasts. Ambiguous CQFE inference together with relatively large standard errors of the optimal combining weights indirectly suggests that the GMM approach may necessitate in this quantile context greater out-of-sample sizes, n, to produce reliable CQFE inference. In fact, the simulations conducted in Giacomini and Komunjer's (2005) to infer acceptable levels for the smoothing parameter τ are based on large $n = \{1000, 2500, 5000\}$ in relation to those typically employed in the empirical forecasting literature.

Figure 2 plots individual 5% VaR forecasts alongside optimal combined forecasts.[21] In all cases, the predicted downside risk is far greater in the earlier part of the out-of-sample period (September 2008 to March 2009) that represents the peak of the late 2000s GFC. For the S&P 500, Nasdaq and FOREX portfolios, the graphs confirm that the ARFIMAX forecasts are closer to the optimal QR-based combination than the GJR forecasts. The remaining graphs are less revealing.

[21] The graphs for the 1% VaR are qualitatively similar and are available from the authors upon request.

5 Conclusions

A distinctive theoretical aspect of this paper is that it extends the forecast combination and encompassing regression literature to a quantile context. Building on Giacomini and Komunjer's (2005) conditional quantile forecast encompassing (CQFE) test, which is GMM-based and necessitates appropriate smoothing of the 'check' loss function, we propose CQFE inference that builds instead on standard quantile regression (QR) theory. The implementation of our novel CQFE test requires the estimation of a combined VaR forecast that is *optimal* because it meets ex post, by construction, the correct out-of-sample conditional VaR coverage criterion. The CQFE test here proposed controls for model risk and estimation uncertainty and is quite general in that it can be applied to any number of VaR forecasts obtained from different (semi/non) parametric approaches, including nested models. Thus the CQFE test serves as a more robust backtesting approach than the unconditional coverage and independence backtesting typically adopted by banks and regulators.

At an empirical level, the main novelty of the paper is to propose conditional quantile forecast combination as an effective device to confront and pool inter-daily and intra-daily information. The empirical analysis is grounded in two VaR models from the location-scale family, namely, the individual quantile forecasts are obtained from an asymmetric GJR-GARCH (GJR) model that exploits daily-recorded closing, high and low prices, and from an asymmetric ARFIMAX realized volatility model that exploits higher frequency 5-minute prices. The techniques are illustrated in a univariate fashion on six datasets that pertain to equity, FOREX, fixed income and commodity trading desks.

Overall the QR-based inference suggests that, especially, for far-tail risks (1% VaR) the contest between inter-daily and intra-daily models is quite tight and calls for optimal quantile forecast combination. The simple equal-weights forecast combination is strongly rejected as optimal. For the less extreme tail risk (5% VaR), the intra-day ARFIMA forecasts encompass inter-day GJR forecasts in various cases despite the fact that the GJR model is helped along with range (high minus low) price information. The FOREX portfolio analysis stands out by unambiguously endorsing the ARFIMA-based tail risk predictions at both the 1% and 5% nominal coverage levels. Our findings based on two well-known inter-day and intra-day models alongside novel CQFE tests indicate that it is worthwhile to consider high frequency intra-day information to set daily VaR limits for different trading desks. Extending our quantile forecast encompassing and optimal combination framework to the entire predictive density is an interesting avenue for further research.

References

[1] Ahoniemi, K., & Lanne, M. (2011). Realized Volatility and Overnight Returns. SSRN Working Paper No. 1945687.

[2] Andersen, T.G., Bollerslev, T. & Huang, X. (2011). A Reduced Form Framework for Modeling Volatility of Speculative Prices based on Realized Variation Measures. *Journal of Econometrics, 160*, 176-189

[3] Angelidis, T., & Degiannakis, S. (2008). Volatility forecasting: Intra-day versus Inter-day models. *Journal of International Financial Markets, Institutions and Money, 18*, 449-465.

[4] Angrist, J., Chernozhukov, V., & Fernández-Val, I. (2006). Quantile Regression under Misspecification, with an Application to the US Wage Structure. *Econometrica, 74*, 539-563.

[5] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., & Sheppard, N. (2008). Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise. *Econometrica, 76*, 1481-1536.

[6] Basel 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, January 1996.

[7] Bauwens, L., Laurent, S. & Rombouts, J. (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics, 21*, 79-109.

