Forecasting Value at Risk and Expected Shortfall using a Model with a Dynamic Omega Ratio

James W. Taylor

Saïd Business School

University of Oxford

25 April 2022

Forthcoming in *Journal of Banking & Finance*

Address for Correspondence:

James W. Taylor
Saïd Business School
University of Oxford
Park End Street
Oxford OX1 1HP, UK
Tel: +44 (0)1865 288927
Email: james.taylor@sbs.ox.ac.uk

Forecasting Value at Risk and Expected Shortfall using a Model with a Dynamic Omega Ratio

Abstract

A joint model for the Value at Risk (VaR) and expected shortfall (ES) can be estimated using a joint scoring function. Previous work has modelled the ES as the product of the VaR and a constant factor. However, this implies the same dynamics for the ES and the VaR. We propose a time-varying multiplicative factor. The ES has been expressed as the product of an expectile and a constant factor that depends on the expectile level. We rewrite this as the product of a quantile and a function of a time-varying expectile level. The expectile level is a function of the Omega ratio, which is the ratio of the expected gain to the expected loss. This leads us to model the ES as the product of the VaR and a factor that is a function of a time-varying Omega ratio. We provide empirical support using stock indices and individual stocks.

JEL classification: C52, C53, C58

Keywords: Expected shortfall; Value at Risk; Expectiles; Autoregressive models; Omega ratio.

1. Introduction

For many years, Value at Risk (VaR) was the established measure of market risk. However, as a conditional tail quantile, the VaR conveys no information regarding potential exceedances of the returns beyond the quantile. Furthermore, the VaR is not *coherent*, meaning that it does not possess a number of attractive properties (Artzner et al., 1999). The Basel Committee on Banking Supervision now recommends the use of the expected shortfall (ES) (Basel Committee, 2016), which is a coherent risk measure (Acerbi, 2002; Acerbi and Tasche, 2002). The ES is defined as the conditional expectation of exceedances beyond the VaR. Although ES is theoretically appealing as a measure of risk, its estimation is typically challenging (Lazar and Zhang, 2019). Indeed, as the ES is defined with respect to the VaR, without reasonable estimation approaches for the ES, the VaR is often easier to estimate, and is therefore a more appealing measure of risk. In this paper, we aim to develop new models for predicting the ES. Although our main aim is to contribute methodologically to the literature on risk management, we have an applied agenda, and so much of the paper is devoted to empirical analysis. In addition to an empirical evaluation of our new models, we also assess the accuracy of a variety of previously proposed models, and so the paper aims to cast light on the relative merits of these different approaches, in terms of both VaR and ES estimation.

A risk measure is *elicitable* if the correct forecast of the measure is the unique minimiser of the expectation of at least one scoring function. If such a scoring function exists, it is called *strictly consistent* for the measure. A strictly consistent scoring function can be used for forecast evaluation and as the loss function in model estimation (Gneiting and Raftery, 2007). Although ES is not elicitable, it has been shown that VaR and ES are jointly elicitable (Fissler and Ziegel, 2016), which enables the estimation of joint models for these two risk measures. Taylor (2019) shows that an example of a joint scoring function is the log likelihood of an asymmetric Laplace distribution, which has been used in the quantile regression literature for statistical inference (see, for example, Komunjer, 2005). Taylor (2019) uses this scoring function to estimate dynamic models for the VaR and ES, and Patton et al. (2019) provide asymptotic results for this proposal. The advantage of directly modelling the VaR and ES is that a distributional assumption is avoided. This is appealing because there is no consensus regarding the distribution of daily returns, and indeed the distribution may vary over time.

In this paper, we develop a new form of dynamic joint model for the VaR and ES. There is an established literature on VaR modelling, with a number of methods available for the direct modelling of the quantile, such as the conditional autoregressive VaR (CAViaR) models of Engle and Manganelli (2004) or the quantile autoregressive models of Koenker and Xiao (2006). In view of this, our emphasis is on modelling the ES. Following Taylor (2019), one simple approach is to treat the ES as a constant multiple of the VaR. Although it seems reasonable that the ES and VaR will, to some extent, vary together, the use of a constant multiplicative factor seems overly restrictive, as it implies that the dynamics of the ES are identical to those of the VaR. As an alternative, Taylor (2019) considers autoregressive models of the difference between the ES and VaR, which allows the dynamics of the ES and VaR to differ. However, this additive autoregressive structure seems inefficient, as the difference between the ES and VaR is likely to be at least partly related to changes in the VaR, with both being driven by changes in the volatility.

In this paper, we model the ES as the product of the VaR and a time-varying factor. Our proposal is based on previous research that shows that if the quantile for a particular probability level is approximated by an expectile with a suitably chosen expectile level, the ES can be expressed as the product of the quantile and a constant factor, which depends on the expectile level (Taylor, 2008). With increasing interest in ES and expectiles over the past decade, this idea has received significant attention in the literature (see, for example, Gerlach and Chen, 2015; Gerlach and Wang, 2022; Jiang et al., 2022; Kim and Lee, 2016). The justification for approximating a quantile by an expectile is that there is a one-to-one mapping between quantiles and expectiles for any distribution (Jones, 1994). However, as the returns distribution will be time-varying, the mapping between expectiles and quantiles will also vary over time. It is, therefore, more reasonable to view the quantile, for a particular probability level, as an expectile with time-varying expectile level. In view of this, we write the ES as the product of a quantile and a factor that is a function of a time-varying expectile level. To model the expectile level, we note that it is a simple function of the Omega ratio, which is the ratio of the expected gain to the expected loss. This leads us to model the ES as the product of the VaR and a factor that is a function of a dynamic Omega ratio, which we model using autoregressive expressions for the gain and loss. Note that, in this paper, the Omega ratio is not used in its traditional way as a measure of portfolio performance (Bernardo and Ledoit, 2000; Bi et al., 2019). Instead, we use it to summarise an important time-varying feature of our proposed model.

Section 2 reviews joint scoring functions for the VaR and ES. Section 3 describes limitations of previously proposed joint dynamic models that have been estimated using joint scoring functions. Section 4 first shows how the ES can be expressed as the product of the VaR and a factor that is a function of a time-varying Omega ratio. We then use this as the basis for our proposed new model. Empirical studies using stock indices and individual stocks are provided in Sections 5 and 6, respectively. Section 7 concludes the paper.

2. Scoring functions for VaR and ES

2.1. Scoring functions for VaR

For a variable y_t, the VaR at level α is defined as $\inf\{y_t \mid F_t(y_t) \geq \alpha\}$ where F_t is the cumulative distribution function of y_t. VaR is, therefore, the quantile with probability level α of F_t. VaR is an elicitable risk measure. Consistent scoring functions are of the form:

$$S\left(q_t(\alpha), y_t\right) = \left(\alpha - I\left(y_t \le q_t(\alpha)\right)\right)\left(H_1\left(y_t\right) - H_1\left(q_t(\alpha)\right)\right)$$

where $q_t(\alpha)$ is the quantile (i.e. the VaR) with probability level α; I is the indicator function; and H_1 is a weakly increasing function. If H_1 is strictly increasing, the scoring function is strictly consistent (Gneiting, 2011). In this paper, we consider daily returns r_t, and adopt the common assumption that their conditional mean is a small constant c, estimated as the mean of the in-sample returns. We define y_t as the residual $y_t = r_t - c$.

The score of expression (1) is typically used for quantiles, due to its simplicity, and its familiarity as the quantile regression loss function. We refer to this as the *quantile score*.

$$S\left(q_t(\alpha), y_t\right) = \left(\alpha - I\left(y_t \le q_t(\alpha)\right)\right)\left(y_t - q_t(\alpha)\right) \tag{1}$$

2.2 Joint scoring functions for VaR and ES

For a variable y_t, the ES at level α is defined as the tail expectation $E(y_t \mid y_t \le q_t(\alpha))$. Although the ES is not elicitable, Fissler and Ziegel (2016) prove it is jointly elicitable with the VaR. They show that expression (2) is a consistent scoring function for the VaR and ES.

$$\begin{aligned} S\left(q_t(\alpha), ES_t(\alpha), y_t\right) = {} & \left(I\left(y_t \le q_t(\alpha)\right) - \alpha\right) H_1\left(q_t(\alpha)\right) - I\left(y_t \le q_t(\alpha)\right) H_1\left(y_t\right) \\ & + H_2\left(ES_t(\alpha)\right)\left(ES_t(\alpha) - q_t(\alpha) + I\left(y_t \le q_t(\alpha)\right)\left(q_t(\alpha) - y_t\right)/\alpha\right) \\ & - \zeta_2\left(ES_t(\alpha)\right) + a\left(y_t\right) \end{aligned} \tag{2}$$

$ES_t(\alpha)$ is the ES for probability level α; and H_1, H_2, ζ_2 and a are functions for which $\zeta_2' = H_2$, H_1 is increasing, and ζ_2 is increasing and convex. The scoring function is strictly consistent if ζ_2 is strictly increasing and strictly convex. The terms involving H_1 constitute a consistent scoring function for a quantile, with the other terms evaluating both quantile and ES accuracy (Fissler et al., 2016).

In our empirical analysis, we use the two versions of the joint score that are considered by Nolde and Ziegel (2017). We present the two scores in Table 1. For each score, ζ_2 is strictly increasing and strictly convex, implying that the scores are strictly consistent.

The *AL score* in Table 1 is so named because it differs by just a constant term from the negative of the log likelihood function of an asymmetric Laplace density with time-varying location and scale parameters. Taylor (2019) points out that using this score for estimation amounts to a simple extension of quantile regression. The AL score is the only version of Fissler and Ziegel's (2016) joint score for which the score differences between competing forecasts is zero-homogeneous (Patton et al., 2019). This means that the score difference is unaffected by multiplying the return and risk measures by a positive value. Patton et al. (2019) explain that zero-homogeneity is an appealing property, which has been shown to have theoretical advantages in the context of volatility forecast evaluation. We refer to the other score in Table 1 as the *NZ score*, as it was proposed by Nolde and Ziegel (2017).

Table 1. Two joint VaR and ES scoring functions from the set of scoring functions in expression (2).

