Dynamic multiple standardised quantiles for asset returns

Gelly Mitrodima
Department of Statistics, London School of Economics
Email: E.Mitrodima@lse.ac.uk

Jaideep Oberoi
School of Finance & Management, SOAS University of London
Email: jaideep.oberoi@soas.ac.uk

June 24, 2025

Abstract

We explore the time variation in the shape of the conditional return distribution using a model of multiple quantiles. We propose a joint model of scale (proxied by the interquartile range) and other quantiles standardised by the scale. The model allows us to estimate the scale and shape of the distribution, which are both time-varying, in one step. We find that once we effectively capture the dynamics of the scale, the time variation in the shape allows a simpler interpretation.

Keywords: Dynamic multiple quantile model; Return decomposition; Robust methods; CAViaR-type models; Elicitability

1 Introduction

The conditional distribution of asset returns has been widely studied and modelled using a range of techniques. This has led to a large amount of evidence that returns of most assets display not only time-dependence in variance, but also in higher moments. In this paper, we model time variation in the conditional return distribution using a semi-parametric approach based on joint models of multiple quantiles.

Using a joint model of conditional quantiles, we separately specify dynamics for a common time-varying scale and the quantiles standardised by this time-varying scale. The joint model is estimated in a single step. Our proposed approach naturally extends GARCH and stochastic volatility models and allows for a richer analysis of the conditional distribution via quantiles. An important advantage of our approach is that the scale acts as a common factor that builds dependence across quantiles in a parsimonious manner. In turn, this makes it easier to estimate the model with the quantiles in their standardised form.

There is a long history of research on the shape of the return distribution, which is of central importance for asset pricing and risk management. Modellers have developed many techniques to capture the fat-tailed, and often skewed nature of return distributions (Bates 2006). By introducing the conditionally heteroskedastic class of time series models Engle (1982) separated the shape of the underlying return surprise from its time-varying and serially dependent scale. An entire literature has followed, introducing more complicated dynamics for the scale and different distributions for the underlying shocks. GARCH (1,1) with standard normal shocks (Bollerslev 1986) captures a great deal of dependence in the conditional variance but requires modification to better fit the tails of many return distributions (French et al. 1987, Bollerslev 1987, Hansen & Lunde 2005). Firstly, a leverage effect (Black 1976) can be incorporated to allow asymmetric responses of volatility to positive and negative returns (Schwert 1989, Nelson 1991, Glosten et al. 1993, Engle & Ng 1993); or, secondly, alternative distributions can be applied for the underlying shocks (Bollerslev 1987, McNeil & Frey 2000). Similarly, models with fat tails and long memory in volatility (Ding et al. 1993) can be obtained through fractionally integrated GARCH models (Baillie et al. 1996) or component models (Ding & Granger 1996, Engle & Lee 1999). Semiparametric and nonparametric models for the conditional return distribution

have also been explored. For example, Engle & Gonzalez-Rivera (1991) and Kalli et al. (2014) consider using GARCH models, while Jensen & Maheu (2010) and Delatola & Griffin (2011) use stochastic volatility models.

Outside the discrete time GARCH framework, continuous time stochastic volatility models are also able to generate fat-tailed return distributions with leverage effects. These models can also incorporate features such as jumps in returns and volatility to better explain the dynamics of returns and to improve their applicability to problems such as option pricing; see *e.g.* Chernov et al. (2003) and Bates (2006) for an overview.

Despite the above improvements to GARCH or stochastic volatility models, fitting the tails of the return distribution remains a difficult task. For instance, Gallant et al. (1991) show that the standardised error term has a time-varying conditional distribution. As it is difficult to model the entire distribution, modelling efforts beyond the variance have focused on higher moments such as skewness and kurtosis. Hansen (1994) proposed an explicit model of time-varying higher moments, while Harvey & Siddique (1999) applied an autoregressive conditional skewness model to returns. Backus et al. (1997) made use of a Gram Charlier expansion around the normal to back out time-varying skewness and kurtosis from the option smirk, while Christoffersen et al. (2013) utilised the inverse Gaussian distribution to obtain time-varying skewness in an option pricing model.

Rather than relying on estimators of higher moments based on averages of past realised moments, Kim & White (2004) proposed the use of quantile based estimators. They showed through simulations how empirical moment based estimators are significantly more sensitive to outliers. Effectively, they question existing beliefs about the extent and the time variation of skewness and kurtosis.

Following Kim & White (2004), White et al. (2010) proposed joint models of conditional quantiles to obtain robust estimates of conditional skewness and kurtosis.[1] This approach develops CAViaR models (Engle & Manganelli 2004) which work well in estimating particular probability levels of the conditional distribution such as the Value at Risk (VaR),

[1]The approach of estimating moments from quantiles has also been applied by others, including Taylor (2005) who estimates volatility from an interquartile range, and Xiao & Koenker (2009) who estimate GARCH models in a two-step procedure. Ghysels et al. (2011) examine the higher moments to form portfolios, while Coroneo & Veredas (2012) to study their interdependence over time.

(see *e.g.* Engle & Manganelli 2001, Kuester et al. 2006, Gerlach et al. 2011). They also offer modelling flexibility as the conditional quantiles can respond to observations without assuming a parametric return distribution. However, multiple quantiles cannot be satisfactory estimated using separate application of single quantile models at different probability levels since these estimates do not guarantee monotonicity of the quantile function, which is often called the crossing problem (see *e.g.* Chernozhukov & Galichon 2008, Chernozhukov et al. 2010, Gourieroux & Jasiak 2008, for explanations and possible solutions). Thus, it is useful to jointly model quantiles and avoid quantiles crossing in the estimation procedure.

The specification of the multi-quantile model (MQ CAViaR) of White et al. (2010) takes the form of a vector autoregression, which they find awkward to estimate. We propose an alternative model for multiple quantiles which builds on the intuition that most of the variation in the shape of the conditional return distribution should be captured by a time-varying volatility process. Our approach builds a dependence between the quantiles via the common scale factor (rather than the linear functions of a vector autoregression). This model is more parsimonious than the MQ CAViaR model and achieves a robust scale and shape decomposition of the conditional return distribution. This allows simpler comparison to commonly used models for financial time series which directly model the conditional variance.

Previous papers where the scale information is separately obtained and then used to enhance quantile models, include Jeon & Taylor (2013) and Chen & Gerlach (2014). Jeon & Taylor (2013) model VaR by incorporating option implied volatility in a CAViaR model. Chen & Gerlach (2014) use intra-day data to capture volatility and tail risk for estimating expected shortfall (ES) in an auto-regressive expectile model (Taylor 2008). Our approach differs by only using the information in the daily returns to model both the scale and the shape dynamics in a single estimation step.

This paper makes four main contributions. Firstly, we present some novel dynamic multiple standardised quantile (DMSQ) models. The models we propose are semiparametric in that they impose parametric structures for the dynamics of the quantiles and the interquantile range (IQR), but avoid making any strong distributional assumptions for the conditional distribution of returns (the usual regularity conditions for estimation and

inference are assumed though). Under this framework estimation and prediction is fast and simple to implement.

The second contribution of the paper is the asymptotic theory for a general class of semiparametric DMSQ models. This theory is an extension of results for VaR presented in Weiss (1991) and Engle & Manganelli (2004), and draws on identification results in Fissler & Ziegel (2016) and results for M-estimators in Newey & McFadden (1994). We discuss the conditions under which the estimated model parameters are consistent and asymptotically normal, and we present a consistent estimator of the asymptotic covariance matrix.

Although elicitability has appeared in the literature for single quantiles and ES there has been no work on multiple quantiles to the best of our knowledge. Drawing on Fissler & Ziegel (2016), our third contribution is to address the problem of elicitability directly. This introduces new directions in multiple quantile modeling and prediction.

Our forth and final contribution is an application of our new models and estimation methods on four equity indices over a long period ranging from January 2001 to November 2024. We compare these new models with existing methods from the literature.

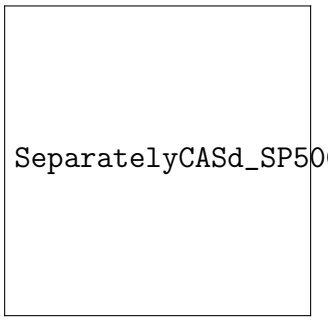
The paper is organised as follows. The following section provides a motivation example, presents the construction of our model and discusses alternative forms for the time evolution of the scale and conditional quantiles. Section 3 describes the estimation of the model. Applications of the models to stock index data are discussed in Section 4. Section 5 concludes.

2 DMSQ

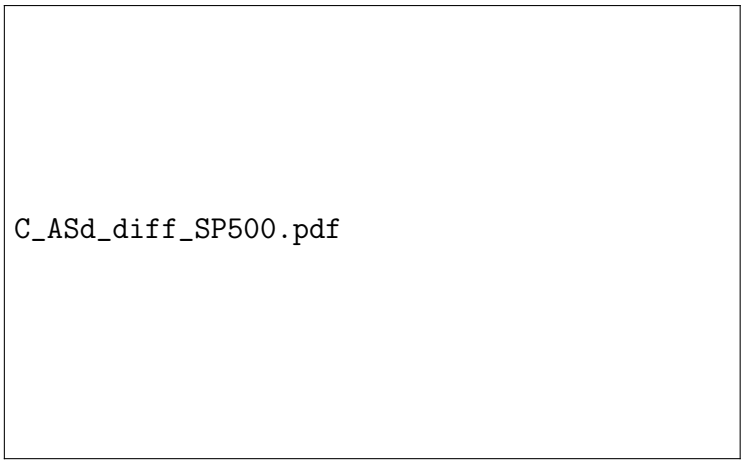
2.1 An alternative approach: Scale and shape decomposition

A reasonable goal for a multi-quantile time series model is to gain a better understanding of the evolution of the conditional quantiles (and, more generally, the conditional distribution), and to improve the prediction of individual conditional quantiles using information from other conditional quantiles. To guide us about the form of the relationship between conditional quantiles without making any strong parametric assumptions about the conditional return distribution, we first examine several separately estimated conditional quan-

tiles. Conditional quantiles at probability levels $1\%, 5\%, 25\%, 50\%, 75\%, 95\%$, and 99% were separately estimated using a single component CAViaR model (see *e.g.* Mitrodima & Oberoi 2023). The data consist of S&P500 index (SPX) returns obtained from CRSP for the period January $01, 2010$ to December $31, 2021$. By plotting the separately estimated conditional quantiles together in Figure 1a, we observe the time-variation in the scale of the distribution. The conditional quantiles increase and decrease together (as it might be expected), but they also appear to do so proportionately to each other.