[8] Black, F. (1976). Studies of Stock Market Volatility Changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

[9] Brandt, M.W., & Jones, C.S. (2006). Volatility Forecasting With Range-Based EGARCH Models. *Journal of Business and Economic Statistics, 24*, 470-486.

[10] Brooks, C., & Persand, G. (2003). Volatility Forecasting for Risk Management. *Journal of Forecasting, 22*, 1-22.

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

[12] Campbell, J., Hentschel, L. (1992). No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns. *Journal of Financial Economics, 31*, 281-318.

[13] Chen, C., Gerlach, R., Hwang, B., & McAleer, M. (2011). Forecasting Value-at-Risk Using Nonlinear Regression Quantiles. *International Journal of Forecasting*, forthcoming.

[14] Chernozhukov, V., & Umantsev, L. (2001). Conditional Value-at-Risk: Aspects of Modeling and Estimation. *Empirical Economics, 26*, 271-292.

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

[16] Clemen, R. T. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting, 5*, 559-581.

[17] 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.

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

[19] Corradi, V., & Swanson, N.R. (2002). A Consistent Test for Nonlinear Out-of-Sample Predictive Accuracy. *Journal of Econometrics 110*, 353-381.

[20] Corrado, C., & Truong, C. (2007). Forecasting Stock Index Volatility: Comparing Implied Volatility and the Intra-Day High-Low Price Range. *Journal of Financial Research, 30*, 201-215.

[21] Corsi, F. (2004). A Simple Long Memory Model of Realized Volatility, University of Southern Switzerland, mimeo.

[22] Diebold, F.X., & Mariano, R.S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, 13*, 253-263.

[23] Engle, R., & Manganelli, S. (2004). CAViaR Conditional Autoregressive Value-at-Risk by Regression Quantiles. *Journal of Business & Economic Statistics, 22*, 367-381.

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

[25] Flemming, M. J., (1997). The Round-the-Clock Market for U.S. Treasury Securities. *FRBNY Economic Policy Review*, July-97.

[26] Füss, R., Adams, Z., & Kaiser, D.G. (2010). The Predictive Power of Value-at-Risk Models in Commodity Futures Markets. *Journal of Asset Management, 11*, 261-285.

[27] Granger, C., Joyeux, R. (1980). An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis 1,* 15 29.

[28] Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004). The MIDAS Touch: Mixed Data Sampling Regression Models, Chapel Hill, mimeo.

[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] Glosten, L., Jagannathan, R. & Runkle, D. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance, 48*, 1779-1801.

[32] Gourieroux, C., & Jasiak, J. (2006). Dynamic Quantile Models. York University Working Paper 2006-4.

[33] Granger, C.W.J. (1989). Combining Forecasts – Twenty Years Later. *Journal of Forecasting, 8*, 167-173.

[34] Guidolin, M., & Timmermann, A. (2006). Term Structure of Risk Under Alternative Econometric Specifications. *Journal of Econometrics, 131*, 285-308.

[35] Haas, M., Mittnik, S., & Paolella, M. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics, 2*, 493-30.

[36] 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.

[37] Koenker, R. (2005). Quantile Regression. Cambridge University Press, New York.

[38] Koenker, R. & Xiao, Z. (2006). Quantile Autoregression. *Journal of the American Statistical Association, 101*, 980-990.

[39] Kuester, K., Mittnik, S., & Paolella, M.S. (2006) Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics, 4*, 1, 53-89.

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

[41] 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.

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

[43] McAleer, M., & da Veiga, B. (2008). Single-Index and Portfolio Models for Forecasting Value-at-Risk Thresholds. *Journal of Forecasting, 27*, 217-235.

[44] Parkinson, M. (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business, 53*, 1, 61-65.

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

[46] Powell, J.L. (1991). Estimation of Monotonic Regression Models Under Quantile Restrictions. In: *Nonparametric and Semiparametric Models in Econometrics* (eds. W. Barnett, J.L. Powell and G. Tauchen). Cambridge University Press.

[47] Rallis, G., J. Miffre, & Fuertes, A.-M. (2012). Strategic and Tactical Roles of Enhanced Commodity Indices. *Journal of Futures Markets*, forthcoming.

[48] Rodríguez, M.J., Ruiz, E. (2012). Revisiting Several Popular GARCH Models with Leverage Effect: Differences and Similarities, *Journal of Financial Econometrics*, forthcoming.