	$H_1(x)$	$H_2(x)$	$\zeta_2(x)$	$a(x)$
AL	0	$-1/x$	$-\ln(-x)$	0
NZ	0	$\frac{1}{2}(-x)^{-\frac{1}{2}}$	$-(-x)^{\frac{1}{2}}$	0

3. Limitations of previously proposed joint dynamic models for VaR and ES

Taylor (2019) uses the AL score to estimate dynamic joint models for the VaR and ES. A CAViaR model is used for the VaR (see Appendix 1). For the ES, Taylor (2019) uses two formulations. The *static multiplicative* ES formulation of expression (3) models the ES as the product of the VaR and a constant factor, ψ_0, which is constrained to be greater than 1.

$$ES_t(\alpha) = \psi_0 \, q_t(\alpha) \tag{3}$$

A limitation of expression (3) is that it does not allow the ES and VaR to evolve with different dynamics. This is addressed by Taylor's (2019) *dynamic additive* ES formulation, which autoregressively models the distance of the ES beyond the VaR. We present it in expressions (4) and (5), where the ω_i are constant non-negative parameters.

$$ES_t(\alpha) = q_t(\alpha) - x_t \tag{4}$$

where $$x_t = \begin{cases} \omega_0 + \omega_1 \left(q_{t-1}(\alpha) - y_{t-1} \right) + \omega_2\, x_{t-1} & if\ \ y_{t-1} \le q_{t-1}(\alpha) \\ x_{t-1} & otherwise \end{cases} \tag{5}$$

A limitation of expressions (4) and (5) is that, for tail quantiles, there will be a small number of exceedances, implying that the formulation will respond only slowly to changing volatility, and certainly more slowly than the CAViaR model used to model the VaR. Furthermore, expressions (4) and (5) seem inefficient, as the difference between the ES and VaR is likely to be at least partly related to changes in the VaR, with changes in the volatility probably being the fundamental driver behind fluctuations in both. The limitations of the static multiplicative and dynamic additive formulations motivate us to develop, in Section 4, a new dynamic multiplicative formulation that expresses the ES as the product of the VaR and a factor that we model autoregressively, with structure based on theoretical reasoning.

4. Modelling ES as the product of VaR and a dynamic factor

In this section, we introduce a formulation that expresses the ES as the product of the VaR and a factor that we model autoregressively. Before describing our proposal, we summarise results from the literature relating expectiles and expected shortfall.

4.1. Expectiles and expected shortfall

An expectile is defined by Newey and Powell (1987) as the solution of an asymmetric least squares minimisation. Just as quantiles generalise the median, expectiles generalise the mean (Nolde and Ziegel, 2017). Expression (6) provides a strictly consistent scoring function for an expectile $e_t(\tau)$ with expectile level τ. We refer to this as the *expectile score*. The general class of strictly consistent scoring functions is provided by Gneiting (2011).

$$S\left(e_t(\tau), y_t\right) = \left|\tau - I\left(y_t \le e_t(\tau)\right)\right| \left(y_t - e_t(\tau)\right)^2 \tag{6}$$

The first order condition for the minimisation of the expected expectile score is:

$$\frac{E\left(I\left(y_t > e_t(\tau)\right)\left|y_t - e_t(\tau)\right|\right)}{E\left(I\left(y_t \le e_t(\tau)\right)\left|y_t - e_t(\tau)\right|\right)} = \frac{1-\tau}{\tau}. \tag{7}$$

This can be rewritten as the following expression (Taylor, 2008):

$$ES_t(\alpha) = \left(1 + \frac{\tau}{\alpha\left(1-2\tau\right)}\right)e_t(\tau) - \frac{\tau}{\alpha\left(1-2\tau\right)}E\left(y_t\right) \tag{8}$$

where α is the probability of exceeding $e_t(\tau)$. Defining y_t to be a zero mean residual term, resulting after the conditional mean is subtracted from the return, expression (8) becomes:

$$ES_t(\alpha) = \left(1 + \frac{\tau}{\alpha\left(1-2\tau\right)}\right)e_t(\tau). \tag{9}$$

As the probability of exceeding $e_t(\tau)$ is α, $e_t(\tau)$ is equal to the quantile with probability level α. Taylor (2008) proposes that expression (9) is used to estimate the ES for probability level α by using the expectile $e_t(\tau)$, with suitably chosen expectile level τ, to approximate the quantile with probability level α. On the face of it, this seems reasonable, because there is a one-to-one mapping between expectiles and quantiles for any distribution (Jones, 1994). However, the approach can be criticised because the conditional distribution of daily returns varies over time, and so the mapping is also likely to vary over time.

In view of this, expression (9) should be rewritten because there is unlikely to be an expectile with constant expectile level τ for which the probability of exceedance is a constant α. Instead, if we wish to keep the constant probability level α in expression (9) we must redefine the expectile as having a time-varying expectile level. In view of this, we rewrite expression (9) as:

$$ES_t(\alpha) = \left(1 + \frac{\tau_t}{\alpha\left(1-2\tau_t\right)}\right)e_t(\tau_t). \tag{10}$$

where α is the probability of exceeding $e_t(\tau_t)$. With $e_t(\tau_t)$ defined in this way, it is equal to the quantile with probability level α, and so we can rewrite expression (10) as:

$$ES_t(\alpha) = \left(1 + \frac{\tau_t}{\alpha\left(1-2\tau_t\right)}\right) q_t(\alpha) . \quad (11)$$

Equating a quantile, with constant probability level, to an expectile, with time-varying expectile level, relates to the work of Schmidt et al. (2021), who view a point forecast as a quantile with dynamic probability level, or expectile with dynamic expectile level.

4.2. A new joint model with ES as the product of VaR and a dynamic factor

Expression (11) is of practical relevance only if we can model the time-varying expectile level τ_t. However, an autoregressive model for τ_t is not immediately apparent, as there is no obvious forcing variable. (We use the term 'forcing variable' for the variable that is lagged and drives changes in an autoregressive model.) To address this, we first turn to Bellini and Di Bernardino (2017), who point out that the left-hand side of expression (7) is the Omega ratio of Keating and Shadwick (2002). This is an extension of the gain/loss ratio originally proposed by Bernardo and Ledoit (2000) as a measure of portfolio performance. The Omega ratio is the ratio of the expected gain to the expected loss, as in expression (12).

$$\Omega\left(e_t(\tau)\right) = \frac{E\left(I\left(y_t > e_t(\tau)\right)\left|y_t - e_t(\tau)\right|\right)}{E\left(I\left(y_t \le e_t(\tau)\right)\left|y_t - e_t(\tau)\right|\right)} . \quad (12)$$

Based on this expression and expression (7), we can write:

$$\Omega\left(e_t(\tau)\right) = \frac{1-\tau}{\tau} .$$

We use this to rewrite expression (11) as expression (13), which shows the ES expressed as the product of the VaR and a factor that is a function of a time-varying Omega ratio Ω_t.

$$ES_t(\alpha) = \left(1 + \frac{1}{\alpha\left(\Omega_t\left(q_t(\alpha)\right) - 1\right)}\right) q_t(\alpha) \quad (13)$$

This is useful because the dynamic Omega ratio $\Omega_t\left(q_t(\alpha)\right)$ is something that we can model. We do this through autoregressions for the gain and loss corresponding to the quantile $q_t(\alpha)$. In view of the expression for the expected gain in the numerator of expression (12), the variable $I\left(y_t > q_t(\alpha)\right)\left|y_t - q_t(\alpha)\right|$ is the natural choice of forcing variable in the autoregression for the gain corresponding to the quantile $q_t(\alpha)$. Similarly, the denominator of expression (12) suggests that the natural choice of forcing variable in the autoregression for the loss is $I\left(y_t \le q_t(\alpha)\right)\left|y_t - q_t(\alpha)\right|$.

We present our proposal in the joint model of expressions (14) to (18), where we use asymmetric slope CAViaR for the VaR, and ES modelled using a dynamic Omega ratio.

$$q_t(\alpha) = \beta_0 + \beta_1 I\left(y_{t-1} > 0\right)\left|y_{t-1}\right| + \beta_2 I\left(y_{t-1} \le 0\right)\left|y_{t-1}\right| + \beta_3\, q_{t-1}(\alpha) \qquad (14)$$

$$ES_t(\alpha) = \left(1 + \frac{1}{\alpha\left(\Omega_t\left(q_t(\alpha)\right) - 1\right)}\right) q_t(\alpha) \qquad (15)$$

where $$\Omega_t\left(q_t(\alpha)\right) = \frac{G_t\left(q_t(\alpha)\right)}{L_t\left(q_t(\alpha)\right)} \qquad (16)$$

$$G_t\left(q_t(\alpha)\right) = \gamma_0 + \gamma_1 I\left(y_{t-1} > q_{t-1}(\alpha)\right)\left|y_{t-1} - q_{t-1}(\alpha)\right| + \gamma_2 G_{t-1}\left(q_{t-1}(\alpha)\right) \qquad (17)$$

$$L_t\left(q_t(\alpha)\right) = \lambda_0 + \lambda_1 I\left(y_{t-1} \le q_{t-1}(\alpha)\right)\left|y_{t-1} - q_{t-1}(\alpha)\right| + \lambda_2 L_{t-1}\left(q_{t-1}(\alpha)\right) \qquad (18)$$

G_t and L_t are the gain and loss. In our implementation of the model, we include the constraints that the constant parameters γ_1, γ_2, λ_1 and λ_2 are non-negative, and that $\gamma_1+\gamma_2<1$ and $\lambda_1+\lambda_2<1$. We set γ_0 to be equal to $(1-\gamma_1-\gamma_2)$ multiplied by the mean of the in-sample values of $I\left(y_t > q_t(\alpha)\right)\left|y_t - q_t(\alpha)\right|$. Similarly, we set λ_0 to be equal to $(1-\lambda_1-\lambda_2)$ multiplied by the mean of the in-sample values of $I\left(y_t \le q_t(\alpha)\right)\left|y_t - q_t(\alpha)\right|$. The static multiplicative ES formulation of expression (3) is a special case of this model, with γ_i and λ_i chosen so that G_t, L_t and Ω_t are constant. In terms of initialisation, we set $q_0(\alpha)$ to be the α quantile of the empirical distribution of the first 300 observations, which was the approach taken by Engle and Manganelli (2004) for their CAViaR models. In our empirical work, there was sometimes no quantile exceedances in the first 300 periods, so we opted to use all in-sample observations to initialise G_0 and L_0. We set these as the mean of the in-sample values of $I\left(y_t > q_t(\alpha)\right)\left|y_t - q_t(\alpha)\right|$ and $I\left(y_t \leq q_t(\alpha)\right)\left|y_t - q_t(\alpha)\right|$, respectively.