(a) Separately estimated conditional quantiles



(b) Differences between adjacent conditional quantiles

Figure 1: SPX from January $01, 2010$ to December $31, 2021$.

To make this point clearer, we plot the differences between adjacent conditional quantiles in Figure 1b. The differences between adjacent conditional quantiles tend to follow very similar patterns over time. This suggests that a substantial part of the joint dynamics of the conditional quantiles could be simply captured in a multi-quantile time series model by introducing a common scale process (a common scale parameter is often assumed in models of returns data such as in the GARCH literature).

2.2 DMSQ models

We assume that $y_t = s_t z_t$, where s_t is the time-varying scale and z_t is a standardised return which follows a time-varying distribution (many standard models for returns assume

a time-invariant distribution for z_t which can be more constrictive). We define s_t to be a robust scale measure defined in terms of the conditional quantiles of y_t (*e.g.* the conditional interquartile range) and assume that a further $K-2$ conditional quantiles of the standardised returns follow independent CAViaR-type models. Cont et al. (2010) propose the interquartile as a risk measure that balances both robustness and coherence. It serves as an interpolation between the highly robust, but non-coherent, VaR and the entirely non-robust, but coherent, ES.

Suppose that $s_t = q_{\theta_1,t} - q_{\theta_2,t}$, where $\theta_1 = 1-\alpha$ and $\theta_2 = \alpha$ for $\alpha < 1/2$ and $\theta_3, \ldots, \theta_K$ are $K-2$ further probability levels. We choose to use the conditional interquartile range, which arises for $\alpha = 0.25$, as our robust scale measure. This choice seems reasonable as we hypothesise that our data are characterised by less time-variation in the body of the distribution compared to the tails of the distribution. Clearly the model can be expanded to include any other interquantile range. In our application, we have found that the results do not significantly change with different values of α.

The model can be expressed in terms of the conditional quantiles of the return distribution at probability levels $\theta_1, \theta_2, \ldots, \theta_K$ using the transformations $q_{\theta_1,t} = q_{\theta_2,t} + s_t$ and $q_{\theta_k,t} = s_t q^{(z)}_{\theta_k,t}$ for $k = 2, \ldots, K$ leading to the following model for $q_{\theta_1,t}, q_{\theta_2,t}, \ldots, q_{\theta_K,t}$,

$$s_t = u^{(s)} + \sum_{i=1}^{M} \beta_i^{(s)} s_{t-i} + \sum_{j=1}^{L} \gamma_j^{(s)} \, l_j \left(y_1, \ldots, y_{t-1}\right) \tag{1}$$

$$q_{\theta_1,t} = q_{\theta_2,t} + s_t \tag{2}$$

$$q_{\theta_k,t} = s_t \left(u^{(z)}_{\theta_k} + \sum_{i=1}^{M} \beta^{(z)}_{\theta_k,i} \frac{q_{\theta_k,t-i}}{s_{t-i}} + \sum_{j=1}^{L} \gamma^{(z)}_{\theta_k,j} \, l_j \left(\frac{y_1}{s_1}, \ldots, \frac{y_{t-1}}{s_{t-1}} \right) \right), \quad k = 2, \ldots, K \tag{3}$$

where M and L are the orders of the model, $u^{(s)}$, $\beta_i^{(s)}$, and $\gamma_i^{(s)}$ are scalars, $u^{(z)}_{\theta_k}$ and $\gamma^{(z)}_{\theta_k,j}$ are $(K-1)$-dimensional vector, $\beta^{(z)}_{\theta_k,i}$ is a $((K-1)\times(K-1))$-dimensional matrix with the non-diagonal equal to zero and $l_j\left(y_1, \ldots, y_{t-1}\right)$ represents a function of the history of the return series.

Choosing $M = 1$, $L = 1$, $\theta_1 = 0.75$, $\theta_2 = 0.25$ and $l_1(y_1, \ldots, y_{t-1}) = |y_{t-1}|$ leads to a generalisation of the symmetric absolute value (SAV) CAViaR model (Engle & Manganelli 2004), which we define to be the *DMSQ-SAV* model and has the form

- DMSQ-SAV

$$s_t = u^{(s)} + \beta^{(s)} s_{t-1} + \gamma^{(s)} |y_{t-1}| \tag{4}$$

$$q_{\theta_1,t} = q_{\theta_2,t} + s_t \tag{5}$$

$$q_{\theta_k,t} = s_t \left(u^{(z)}_{\theta_k} + \beta^{(z)}_{\theta_k} \frac{q_{\theta_k,t-1}}{s_{t-1}} + \gamma^{(z)}_{\theta_k} \frac{|y_{t-1}|}{s_{t-1}} \right), \quad k = 3, \ldots, K \tag{6}$$

where we drop 1 from $\beta^{(z)}_{\theta_k,1}$ and $\gamma^{(z)}_{\theta_k,1}$ for ease of reading.

We consider one more specification. We choose $l_1(y_1, \ldots, y_{t-1}) = y^+_{t-1}$ and $l_1(y_1, \ldots, y_{t-1}) = y^-_{t-1}$ where the notation employed is: $y^+ = \max(y, 0)$ and $y^- = -\min(y, 0)$. This leads to a generalisation of the asymmetric slope (AS) CAViaR model (Engle & Manganelli 2004), which we define to be the *DMSQ-AS* model which has the form

- DMSQ-AS

$$s_t = u^{(s)} + \beta^{(s)} s_{t-1} + \gamma^{(s)} y^+_{t-1} + \delta^{(s)} y^-_{t-1} \tag{7}$$

$$q_{\theta_1,t} = q_{\theta_2,t} + s_t \tag{8}$$

$$q_{\theta_k,t} = s_t \left(u^{(z)}_{\theta_k} + \beta^{(z)}_{\theta_k} \frac{q_{\theta_k,t-1}}{s_{t-1}} + \gamma^{(z)}_{\theta_k} \frac{y^+_{t-1}}{s_{t-1}} + \delta^{(z)}_{\theta_k} \frac{y^-_{t-1}}{s_{t-1}} \right), \quad k = 2, \ldots, K \tag{9}$$

The DMSQ models have some interesting features. They directly model the extent of time variation in the shape of the conditional return distribution after controlling for the scale. They allow us to robustly estimate the scale (without separately estimating a volatility model) and the conditional return distribution in a single estimation step. The quantile crossing problem can be avoided during estimation, without assuming an explicit model for the conditional return distribution.

2.3 Consistent scoring rule for dynamic multiple quantiles standardised by IQR

We show that $T = (q_{\theta_1}, \cdots, q_{\theta_6}, s)$ is elicitable and identifiable. To do so, we follow similar steps to Fissler & Ziegel (2021) who show that $T' = (q_{\theta_1}, q_{\theta_2}, \text{RVaR}_{\theta_1,\theta_2})$ ($\text{RVaR}_{\theta_1,\theta_2}$ is the range VaR) is elicitable under weak regularity conditions. The tick

loss function $L_{\theta_r} = (\mathbf{1}_{\{y \leq q_r\}} - \theta_r)(q_r - y)$ is $\mathcal{F}$-consistent for q_r when $\int_{-\infty}^{0} |y| dF(y) < \infty$, for all $F \in \mathcal{F}_{\theta_r}$, and strictly $\mathcal{F}$-consistent if furthermore $\mathcal{F} \subset \mathcal{F}_{\theta_r}$.

Proposition 1 *For $0 < \theta_3 < \theta_4 < \theta_2 < \theta_1 < \theta_5 < \theta_6 < 1$, the map $V : \mathbb{R}^7 \times \mathbb{R} \to \mathbb{R}^7$ defined by*

$$V(q_{\theta_1}, \cdots, q_{\theta_6}, s, y) = \begin{pmatrix} \mathbf{1}\{y \leq q_{\theta_1}\} - \theta_1 \\ \vdots \\ \mathbf{1}\{y \leq q_{\theta_6}\} - \theta_6 \\ s + (\mathbf{1}\{y \leq q_{\theta_1}\} - \theta_1) - (\mathbf{1}\{y \leq q_{\theta_2}\} - \theta_2) \end{pmatrix}$$

is an $\mathcal{F}(\theta_1) \cap \cdots \cap \mathcal{F}(\theta_6)$-identification function for $(q_{\theta_1}, \cdots, q_{\theta_6}, s)$, which is strict on $\mathcal{F}_{\theta_1} \cap \mathcal{F}(\theta_1) \cap \cdots \cap \mathcal{F}_{\theta_6} \cap \mathcal{F}(\theta_6)$.

Proof 1 *Note that*

$$\bar{V}(q_{\theta_1}(F), \cdots, q_{\theta_6}(F), s, F) = s - s(F)$$

which concludes the proof.

The following result establishes a (strictly) consistent loss functions $L_0 : \mathbb{R}^7 \times \mathbb{R} \to \mathbb{R}$ for $(q_{\theta_1}, \cdots, q_{\theta_6}, s)$. The values are bounded in $[c_{\min}, c_{\max}]^7$, $-\infty \leq c_{\min} < c_{\max} \leq \infty$.

We expand the consistent loss functions results in Fissler & Ziegel (2016) (Theorem 5.2) and Fissler & Ziegel (2021) for $K + 1 = 7$ and $T = (q_{\theta_1,t}, \cdots, q_{\theta_6,t}, s_t)$.

Theorem 1 *The maximal sensible action domain is $\boldsymbol{A}_0 := \{x \in R^7 : q_{\theta_1} \geq s \geq q_{\theta_2} \geq q_{\theta_3} \leq q_{\theta_4} \leq q_{\theta_5} \leq q_{\theta_6}\}$ as we always have $q_{0.75} \geq s \geq q_{0.25} \geq q_{0.01} \leq q_{0.05} \leq q_{0.95} \leq q_{0.99}$.*

$$L_0(y, Q, s; \theta, G_r, G) = \sum_{r=1}^{(K+1)-1} (\mathbf{1}_{\{y \leq q_r\}} - \theta_r) G_r(q_r) - \mathbf{1}_{\{y \leq q_r\}} G_r(y) \tag{10}$$

$$+ H'(s)\Big(s + \sum_{m=1}^{(K+1)-1} \frac{p_m}{\theta_m}(\mathbf{1}_{\{y \leq q_m\}}(q_m - y) - \theta_m q_m\Big) \tag{11}$$

$$- H(s) + \alpha(y) \tag{12}$$

$$= \sum_{r=1}^{(K+1)-1} (\mathbf{1}_{\{y \leq q_r\}} - \theta_r) G_r(q_r) - \mathbf{1}_{\{y \leq q_r\}} G_r(y) \tag{13}$$

$$+ H'(s)\left(s + L_{\theta_2}(q_2, y) - L_{\theta_1}(q_1, y)\right) \tag{14}$$

$$- H(s) + a(y) \tag{15}$$

where $Q = [q_{\theta_1}, \cdots, q_{\theta_6}]$, $G_r(q_r)$ *and* $G_r(y)$ *cannot identically vanish for strict consistency of* L_0 *in our case, whereas the corresponding functions in* L_0 *when jointly estimating VaR and ES may be set to zero (see for an example Mitrodima & Oberoi (2023)).* $G_r, H, \theta : \mathbb{R} \to \mathbb{R}, \theta$ *is* $\mathcal{F}-$ *integrable and* $\mathbf{1}_{(-\infty, q_r)} G_r$ *is* $\mathcal{F}-$ *integrable for all* $q_r \in \mathbb{R}$*, is* $\mathcal{F}-$ *consistent for* T*, if* G_r *is increasing and* H' *is increasing and convex. If* H' *is strictly increasing and strictly convex, then* L_0 *is strictly* $\mathcal{F}-$*consistent for* T*. Finally, for all* $s \in [c_{min}, c_{max}]$*, the functions*

$$G_{1,s} : [c_{min}, c_{max}] \to \mathbb{R}, \quad q_{\theta_1} \mapsto g_1(q_{\theta_1}) - q_{\theta_1} H'(s),$$

$$G_{2,s} : [c_{min}, c_{max}] \to \mathbb{R}, \quad q_{\theta_2} \mapsto g_2(\theta_2) + q_{\theta_2} H'(s),$$

are increasing.

