The two-sided Weibull distribution and forecasting financial tail risk

Qian Chen [a,*], Richard H. Gerlach [b]

[a] *Peking University HSBC Business School, Shenzen Campus, China*
[b] *The University of Sydney Business School, Australia*

A R T I C L E   I N F O

A B S T R A C T

A two-sided Weibull is developed for modelling the conditional financial return distribution, for the purpose of forecasting tail risk measures. For comparison, a range of conditional return distributions are combined with four volatility specifications in order to forecast the tail risk in seven daily financial return series, over a four-year forecast period that includes the recent global financial crisis. The two-sided Weibull performs at least as well as other distributions for Value at Risk (VaR) forecasting, but performs most favourably for conditional VaR forecasting, prior to the crisis as well as during and after it.

© 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

The Global Financial Crisis (GFC) has highlighted the fact that international financial markets can be subject to very quickly changing volatility and risk levels, and has called into question risk measurement, and risk management practices in general. In the academic literature, there has been a considerable amount of interest in conditional asset return distributions, which could help with these issues, if well specified, and with two aspects in particular: (i) the time-varying nature of the distribution, e.g. volatility; and (ii) the shape and form of the standardised conditional distribution itself, e.g. Gaussian.

Two well-known market risk measures are the Value-at-Risk (VaR), pioneered by JP Morgan in 1993, and the conditional VaR, or expected shortfall (ES), proposed by Artzner, Delbaen, Eber, and Heath (1997, 1999). VaR is the minimum loss expected on an investment, over a given time period and at a specific quantile level. It is an important regulatory tool, and was recommended by the Basel Committee on Banking Supervision in Basel II

for controlling the risk levels of financial institutions, by helping to set minimum capital requirements in order to protect against large unexpected losses. However, VaR has been criticised, for example when the Bank of International Settlements (BIS) Committee pointed out that extreme market movements "were in the 'tail' of distributions, and that VaR models were useless for measuring and monitoring market risk" (Committee on the Global Financial System, 1999). VaR does not measure the magnitude of the loss for violating returns; however, ES does give the expected loss (magnitude), conditional on exceeding a given VaR threshold. Artzner et al. (1999) found that the VaR is not a 'coherent' measure, i.e., it is not sub-additive, while ES, the measure they proposed, is. Consequently, the use of VaR can lead to portfolio concentration rather than diversification, while ES cannot. Thus, these two tail risk measures, VaR and ES, are both considered here.

Chen, Gerlach, and Lu (2012) employed an asymmetric Laplace distribution (Hinkley & Revankar, 1977) and combined this with a GJR-GARCH (Glosten, Jagannathan, & Runkle, 1993) model; they found this to be the only choice (Student-*t* and Gaussian distributions were also considered) that consistently over-estimated tail risk levels, and thus was conservative, during the GFC period. This paper aims to estimate the risk levels more

---

* Corresponding author.
   E-mail addresses: qianchen@phbs.pku.edu.cn (Q. Chen),
richard.gerlach@sydney.edu.au (R.H. Gerlach).

accurately by employing a more flexible extension of the Laplace distribution, namely the Weibull distribution, and ultimately developing a two-sided Weibull distribution. Sornette, Simonetti, and Andersen (2000) developed a symmetric, two-sided 'modified' Weibull, which was used subsequently by Malevergne and Sornette (2004), as an unconditional asset return distribution; an asymmetric 'modified' Weibull was also discussed briefly. A more flexible asymmetric two-sided Weibull is proposed as a conditional return distribution in this paper.

An under-estimation of VaR (and ES) levels can result in insufficient regularity capital amounts being set aside, and thus in fatal losses being suffered during extreme market movements. Ewerhart (2002) argued that prudent financial institutions tend to hold unnecessary, excessive amounts of regulatory capital in order to ensure their reputation and quality, while Bakshi and Panayotov (2007) called this the 'Capital Charge Puzzle'. Intuitively, an overstated VaR will lead financial institutions to allocate excessive amounts of capital, which may be attractive in the post-GFC market. However, as the goals of financial institutions are to meet the regulatory and capital requirements, and to maximize profits and attract investors, such a capital over-allocation represents an investment opportunity cost. Thus, although the regulators may prefer smaller violation numbers in order to avoid excessive losses, investors favour models which predict the risk adequately, instead of either over- or under-predicting it. The goal of our paper is to find a model which achieves that prior to, during and after the recent GFC.

With regard to volatility, the focus here is on four widely used specifications: GARCH (Bollerslev, 1986), GJR-GARCH, the threshold T-GARCH (Brooks, 2001) and a smooth threshold ST-GARCH (Gonzalez-Rivera, 1998). Estimation and inference employ a Bayesian approach, via an adaptive Markov chain Monte Carlo (MCMC) method, adapted from Chen et al. (2012).

The rest of the paper is structured as follows: Section 2 introduces the two-sided Weibull distribution, and Section 3 specifies the volatility models considered. Section 4 briefly describes the Bayesian approach and the MCMC methods, and Section 5 presents the empirical studies from four international stock markets, two exchange rates and one individual asset return series, back-testing and comparing a range of models for VaR and ES. Finally, Section 6 summarizes our conclusions.

2. A standardised two-sided Weibull distribution

Mittnik and Ratchev (1989) applied the Weibull distribution to positive and negative returns separately for the US S&P500 index, while various other authors have employed it as an error distribution in price range data modeling (see Chen, Gerlach, & Lin, 2008) and in trading duration (ACD) models (see for example Engle & Russell, 1998). Sornette et al. (2000) proposed a symmetric modified (two-sided) Weibull as an unconditional return distribution, and also mentioned a two-sided Weibull but did not explore its properties.