The parameters in the VaR and ES parts of the model are estimated jointly by optimising one of Fissler and Ziegel's (2016) joint scoring functions of expression (2). Hypothesis testing of the model parameters can be performed using the bootstrapping approach of Taylor (2019) or, when the AL score is used for estimation, the asymptotic results of Patton et al. (2019).

We should point out that, in its traditional use as a measure of portfolio performance, the Omega ratio is not time-varying and has gain and loss defined with respect to a constant threshold. By contrast, to capture time-variation in the relationship between the ES and the VaR, in our model of expressions (14) to (18), the Omega ratio is time-varying due to the gain and loss being modelled autoregressively. Allowing time-variation in the gain and loss is an important aspect of the model, with the dynamic Omega ratio simply being a convenient way to summarise this. Indeed, it is worth noting that the Omega ratio is not in itself of great interest, and that substituting expression (16) into expression (15) removes the Omega ratio from the formulation.

5. Empirical analysis with stock indices

In this section, we compare the accuracy of day-ahead VaR and ES forecasts produced by our proposed approach and a set of benchmark methods. Our empirical study considered the following three probability levels: 1%, 2.5% and 5%. The levels 1% and 5% have been very widely considered in studies of VaR estimation, and 2.5% has relatively recently been proposed, particularly when estimating the ES (Basel Committee, 2016).

We used daily log returns for the following five stock indices: CAC 40, DAX 30, FTSE 100, NIKKEI 225 and S&P 500. Each series consisted of the 4000 daily observations ending on 29 June 2018. Due to different holiday periods in each country, the start dates differed for the five indices. The start dates were 14 November 2002, 2 October 2002, 3 September 2002, 14 March 2002, and 12 August 2002 for the CAC 40, DAX 30, FTSE 100, NIKKEI 225 and S&P 500, respectively. In our forecasting study, we used a rolling window of 2,000 observations for the repeated re-estimation of parameters for a variety of methods. This delivered out-of-sample forecasts for the final 2,000 periods of each series. We used these forecasts to compare the accuracy of the various methods. As we stated in Section 2.1, our modelling focused on the variable $y_t = r_t - c$, where r_t is the daily log return and c is a constant. We estimated c using the mean of each rolling window of 2,000 in-sample returns.

5.1. Benchmark methods

We implemented a variety of time series methods for predicting the VaR and ES. These methods can be classed as nonparametric, parametric and semiparametric. As a simple nonparametric benchmark, we used historical simulation based on the most recent 250 observations. The quantile of the empirical distribution of the 250 values was used as the VaR forecast for the next day, and the ES forecast was computed as the mean of the values exceeding the VaR forecast within the sample of 250.

In terms of parametric methods, we fitted the GARCH(1,1) and asymmetric GJR-GARCH(1,1) models using maximum likelihood based on the Student *t* distribution. For each fitted model, we produced VaR and ES forecasts using three different approaches. Our first approach simply used the fitted Student *t* distribution. The second approach constructed VaR and ES forecasts as the product of the volatility forecast and the corresponding VaR or ES forecast of the empirical distribution of in-sample residuals y_t standardised by the estimated volatility (see Christoffersen, 2012, Section 6.4). We term this the *filtered* approach. The third approach was the method of McNeil and Frey (2000), which applies peaks-over-threshold extreme value theory (EVT) to the standardised residuals. The filtered and EVT approaches can be classed as semiparametric methods, as they involve a model for the volatility, but avoid an assumption for the type of conditional distribution. The remaining methods that we describe in this section, and also the methods in Section 5.2, are also classed as semiparametric methods because they also involve model formulations, but avoid a distributional assumption.

We implemented the method of Taylor (2008), which uses expression (9) to estimate the ES for probability level α with the expectile $e_t(\tau)$ used to approximate the quantile with probability level α. Taylor (2008) introduces conditional autoregressive expectile (CARE) models for this purpose. The value of τ is chosen so that the proportion of in-sample observations that exceed the fitted expectile values is close to α, the probability level. To optimise τ, a CARE model is repeatedly re-estimated, reducing the τ by 0.0001 each time, until the proportion of in-sample exceedances beyond the fitted expectile is closer to α than a predefined tolerance. We initialised the procedure with values of τ=0.0018, τ=0.0055, and τ=0.0167 for the 1%, 2.5% and 5% probability levels, respectively. These values were chosen after initial experimentation. In our implementation of the approach, we used the symmetric absolute value CARE and asymmetric slope CARE models, which are presented in expressions (19) and (20), respectively.

$$e_t(\tau) = \phi_0 + \phi_1 |y_{t-1}| + \phi_2 e_{t-1}(\tau) \tag{19}$$

$$e_t(\tau) = \phi_0 + \phi_1 I\left(y_{t-1} > 0\right)|y_{t-1}| + \phi_2 I\left(y_{t-1} \le 0\right)|y_{t-1}| + \phi_3\, e_{t-1}(\tau) \tag{20}$$

The ϕ_i are constant parameters, optimised using the procedure of Taylor (2008), which was closely based on the approach of Engle and Manganelli (2004) for CAViaR models. First, 10^4

candidate parameter vectors are sampled from uniform distributions with lower and upper bounds based on initial experimentation. The optimised parameter vector from the previous window of observations was also included as a candidate vector. Of all the candidate vectors, the ten giving the lowest value of the expectile score of expression (6) were then used in turn as the initial parameter vectors in a quasi-Newton algorithm. The resulting parameter vector with lowest expectile score was chosen as the final parameter vector.

Patton et al. (2019) propose joint dynamic models for the VaR and ES, based on the generalised autoregressive score (GAS) models of Creal et al. (2013) and Harvey (2013), which are autoregressive with forcing variable specified as the lagged score of the log-likelihood. The negative of the AL scoring function is used as a quasi-log-likelihood. In their work, the most accurate of their models was the following *one-factor GAS model*:

$$q_t(\alpha) = a\exp(\kappa_t) \tag{21}$$

$$ES_t(\alpha) = b\exp(\kappa_t), \quad \text{where } b < a < 0 \tag{22}$$

and $$\kappa_t = \beta\kappa_{t-1} + \gamma\frac{-1}{b\exp(\kappa_{t-1})}\left(\frac{1}{\alpha}I\left(y_{t-1} \le a\exp(\kappa_{t-1})\right)y_{t-1} - b\exp(\kappa_{t-1})\right). \tag{23}$$

where a, b, β and γ are constant parameters. In this model, the ES is a constant multiple of the VaR, and so it can be viewed as using a static multiplicative formulation for the ES, as in expression (3). We fitted the model of expressions (21) to (23) to each of the five stock indices. In Appendix C of their paper, Patton et al. (2019) describe a parameter estimation approach that aims to overcome sensitivity to initial parameter values. Their approach involves initially estimating models using a "smoothed" form of the AL score that avoids the use of the indicator function in the score and in the model expressions. We did not find this approach to be useful in our empirical work, which differs from the study of Patton et al. (2019), in that we repeatedly re-estimate parameters, while they estimate the model just once for each stock index. To estimate the one-factor GAS model, we employed the same approach used by Taylor (2019) for joint VaR and ES models. This is the same as the parameter optimisation procedure that we described for CARE models in the previous paragraph, with the expectile score replaced by the AL joint score.

5.2. Joint models with CAViaR for VaR and four specifications for ES

We initially implemented six different joint models for VaR and ES. The different model specifications involved either the symmetric or asymmetric CAViaR models of Appendix 1, with one of the following three different specifications for the ES: the static multiplicative formulation of expression (3), the dynamic additive formulation of expressions (4) and (5), and our new dynamic Omega formulation of expressions (15) to (18).

We estimated the six joint VaR and ES models using the AL and NZ scores. We used the same estimation approach as Taylor (2019). This is similar to the parameter optimisation procedure that we described for CARE models in Section 5.1 with two notable differences. First, the expectile score was replaced by either the AL or NZ joint scores. Second, for the 10^4 candidate parameter vectors, we set the parameters in the CAViaR component of the joint model to be the values optimised separately by minimising the quantile score, while the other model parameters were randomly sampled. We did this to assist the optimization, and we found it to be preferable to the alternative of using a much larger number of candidate parameter vectors, each consisting of randomly sampled values for all the model parameters. In using, as starting values, CAViaR model parameters, estimated separately by minimising the quantile score, we have followed the approach of White et al. (2015) for their multi-equation models. Note, however, that this approach was used only to get starting parameters, with the final optimised parameters estimated jointly by minimising the AL or NZ score.

We now graphically present estimates from the joint model with asymmetric slope CAViaR for the VaR and our proposed new dynamic Omega formulation for the ES, with parameters estimated by minimising the in-sample AL score for the 2.5% probability level of the S&P 500 returns. This is the model presented in expressions (14) to (18). For the out-of-sample period, Figure 1 plots the minimised in-sample AL score values, and the parameters of the gain and loss expressions in the ES part of the model. The minimised AL score values change relatively smoothly over time, giving some reassurance that the minimisation is reasonably stable. In a GARCH volatility model, the coefficient of the forcing variable is typically quite small, while the autoregressive coefficient is quite close to 1. Figure 1 shows that the same is true for the forcing variable coefficients, γ_1 and λ_1, and the autoregressive coefficients, γ_2 and λ_2, of expressions (17) and (18). γ_1 is larger than λ_1 indicating that the gain is more rapidly evolving than the loss, which is perhaps due to the loss being more challenging to model using an autoregressive formulation, as the forcing variable for the loss relies on quantile exceedances, which occur just 2.5% of the time for this model (because the probability level has been chosen to be 2.5%).

Figure 2 shows the estimated values of the gain and loss. The ratio of expected gain to expected loss is the Omega ratio, and this is plotted in Figure 3, along with the factor that is multiplied by the VaR forecast to give the ES forecast. The formula for this dynamic multiplicative factor is given within the large parentheses of expression (15). Although, as shown in Figure 2, the estimated loss was always substantially smaller than the gain, we found that very occasionally the loss would be relatively large, leading to a relatively small Omega ratio, with the result that the multiplicative factor became unreasonably large. This occurs in two periods in Figure 3, and each can be attributed to sudden large negative values of the return in the previous period (-6.9% on 17 August 2011 and -4.2% on 2 February 2018), which caused the loss to spike in Figure 2. Although we found that the very large values for the multiplicative factor had negligible effect on the out-of-sample evaluation summary results, we felt it was unappealing, and so we took the pragmatic step of including a cap on the multiplicative factor. For the 1% and 2.5% probability levels, we set the cap as 2, and for the 5% probability level, we used a cap of 3. Each of these values was the smallest integer that exceeded the largest of the ratios of ES to VaR for any other method in our empirical analysis. Figure 4 plots the resulting capped multiplicative factor, along with the returns and the VaR and ES out-of-sample forecasts, which can be seen to vary with the volatility in the series. Our decision to cap the multiplicative factor amounts to an acknowledgement that the factor is exposed to the instability inherent in the Omega ratio due to occasional extreme values for the loss. This links to the concerns expressed about the Omega ratio as a measure of portfolio performance (see, for example, Caporin et al., 2018).