Proof 2

$$[c_{\min}, c_{\max}] \times \mathbb{R} \ni (q'_{\theta_1}, y) \mapsto L_0(q'_{\theta_1}, \cdots, q_{\theta_6}, s, y)$$

is F*-consistent for* q_{θ_1}*, and it is strictly* F_{θ_1}*-consistent if* $G_{1,s}$ *is strictly increasing. Similarly,*

$$[c_{\min}, c_{\max}] \times \mathbb{R} \ni (q'_{\theta_2}, y) \mapsto L_0(t_1, q'_{\theta_2}, \cdots, q_{\theta_6} s, y).$$

Hence,

$$\begin{aligned}0 \leq & \bar{L}_0(q_{\theta_1}, \cdots, q_{\theta_6}, s, F) - \bar{L}_0(t_1, q_{\theta_2}, \cdots, q_{\theta_6}, s, F) \\ & + \bar{L}_0(t_1, , q_{\theta_2}, \cdots, q_{\theta_6}, s, F) - \bar{L}_0(t_1, t_2, q_{\theta_3}, \cdots, q_{\theta_6}, s, F) \\ & = \bar{L}_0(q_{\theta_1}, \cdots, q_{\theta_6}, s, F) - \bar{L}_0(t_1, t_2, q_{\theta_3}, \cdots, q_{\theta_6}, s, F),\end{aligned}$$

with a strict inequality under the conditions for strict consistency and if $(q_{\theta_1}, q_{\theta_2}) \neq (t_1, t_2)$. *Finally,*

$$\bar{L}_0(t_1, \cdots, t_6, s, F) - \bar{L}_0(t_1, \cdots, t_7, F) = H'(s)(s - t_7) - H(s) + H(t_7) \geq 0$$

If H *is strictly convex and if* $s \neq t_7$, *the inequality is strict.*

Let $(Q, s) \in A$, $F \in \mathcal{F}$, and $(t_1, \cdots, t_7) := T(F)$. Then, since $G_{1,s}$ is increasing, the map

Proposition 2 *(Normalised distance loss function) We choose* $G_r(q_r) = q_r$ *and* $H(s) = \big|s - c|y|\big|$, *where* $c = 1.349\sqrt{(\pi/2)}$. *Thus, the normalised distance loss function,* L_{ND}, *is a consistent loss function for* (Q, s).

$$\begin{aligned}L_{ND}(y, Q, s; \theta) = & \sum_{r=1}^{6} (\mathbf{1}_{\{y \leq q_r\}} - \theta_r)(q_r - y) && (16) \\ & + \frac{s - c|y|}{\big|s - c|y|\big|}(s + L_{\theta_2}(q_2, y) - L_{\theta_1}(q_1, y)) && (17) \\ & - \big|s - c|y|\big| && (18)\end{aligned}$$

Proof 3 *The result follows directly from Theorem 5.1 of Fissler & Ziegel (2021) and mixture representation of scoring functions by letting* $0 < \theta_3 < \theta_4 < \theta_2 < \theta_1 < \theta_5 < \theta_6 < 1$ *and any loss function* $L_0 : [cmin, cmax]^7 \times \mathbb{R} \to \mathbb{R}$ *of the form of Equations* (10)-(12)

We choose to work with the absolute distances of the scale s from the interquantile of the (relevant) standard normal distribution.

Positive homogeneity of the loss function is essential in regression applications and forecasting exercises. This is because when a loss function is characterised by positive homogeneity the ordering of backtesting forecasts assessed by this is not affected when monetary units change making it convenient for financial applications. Here, we require zero homogeneous loss functions as this has been found to work well in forecasting applications, see Patton (2006) and Patton & Sheppard (2009). We also require that score differences are homogeneous as we are interested in differences of expected scores in our backtesting exercise.

Nolde & Ziegel (2017) show that there does not generally exist a loss function that generates loss differences that are homogeneous of degree zero, however, Patton show that zero-degree homogeneity may be attained by exploiting the fact that, for θ ranging from around 0.01 to 0.10, we may assume the ES<0 and VaR<0 then, up to irrelevant location and scale factors, there is only one loss function that satisfies the requirement.

Here, we assume that IQR is strictly positive, and so the loss differences generated by the loss function are homogeneous of degree zero.

Proposition 3 *Under the assumption that the conditional quantiles of interest can be negative and positive while the IQR is strictly positive, the loss differences generated by the proposed loss function are homogeneous of degree zero iff* $G_r(q_r) = q_r$ *and* $H(s) = \big|s - 1.349\sqrt{(\pi/2)}|y|\big|$.

Proof 4 *A loss function* $L_{ND}(x, y)$ *is positively homogeneous if, for any non negative constant* φ, $L_{ND}(\varphi x, \varphi y) = \varphi L_{ND}(x, y)$. *Denote the resulting loss function as* $L^*_{ND}(y, Q, s; \theta, \varphi_1, \cdots, \varphi_7)$, *and notice that:*

$$L^*_{ND}(y, Q, s; \theta, \varphi_1, \cdots, \varphi_7) = \sum_{r=1}^{6} \varphi_r(\mathbf{1}\{y \le q_r\} - \theta_r)(\mathbb{1}\{q_r \ge 0\} - \mathbb{1}\{y \ge 0\}) \quad (19)$$

$$+ \varphi_7 \frac{s - c|y|}{\big|s - c|y|\big|}(s + L_{\theta_2}(q_2, y) - L_{\theta_1}(q_1, y)) \quad (20)$$

$$- \varphi_7\big|s - c|y|\big| \quad (21)$$

– *The indicator function* $\mathbf{1}\{\varphi_r y \leq \varphi_r q_r\}$ *remains the same as* $\mathbf{1}\{y \leq q_r\}$ *because the inequality is invariant under scaling by a positive* φ_r*.*

– $G_r(q_r) = q_r$ *is homogeneous of degree* 1 *(VaR), then:*

$$G_r(\varphi_r q_r) = \varphi_r G_r(q_r) \quad \textit{and} \quad G_r(\varphi_r y) = \varphi_r G_r(y).$$

– $H(s)$ *and* $H'(s)$ *are homogeneous of degree* 1*, then:*

$$H(\varphi_7 s) = \varphi_7 H(s), \quad H'(\varphi_7 s) = H'(s).$$

$$H(\varphi_7 s) = \varphi_7 \big| s - 1.349\sqrt{(\pi/2)}|y| \big|, \quad H'(\varphi_7 s) = \frac{s - c|y|}{\big|s - c|y|\big|}$$

– $(s + L_{\theta_2}(q_2, y) - L_{\theta_1}(q_1, y))$ *also scales by* φ_7*, but since* $H'(s)$ *multiplies the whole term, the homogeneity of degree* 1 *for* $H(s)$ *ensures the overall scaling factor cancels out.*

3 Estimation of DMSQ models

The parameters of the DMSQ models are estimated by minimising the normalised distance loss function, L_{ND}. Let a sample of observations $y_1, \ldots, y_n$ generated by the model

$$y_t = Q_t(\boldsymbol{\beta}^0) s_t(\boldsymbol{\beta}^0), t = 1, \ldots, n \tag{22}$$

The vector of unknown parameters to be estimated is $\boldsymbol{\beta}^0 \in B \subset \mathbb{R}^p$. Let

$$L_{ND,n}(\boldsymbol{\beta}) = n^{-1} \sum_{t=1}^{n} L_{ND}(y_t, Q_t(\boldsymbol{\beta}), s_t(\boldsymbol{\beta}); \theta)$$

then the unknown parameters are estimated by minimising

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\beta \in B} L_{ND,n}(\boldsymbol{\beta}) \tag{23}$$

at fixed probability levels $\theta = [75\%, 25\%, 1\%, 5\%, 95\%, 99\%]$. We are interested in estimating and forecasting the conditional, on information set $\mathcal{F}_{t-1}$, θ-quantiles and

IQR of y_t, where the conditional distribution function $F_t(\cdot|\mathcal{F}_{t-1})$ and the corresponding density function $f_t(\cdot|\mathcal{F}_{t-1})$. Consider the model in Eq.(22) for the conditional θ-quantiles and IQR of y_t at θ probability levels, that is $q_{t,\theta}(y_t|\mathcal{F}_{t-1})$ and $s_t(y_t|\mathcal{F}_{t-1})$.

$$y_t = Q(y_{t-1}, \boldsymbol{x}_{t-1}, \cdots, y_1, \boldsymbol{x}_1; \boldsymbol{\beta}^0) s(y_{t-1}, \boldsymbol{x}_{t-1}, \cdots, y_1, \boldsymbol{x}_1; \boldsymbol{\beta}^0) \tag{24}$$

$$\equiv Q_t(\boldsymbol{\beta}^0) s_t(\boldsymbol{\beta}^0),\ t = 1, \cdots, n \tag{25}$$

where $\boldsymbol{x}_t$ is a vector of exogenous or predetermined variables, $\boldsymbol{\beta}^0 \in \mathbb{R}^p$ is the vector of true but unknown parameters that we are interested in estimating. Also, $\mathcal{F}_{t-1} = \sigma(y_{t-1}, \boldsymbol{x}_{t-1}, \cdots, y_1, \boldsymbol{x}_1, Q_1(\boldsymbol{\beta}^0), s_1(\boldsymbol{\beta}^0))$ is the information set available at time t, where $Q_1(\boldsymbol{\beta}^0)$ and $s_1(\boldsymbol{\beta}^0)$ are some given initial conditions for the quantiles and IQR.