[49] Sajjad, R., Coakley, J., & Nankervis, J. (2008). Markov-Switching GARCH Modeling of Value-at-Risk. *Studies in Nonlinear Dynamics and Econometrics, 12*, Article 7.

[50] Santos, A., Nogales, F.J., & Ruiz, E. (2009). Comparing Univariate and Multivariate Models to Forecast Portfolio Value-at-Risk. Universidad Carlos III Working Paper 09-72.

[51] 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.

[52] Su, L., & Xiao, Z. (2008). Testing for Parameter Stability in Quantile Regression Models. *Statistics and Probability Letters, 78*, 2768-2775.

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

[54] Žikěs, F. (2009). Semiparametric Conditional Quantile Models for Financial Returns and Realized Volatility, Imperial College London, mimeo.

Table 1. Descriptive statistics for daily return and volatility data.

	Mean	**StDev**	**Skew**	**Kurt**	**Min**	**Max**	**ACF$_1$**	**ACF$_5$**	Q_{20} (p-value)
					S&P 500				
r_t	0.0116	1.3309	-0.1377	9.9481	-9.4819	10.771	-0.069	-0.034	83.115 (0.00)
r_t^2	1.7709	5.2965	10.777	172.34	0.0000	116.02	0.187	0.330	4596.7 (0.00)
$\log HL_t$	-1.3019	1.1672	0.2458	3.2227	-4.5460	3.0255	0.570	0.544	17876 (0.00)
$\log RV_t$	-0.1657	1.0538	0.4680	3.5301	-3.0380	4.1504	0.736	0.671	26060 (0.00)
					Russell				
r_t	0.0204	1.5864	-0.2868	7.3629	-12.578	8.7865	-0.035	-0.037	64.783 (0.00)
r_t^2	2.5162	6.3400	9.5534	156.69	0.0000	158.20	0.277	0.331	5606.3 (0.00)
$\log HL_t$	-1.0388	1.1386	0.1649	3.2985	-4.4097	3.3200	0.528	0.469	13049 (0.00)
$\log RV_t$	-0.1443	1.1102	0.2009	3.5498	-3.6690	4.1355	0.658	0.593	20497(0.00)
					Nasdaq				
r_t	0.0179	1.8212	-0.0070	7.1797	-10.207	13.232	-0.023	-0.024	70.738 (0.00)
r_t^2	3.3161	8.2444	8.0388	105.99	0.0000	175.09	0.207	0.245	3078.8 (0.00)
$\log HL_t$	-0.8347	1.1883	0.1949	2.8697	-4.5104	3.6691	0.642	0.594	21670 (0.00)
$\log RV_t$	0.3433	1.1827	0.3144	2.9255	-3.0617	4.6993	0.804	0.718	24225(0.00)
					USDX				
r_t	-0.0049	0.6590	-0.2775	4.3368	-3.0516	2.4193	0.063	0.008	13.706 (0.85)
r_t^2	0.4337	0.7937	4.7349	36.860	0.0000	9.3124	0.129	0.070	523.83 (0.00)
$\log HL_t$	-1.9503	0.8710	0.0396	3.1375	-5.0518	0.6938	0.454	0.277	5057.6 (0.00)
$\log RV_t$	-0.8613	0.6796	0.3127	3.5508	-3.1346	1.4571	0.641	0.544	12892 (0.00)
					10Y T-Notes				
r_t	0.0054	0.4238	-0.2189	5.9109	-2.3694	3.5163	-0.003	-0.005	33.311 (0.03)
r_t^2	0.1915	0.4376	11.059	245.64	0.0000	12.364	0.090	0.103	449.44 (0.00)
$\log HL_t$	-3.4765	1.0421	-0.0824	3.2074	-6.8621	0.7472	0.277	0.321	4752.1 (0.00)
$\log RV_t$	-2.1023	0.9397	0.3054	3.5066	-5.3819	2.2575	0.458	0.403	7462.5 (0.00)
					Gold				
r_t	0.0475	1.1381	0.2116	10.159	-7.3642	11.452	0.014	0.036	49.446 (0.00)
r_t^2	1.2972	3.9287	16.575	441.34	0.0000	131.14	0.204	0.121	593.32 (0.00)
$\log HL_t$	-1.6352	1.0831	0.1814	3.0887	-4.5856	2.7007	0.413	0.384	7788.4 (0.00)
$\log RV_t$	-0.0331	0.9287	0.3450	3.5456	-3.0847	4.0269	0.630	0.502	14275 (0.00)