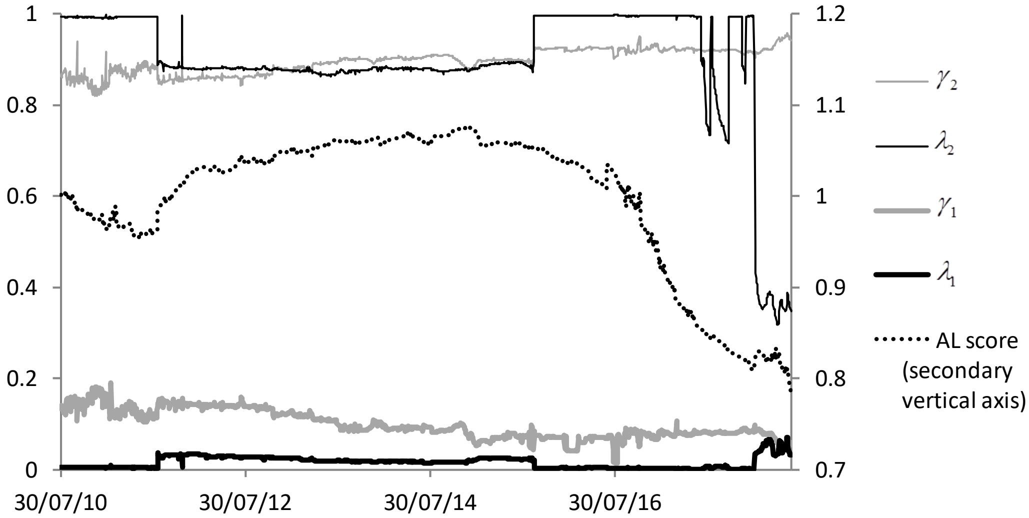

Figure 1. For the S&P 500 and 2.5% probability level, parameters and in-sample AL score from asymmetric slope CAViaR with dynamic Omega for ES, estimated by minimising the AL score.

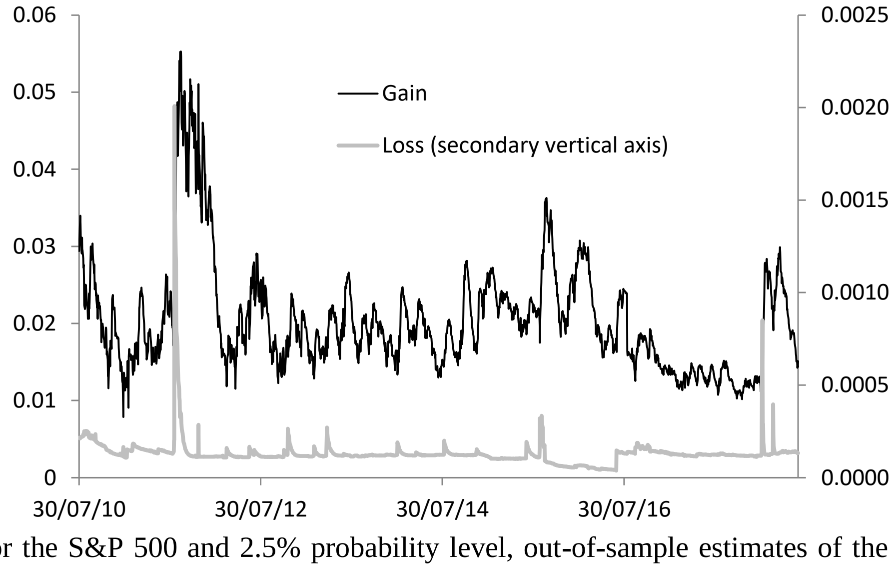

Figure 2. For the S&P 500 and 2.5% probability level, out-of-sample estimates of the gain and loss from asymmetric slope CAViaR with dynamic Omega for ES, estimated by minimising the AL score.

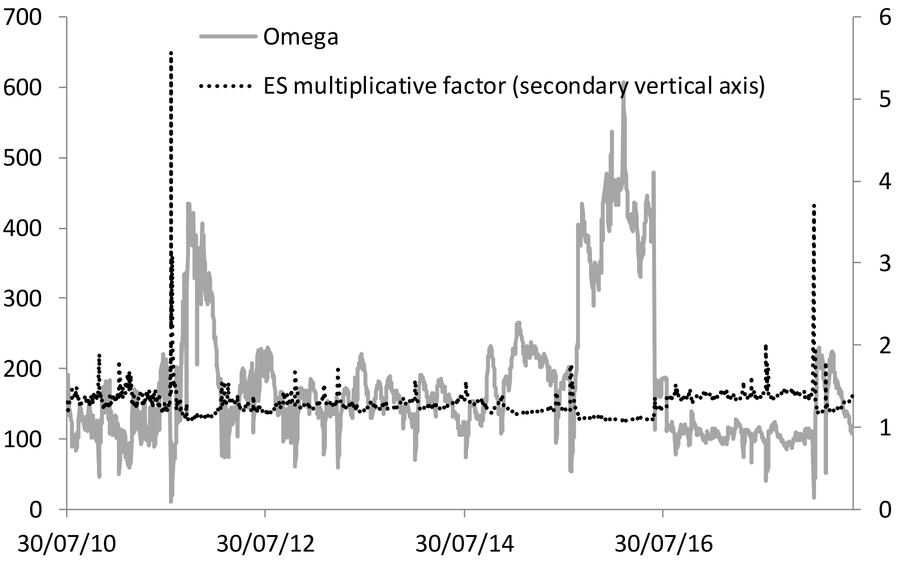

Figure 3. For the S&P 500 and 2.5% probability level, out-of-sample estimates of the Omega ratio and ES multiplicative factor from asymmetric slope CAViaR with dynamic Omega for ES, estimated by minimising the AL score.

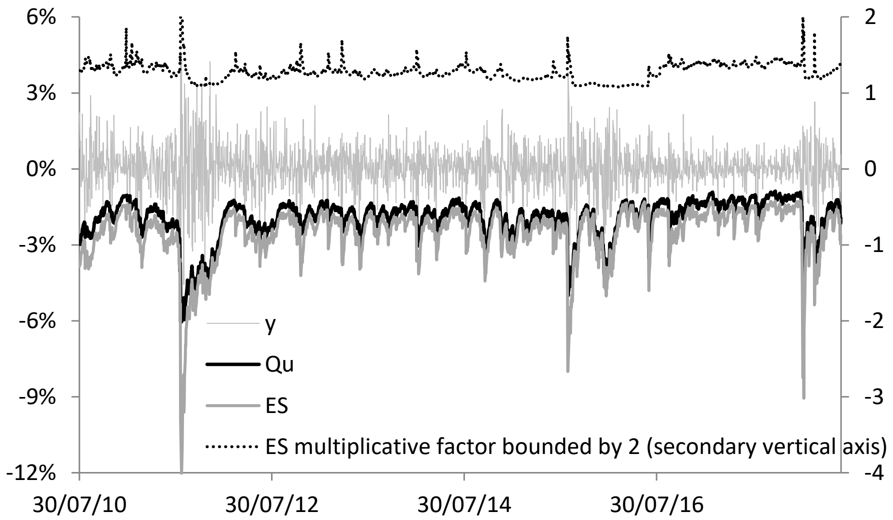

Figure 4. For the S&P 500 and 2.5% probability level, capped ES multiplicative factor and out-of-sample VaR and ES forecasts from asymmetric slope CAViaR with dynamic Omega for ES, estimated by minimising the AL score.

In the dynamic Omega formulation, the loss is challenging to model autoregressively because it relies on a relatively small number of quantile exceedances. The simple alternative is to assume the loss is constant, with one appeal of this being that it avoids the need to cap the multiplicative factor. To test whether the loss should be treated as constant requires a test of the significance of the parameter λ_1, which is the coefficient of the forcing variable in the autoregressive model for the loss. The low estimates for λ_1 in Figure 1 also raise the issue of significance testing. For the test, we produced bootstrapped standard errors using the first estimation window of 2,000 observations for each of the five stock indices (see Taylor, 2019). We applied the test to the models involving the dynamic Omega ratio with either symmetric or asymmetric CAViaR, and estimation based on the AL and NZ scores. As CAViaR models are well established, our interest here is in the significance of the parameters in the equations of expressions (17) and (18), which drive the dynamic omega ratio.

In Table 2, we summarise significance of γ_1 and λ_1, which are the coefficients of the forcing variables in these two equations. The table shows the number of indices for which γ_1 and λ_1 were significant in the different versions of the model. As we are considering five indices, the maximum value for any entry in the table is 5. We have three comments on the results. Firstly, the results are reasonably similar for the two different approaches to estimation (minimising the AL and NZ scores). Secondly, the parameters were significant more often for the symmetric CAViaR model. Thirdly, γ_1 was significant notably more often than λ_1. Indeed, for the formulation involving asymmetric CAViaR, the loss equation forcing parameter λ_1 was generally not significant. Although this may not be detrimental to the model, it seems worth considering an additional version of the dynamic Omega formulation in which the autoregressive model of the loss in expression (18) is replaced by a constant value, while the autoregressive model for the gain in expression (17) is kept in the formulation. We implemented this as a fourth form of joint model for the VaR and ES. In the tables of Section 5.3 in which we report out-of-sample results, we include the text "cst loss" to differentiate this version of the dynamic Omega formulation from the full version presented in expressions (15) to (18). Note that, even if the loss is treated as constant, expression (15) remains a multiplicative formulation in which the ES is expressed as the product of the VaR and a dynamic factor. With constant loss, this factor is modelled as time-varying due to the autoregressive modelling of the gain in expression (17). For tail quantiles, there are few quantile exceedances, implying that the gain, and hence the dynamic factor, is largely dictated by the absolute differences between the observation and the quantile, which will be related to the volatility. This is actually evident in Figure 2, which shows the gain varying over time in a similar fashion to the volatility in the returns in Figure 4.