In Appendix A we provide sufficient and technical conditions on Q and s in Eq.(25) for consistent and asymptotically normal estimators of the parameters in Eq.(23).

Theorem 2 *(Consistency) Under Assumption 1,* $\hat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta^0}$ *as* $n \to \infty$

We provide the proof of Theorem 2 in Appendix B, where we use Theorem 2.1 of Newey & McFadden (1986) and Theorem 5.2 of Fissler & Ziegel (2016). For the financial time series that we consider here we can use a variety of uniform laws of large numbers to satisfy Assumption 1A. As for Assumption 1B, this is standard in parameter inference in time series.

Theorem 3 *(Asymptotic normality) Under Assumptions 1 and 2,*

$$\sqrt{n}\boldsymbol{A}_n^{-1/2}\boldsymbol{D}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta^0}) \xrightarrow{d} N(0, I) \text{ as } n \to \infty$$

where

$$\boldsymbol{D}_n = \mathbb{E}\Big[f_t(Q_t(\boldsymbol{\beta}^0)|\mathcal{F}_{t-1}) \nabla' Q_t(\boldsymbol{\beta}^0) \nabla Q_t(\boldsymbol{\beta}^0) + \frac{s_t(\boldsymbol{\beta}^0) - c|y_t|}{\Big|s_t(\boldsymbol{\beta}^0) - c|y_t|\Big|} \nabla' s_t(\boldsymbol{\beta}^0) \nabla s_t(\boldsymbol{\beta}^0) \Big] \tag{26}$$

$$\boldsymbol{A}_n = \mathbb{E}[h_t(\boldsymbol{\beta}^0) h_t(\boldsymbol{\beta}^0)'] \tag{27}$$

$$h_t(\boldsymbol{\beta}) = \frac{\partial L_{ND}(y_t, Q_t(\boldsymbol{\beta}), s_t(\boldsymbol{\beta}); \theta)}{\partial \boldsymbol{\beta}} \tag{28}$$

$$= \nabla' Q_t(\boldsymbol{\beta}^0) \sum_{r=1}^{6} \big(\mathbf{1}_{\{y_t \leq q_{r,t}\}} - \theta_r \big) + \sum_{r=1}^{6} \big(\mathbf{1}_{\{y_t \leq q_{r,t}\}} - \theta_r \big) (q_{r,t}(\boldsymbol{\beta}) - y_t) \tag{29}$$

$$+ \frac{s_t(\boldsymbol{\beta}^0) - c|y_t|}{\Big|s_t(\boldsymbol{\beta}^0) - c|y_t|\Big|} (\nabla' s_t(\boldsymbol{\beta}^0) + \nabla' L_{\theta_2} - \nabla' L_{\theta_1}) \tag{30}$$

$$+ \frac{s_t(\boldsymbol{\beta}^0) - c|y_t|}{\Big|s_t(\boldsymbol{\beta}^0) - c|y_t|\Big|} \nabla' s_t(\boldsymbol{\beta}^0) \tag{31}$$

$$\nabla' L_{\theta_2} = \nabla' q_{2,t}(\boldsymbol{\beta}^0) \big(\mathbf{1}_{\{y_t \leq q_{2,t}\}} - \theta_2 \big) + \big(\mathbf{1}_{\{y_t \leq q_{2,t}\}} - \theta_2 \big) (q_{2,t}(\boldsymbol{\beta}) - y_t)$$
$$\nabla' L_{\theta_1} = \nabla' q_{1,t}(\boldsymbol{\beta}^0) \big(\mathbf{1}_{\{y_t \leq q_{1,t}\}} - \theta_1 \big) + \big(\mathbf{1}_{\{y_t \leq q_{1,t}\}} - \theta_1 \big) (q_{1,t}(\boldsymbol{\beta}) - y_t)$$

The proof for this theorem is based on work by Huber (1967), Weiss (1991), Engle & Manganelli (2004). See Appendix for a sketch of the proof.

Assumptions 2 are standard in the literature. More specifically, Assumption 2(A)(ii) imposes the correct ordering for the quantiles (here we focus on both the left and the right conditional distribution of asset returns). Assumptions 2(B)-(E) are from Engle & Manganelli (2004) and pretty standard in quantile estimation. We require fat tailed distributions in Assumption 2(B)(ii) and we set $4 + \delta$ moments of returns to exist. Assumptions 2(C)-(D) concern conditions on the magnitude of the quantile and IQR series. Patton et al. (2019) show that these reduce to moment conditions on the observed returns for GARCH processes. Assumption 2(F) is a standard condition on time series dependence for the CLT. Assumption 2(G) limits the number of exact equalities of realised returns and fitted quantile values.

Theorem 4 *(Asymptotic covariance matrix) Under Assumptions 1-3,* $\hat{\boldsymbol{A}}_n - \boldsymbol{A}_n \xrightarrow{p} \mathbf{0}$ *and* $\hat{\boldsymbol{D}}_n - \boldsymbol{D}_n \xrightarrow{p} \mathbf{0}$ *where*

$$\hat{\boldsymbol{D}}_n = n^{-1} \sum_{t=1}^{n} \Big[\frac{1}{2c_n} \mathbf{1}\Big(\big| y_t - Q_t(\hat{\boldsymbol{\beta}}_n) \big| < c_n \Big) \nabla' Q_t(\hat{\boldsymbol{\beta}}_n) \nabla Q_t(\hat{\boldsymbol{\beta}}_n) \tag{32}$$

$$+ \frac{s_t(\hat{\boldsymbol{\beta}}_n) - c|y_t|}{\Big| s_t(\hat{\boldsymbol{\beta}}_n) - c|y_t| \Big|} \nabla' s_t(\hat{\boldsymbol{\beta}}_n) \nabla s_t(\hat{\boldsymbol{\beta}}_n) \Big] \tag{33}$$

$$\hat{\boldsymbol{A}}_n = n^{-1} \sum_{t=1}^{n} h_t(\hat{\boldsymbol{\beta}}_n) h_t(\hat{\boldsymbol{\beta}}_n)' \tag{34}$$

Theorem 4 is extension of work by Engle & Manganelli (2004) for single quantile models and Patton et al. (2019) for joint VaR and ES models.

3.1 Algorithm

The proposed loss function can be highly multi-modal for multi-quantile time series models such as the models developed in this paper. The estimation issue of quantiles (single and multiple) is well known in the literature see *e.g.* Patton et al. (2019), Engle & Manganelli (2004), White et al. (2012).

To address this issue we apply a combination of estimation strategies proposed by Mitrodima & Oberoi (2023) and Patton et al. (2019). This is to obtain "warm" starting values by considering the properties of the conditional quantiles (stationarity and quantile ordering) (Mitrodima & Oberoi 2023) as well as allowing for some smoothing (Patton et al. 2019). This procedure leads to improved estimates and is particularly well suited to our model.

Here we consider the DMSQ-SAV specification for the sake of presenting the procedure. The procedure begins by choosing a small set of possible values for $\beta^{(z)}_{\theta_k,i}$ and $\beta^{(s)}_i$ between 0.5 and 0.9. For instance, one may start off by setting all the $\beta^{(z)}_{\theta_k,i}$ at 0.5 and then at 0.7, and $\beta^{(s)}_i$ at 0.5 and 0.9. This initially gives four sets of starting β values. Associated with each set of β values, a set of five possible values for $\gamma^{(z)}_{\theta_k,i}$ or $\gamma^{(s)}_i$ is chosen. The possible values for $\gamma^{(z)}_{\theta_k,i}$ are $\{-0.2, -0.1, -0.02, -0.01, 0\}$ if $\theta_k < 0.5$ or $\{0, 0.01, 0.02, 0.1, 0.2\}$ if $\theta_k > 0.5$. For $\gamma^{(s)}_i$ and $\gamma^{(z)}_{0.5,i}$, both positive and negative values

from the above sets are chosen. A starting value for $u_{\theta_k}^{(z)}$ can be chosen so that the unconditional mean of $q_{\theta_k,t}/s_t$ is equal to the corresponding quantile of a standard normal distribution (although, other distribution could be used). The initial value for $u_{\theta_k}^{(z)}$ is

$$u_{\theta_k}^{(z)} = (1 - \beta_{\theta_k,1}^{(z)})\overline{q_{\theta_k}} - \gamma_{\theta_k,1}^{(z)} E\left[\frac{|y_t|}{s_t}\right],$$

where $\overline{q_{\theta_k}}$ is the quantile of standard normal distribution at probability level θ_k. We also set $E\left[\frac{|y_t|}{s_t}\right] = \frac{1}{1.347}\sqrt{\frac{2}{\pi}}$. A similar strategy is applied to the IQR, whereby $u_t^{(z)}$ is chosen so that the unconditional mean of s_t is equal to the corresponding sample IQR. This procedure can easily be extended to the other specification that we propose here.

The procedure for finding initial values for the proposed specifications exploits the use of standardised quantiles. Our estimation approach is similar in spirit to the variance targeting strategy explained in Christoffersen (2003), but it is only used to initialise the optimisation, not as a constraint. Following Engle & Manganelli (2004), we initialise the quantile and interquartile range series at their empirical counterparts (using the first 300 observations).

As for smoothing the indicator function we replace this by a Logistic function, since the smoothing function of the Logistic function converges to the indicator function as the smoothing parameter $\nu \to \infty$. In our empirical exercise we choose $\nu = 5$ and $\nu = 20$ following Patton et al. (2019).

As the smoothed objective function is differentiable, we first use gradient-based numerical search algorithms (*e.g.* fminunc or fmincon) for the set of initial parameter values described above first for $\nu = 5$ and then for $\nu = 20$. Separate optimisers of the loss function are then run for the obtained initial parameter values (the Matlab optimiser "fminsearch" was used in our examples), and the outputted parameter value with the largest value is reported as the estimate.

4 Analysis of stock index data

We use data provided by Datastream. We analyse the S&P500 index (SPX), along with Russell 2000 (RUT), Russell 1000 (RUI), and NASDAQ 100 (NDX). The complete data set consists of daily log returns from January 01, 2001 to November 20, 2024. We will initially consider inference to understand the underlying dynamics of the return distribution. For comparison, we estimate a GARCH(1,1) and the autoregressive conditional density model (ARCD) by Hansen (1994) as well as our proposed scale-shape models DMSQ-SAV and DMSQ-AS for the $1\%, 5\%, 25\%, 50\%, 75\%, 95\%$ and 99% probability levels.

The results for RUT, RUI, and NDX can be found in the Appendix C as here we only present the SPX results.