r_t are daily log close-to-close returns, HL_t is the intraday high-low price range volatility, RV is the realized variance defined as the daily integrated sum of intraday (5min) squared returns, adjusted to include the overnight return using Hansen and Lunde's (2005) approach. ACF is autocorrelation function. Q_k is the Ljung-Box test. Returns are in percentages throughout the analysis. All samples end May 2011. Start date is November 1997 for equities (T=3404 days), April 1998 for US$ index (USDX; T=3361), September 1998 for T-Notes (T=3134) and December 1997 for gold (T=3395).

Table 2. Conditional volatility model estimates and diagnostics.

	S&P500	Russell	Nasdaq	USDX	10Y T-Notes	Gold
Panel A: Student GJR-GARCH model						
α_0	0.050 (0.012)	0.119 (0.020)	0.081 (0.017)	0.014 (0.005)	0.003 (0.002)	0.042 (0.040)
α_1	-0.007 (0.008)	-0.008 (0.011)	-0.002 (0.010)	0.021 (0.007)	0.041 (0.009)	0.055 (0.021)
β_1	0.937 (0.008)	0.888 (0.013)	0.927 (0.009)	0.957 (0.009)	0.958 (0.007)	0.956 (0.015)
γ_1	0.150 (0.013)	0.150 (0.018)	0.105 (0.013)	-0.001 (0.008)	-0.010 (0.010)	-0.052 (0.025)
δ	0.015 (0.004)	0.031 (0.007)	0.029 (0.007)	0.003 (0.001)	0.004 (0.002)	0.014 (0.018)
υ	10.689 (1.564)	50.471 (29.774)	17.487 (3.762)	13.292 (2.788)	10.012 (1.518)	5.660 (0.536)
$Q_{\mathbf{20}}$	25.410 [0.186]	28.161 [0.106]	21.050 (0.394)	11.325 [0.937]	26.621 (0.146)	31.256 (0.052)
$Q^2_{\mathbf{20}}$	24.643 [0.215]	28.264 [0.103]	17.280 (0.635)	26.491 [0.150]	12.319 (0.905)	33.477 (0.030)
Panel B: ARFIMAX model						
ω_0	-0.595 (0.022)	-0.595 (0.025)	-0.155 (0.026)	-1.309 (0.020)	-1.041 (0.030)	-0.351 (0.021)
ω_1	0.356 (0.021)	0.299 (0.020)	0.302 (0.017)	0.065 (0.039)	0.179 (0.073)	0.394 (0.022)
ω_2	-0.943 (0.033)	-0.790 (0.032)	-0.779 (0.028)	-0.091 (0.067)	-0.372 (0.128)	0.792 (0.037)
θ_1	-0.540 (0.013)	-0.550 (0.015)	-0.207 (0.024)	-0.233 (0.033)	-0.397 (0.045)	-0.399 (0.037)
d	0.496 (0.005)	0.493 (0.009)	0.488 (0.013)	0.449 (0.024)	0.424 (0.035)	0.457 (0.029)
υ	10.621 (1.531)	47.623 26.436	18.336 (4.128)	15.276 (3.923)	10.549 (1.672)	6.085 (0.591)
$Q_{\mathbf{20}}$	25.384 [0.187]	24.783 [0.210]	20.557 [0.424]	11.229 [0.940]	25.006 [0.201]	30.929 [0.056]
$Q^2_{\mathbf{20}}$	23.734 [0.254]	24.665 [0.215]	16.893 [0.660]	33.450 [0.030]	12.777 [0.887]	19.494 [0.490]
σ^2_u	0.513	0.626	0.539	0.253	0.635	0.523

Panels A and B correspond to equations (5) and (6), respectively. Standard errors are reported in parentheses. $Q_{\mathbf{20}}$ and $Q^2_{\mathbf{20}}$ are the Ljung-Box test statistics (p-values in square brackets) for the demeaned and standardized returns and their squares, respectively, according to an AR(1) model and the pertinent conditional volatility specification. All datasets end on 31 May 2011 and begin as noted in Table 1.

Table 3. VaR forecast encompassing regressions and tests.