Table 2. The number of indices for which the parameters γ_1 and λ_1 were significant in the dynamic Omega formulation of expressions (15) to (18), fitted to the first estimation window of 2,000 observations. Significance testing is based on bootstrapped standard errors. The four models correspond to two different forms of CAViaR, and estimation based on two different scores.

	γ_1				λ_1			
	1%	2.5%	5%	Mean	1%	2.5%	5%	Mean
AL score & symmetric CAViaR	3	3	3	3.00	2	3	4	3.00
AL score & asymmetric CAViaR	2	1	1	1.33	0	0	0	0.00
NZ score & symmetric CAViaR	3	4	4	3.67	0	1	5	2.00
NZ score & asymmetric CAViaR	1	2	1	1.33	0	0	2	0.67
Mean	2.25	2.5	2.25		0.5	1	2.75	

5.3. Out-of-sample results

We now describe the results of calibration tests and scoring functions that we used to evaluate the 2,000 out-of-sample VaR and ES forecasts for the five indices. Our discussion focuses on Tables 3 to 5, which summarise the results for the 1%, 2.5% and 5% probability levels, respectively. In each column of the tables, the best result is indicated in bold.

If a quantile forecast $\hat{q}_t(\alpha)$ is unconditionally calibrated, the percentage of VaR exceedances will be equal to the probability level α. The first column of values in Tables 3 to 5 presents this 'VaR violation percentage' averaged across the five indices. To test unconditional calibration, we investigated whether the variable $Hit_t = \alpha - I(y_t \leq \hat{q}_t(\alpha))$ has zero unconditional expectation. We tested this by applying a standard test for a proportion to the mean of the Hit_t variable. In Tables 3 to 5, the second column of values show the number of indices for which unconditional calibration was rejected for each method at the 5%

significance level. The only notably poor unconditional calibration results were for historical simulation for the 1% probability level, and asymmetric CARE for the 5% probability level.

We tested for quantile conditional calibration using Engle and Manganelli's (2004) dynamic quantile (DQ) test, which tests the variable Hit_t for zero conditional expectation. We included four lags in the test's regression. In Tables 3 to 5, the third column of values shows the number of indices for which conditional calibration was rejected by this test, using a 5% significance level. For all three probability levels, historical simulation performed poorly, and conditional calibration tended to be better for the asymmetric GARCH and CAViaR models than for the symmetric versions of these models. The benefit from using asymmetric models is often reported in the literature (see, for example, Ning et al., 2015). We also assessed quantile conditional calibration using the VQR test of Gaglianone et al. (2011), which tests for zero intercept and unit coefficient in a quantile regression of the returns on the quantile forecast. For this test, the number of test rejections are presented in the fourth column of values in Tables 3 to 5. The only clear conclusion from these values is that historical simulation performed poorly.

To evaluate calibration of the ES forecasts, we used the bootstrap test of McNeil and Frey (2000), which focuses on the discrepancy between a VaR exceedance and the ES forecast. We standardised by dividing each discrepancy by the VaR forecast. The standardised discrepancies were tested for zero mean. In Tables 3 to 5, the fifth column of values shows the number of indices for which calibration was rejected at the 5% significance level. The results show that for all methods involving asymmetric models, calibration was rejected for none or just one of the five indices.

We used the AL and NZ scores of Table 1 for the joint evaluation of the out-of-sample VaR and ES forecasts. To summarise performance across the five stock indices, we calculated an overall skill score, which assessed performance relative to historical simulation. To obtain the skill score for each method, we calculated the geometric mean of the ratios of the (arithmetic) mean score for the method to the (arithmetic) mean score for historical simulation, then subtracted this from 1, and multiplied the result by 100. The AL and NZ skill scores are presented in the final two columns in Tables 3 to 5. Note that higher values of the skill scores are better. In addition to using bold to indicate the best performing method in each column, we use underlining to highlight the best performing ES formulation within each group of four joint models estimated using the AL or NZ scores in the bottom 16 rows of each table. The underlining is of interest because our focus in this paper has been to propose the dynamic Omega formulation for ES as an alternative to the static multiplicative and dynamic additive ES formulations previously proposed.

In Table 3, the bold indicates that, overall, the best AL and NZ skill scores for the 1% probability level correspond to the use of asymmetric CAViaR for the VaR with the dynamic Omega modelling of the ES, with parameters estimated by minimising either the AL or NZ scores. There was no clear superiority between the version with the loss modelled autoregressively and with the loss assumed constant. The same comment can be made for the 2.5% and 5% probability levels in Tables 4 and 5. In these two tables, we see support for the two dynamic Omega ES formulations, regardless of whether the symmetric or asymmetric CAViaR models were used for the VaR, but particularly when the symmetric CAViaR model was used. These tables show that, overall, for the 2.5% and 5% probability levels, the best AL and NZ skill scores were produced by the joint VaR and ES models based on asymmetric CAViaR for the VaR with either of the two dynamic Omega formulations for the ES, estimated using either the AL or NZ scores.

The joint VaR and ES score results in Tables 3 to 5 for the one-factor GAS model are poor. We feel that this is due to the choice of the forcing variable. Expression (23) shows that this forcing variable is such that the dynamics of the model are affected by the magnitude of the observations y_{t-1} only when they exceed the VaR, which limits the ability of the model to capture time-varying volatility. This is reminiscent of the adaptive CAViaR model of Engle and Manganelli (2004), which is notably less accurate than the other CAViaR models. The similarity of the GAS model with the adaptive CAViaR model is evident in Figure 5 of Patton et al. (2019), which shows that the GAS model forecasts form a sequence of monotonic rising functions between periods when the observation exceeds the VaR. This behaviour is also found in the quantile forecasts from the adaptive CAViaR model.

For each of the five stock indices, Table 6 presents the AL scores for the out-of-sample period for the 2.5% probability level. Lower values are better. The best scores are for the asymmetric models, which is also the case for the corresponding averaged skill score values presented in the penultimate columns of Tables 3 to 5. For the 2,000 out-of-sample periods for the S&P 500, Figure 1 shows the minimised in-sample AL scores for the model that used asymmetric CAViaR for the VaR and dynamic Omega modelling of the ES. The final value in the plot is approximately 0.8, and this corresponds to the parameter optimisation being performed on the moving window of 2,000 observations that almost entirely coincides with the out-of-sample period. From Table 6, we can see that the out-of-sample value was 0.827 for this model applied to the S&P 500, which is only a little higher than the in-sample minimised value.

Although we have highlighted differences among the AL and NZ scores and skill scores, in many cases the differences are quite small, which motivates consideration of significance testing. The model confidence set (MCS) testing framework of Hansen et al. (2011) enables a set of models to be obtained for which there is a pre-specified probability that the set contains the best model, when judged by a chosen loss function. If a model is not contained in the MCS, it is considered less likely to be the best model than those that are included in the MCS. We implemented MCS testing separately based on the AL and NZ joint scores. In each MCS test, we used the *equivalence* test based on the Diebold-Mariano test, and the one-sided *elimination* rule described as $T_{max,M}$ by Hansen et al. (2011). We followed Hansen et al. (2011) by considering 75% and 90% confidence levels. For conciseness, we report the results for only the 75% confidence level, as they show greater differences between the methods. For each of the three probability levels (1%, 2.5% and 5%), and for each scoring function (AL and NZ score), Table 7 reports the number of indices for which each method was included in the MCS. As we have five indices in our study, the best possible value in each table is 5. Historical simulation and the GAS model were clearly the worst methods. For the other methods, the table shows that it is better to use the asymmetric versions of the GARCH, CARE and CAViaR models, although this is more clearly the case for the 5% probability level than the 1% and 2.5% probability levels. The final 16 rows of values in Table 7 do not show clear preference for any one of the four ES formulations (static multiplicative, dynamic additive, dynamic Omega or dynamic Omega with constant loss). Nevertheless, it is interesting to see that joint models, with asymmetric CAViaR for the VaR and any of the four formulations for the ES, were included in the MCS for all five indices, for all three probability levels, and for estimation based on either the AL or NZ score.

Table 3. Summary of results for 1% VaR and ES estimated for the five stock indices. The first column of values is the percentage of VaR exceedances averaged across the indices; the ideal would be 1%. The next four columns present the number of indices for which calibration was rejected; lower values are better. The final two columns present skill scores; higher values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	VaR violation %	VaR hit % test	VaR DQ test	VaR VQR test	ES bootstrap test	VaR & ES AL skill score	VaR & ES NZ skill score
Historical simulation and GAS							
Historical simulation	1.6	4	5	5	2	0.0	0.0
1-factor GAS	1.1	**0**	2	1	1	10.5	5.9
Symmetric GARCH and CARE							
GARCH-t	1.3	1	3	2	1	17.5	10.3
GARCH-filtered	**1.0**	**0**	4	2	**0**	17.8	10.4
GARCH-EVT	0.8	**0**	3	3	**0**	17.8	10.3
Symmetric-CARE	0.9	**0**	1	2	**0**	17.9	10.5
Asymmetric GARCH and CARE							
GJR-GARCH-t	1.2	**0**	1	3	1	18.7	11.3
GJR-GARCH-filtered	**1.0**	**0**	**0**	2	**0**	18.6	11.2
GJR-GARCH-EVT	0.8	**0**	1	2	1	18.5	10.9
Asymmetric-CARE	0.8	**0**	2	**0**	**0**	18.0	10.8
AL score for estimation with symmetric CAViaR							
Static multiplicative for ES	**1.0**	**0**	2	1	**0**	17.5	10.3
Dynamic additive for ES	**1.0**	**0**	2	2	**0**	<u>17.9</u>	<u>10.5</u>
Dynamic Ω for ES	**1.0**	**0**	2	1	**0**	17.6	10.2
Dynamic Ω for ES: cst loss	**1.0**	**0**	2	1	**0**	17.7	10.3
AL score for estimation with asymmetric CAViaR							
Static multiplicative for ES	0.9	**0**	1	**0**	**0**	19.1	11.5
Dynamic additive for ES	0.9	**0**	1	**0**	**0**	19.2	11.5
Dynamic Ω for ES	**1.0**	**0**	1	**0**	**0**	**<u>19.6</u>**	**<u>11.7</u>**
Dynamic Ω for ES: cst loss	0.9	**0**	1	1	**0**	19.4	11.6
NZ score for estimation with symmetric CAViaR							
Static multiplicative for ES	**1.0**	**0**	2	1	**0**	17.7	10.4
Dynamic additive for ES	**1.0**	**0**	3	**0**	**0**	17.8	10.4
Dynamic Ω for ES	**1.0**	**0**	2	1	**0**	<u>18.0</u>	10.4
Dynamic Ω for ES: cst loss	**1.0**	**0**	2	3	**0**	<u>18.0</u>	<u>10.5</u>
NZ score for estimation with asymmetric CAViaR							
Static multiplicative for ES	0.9	**0**	1	1	**0**	19.2	11.6
Dynamic additive for ES	0.9	**0**	1	**0**	**0**	19.2	11.5
Dynamic Ω for ES	**1.0**	**0**	1	**0**	**0**	19.5	**<u>11.7</u>**
Dynamic Ω for ES cst loss	**1.0**	**0**	1	**0**	**0**	**<u>19.6</u>**	**<u>11.7</u>**

Notes: Bold indicates the best performing method in each column. For each group of four models, estimated using either the AL or NZ score, underlining indicates the best skill score value.