4.1 Results

The estimated conditional quantiles and their values standardised by the conditional IQR for SPX are presented in Figures 2 and 3 along with the relevant quantiles for the benchmarks (GARCH and ARCD).

The results show that there is substantial time-variation in the conditional quantiles captured by our models and ARCD, but not by GARCH. The estimated standardised conditional quantiles for our models show much less variation over time and illustrate that a lot of the time-variation in the conditional quantiles can be captured through the conditional IQR. The smoothness of the estimated standardised conditional quantiles varies with the probability level with the paths becoming rougher in the tails of the distribution (DMSQ-SAV). The estimated standardised conditional quantiles also show clear differences between our models and GARCH and ARCD.

GARCH assumes that the standardised conditional quantiles are constant over time. The estimates of the conditional quantiles are similar. This is because GARCH cannot accommodate the heavier tails following a shock leading to overestimation of the volatility. This emphasises the importance of allowing the shape of the conditional

Figure 2: Conditional density: quantile estimates for SPX
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD

Figure 3: Conditional density: standardised quantile estimates for SPX
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD

distribution of returns to change over time.

The estimates of the conditional skewness and kurtosis can be calculated from the estimated conditional quantiles using the following formulae suggested by White et al. (2010) and Kim & White (2004), who follow the statistics literature on robust estimators of higher moments. The daily conditional skewness is estimated as

$$\frac{q_{0.75,t} + q_{0.25,t} - 2 \times q_{0.50,t}}{q_{0.75,t} - q_{0.25,t}}$$

and the kurtosis as

$$\frac{q_{0.99,t} - q_{0.01,t}}{q_{0.75,t} - q_{0.25,t}} - 3.45.$$

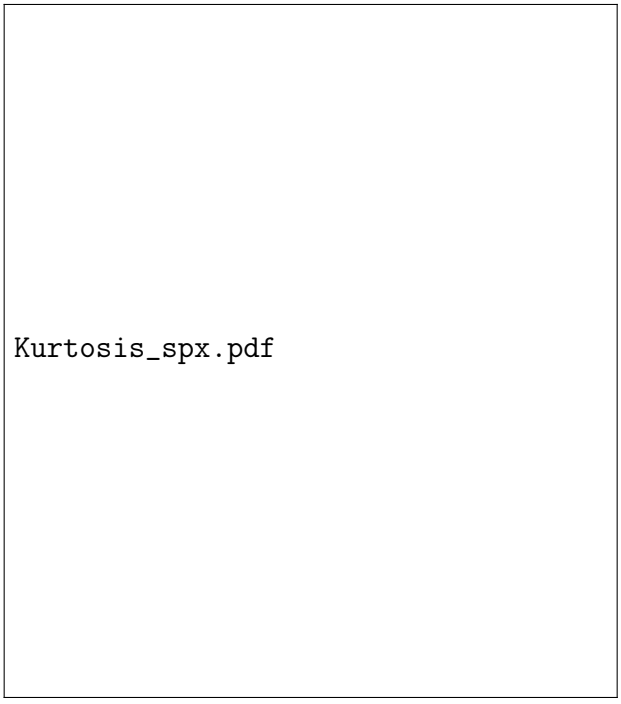
In Figures 4a and 4b, we plot the series of the conditional skewness and kurtosis for SPX, estimated by the four models and an empirical rolling one against the data.

The conditional skewness results for SPX using the DMSQ models (in Figure 4a) show a clear pattern with periods of both positive and negative skewness. These estimates are consistent with what we might expect from a stock, with positive skewness in better times and negative skewness in times of stress. The ARCD model also shows a similar pattern over time but smooths out some of the features. The DMSQ models also generate much greater range in the conditional kurtosis over time (Figure 4b). In contrast, the estimated conditional skewness for SPX is negative at all times but with a much wider range of values for the DMSQ-SAV and DMSQ-AS models. Once again, the negative conditional skewness of the index is consistent with the stylised facts of returns data. Although the causes of this effect may be disputed, it is well known that indices display a more pronounced leverage effect than their constituent stocks individually. One explanation for this is that correlations between returns are higher in bad times than in good ones.

4.2 Hit ratios

One way to evaluate the fit or the performance of quantile models is to compare their in-sample hit ratios (coverage). We report in-sample hit ratios ($\frac{1}{T}\sum_{t=1}^{T} I(y_t < q_{\theta,t})$) for each of the joint quantile models and their benchmarks for SPX.

(a) Skewnes



(b) Kurtosis

Figure 4: Conditional skewness and kurtosis for SPX

Table 1: In-sample hit ratios ($\frac{1}{T}\sum_{t=1}^{T} I(y_t < q_{\theta,t})$) for SPX

θ	**0.99**	**0.95**	**0.50**	**0.05**	**0.01**	**0.25**	**0.75**
DMSQ SAV	**0.9903**	**0.9499**	0.4963	**0.0511**	**0.0117**	0.2422	**0.7505**
DMSQ AS	0.9888	**0.9501**	0.4846	0.0519	0.0120	0.2397	0.7508
ARCD	0.9752	0.9176	0.4962	0.0661	0.0195	0.2287	0.7485
GARCH	0.9842	0.9441	**0.4993**	0.0588	0.0145	**0.2543**	0.7457

Note: The sample ranges from January 01, 2002, to December 31, 2014. With bold we indicate the best performing model. The winner here is DMSQ SAV.

All models which account for a time-varying conditional density perform well overall, with the two new models outperforming ARCD. On the other hand, GARCH doens't seem to be able to capture potential short-term shifts in the location of the return distribution. As a result, the in-sample hit ratios of the outer conditional quantiles appear to be biased. This may support the case for allowing the returns themselves to have some dependence, as suggested by Kuester et al. (2006). An alternative explanation is that our proxy for the scale does a good job in capturing the long-range dependence in scale.

However, the objective of this paper is not only to propose a more parsimonious and intuitive model for multiple conditional quantiles, but also to examine the implications for time variation in the shape of the underlying distribution.

5 Conclusion

In this paper, we propose a multi-quantile time series model for the conditional distribution of index returns. The model is constructed by specifying a process for the scale (which is taken to be the IQR) and the standardised conditional quantiles (which control the overall shape of the distribution). This mimics the standard approach to modelling financial time series, where the returns are modelled as the product of a scale and a random shock drawn from a time-invariant distribution (*e.g.* GARCH or stochastic volatility modelling). Our approach builds a single model for several conditional quantiles estimated jointly and so avoids the need for two-stage estimation. The decomposition also assists with the crossing problem, as does our estimation strategy of targeting the level of the quantile process when setting initial values.

A major benefit of this approach is that we have both a robust estimate of the time-varying scale of return distribution whilst also having time-varying estimates of the shape of the conditional distribution. This allows us to better understand where the largest changes in the conditional distribution of returns occur. Our results suggest that a robustly estimated scale can capture most of the dynamics of the conditional return distribution which leaves a relatively stable underlying shape that can be

captured by fairly simple dynamics.

The estimated conditional return distribution offers useful insights into the dynamics of asset returns. While each quantile, when individually modelled, requires long memory and asymmetry to be captured adequately, this effect is subsumed in the scale dynamics. Once the scale dynamics are accounted for, the standardised tails are suitably represented by a very simple time series process, with very occasional spikes and a little structure still remaining. This supports the use of a sufficiently complicated volatility structure to model fat-tailed and skewed return distributions, but not necessarily the introduction of complicated distributional assumptions for the underlying shocks.

References

Backus, D., Foresi, S., Li, K. & Wu, L. (1997), Accounting for Biases in Black-Scholes, Technical Report 30, CRIF Working Paper series.

Baillie, R. T., Bollerslev, T. & Mikkelsen, H. O. (1996), 'Fractionally integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **74**(1), 3–30.

Bates, D. S. (2006), 'Maximum likelihood estimation of latent affine processes', *Review of Financial Studies* **19**(3), 909–965.

Black, F. (1976), 'The pricing of commodity contracts', *Journal of Financial Economics* **3**, 167–179.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (1987), 'A conditionally heteroskedastic time series model for speculative prices and rates of return', *The Review of Economics and Statistics* **69**(3), 542–47.

Chen, C. & Gerlach, R. (2014), 'Bayesian expected shortfall forecasting incorporating the intraday range', *Journal of Financial Econometrics* **14**, 128–158.

Chernov, M., Gallant, R. A., Ghysels, E. & Tauchen, G. (2003), 'Alternative models of stock price dynamics', *Journal of Econometrics* **116(1-2)**, 225–257.

Chernozhukov, V., Fernández-Val, I. & Galichon, A. (2010), 'Quantile and probability curves without crossing', *Econometrica* **78**(3), 1093–1125.

Chernozhukov, V. & Galichon, A. (2008), 'Improving point and interval estimates of monotone functions by rearrangement', *Biometrika* **96**, 559–575.

Christoffersen, P. (2003), *Elements of Financial Risk Management*, Academic Press, London.

Christoffersen, P., Heston, S. & Jacobs, K. (2013), 'Capturing option anomalies with a variance-dependent pricing kernel', *Review of Financial Studies* **26**(8), 1963–2006.

Cont, R., Deguest, R. & Scandolo, G. (2010), 'Robustness and sensitivity analysis of risk measurement procedures', *Quantitative Finance* **10**(6), 593–606.
URL: *https://EconPapers.repec.org/RePEc:taf:quantf:v:10:y:2010:i:6:p:593-606*

Coroneo, L. & Veredas, D. (2012), 'A simple two-component model for the distribution of intraday returns', *The European Journal of Finance* **18**(9), 775–797.

Delatola, E.-I. & Griffin, J. E. (2011), 'Bayesian nonparametric modelling of the return distribution with stochastic volatility', *Bayesian Analysis* **6**(4), 901–926.

Ding, Z. & Granger, C. W. (1996), 'Modeling volatility persistence of speculative returns: A new approach', *Journal of Econometrics* **73**(1), 185 – 215.

Ding, Z., Granger, C. W. & Engle, R. F. (1993), 'A long memory property of stock market returns and a new model', *Journal of Empirical Finance* **1**(1), 83 – 106.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of uk inflation', *Econometrica* **50**, 987–1008.

Engle, R. F. & Gonzalez-Rivera, G. (1991), 'Semiparametric ARCH models', *Journal of Business & Economic Statistics* **9**(4), 345–359.

Engle, R. F. & Lee, G. (1999), A permanent and transitory component model of stock return volatility, *in* 'Cointegration, Causality and Forecasting: A festschrift in honor of Clive W.J. Granger', Oxford University Press, pp. 475–497.

Engle, R. F. & Manganelli, S. (2001), Value At Risk models in finance, *in* G. P. Szegő, ed., 'Risk Measures for the 21st Century', Wiley finance series, John Wiley and Sons, West Sussex.