	Indiv. VaRs		QR framework							GMM framework						
	$fr_1\%$	$fr_2\%$	$fr_c\%$	$\widehat{\lambda}^*_{0n}$	$\widehat{\lambda}^*_{1n}$	$\widehat{\lambda}^*_{2n}$	ENC_1	ENC_2	ENC_e	$fr_c\%$	$\widetilde{\lambda}^*_{0n}$	$\widetilde{\lambda}^*_{1n}$	$\widetilde{\lambda}^*_{2n}$	ENC_1	ENC_2	ENC_e
$\alpha = 0.05$																
S&P 500	7.37	7.80	5.20	−0.1030 (0.0361)	0.0896 (0.0530)	1.1046 (0.0433)	236.1**	46.62**	120.7**	5.06	−0.0014 (0.0024)	0.6139 (0.0335)	0.5917 (0.0335)	11.30*	11.60**	11.03*
Russell 2000	7.53	9.55	5.21	0.2936 (0.0197)	0.1760 (0.0331)	1.2385 (0.0349)	28.83**	56.32**	40.84**	5.07	−0.0017 (0.0033)	0.5917 (0.0470)	0.6000 (0.0486)	6.254	9.989*	7.908
Nasdaq	6.07	7.23	5.06	−0.1551 (0.0227)	0.2027 (0.0221)	0.8609 (0.0175)	139.3**	2.875	37.84**	5.06	0.0000 (0.0029)	0.5250 (0.0337)	0.5250 (0.0342)	0.625	0.877	0.336
USDX	6.01	3.49	5.03	0.2490 (0.0183)	0.0395 (0.0256)	1.1437 (0.0227)	11.23*	2.907	4.803	5.03	0.0000 (0.0034)	0.5083 (0.0527)	0.4875 (0.0443)	0.195	0.497	0.034
10Y T-Notes	5.24	7.24	5.08	0.2656 (0.0173)	−0.1350 (0.0198)	1.6739 (0.0210)	41.33**	8.580*	17.41**	4.93	0.0000 (0.0032)	0.5463 (0.0397)	0.5046 (0.0432)	0.317	0.716	0.244
Gold	5.23	3.54	4.95	−0.1094 (0.0221)	0.0679 (0.0248)	0.7809 (0.0240)	5.119	10.64*	4.910	4.95	0.0000 (0.0028)	0.5016 (0.0217)	0.4669 (0.0175)	30.00**	155.7**	13.01**
$\alpha = 0.01$																
S&P500	2.60	2.46	0.87	−0.4501 (0.0174)	0.9172 (0.0084)	0.0680 (0.0072)	28.98**	732.3**	150.8**	1.01	−0.0014 (0.0042)	0.5639 (0.0430)	0.5722 (0.0517)	1.118	13.66**	4.454
Russell 2000	2.46	3.18	1.01	0.0428 (0.0238)	1.5273 (0.0284)	−0.4129 (0.0276)	18.18**	30.42**	23.23**	0.87	−0.0014 (0.0048)	0.5639 (0.0690)	0.5722 (0.0685)	2.064	4.288	3.037
Nasdaq	1.88	1.73	1.01	−0.7008 (0.0289)	0.7368 (0.0183)	0.1157 (0.0178)	2.718	16.43**	3.989	0.90	0.0603 (0.1935)	1.3531 (0.1768)	0.0107 (0.1768)	0.384	0.101	0.218
USDX	1.26	0.42	0.98	0.6451 (0.0313)	0.3224 (0.0271)	1.0370 (0.0170)	25.44**	5.381	7.556*	0.98	0.0000 (0.0056)	0.5000 (0.0505)	0.5000 (0.0468)	0.330	0.330	0.000
10Y T-Notes	1.23	0.92	0.92	0.2492 (0.0287)	−0.5232 (0.0205)	1.9607 (0.0168)	533.4**	43.04**	208.6**	0.92	0.0000 (0.0050)	0.5125 (0.0583)	0.5125 (0.0684)	0.070	0.090	0.014
Gold	1.13	0.71	0.99	−0.7814 (0.0263)	0.4749 (0.0173)	0.2235 (0.0193)	55.86**	205.3**	83.21**	0.99	0.0000 (0.0130)	0.5031 (0.0696)	0.4537 (0.0741)	0.063	0.890	0.139**

Subscripts $i = 1, 2, c$ denote the GJR-GARCH model, ARFIMAX model and optimal combination, respectively. fr_i is the out-of-sample failure rate at an α coverage probability. $\widehat{\lambda}_n$ and $\widetilde{\lambda}_n$ are the QR and GMM weight estimates. ENC_i with $i = 1, 2, e$ denote the three encompassing Wald tests for $H_{10} : (0, 1, 0)$, $H_{20} : (0, 0, 1)$ and $H_{e0} : (0, 0.5, 0.5)$, respectively. * and ** denote rejection at the 5% and 1% levels, respectively. Standard errors are reported in parentheses.