Table 4 Summary of results for 2.5% VaR and ES estimated for the five stock indices. The first column of values is the percentage of VaR exceedances averaged across the indices; the ideal would be 2.5%. The next four columns present the number of indices for which calibration was rejected; lower values are better. The final two columns present skill scores; higher values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	VaR violation %	VaR hit % test	VaR DQ test	VaR VQR test	ES bootstrap test	VaR & ES AL skill score	VaR & ES NZ skill score
Historical simulation and GAS							
Historical simulation	2.9	**0**	5	3	3	0.0	0.0
1-factor GAS	2.2	**0**	2	**0**	**0**	10.8	5.1
Symmetric GARCH and CARE							
GARCH-t	3.0	2	1	**0**	1	13.9	6.7
GARCH-filtered	**2.3**	**0**	1	1	1	14.3	6.9
GARCH-EVT	**2.3**	**0**	2	**0**	**0**	14.4	6.9
Symmetric-CARE	2.1	1	1	1	2	14.3	6.8
Asymmetric GARCH and CARE							
GJR-GARCH-t	3.0	1	**0**	**0**	1	16.1	8.0
GJR-GARCH-filtered	**2.3**	**0**	**0**	**0**	**0**	16.2	7.9
GJR-GARCH-EVT	**2.3**	**0**	**0**	1	**0**	16.2	7.9
Asymmetric-CARE	2.0	1	1	**0**	**0**	16.4	**8.1**
AL score for estimation with symmetric CAViaR							
Static multiplicative for ES	2.2	1	1	**0**	2	14.1	6.8
Dynamic additive for ES	2.2	**0**	1	**0**	**0**	14.0	6.7
Dynamic Ω for ES	**2.3**	**0**	2	2	2	<u>14.5</u>	6.8
Dynamic Ω for ES: cst loss	**2.3**	**0**	2	**0**	1	14.4	<u>6.9</u>
AL score for estimation with asymmetric CAViaR							
Static multiplicative for ES	2.1	1	1	**0**	1	16.3	8.0
Dynamic additive for ES	2.2	**0**	1	**0**	**0**	**<u>16.5</u>**	**<u>8.1</u>**
Dynamic Ω for ES	2.2	**0**	1	1	**0**	16.4	**<u>8.1</u>**
Dynamic Ω for ES: cst loss	2.2	**0**	1	**0**	**0**	**<u>16.5</u>**	**<u>8.1</u>**
NZ score for estimation with symmetric CAViaR							
Static multiplicative for ES	2.2	1	3	**0**	2	14.0	6.7
Dynamic additive for ES	2.2	**0**	2	**0**	**0**	14.0	6.7
Dynamic Ω for ES	2.2	**0**	2	**0**	**0**	<u>14.6</u>	<u>6.9</u>
Dynamic Ω for ES: cst loss	2.2	**0**	2	1	**0**	14.4	<u>6.9</u>
NZ score for estimation with asymmetric CAViaR							
Static multiplicative for ES	2.1	1	1	**0**	1	16.2	8.0
Dynamic additive for ES	2.2	**0**	1	**0**	**0**	16.3	8.0
Dynamic Ω for ES	2.1	**0**	1	**0**	**0**	16.4	8.0
Dynamic Ω for ES: cst loss	2.1	**0**	1	**0**	**0**	**<u>16.5</u>**	**<u>8.1</u>**

Notes: Bold indicates the best performing method in each column. For each group of four models, estimated using either the AL or NZ score, underlining indicates the best skill score value.

Table 5. Summary of results for 5% VaR and ES estimated for the five stock indices. The first column of values is the percentage of VaR exceedances averaged across the indices; the ideal would be 5%. The next four columns present the number of indices for which calibration was rejected; lower values are better. The final two columns present skill scores; higher values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	VaR violation %	VaR hit % test	VaR DQ test	VaR VQR test	ES bootstrap test	VaR & ES AL skill score	VaR & ES NZ skill score
Historical simulation and GAS							
Historical simulation	**5.2**	**0**	5	3	1	0.0	0.0
1-factor GAS	4.4	1	2	1	1	9.8	3.8
Symmetric GARCH and CARE							
GARCH-t	5.5	1	1	**0**	1	12.3	4.7
GARCH-filtered	4.3	1	3	**0**	1	12.2	4.6
GARCH-EVT	4.5	1	3	**0**	**0**	12.2	4.6
Symmetric-CARE	4.2	2	1	**0**	**0**	12.1	4.6
Asymmetric GARCH and CARE							
GJR-GARCH-t	5.4	**0**	**0**	**0**	1	15.1	**5.9**
GJR-GARCH-filtered	4.2	1	1	1	1	14.8	5.7
GJR-GARCH-EVT	4.5	**0**	**0**	**0**	1	15.0	5.8
Asymmetric-CARE	4.0	3	1	2	1	14.8	5.8
AL score for estimation with symmetric CAViaR							
Static multiplicative for ES	4.4	2	2	1	1	12.1	4.5
Dynamic additive for ES	4.4	1	1	1	**0**	11.9	4.5
Dynamic Ω for ES	4.6	1	3	**0**	3	<u>12.9</u>	<u>4.8</u>
Dynamic Ω for ES: cst loss	4.6	**0**	2	**0**	3	12.6	4.7
AL score for estimation with asymmetric CAViaR							
Static multiplicative for ES	4.1	1	1	1	1	15.1	<u>**5.9**</u>
Dynamic additive for ES	4.2	1	1	1	**0**	<u>**15.2**</u>	<u>**5.9**</u>
Dynamic Ω for ES	4.2	1	2	1	**0**	15.1	5.8
Dynamic Ω for ES: cst loss	4.1	1	2	1	**0**	15.1	<u>**5.9**</u>
NZ score for estimation with symmetric CAViaR							
Static multiplicative for ES	4.4	1	2	1	**0**	12.1	4.5
Dynamic additive for ES	4.4	**0**	2	**0**	**0**	12.0	4.5
Dynamic Ω for ES	4.5	**0**	3	**0**	2	<u>12.8</u>	<u>4.7</u>
Dynamic Ω for ES: cst loss	4.5	**0**	1	1	3	12.6	<u>4.7</u>
NZ score for estimation with asymmetric CAViaR							
Static multiplicative for ES	4.1	2	1	1	1	15.0	<u>**5.9**</u>
Dynamic additive for ES	4.2	1	1	1	**0**	15.0	<u>**5.9**</u>
Dynamic Ω for ES	4.1	1	1	1	**0**	<u>15.1</u>	<u>**5.9**</u>
Dynamic Ω for ES: cst loss	4.1	1	1	1	**0**	<u>15.1</u>	<u>**5.9**</u>

Notes: Bold indicates the best performing method in each column. For each group of four models, estimated using either the AL or NZ score, underlining indicates the best skill score value.

Table 6. AL score for 2.5% VaR and ES for each of the five stock indices. Lower values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	CAC 40	DAX 30	FTSE 100	NIKKEI 225	S&P 500
Historical simulation and GAS					
Historical simulation	1.266	1.256	0.994	1.457	1.067
1-factor GAS	1.173	1.135	0.871	1.333	0.892
Symmetric GARCH and CARE					
GARCH-t	1.115	1.088	0.848	1.283	0.876
GARCH-filtered	1.115	1.080	0.843	1.278	0.872
GARCH-EVT	1.115	1.082	0.841	1.278	0.871
Symmetric-CARE	1.113	1.080	0.839	1.285	0.875
Asymmetric GARCH and CARE					
GJR-GARCH-t	1.095	1.078	0.820	1.267	0.827
GJR-GARCH-filtered	1.095	1.077	0.819	1.263	0.827
GJR-GARCH-EVT	1.099	1.079	0.818	1.262	**0.822**
Asymmetric-CARE	1.091	**1.061**	0.818	**1.256**	0.838
AL score for estimation with symmetric CAViaR					
Static multiplicative for ES	1.117	<u>1.080</u>	0.843	1.279	0.884
Dynamic additive for ES	1.115	1.082	0.844	1.282	0.886
Dynamic Ω for ES	1.114	1.085	<u>0.835</u>	<u>1.269</u>	<u>0.874</u>
Dynamic Ω for ES: cst loss	<u>1.109</u>	<u>1.080</u>	0.840	<u>1.269</u>	0.882
AL score for estimation with asymmetric CAViaR					
Static multiplicative for ES	1.096	<u>1.064</u>	0.816	1.272	0.825
Dynamic additive for ES	<u>1.088</u>	<u>1.064</u>	<u>**0.815**</u>	<u>1.268</u>	0.827
Dynamic Ω for ES	1.092	1.065	0.816	<u>1.268</u>	0.827
Dynamic Ω for ES: cst loss	1.092	1.065	<u>**0.815**</u>	1.269	<u>0.823</u>
NZ score for estimation with symmetric CAViaR					
Static multiplicative for ES	1.119	1.081	0.842	1.281	0.886
Dynamic additive for ES	1.114	1.082	0.840	1.283	0.885
Dynamic Ω for ES	<u>1.113</u>	<u>1.077</u>	<u>0.836</u>	<u>1.272</u>	<u>0.871</u>
Dynamic Ω for ES: cst loss	<u>1.113</u>	1.078	0.840	<u>1.272</u>	0.878
NZ score for estimation with asymmetric CAViaR					
Static multiplicative for ES	1.094	1.066	0.819	1.276	0.828
Dynamic additive for ES	<u>**1.087**</u>	1.064	0.819	1.271	0.832
Dynamic Ω for ES	1.090	<u>1.063</u>	<u>0.817</u>	1.271	<u>0.826</u>
Dynamic Ω for ES: cst loss	1.090	<u>1.063</u>	0.818	<u>1.267</u>	<u>0.826</u>

Notes: Bold indicates the best performing method in each column. For each group of four models, estimated using either the AL or NZ score, underlining indicates the best score value.