Engle, R. F. & Manganelli, S. (2004), 'CAViaR: Conditional autoregressive value at risk by regression quantiles', *Journal of Business & Economic Statistics* **22**, 367–381.

Engle, R. F. & Ng, V. K. (1993), 'Measuring and testing the impact of news on volatility', *The Journal of Finance* **48**(5), 1749–1778.

Fissler, T. & Ziegel, J. F. (2016), 'Higher order elicitability and Osband's principle', *The Annals of Statistics* **44**(4), 1680–1707.

Fissler, T. & Ziegel, J. F. (2021), 'On the elicitability of range value at risk', *Statistics & Risk Modeling* **38**(1-2), 25–46.

French, K. R., Schwert, G. W. & Stambaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**(1), 3–29.

Gallant, R. A., Hsieh, D. A. & Tauchen, G. E. (1991), On fitting a recalcitrant series: The pound/dollar exchange rate, 1974-1983, *in* W. A. Barnett, J. Powell & G. E. Tauchen, eds, 'Nonparametric and Semiparametric Methods in Economics and Statistics', Cambridge University Press, Cambridge, pp. 199–2401.

Gerlach, R. H., Chen, C. W. & Chan, N. Y. (2011), 'Bayesian time-varying quantile forecasting for value-at-risk in financial markets', *Journal of Business & Economic Statistics* **29**(4).

Ghysels, E., Plazzi, A. & Valkanov, R. I. (2011), Conditional skewness of stock market returns in developed and emerging markets and its economic fundamentals, Swiss Finance Institute Research Paper Series 11-06, Swiss Finance Institute.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The Journal of Finance* **48**(5), 1779–1801.

Gourieroux, C. & Jasiak, J. (2008), 'Dynamic quantile models', *Journal of Econometrics* **147**, 198–205.

Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Academic Press, New York.

Hansen, B. E. (1994), 'Autoregressive conditional density estimation', *International Economic Review* **35**, 705–730.

Hansen, P. R. & Lunde, A. (2005), 'A forecast comparison of volatility models: does anything beat a garch(1,1)?', *Journal of Applied Econometrics* **20**(7), 873–889. **URL:** *http://dx.doi.org/10.1002/jae.800*

Harvey, C. R. & Siddique, A. (1999), 'Autoregressive conditional skewness', *Journal of Financial and Quantitative Analysis* **34**(4), 465–487.

Huber, P. J. (1967), The behavior of maximum likelihood estimates under nonstandard conditions, *in* 'Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability', University of California Press, Berkeley.

Jensen, M. J. & Maheu, J. M. (2010), 'Bayesian semiparametric stochastic volatility modeling', *Journal of Econometrics* **157**, 306–316.

Jeon, J. & Taylor, J. W. (2013), 'Using CAViaR models with implied volatility for value-at-risk estimation', *Journal of Forecasting* **32**(1), 62–74.

Kalli, M., Walker, S. G. & Damien, P. (2014), 'Modelling the conditional distribution of daily stock index returns: an alternative bayesian semi-parametric model', *Journal of Business and Economic Statistics* **31**, 371–383.

Kim, T.-H. & White, H. (2004), 'On more robust estimation of skewness and kurtosis', *Finance Research Letters* **1**(1), 56–73.

Kuester, K., Mittnik, S. & Paolella, M. (2006), 'Value-at-risk prediction: A comparison of alternative strategies', *Journal of Financial Econometrics* **4(1)**, 53–89.

McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* **7**(3â€“4), 271–300. Special issue on Risk Management.

Mitrodima, G. & Oberoi, J. (2023), 'CAViaR models for Value-at-Risk and Expected Shortfall with long range dependency features', *Journal of the Royal Statistical Society Series C: Applied Statistics* .

Nelson, D. B. (1991), 'Conditional Heteroskedasticity in Asset Returns: A New Approach', *Econometrica* **59**(2), 347–70.

Newey, W. K. & McFadden, D. (1986), Large sample estimation and hypothesis testing, *in* R. Engle & D. McFadden, eds, 'Handbook of Econometrics', Vol. IV, Elsevier, Amsterdam, The Netherlands, chapter 36, pp. 2111–2245.

Newey, W. K. & McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle & D. L. McFadden, eds, 'Handbook of Econometrics', Vol. 4, Elsevier, pp. 2111–2245.

Nolde, N. & Ziegel, J. F. (2017), 'Elicitability and backtesting: Perspectives for banking regulation', *Annals of Applied Statistics* **11**(4), 1833–1874.

Patton, A. J. (2006), 'Modelling asymmetric exchange rate dependence', *International Economic Review* **47**(2), 527–556.

Patton, A. J. & Sheppard, K. (2009), 'Evaluating volatility and correlation forecasts', *Handbook of Financial Time Series* pp. 801–838.

Patton, A. J., Ziegel, J. F. & Chen, R. (2019), 'Dynamic semiparametric models for expected shortfall (and value-at-risk)', *Journal of Econometrics* **211**(2), 388–413.

Schwert, G. W. (1989), 'Why does stock market volatility change over time?', *The Journal of Finance* **44**, 1115–1153.

Taylor, J. (2008), 'Estimating value at risk and expected shortfall using expectiles', *Journal of Financial Econometrics* **6**, 231–252.

Taylor, J. W. (2005), 'Generating volatility forecasts from value at risk estimates', *Management Science* **51**(5), 712–725.

Weiss, A. A. (1991), Asymmetric garch models and financial time series, *in* T. B. Fomby & T. K. Seo, eds, 'Nonlinear Models', Vol. 11 of *Advances in Econometrics*, JAI Press, pp. 95–109.

White, H., Kim, T.-H. & Manganelli, S. (2010), Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR, *in* 'Volatility and Time Series Econometrics: Essays in Honour of Robert Engle, ed. by M. Watson, T. Bollerslev and J. Russell', Oxford University Press.

White, H., Kim, T.-H. & Manganelli, S. (2012), Var for var: Measuring tail dependence using multivariate regression quantiles, Working papers 2012rwp-45, Yonsei University, Yonsei Economics Research Institute.

Xiao, Z. & Koenker, R. (2009), 'Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models', *Journal of the American Statistical Association* **104**(488), 1696–1712.

Appendix

A Assumptions

Assumption 1 *(A)* $L(y_t, Q_t(\boldsymbol{\beta}), s_t(\boldsymbol{\beta}); \theta)$ *obeys the uniform law of large numbers*

(B) (i) The parameter space B *is a compact subset of* $\mathbb{R}^p$ *for* $p < \infty$

(ii) $\{y_t\}_{t=1}^{\infty}$ *is a strictly stationary process. Conditional on all information set in the past,* $\mathcal{F}_{t-1}$*, the distribution of* y_t *is* $F_t(\cdot|\mathcal{F}_{t-1})$*,* $\forall t$ *and belongs to a class of distribution functions on* $\mathbb{R}$ *with finite first moments and unique* θ*-quantiles*

(iii) Both $Q_t(\boldsymbol{\beta})$*,* $\forall t$ *and* $s_t(\boldsymbol{\beta})$*,* $\forall t$ *are* $\mathcal{F}_{t-1}$*-measurable and a.s. continuous in* $\boldsymbol{\beta}$

(iv) If $Pr[Q_t(\boldsymbol{\beta}) = Q_t(\boldsymbol{\beta}^0) \cap s_t(\boldsymbol{\beta}) = s_t(\boldsymbol{\beta}^0)] = 1$ $\forall t$*,then* $\boldsymbol{\beta} = \boldsymbol{\beta}^0$

Assumption 2 *(A) We have* $\forall t$*:*

(i) $Q_t(\boldsymbol{\beta})$ *and* $s_t(\boldsymbol{\beta})$ *are a.s. twice continuously differentiable in* $\boldsymbol{\beta}$

(ii) $0 < q_{\theta_3,t}(\boldsymbol{\beta}^0) < q_{\theta_4,t}(\boldsymbol{\beta}^0) < q_{\theta_2,t}(\boldsymbol{\beta}^0) < s_t(\boldsymbol{\beta}^0) < q_{\theta_1,t}(\boldsymbol{\beta}^0) < q_{\theta_5,t}(\boldsymbol{\beta}^0) < q_{\theta_6,t}(\boldsymbol{\beta}^0) < 0$

(B) We have $\forall t$*:*

(i) conditional on all the past information $\mathcal{F}_{t-1}$*,* y_t *has a continuous density* $f_t(\cdot|\mathcal{F}_{t-1})$ *that satisfies* $f_t(y|\mathcal{F}_{t-1} \leq K_0 < \infty$ *and* $|f_t(y'|\mathcal{F}_{t-1}) - f_t(y''|\mathcal{F}_{t-1})| \leq K_0|y' - y''|$

(ii) $\mathbb{E}[|y_t|^{4+\delta}] \leq K_0 < \infty$*, for some* $0 < \delta < 1$

(C) There exists a neighbourhood of $\boldsymbol{\beta}^0, \mathcal{N}(\boldsymbol{\beta}^0)$*, such that* $\forall t$ *we have:*

(i) $\frac{s - c|y|}{\left|s - c|y|\right|} \leq K_0 < \infty, \forall \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}^0)$

(ii) there exist some (possibly stochastic) $\mathcal{F}_{t-1}$*-measurable functions*

$$V(\mathcal{F}_{t-1}), V_1(\mathcal{F}_{t-1}), H_1(\mathcal{F}_{t-1}), V_2(\mathcal{F}_{t-1}), H_2(\mathcal{F}_{t-1})$$

that satisfy $\forall \boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\beta}^0) : |Q_t(\boldsymbol{\beta})| \leq V(\mathcal{F}_{t-1}), ||\nabla Q_t(\boldsymbol{\beta})|| \leq V_1(\mathcal{F}_{t-1}), ||\nabla s_t(\boldsymbol{\beta})|| \leq H_1(\mathcal{F}_{t-1}), ||\nabla^2 Q_t(\boldsymbol{\beta})|| \leq V_2(\mathcal{F}_{t-1})$*, and* $||\nabla^2 s_t(\boldsymbol{\beta})|| \leq H_2(\mathcal{F}_{t-1})$

(D) For some $0 < \delta < 1$ *and* $\forall t$ *we have*

(i) $\mathbb{E}[V_1(\mathcal{F}_{t-1})^{3+\delta}], \mathbb{E}[H_1(\mathcal{F}_{t-1})^{3+\delta}], \mathbb{E}[V_2(\mathcal{F}_{t-1})^{\frac{3+\delta}{2}}], \mathbb{E}[H_2(\mathcal{F}_{t-1})^{\frac{3+\delta}{2}}] \leq K_0$