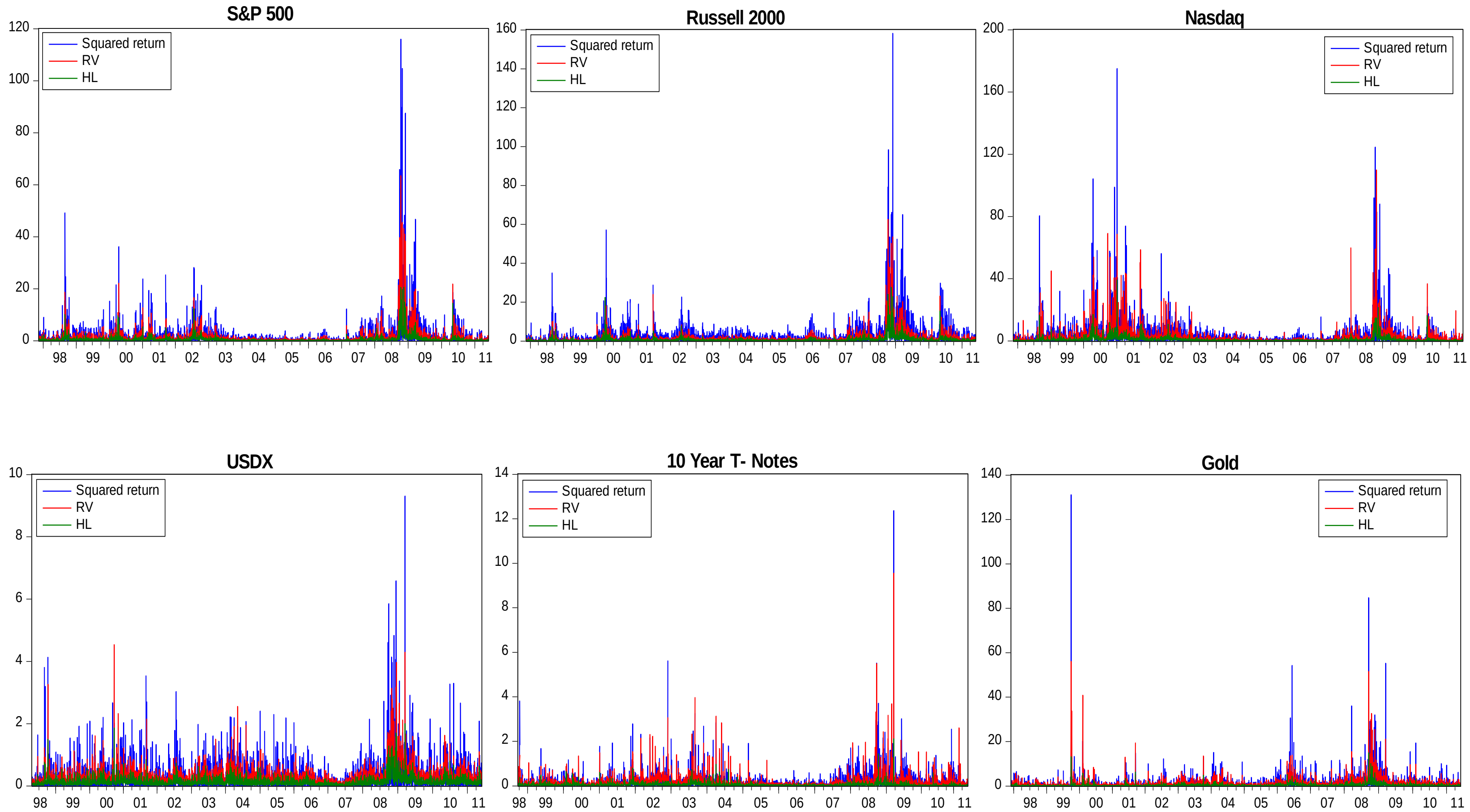

Figure 1. Daily volatility measures. The graphs depict the dynamics of the daily squared logarithmic close-to-close return, realized variance (RV) and intraday high-low range volatility (HL) over the entire 14-year sample period ending 31 May 2011.