Table 7. Summary of model confidence set results, based on AL and NZ scores, for models applied to the five stock indices. Values presented are the number of indices for which each method is within the model confidence set for the 75% confidence level. Higher values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	1%		2.5%		5%	
	AL	NZ	AL	NZ	AL	NZ
Historical simulation and GAS						
Historical simulation	1	1	0	0	0	0
GAS	2	1	0	0	0	0
Symmetric GARCH and CARE						
GARCH-t	5	5	4	5	4	4
GARCH-filtered	5	5	4	4	4	2
GARCH-EVT	5	5	4	4	4	3
Symmetric-CARE	5	5	4	4	3	1
Asymmetric GARCH and CARE						
GJR-GARCH-t	5	5	5	5	5	5
GJR-GARCH-filtered	5	5	5	5	5	5
GJR-GARCH-EVT	4	4	5	5	5	5
Asymmetric-CARE	5	5	5	5	5	5
AL score for estimation with symmetric CAViaR						
Static multiplicative for ES	5	5	4	4	3	1
Dynamic additive for ES	5	5	4	4	3	1
Dynamic Ω for ES	5	5	4	4	4	2
Dynamic Ω for ES: cst loss	5	5	4	4	4	3
AL score for estimation with asymmetric CAViaR						
Static multiplicative for ES	5	5	5	5	5	5
Dynamic additive for ES	5	5	5	5	5	5
Dynamic Ω for ES	5	5	5	5	5	5
Dynamic Ω for ES: cst loss	5	5	5	5	5	5
NZ score for estimation with symmetric CAViaR						
Static multiplicative for ES	5	5	4	3	4	1
Dynamic additive for ES	5	5	4	4	4	2
Dynamic Ω for ES	5	5	4	4	3	2
Dynamic Ω for ES: cst loss	5	5	4	4	4	3
NZ score for estimation with asymmetric CAViaR						
Static multiplicative for ES	5	5	5	5	5	5
Dynamic additive for ES	5	5	5	5	5	5
Dynamic Ω for ES	5	5	5	5	5	5
Dynamic Ω for ES: cst loss	5	5	5	5	5	5

6. Empirical analysis with individual stocks

As stock indices are portfolios, their extreme behaviour is somewhat limited. In view of this, we also analysed a set of individual stocks[1]. We chose five individual U.S. stocks that had been components of the S&P 500 for the full 4000-day period for which we had analysed the S&P 500 returns in Section 5. We chose the five stocks that had highest market capitalisation on 29 June 2018, which was the final day of our dataset of index returns. These stocks were Apple, Microsoft, Amazon, Berkshire Hathaway and JP Morgan. We performed the same analysis as in Section 5 for the indices, with rolling windows of 2,000 observations for estimation, and an out-of-sample period of the same length. We summarise the results for the 1%, 2.5% and 5% probability levels in Tables 8 to 10, respectively.

For brevity, we focus our comments on the AL and NZ skill scores in the final two columns of each table. For all three probability levels, the skill scores show that each asymmetric model was preferable to the corresponding symmetric model. For the asymmetric models, the CAViaR-based joint models were more accurate than the GARCH and CARE models. The same was true for the symmetric models.

It is interesting to see the relative performances of the four joint models estimated using the AL or NZ scores in the bottom 16 rows of each table. The underlining in these rows shows that, for each block of four models, the most accurate ES formulation was one of the two dynamic Omega models. This is the case for all three probability levels. For the 1% level in Table 8, modelling the ES using the dynamic Omega formulation with constant loss was preferable, while for the 5% level, the dynamic Omega formulation with loss not constrained to be constant was more accurate. Both performed well for the 2.5% level. It is reasonable that a constant loss is more advisable for the 1% level than the 5% level, because the autoregressive modelling of the loss in expression (18) relies on exceedances beyond the VaR, which occur much less often for the 1% level than the 5% level.

---

[1] We are grateful to a reviewer for suggesting this additional empirical analysis.

Table 8. Summary of results for 1% VaR and ES estimated for the five individual stocks. The first column of values is the percentage of VaR exceedances averaged across the stocks; the ideal would be 1%. The next four columns present the number of stocks for which calibration was rejected; lower values are better. The final two columns present skill scores; higher values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	VaR violation %	VaR hit % test	VaR DQ test	VaR VQR test	ES bootstrap test	VaR & ES AL skill score	VaR & ES NZ skill score
Historical simulation and GAS							
Historical simulation	1.5	3	4	5	1	0.0	0.0
1-factor GAS	**1.0**	**0**	4	5	**0**	4.7	2.2
Symmetric GARCH and CARE							
GARCH-t	1.0	**0**	1	4	**0**	10.3	6.5
GARCH-filtered	1.1	**0**	3	4	**0**	10.2	6.5
GARCH-EVT	**1.0**	**0**	2	4	**0**	10.1	6.3
Symmetric-CARE	0.9	**0**	3	5	**0**	9.6	6.2
Asymmetric GARCH and CARE							
GJR-GARCH-t	0.9	**0**	1	3	**0**	11.0	7.1
GJR-GARCH-filtered	**1.0**	**0**	1	3	**0**	10.9	7.0
GJR-GARCH-EVT	**1.0**	**0**	**0**	2	**0**	11.0	7.0
Asymmetric-CARE	0.7	1	2	4	**0**	10.3	6.6
AL score for estimation with symmetric CAViaR							
Static multiplicative for ES	0.9	**0**	1	2	**0**	10.6	7.0
Dynamic additive for ES	**1.0**	**0**	2	2	**0**	10.5	6.9
Dynamic Ω for ES	**1.0**	**0**	2	3	**0**	10.5	6.9
Dynamic Ω for ES: cst loss	0.9	**0**	1	3	**0**	<u>11.1</u>	<u>7.2</u>
AL score for estimation with asymmetric CAViaR							
Static multiplicative for ES	0.9	**0**	2	2	**0**	11.2	7.5
Dynamic additive for ES	0.9	**0**	1	4	**0**	11.3	7.5
Dynamic Ω for ES	**1.0**	**0**	2	**1**	**0**	11.4	7.5
Dynamic Ω for ES: cst loss	0.9	**0**	**0**	2	**0**	<u>11.7</u>	<u>7.7</u>
NZ score for estimation with symmetric CAViaR							
Static multiplicative for ES	0.9	**0**	2	4	**0**	10.2	6.7
Dynamic additive for ES	**1.0**	**0**	2	2	**0**	10.4	6.8
Dynamic Ω for ES	**1.0**	**0**	2	2	**0**	<u>10.8</u>	<u>7.0</u>
Dynamic Ω for ES: cst loss	0.9	**0**	2	2	**0**	<u>10.8</u>	<u>7.0</u>
NZ score for estimation with asymmetric CAViaR							
Static multiplicative for ES	0.9	**0**	1	2	**0**	11.5	7.7
Dynamic additive for ES	0.9	**0**	1	2	**0**	11.5	7.6
Dynamic Ω for ES	0.9	**0**	1	2	**0**	11.7	7.6
Dynamic Ω for ES cst loss	0.9	**0**	1	2	**0**	**<u>11.9</u>**	**<u>7.8</u>**

Notes: Bold indicates the best performing method in each column. For each group of four models, estimated using either the AL or NZ score, underlining indicates the best skill score value.

Table 9 Summary of results for 2.5% VaR and ES estimated for the five individual stocks. The first column of values is the percentage of VaR exceedances averaged across the stocks; the ideal would be 2.5%. The next four columns present the number of stocks for which calibration was rejected; lower values are better. The final two columns present skill scores; higher values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	VaR violation %	VaR hit % test	VaR DQ test	VaR VQR test	ES bootstrap test	VaR & ES AL skill score	VaR & ES NZ skill score
Historical simulation and GAS							
Historical simulation	3.0	1	5	4	**0**	0.0	0.0
1-factor GAS	2.1	1	4	**0**	**0**	6.5	3.4
Symmetric GARCH and CARE							
GARCH-t	2.2	**0**	3	**0**	1	8.0	4.3
GARCH-filtered	**2.3**	**0**	3	**0**	**0**	8.0	4.4
GARCH-EVT	**2.3**	**0**	3	2	**0**	8.0	4.4
Symmetric-CARE	2.1	1	2	**0**	**0**	8.7	4.8
Asymmetric GARCH and CARE							
GJR-GARCH-t	2.1	**0**	2	**0**	1	8.9	4.8
GJR-GARCH-filtered	**2.3**	**0**	4	1	**0**	8.9	4.9
GJR-GARCH-EVT	**2.3**	**0**	3	1	**0**	8.9	4.8
Asymmetric-CARE	1.9	2	2	2	**0**	9.4	5.2
AL score for estimation with symmetric CAViaR							
Static multiplicative for ES	2.2	**0**	2	1	**0**	9.2	5.1
Dynamic additive for ES	**2.3**	**0**	**1**	3	**0**	9.1	5.1
Dynamic Ω for ES	**2.3**	**0**	2	1	**0**	9.6	5.3
Dynamic Ω for ES: cst loss	**2.3**	**0**	**1**	**0**	**0**	9.5	5.2
AL score for estimation with asymmetric CAViaR							
Static multiplicative for ES	2.2	**0**	2	**0**	**0**	10.0	5.5
Dynamic additive for ES	**2.3**	**0**	**1**	1	**0**	9.9	5.5
Dynamic Ω for ES	**2.3**	**0**	2	2	**0**	10.2	**5.6**
Dynamic Ω for ES: cst loss	**2.3**	**0**	2	**0**	**0**	10.2	**5.6**
NZ score for estimation with symmetric CAViaR							
Static multiplicative for ES	2.2	**0**	2	1	**0**	9.4	5.2
Dynamic additive for ES	**2.3**	**0**	2	3	**0**	9.3	5.2
Dynamic Ω for ES	2.2	**0**	2	2	**0**	9.6	5.3
Dynamic Ω for ES: cst loss	2.2	**0**	2	3	**0**	9.6	5.3
NZ score for estimation with asymmetric CAViaR							
Static multiplicative for ES	2.2	**0**	2	**0**	**0**	10.1	**5.6**
Dynamic additive for ES	2.2	**0**	**1**	1	**0**	9.9	5.5
Dynamic Ω for ES	2.2	**0**	3	2	**0**	**10.3**	**5.6**
Dynamic Ω for ES: cst loss	2.2	**0**	2	2	**0**	10.2	**5.6**

Notes: Bold indicates the best performing method in each column. For each group of four models, estimated using either the AL or NZ score, underlining indicates the best skill score value.