(ii) $\mathbb{E}[V(\mathcal{F}_{t-1})^2 + \delta V_1(\mathcal{F}_{t-1}) H_1(\mathcal{F}_{t-1})^2 + \delta] \leq K_0$

(iii) $\mathbb{E}[H_1(\mathcal{F}_{t-1})^{1+\delta} H_2(\mathcal{F}_{t-1}) |y_t|^{2+\delta}], \mathbb{E}[H_1(\mathcal{F}_{t-1})^{3+\delta} |y_t|^{2+\delta}] \leq K_0$

(E) The matrix $\boldsymbol{D}_n$ *is (strictly) positive definite for* n *sufficiently large*

(F) $\{[y_t, Q_t(\boldsymbol{\beta}^0), s_t(\boldsymbol{\beta}^0), \nabla' Q_t(\boldsymbol{\beta}^0), \nabla' s_t(\boldsymbol{\beta}^0)]\}$ *is* α*-mixing with* $\sum_{m=1}^{\infty} a(m)^{(r-2)/r} < \infty$ *for some* $r > 2$

(G) For any $n, \sup_{\boldsymbol{\beta} \in B} \sum_{t=1}^{n} I_{[y_t = Q_t(\boldsymbol{\beta})]} \leq K_0$ *a.s.*

Assumption 3 *(A) Let a deterministic positive sequence* c_n *that satisfies* $c_n = o(1)$ *and* $c_n^{-1} = o(n^{1/2})$

(B) (i) $\frac{1}{n} \sum_{t=1}^{n} h_t(\boldsymbol{\beta}^0) h_t(\boldsymbol{\beta}^0)' - \boldsymbol{A_n} \xrightarrow{p} \mathbf{0}$

(ii) $\frac{1}{n} \sum_{t=1}^{n} \frac{1}{s_t(\boldsymbol{\beta}^0)^2} \nabla' s_t(\boldsymbol{\beta}^0) \nabla s_t(\boldsymbol{\beta}^0) - \mathbb{E}[\frac{1}{s_t(\boldsymbol{\beta}^0)^2} \nabla' s_t(\boldsymbol{\beta}^0) \nabla s_t(\boldsymbol{\beta}^0)] \xrightarrow{p} \mathbf{0}$

(iii) $\frac{1}{n} \sum_{t=1}^{T} \frac{f_t(Q_t(\boldsymbol{\beta}^0)|\mathcal{F}_{t-1})}{-s_t(\boldsymbol{\beta}^0)\theta} \nabla' Q_t(\boldsymbol{\beta}^0) \nabla Q_t(\boldsymbol{\beta}^0) - \mathbb{E}[\frac{f_t(Q_t(\boldsymbol{\beta}^0)|\mathcal{F}_{t-1})}{-s_t(\boldsymbol{\beta}^0)\theta} \nabla' Q_t(\boldsymbol{\beta}^0) \nabla Q_t(\boldsymbol{\beta}^0)] \xrightarrow{p} \mathbf{0}$

B Proofs

Theorem 2 proof. To prove Theorem 2 we use the results in Newey & McFadden (1986) (Theorem 2.1). Assumptions (ii) to (iv) of Theorem 2.1 (Newey & McFadden 1986) are clearly satisfied for the time series data that we consider here. Thus, we need to show only that assumption (i) holds *i.e.* that $\mathbb{E}[L_{ND,n}(\cdot)]$ is uniquely minimised at $\boldsymbol{\beta}^0$.

Under Assumption 1(B)(iii) and the fact that our choice of the objective function L_{ND} satisfies the condition as in Theorem 5.2 (Fissler & Ziegel 2016) it holds that $\mathbb{E}[L_{ND}(y_t, Q_t(\boldsymbol{\beta}), s_t(\boldsymbol{\beta}); \theta)|\mathcal{F}_{t-1}]$ is uniquely minimised at $(Q(y_t|\mathcal{F}_{t-1}), s(y_t|\mathcal{F}_{t-1}))$, which equals $(Q_t(\boldsymbol{\beta}^0), s_t(\boldsymbol{\beta}^0))$ under the correct specification. Combining this assumption and Assumption 1(B)(iv), $\boldsymbol{\beta}^0$ is a unique minimiser of $\mathbb{E}[L_{ND,n}(\cdot)]$ which completes the proof.

Theorem 3 proof. The result is obtained directly by drawing from Lemma A.1 of Weiss (1991) and verifying that assumptions N1–N5 are satisfied. Assumptions N1, N2, N5 are satisfied given our Assumptions 1–2, while for assumptions N3 and N4 we need to show:

- Under Assumptions 1-2, Assumption N3(i) of Weiss (1991) holds:

$$\|\lambda_n(\boldsymbol{\beta})\| \geq a\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\|, \quad \text{for } \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \leq d_0,$$

  for n sufficiently large, where a and d_0 are strictly positive numbers.

- Define

$$\mu_t(\boldsymbol{\beta}; d) = \sup_{\|\boldsymbol{\xi} - \boldsymbol{\beta}\| \leq d} \|h_t(\boldsymbol{\xi}) - h_t(\boldsymbol{\beta})\|.$$

  Then, under Assumptions 1-2, Assumption N3(ii) of Weiss (1991) holds:

$$\mathbb{E}[\mu_t(\boldsymbol{\beta}; d)] \leq bd, \quad \text{for } \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| + d \leq d_0,\ d \geq 0,$$

  for n sufficiently large, where b, d, and d_0 are strictly positive numbers.

- Under Assumptions 1-2, Assumption N3(iii) of Weiss (1991) holds:

$$\mathbb{E}[\mu_t(\boldsymbol{\beta}; d)^p] \leq bd, \quad \text{for } \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| + d \leq d_0, \text{ and some } p > 2,$$

  for n sufficiently large, and where $b > 0$, $d \geq 0$, and $d_0 > 0$.

- Under Assumptions 1–2, $\mathbb{E}\|h_t(\boldsymbol{\beta}^0)\|^{2+\delta} \leq N$, for all t and some $N > 0$.

The above results follow directly from the lemmas in Patton et al. (2019) and for the proposed models here.

We also need to show that a CLT applies for the sequence

$$\left\{ n^{-1/2} \sum_{t=1}^{n} h_t(\boldsymbol{\beta}^0) \right\},$$

with asymptotic covariance matrix $\boldsymbol{A}_n = \mathbb{E}[h_t(\boldsymbol{\beta}^0) h_t(\boldsymbol{\beta}^0)']$. We denote $\Lambda(\boldsymbol{\beta}^0)$ as $\boldsymbol{D}_n$, leading to the stated result. Again, the result follows from Lemma 7 in Patton et al. (2019):

Under Assumptions 1-2:

$$n^{-1/2}\sum_{t=1}^{n} h_t(\boldsymbol{\beta}^0) \xrightarrow{d} N(0, \boldsymbol{A}_n) \quad \text{as } n \to \infty,$$

The sequence $\{h_t(\boldsymbol{\beta}^0)\}$ is stationary by Assumption 1(B)(ii), and has zero mean. Corollary 5.1 of Hall & Heyde (1980) and the Cramer-Wold result along with Assumption 2(F) and the fact that under Assumptions 1-2, $\mathbb{E}\|h_t(\boldsymbol{\beta}^0)\|^{2+\epsilon} \leq N$, for all t and some $N > 0$ give the result.

Outline of Theorem 4 proof. The proof that $\hat{\boldsymbol{A}}_n - \boldsymbol{A}_n \xrightarrow{p} \mathbf{0}$ (Assumption 3B(i), Theorem 2) is standard.

Let

$$\tilde{\boldsymbol{D}}_n = n^{-1}\sum_{t=1}^{n}\left[\frac{1}{2c_n}\mathbf{1}\Big(\big|y_t - Q_t(\boldsymbol{\beta}^0)\big| < c_n\Big)\nabla' Q_t(\boldsymbol{\beta}^0)\nabla Q_t(\boldsymbol{\beta}^0) + \frac{s_t(\boldsymbol{\beta}^0) - c|y_t|}{|s_t(\boldsymbol{\beta}^0) - c|y_t||}\nabla' s_t(\boldsymbol{\beta}^0)\nabla s_t(\boldsymbol{\beta}^0)\right]$$

We show that $\hat{\boldsymbol{D}}_n - \tilde{\boldsymbol{D}}_n = o_p(1)$ and $\tilde{\boldsymbol{D}}_n - \boldsymbol{D}_n = o_p(1)$.

Let:

$$\begin{aligned}
\|\hat{\boldsymbol{D}}_n - \tilde{\boldsymbol{D}}_n\| \leq \| & (2nc_n)^{-1} \\
& \times \sum_{t=1}^{n}\left[\mathbf{1}\Big(\big|y_t - Q_t(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}})\big| < c_n\Big) - \mathbf{1}\Big(\big|y_t - Q_t(\boldsymbol{\beta}^{\mathbf{0}})\big| < c_n\Big)\right]\nabla' Q_t(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}})\nabla Q_t(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}) \\
& + \mathbf{1}\Big(\big|y_t - Q_t(\boldsymbol{\beta}^{\mathbf{0}})\big| < c_n\Big)\big(\nabla Q_t(\hat{\boldsymbol{\beta}}_n) - \nabla Q_t(\boldsymbol{\beta}^0)\big)'\nabla Q_t(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}) \\
& + \mathbf{1}\Big(\big|y_t - Q_t(\boldsymbol{\beta}^{\mathbf{0}})\big| < c_n\Big)\nabla Q_t(\boldsymbol{\beta}^0)'\big(\nabla Q_t(\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}) - \nabla Q_t(\boldsymbol{\beta}^{\mathbf{0}})\big)\| \\
& + n^{-1}\sum_{t=1}^{n}\Big\| \frac{s_t(\hat{\boldsymbol{\beta}}_n) - c|y_t|}{\Big|s_t(\hat{\boldsymbol{\beta}}_n) - c|y_t|\Big|}\nabla' s_t(\hat{\boldsymbol{\beta}}_n)\nabla s_t(\hat{\boldsymbol{\beta}}_n) \\
& - \frac{s_t(\boldsymbol{\beta}^{\mathbf{0}}) - c|y_t|}{\Big|s_t(\boldsymbol{\beta}^{\mathbf{0}}) - c|y_t|\Big|}\nabla' s_t(\boldsymbol{\beta}^{\mathbf{0}})\nabla s_t(\boldsymbol{\beta}^{\mathbf{0}})\Big\|
\end{aligned}$$

In Theorem 2 we show that the last two lines are $o_p(1)$ in the proof of Theorem 2. Engle & Manganelli (2004) also show that the lines with the indicator functions are also $o_p(1)$.