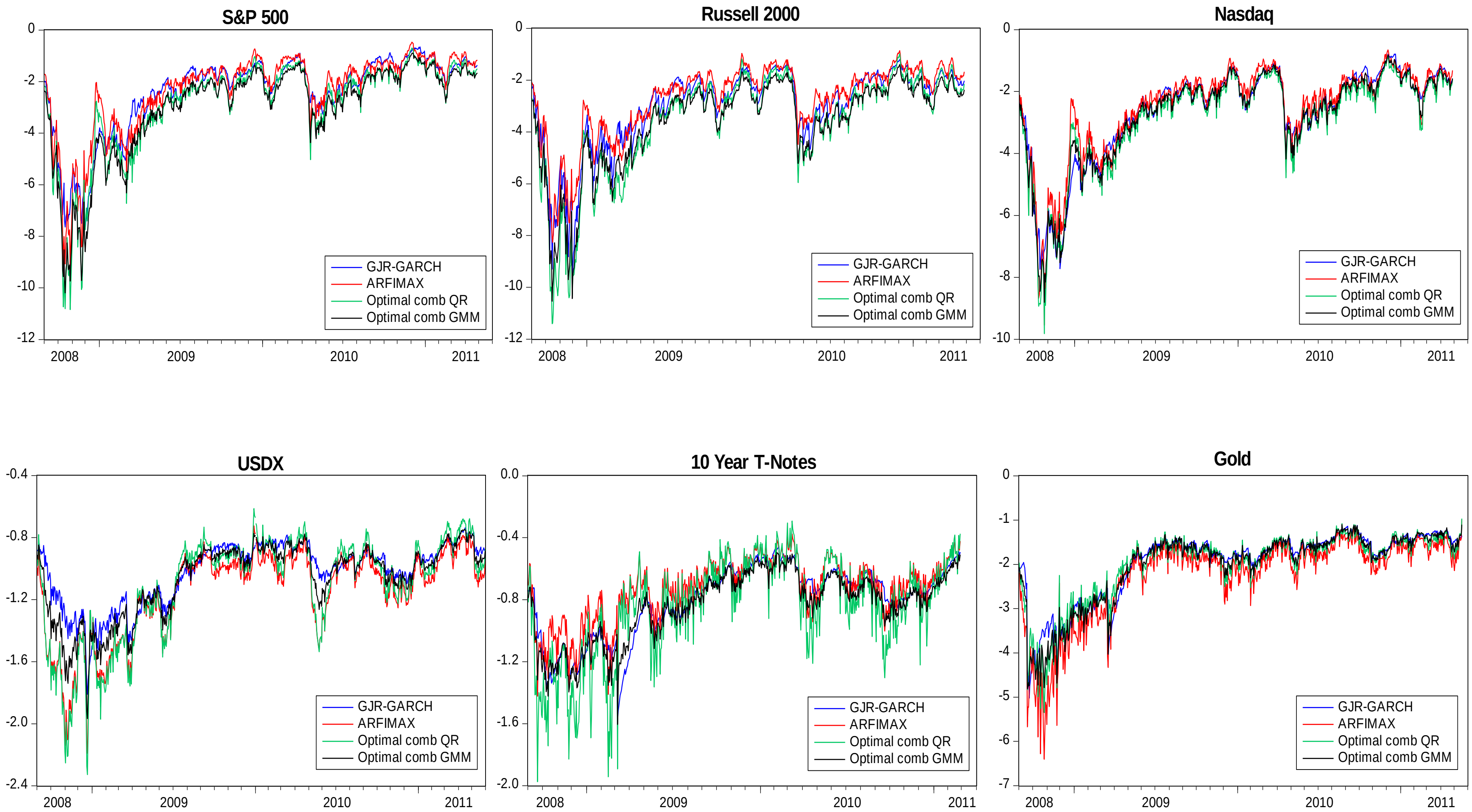

Figure 2. Out-of-sample 5% VaR forecasts. The graphs depict the dynamics of the individual inter-day (GJR-GARCH) and intra-day (ARFIMAX) VaR forecasts and the optimally combined forecasts using the QR and GMM frameworks. The out-of-sample period starts 1 September 2008 and ends 31 May 2011.