Table 10. Summary of results for 5% VaR and ES estimated for the five individual stocks. The first column of values is the percentage of VaR exceedances averaged across the stocks; the ideal would be 5%. The next four columns present the number of stocks for which calibration was rejected; lower values are better. The final two columns present skill scores; higher values are preferable. Horizontal lines separate different types of model, with hashed lines separating symmetric and asymmetric versions of the models.

	VaR violation %	VaR hit % test	VaR DQ test	VaR VQR test	ES bootstrap test	VaR & ES AL skill score	VaR & ES NZ skill score
Historical simulation and GAS							
Historical simulation	**5.3**	**0**	5	4	1	0.0	0.0
1-factor GAS	4.0	2	5	3	**0**	6.4	3.1
Symmetric GARCH and CARE							
GARCH-t	4.2	2	4	2	1	7.6	3.5
GARCH-filtered	4.5	**0**	4	1	**0**	7.8	3.7
GARCH-EVT	4.4	**0**	3	1	**0**	7.8	3.7
Symmetric-CARE	4.2	1	3	**0**	**0**	8.3	3.9
Asymmetric GARCH and CARE							
GJR-GARCH-t	3.9	2	4	1	**0**	8.6	4.0
GJR-GARCH-filtered	4.3	1	3	1	**0**	8.7	4.1
GJR-GARCH-EVT	4.3	1	3	1	**0**	8.7	4.1
Asymmetric-CARE	4.1	2	4	3	1	8.8	4.1
AL score for estimation with symmetric CAViaR							
Static multiplicative for ES	4.5	**0**	3	**0**	**0**	8.8	4.2
Dynamic additive for ES	4.6	**0**	3	1	1	8.5	4.1
Dynamic Ω for ES	4.5	**0**	4	**0**	**0**	<u>9.2</u>	<u>4.3</u>
Dynamic Ω for ES: cst loss	4.6	**0**	4	**0**	**0**	8.9	4.2
AL score for estimation with asymmetric CAViaR							
Static multiplicative for ES	4.2	2	**1**	1	**0**	9.5	4.5
Dynamic additive for ES	4.5	1	**1**	2	**0**	9.2	4.4
Dynamic Ω for ES	4.4	**0**	3	1	**0**	**<u>9.8</u>**	**<u>4.6</u>**
Dynamic Ω for ES: cst loss	4.3	**0**	**1**	1	**0**	9.6	4.5
NZ score for estimation with symmetric CAViaR							
Static multiplicative for ES	4.5	**0**	4	**0**	**0**	8.8	4.2
Dynamic additive for ES	4.6	**0**	4	**0**	**0**	8.5	4.1
Dynamic Ω for ES	4.6	**0**	4	**0**	**0**	<u>9.2</u>	<u>4.3</u>
Dynamic Ω for ES: cst loss	4.6	**0**	4	**0**	**0**	8.9	4.2
NZ score for estimation with asymmetric CAViaR							
Static multiplicative for ES	4.3	1	**1**	1	**0**	9.5	<u>4.5</u>
Dynamic additive for ES	4.5	1	**1**	2	**0**	9.1	4.3
Dynamic Ω for ES	4.3	1	3	1	**0**	<u>9.7</u>	<u>4.5</u>
Dynamic Ω for ES: cst loss	4.3	1	2	1	**0**	9.6	<u>4.5</u>

Notes: Bold indicates the best performing method in each column. For each group of four models, estimated using either the AL or NZ score, underlining indicates the best skill score value.

7. Summary and concluding remarks

We have presented a new joint model for the VaR and ES, in which the ES is modelled as the product of the VaR and a factor that is a simple function of a dynamic Omega ratio. We model the VaR using the autoregressive quantile models of Engle and Manganelli (2004), and we model the Omega ratio using autoregressive expressions for the gain and loss. Parameters are estimated jointly by optimising a joint scoring function from the class proposed by Fissler and Ziegel (2016). The new model extends the work of Taylor (2019), who obtained promising results by using this estimation approach when modelling the ES as the product of the VaR and a constant factor. Our use of a factor that is a function of the Omega ratio is a development of the proposal of Taylor (2008) to approximate the ES as the product of an expectile and a factor that depends on the expectile level. Our empirical analyses considered the 1%, 2.5% and 5% probability levels. For five stock indices, the out-of-sample results showed that the dynamic Omega formulation for the ES produced slightly better forecast accuracy than previously proposed ES formulations. We obtained similar accuracy with the dynamic Omega formulation when treating the loss as a constant, rather than modelling it autoregressively. We also considered five individual stocks, and found that the dynamic Omega ES formulation produced the best results, with benefit from modelling the loss autoregressively for the 5% probability level. In terms of future research, it would be interesting to extend the ideas in this paper to the multivariate context (see Merlo et al., 2021).

Acknowledgements

We are very grateful to the review team for providing comments that helped greatly to improve the paper.

Appendix 1

The symmetric absolute value and asymmetric slope CAViaR models of Engle and Manganelli (2004) are presented in expressions (24) and (25), respectively. The β_i in these models are constant parameters. Asymmetric slope CAViaR aims to capture the asymmetric leverage effect that is typically observed in stock indices.

$$q_t(\alpha) = \beta_0 + \beta_1 |y_{t-1}| + \beta_2\, q_{t-1}(\alpha) \tag{24}$$

$$q_t(\alpha) = \beta_0 + \beta_1 I\left(y_{t-1} > 0\right)|y_{t-1}| + \beta_2 I\left(y_{t-1} \le 0\right)|y_{t-1}| + \beta_3\, q_{t-1}(\alpha) \tag{25}$$

References

Acerbi, C. 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance* 26(7), 1505-1518.

Acerbi, C., Tasche, D. 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26(7), 1487-1503.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D. 1999. Coherent measures of risk. *Mathematical Finance* 9(3), 203-228.

Basel Committee 2016. Minimum Capital Requirements for Market Risk. Technical report, Basel Committee on Banking Supervision.

Bellini, F., Di Bernardino, E. 2017. Risk management with expectiles. *European Journal of Finance* 23(6), 487-506.

Bernardo, A.E., Ledoit, O. 2000. Gain, loss, and asset pricing. *Journal of Political Economy* 108(1), 144-172.

Bi, H., Huang, R.J., Tzeng, L.Y., Zhu, W. 2019. Higher-order Omega: A performance index with a decision-theoretic foundation. *Journal of Banking & Finance* 100, 43-57.

Caporin, M., Costola, M., Jannin, G., Maillet, B. 2018. On the (Ab)use of Omega?. *Journal of Empirical Finance* 46(March), 11-33.

Christoffersen, P. 2012. *Elements of Financial Risk Management*. Academic Press.

Creal, D. D., Koopman, S.J., Lucas, A. 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28(5), 777-795.

Engle, R.F., Manganelli, S. 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22(4), 367-381.

Fissler, T., Ziegel, J.A. 2016. Higher order elicitability and Osband's Principle. *Annals of Statistics* 44(4), 1680-1707.

Fissler, T., Ziegel, J.A., Gneiting, T. 2016. Expected shortfall is jointly elicitable with value at risk - implications for backtesting. *Risk* January, 58-61.

Gaglianone, W.P., Lima, L.R. Linton, O., Smith, D.R. 2011. Evaluating value-at-risk Models via quantile regression. *Journal of Business & Economic Statistics* 29(1), 150-160.

Gerlach, R.H., Chen, C.W. S. 2015. Bayesian expected shortfall forecasting incorporating the intraday range. *Journal of Financial Econometrics* 14(1), 128-158.

Gerlach, R.H., Wang, C. 2022. Bayesian semi-parametric realized conditional autoregressive expectile models for tail risk forecasting. *Journal of Financial Econometrics* 20(1), 105-138.

Gneiting, T. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106(494), 746-762.

Gneiting, T., Raftery, A.E. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477), 359-378.

Hansen, P.R., Lunde, A., Nason, J.M. 2011. The model confidence set. *Econometrica* 79(2), 453-497.

Harvey, A.C. 2013. *Dynamic Models for Volatility and Heavy Tails*, Econometric Society Monograph 52 Cambridge University Press, Cambridge.

Jiang, R., Hu, X., Yu, K. 2022. Single-index expectile models for estimating conditional value at risk and expected shortfall. *Journal of Financial Econometrics* 20(2), 345-366.

Jones, M. C. 1994. Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters* 20(2), 149-153.

Keating, C., Shadwick, W. 2002. A universal performance measure. *Journal of Performance Measurement* 6(3), 59-84.

Kim, M., Lee, S. 2016. Nonlinear expectile regression with application to value-at-risk and expected shortfall estimation. *Computational Statistics & Data Analysis* 94, 1-19.

Koenker, R., Xiao, Z. 2006. Quantile autoregression. *Journal of the American Statistical Association* 101(475), 980-990.

Komunjer, I. 2005. Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics* 128(1), 137-164.

Lazar, E., Zhang, N. 2019. Model risk of expected shortfall. *Journal of Banking & Finance* 105, 74-93.

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(3-4), 271-300.

Merlo, L., Petrella, L., Raponi, V. 2021. Forecasting VaR and ES using a joint quantile regression and its implications in portfolio allocation. *Journal of Banking & Finance* 133, 106248.

Newey, W.K., Powell, J.L. 1987. Asymmetric least squares estimation and testing. *Econometrica* 55, 819-847.

Ning, C., Xu, D., Wirjanto, T.S. 2015. Is volatility clustering of asset returns asymmetric?. *Journal of Banking & Finance* 52, 62-76.

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

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.

Schmidt, P., Katzfuss, M., Gneiting, T. 2021. Interpretation of point forecasts with unknown directive. *Journal of Applied Econometrics* 36(6), 728-743.

Taylor, J.W. 2008. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics* 6(2), 231-252.

Taylor, J.W. 2019. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business and Economic Statistics* 37(1), 121-133.

White, H., Kim, T.-H., Manganelli, S. 2015. VAR for VaR: measuring tail dependence using multivariate regression quantiles, *Journal of Econometrics* 187(1), 169-188.