Let's now consider:

$$\begin{aligned}\tilde{\boldsymbol{D}}_n - \boldsymbol{D}_n = \frac{1}{2nc_n}\sum_{t=1}^{n}\Big[\mathbf{1}\big(\big|y_t - Q_t(\boldsymbol{\beta^0})\big| < c_n\big) - \mathbb{E}\big[\mathbf{1}\big(\big|y_t - Q_t(\boldsymbol{\beta^0})\big| < c_n\big)|\mathcal{F}_{t-1}\big]\Big]\nabla' Q_t(\boldsymbol{\beta^0})\nabla Q_t(\boldsymbol{\beta^0}) \\ + \frac{1}{n}\sum_{t=1}^{n}\Big[\frac{1}{2c_n}\mathbb{E}\big[\mathbf{1}\big(\big|y_t - Q_t(\boldsymbol{\beta^0})\big| < c_n\big)|\mathcal{F}_{t-1}\big]\nabla' Q_t(\boldsymbol{\beta^0})\nabla Q_t(\boldsymbol{\beta^0}) \\ - \mathbb{E}\big[f_t(Q_t(\boldsymbol{\beta^0}))\nabla' Q_t(\boldsymbol{\beta^0})\nabla Q_t(\boldsymbol{\beta^0})\big]\Big] \\ + o_p(1)\end{aligned}$$

Following Engle & Manganelli (2004), Assumptions 1–3 are sufficient to show $\tilde{\boldsymbol{D}}_n - \boldsymbol{D}_n = o_p(1)$ and the result follows.

C Additional Figures

This sectiom contains figures and tables for RUT, RUI, and NDX corresponding to those reported in the paper for SPX.

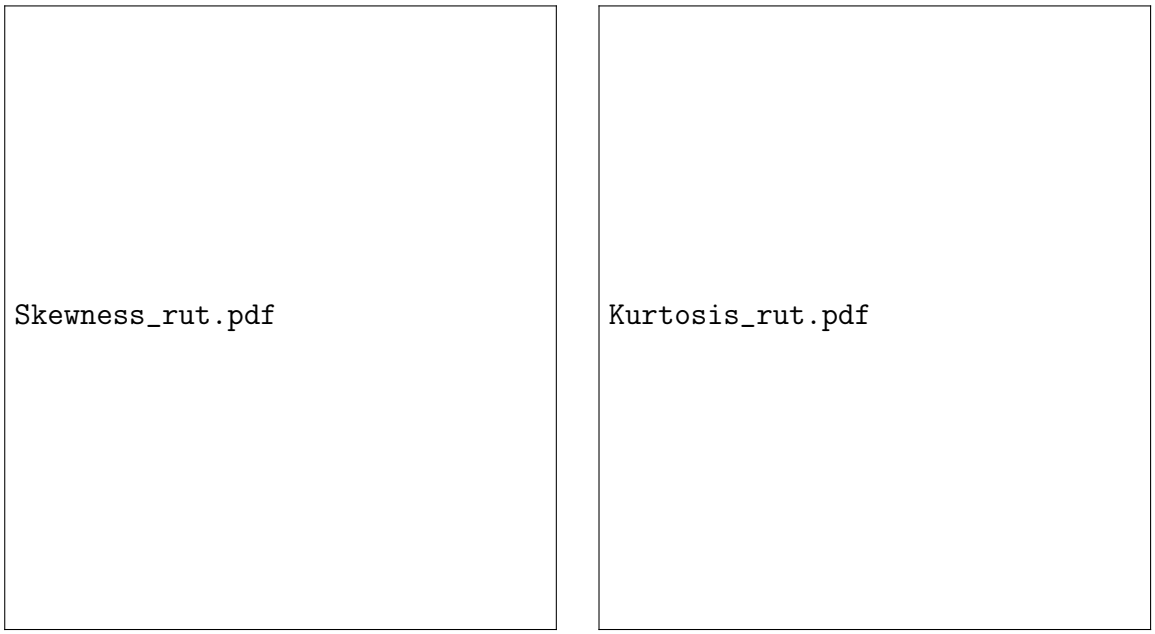
The estimated conditional quantiles and their values standardised by the conditional IQR for RUT are presented in Figures A1 and A2 along with the relevant quantiles for the benchmarks (GARCH and ARCD).

In Figures A3a and A3b, we plot the series of the conditional skewness and kurtosis for RUT, estimated by the four models and an empirical rolling one against the data.

We also report the violation hits in Table 2.

Figure A1: Conditional density: quantile estimates for RUT
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD

Figure A2: Conditional density: standardised quantile estimates for RUT
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD



(a) Skewnes (b) Kurtosis

Figure A3: Conditional skewness and kurtosis for RUT

Table 2: In-sample hit ratios ($\frac{1}{T}\sum_{t=1}^{T} I(y_t < q_{\theta,t})$) for RUT

θ	**0.9**	**0.95**	**0.50**	**0.05**	**0.01**	**0.25**	**0.75**
DMSQ SAV	0.9892	**0.9490**	**0.4999**	**0.0510**	**0.0102**	0.2471	**0.7494**
DMSQ AS	0.9885	**0.9490**	0.5007	0.0518	0.0097	0.2445	0.7512
ARCD	**0.9923**	0.9479	0.4923	0.0456	0.0083	0.2414	0.7463
GARCH	0.9847	0.9414	0.4990	0.0548	0.0122	**0.2511**	0.7474

Note: The sample ranges from January 01, 2002, to December 31, 2014. With bold we indicate the best performing model.

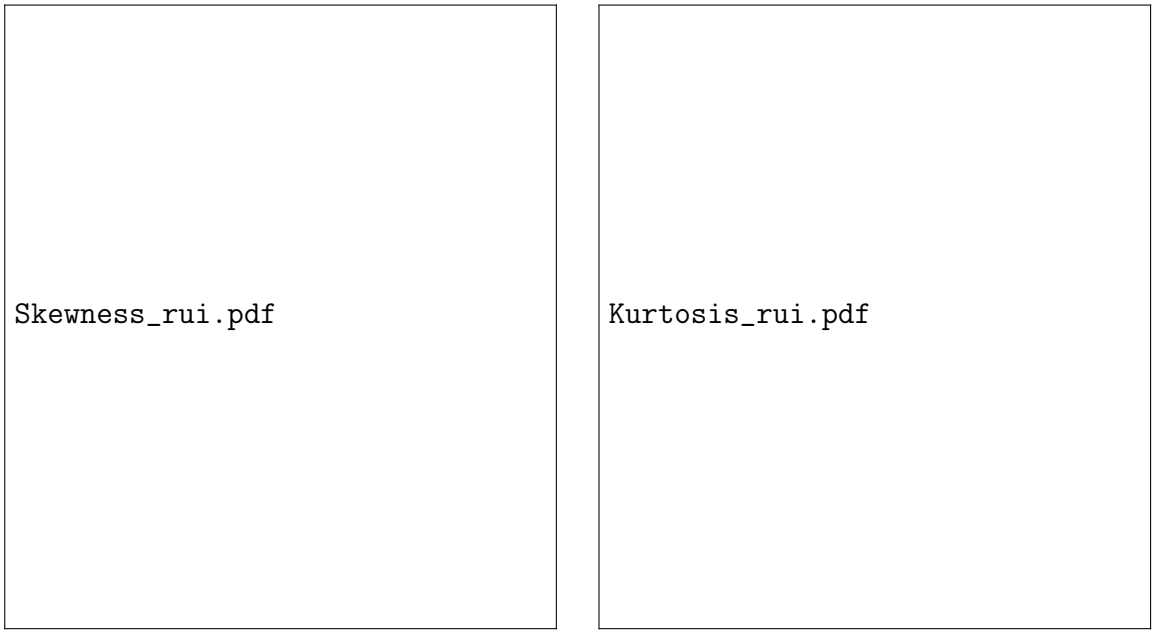
The estimated conditional quantiles and their values standardised by the conditional interquartile range for RUI are presented in Figures A4 and A5 along with the relevant quantiles for the benchmarks (GARCH and ARCD).

In Figures A6a and A6b, we plot the series of the conditional skewness and kurtosis for RUT, estimated by the four models and an empirical rolling one against the data.

We also report the violation hits in Table 3.

Figure A4: Conditional density: quantile estimates for RUI
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD

Figure A5: Conditional density: standardised quantile estimates for RUI
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD



(a) Skewnes (b) Kurtosis

Figure A6: Conditional skewness and kurtosis for RUI

Table 3: In-sample hit ratios ($\frac{1}{T}\sum_{t=1}^{T} I(y_t < q_{\theta,t})$) for RUI

θ	**0.99**	**0.95**	**0.50**	**0.05**	**0.01**	**0.25**	**0.75**
DMSQ SAV	**0.9903**	**0.9499**	0.4950	0.0515	0.0118	0.2430	0.7508
DMSQ AS	0.9895	0.9484	0.4917	0.0513	**0.0117**	0.2375	0.7547
ARCD	0.9940	0.9462	0.4931	**0.0496**	0.0080	0.2320	0.7474
GARCH	0.9852	0.9429	**0.5014**	0.0590	0.0128	**0.2508**	0.7479

Note: The sample ranges from January 01, 2002, to December 31, 2014. With bold we indicate the best performing model.

The estimated conditional quantiles and their values standardised by the conditional interquartile range for NDX are presented in Figures A7 and A8 along with the relevant quantiles for the benchmarks (GARCH and ARCD).

In Figures A9a and A9b, we plot the series of the conditional skewness and kurtosis for NDX, estimated by the four models and an empirical rolling one against the data.

We also report the violation hits in Table 4.

Figure A7: Conditional density: quantile estimates for NDX
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD

Figure A8: Conditional density: standardised quantile estimates for NDX
(a) DMSQ-SAV, (b) DMSQ-AS, (c)GARCH, (d)ACRD

(a) Skewnes (b) Kurtosis

Figure A9: Conditional skewness and kurtosis for NDX

Table 4: In-sample hit ratios ($\frac{1}{T}\sum_{t=1}^{T} I(y_t < q_{\theta,t})$) for NDX

θ	**0.99**	**0.95**	**0.50**	**0.05**	**0.01**	**0.25**	**0.75**
DMSQ SAV	0.9907	**0.9502**	0.4901	0.0519	**0.0118**	0.2371	0.7473
DMSQ AS	**0.9900**	**0.9502**	0.4856	0.0518	0.0127	0.2342	0.7460
ARCD	0.9907	0.9454	0.4850	**0.0503**	0.0080	0.2295	0.7477
GARCH	0.9844	0.9417	**0.4980**	0.0644	**0.0118**	**0.2562**	**0.7487**

Note: The sample ranges from January 01, 2002, to December 31, 2014. With bold we indicate the best performing model.