We introduce a more flexible, two-sided Weibull distribution (TW). Since a conditional error in a GARCH-type model should have a mean 0 and variance 1, we develop the standardized two-sided Weibull distribution (STW). The aim is to capture the empirical traits of conditional return distributions, such as fat-tails and skewness, for the purposes of more accurate tail risk measure forecasting, meaning that the tails are the regions which it is most important to model accurately. The idea, as in the work of Malevergne and Sornette (2004) and Mittnik and Ratchev (1989), is to allow different Weibull distributions for positive and negative returns. We subsequently derive the pdf, cdf, quantile function and conditional expectation functions.

2.1. Details of STW distribution

The pdf for an STW random variable $X \sim STW(\lambda_1, k_1, \lambda_2, k_2)$ is:

$$f(x|\lambda_1, k_1, k_2)$$

$$= \begin{cases} b_p \left(\dfrac{-b_p x}{\lambda_1} \right)^{k_1-1} \exp\left[-\left(\dfrac{-b_p x}{\lambda_1} \right)^{k_1} \right]; & x < 0 \\ b_p \left(\dfrac{b_p x}{\lambda_2} \right)^{k_2-1} \exp\left[-\left(\dfrac{b_p x}{\lambda_2} \right)^{k_2} \right]; & x \geq 0, \end{cases} \quad (1)$$

where the shape parameters satisfy $k_1, k_2 > 0$, the scale parameters satisfy $\lambda_1, \lambda_2 > 0$, and

$$b_p^2 = \frac{\lambda_1^3}{k_1} \Gamma\left(1 + \frac{2}{k_1} \right) + \frac{\lambda_2^3}{k_2} \Gamma\left(1 + \frac{2}{k_2} \right)$$
$$- \left[-\frac{\lambda_1^2}{k_1} \Gamma\left(1 + \frac{1}{k_1} \right) + \frac{\lambda_2^2}{k_2} \Gamma\left(1 + \frac{1}{k_2} \right) \right]^2.$$

To ensure that the pdf integrates to 1:

$$\frac{\lambda_1}{k_1} + \frac{\lambda_2}{k_2} = 1. \quad (2)$$

In this formulation, there are three free parameters, so we write $X \sim STW(\lambda_1, k_1, k_2)$, where λ_2 is fixed by Eq. (2). In this parametrization, $Pr(X < 0) = \frac{\lambda_1}{k_1}$; thus, if $\frac{\lambda_1}{k_1} < 0.5$, the density is skewed positively (to the right), while negative (or left) skewness occurs when $\frac{\lambda_1}{k_1} > 0.5$; symmetry occurs when $\frac{\lambda_1}{k_1} = 0.5$. The $STW(\lambda_1, k_1, k_2)$ has a cdf, obtained by direct integration, of

$$F(x|\lambda_1, k_1, k_2) = \begin{cases} \dfrac{\lambda_1}{k_1} \exp\left[-\left(\dfrac{-b_p x}{\lambda_1} \right)^{k_1} \right]; & x < 0 \\ 1 - \dfrac{\lambda_2}{k_2} \exp\left[-\left(\dfrac{b_p x}{\lambda_2} \right)^{k_2} \right]; \\ \qquad x \geq 0. \end{cases} \quad (3)$$

The inverse cdf or quantile function of an STW is:

$$F^{-1}(\alpha|\lambda_1, k_1, k_2)$$

$$= \begin{cases} -\dfrac{\lambda_1}{b_p} \left[-\ln\left(\dfrac{k_1}{\lambda_1}\alpha \right) \right]^{\frac{1}{k_1}}; & 0 \leq \alpha < \dfrac{\lambda_1}{k_1} \\ \dfrac{\lambda_2}{b_p} \left[-\ln\left(\dfrac{k_2}{\lambda_2}(1-\alpha) \right) \right]^{\frac{1}{k_2}}; & \dfrac{\lambda_1}{k_1} \leq \alpha < 1. \end{cases} \quad (4)$$

The mean of an STW, μ_X, is given in Appendix A. Thus, $Z = X - \mu_X$ has a shifted $STW(\lambda_1, k_1, k_2)$ distribution with mean 0 and variance 1. To save space, other relevant characteristics, such as skewness and kurtosis, are summarized in Appendix A.

For the sake of parsimony and simplification, and since the real return data support this choice, we consider only the case $k_1 = k_2$. Thus, we simply write $STW(\lambda_1, k_1)$ with only two parameters. The fact that $Pr(X < 0) = \frac{\lambda_1}{k_1}$ means that $0 < \lambda_1 \leq k_1$, and $\lambda_2 = k_1 - \lambda_1$. Chen et al. (2012) considered the asymmetric Laplace (AL) distribution, whose skewness and kurtosis were in the ranges $[-2, 2]$ and $[6, 9]$, respectively; the STW is a direct extension of the AL. When $k_1 = k_2$, the skewness in the STW is in the range $[-2.4, 2.4]$, and the kurtosis is in the range $[2.5, 11.5]$, which illustrates a considerable degree of flexibility in the shapes and tail properties of the STW distribution. Malevergne and Sornette (2004) consider only $k_1 \leq 1$, which preserves a single mode of the density. However, when $k_1 < 1$, the tails of the STW are fatter than for $k_1 = 1$ (AL), though Chen et al. (2012) found even $k_1 = 1$ to be too fat-tailed during the GFC period. Thus, we do not restrict k_1 in our estimation. This allows the conditional distribution to be potentially bi-modal, which may not be a good fit in the centre of the return distribution, but will potentially allow the tails, and thus the VaR and ES, to be estimated more accurately, which is our explicit goal. This result is confirmed in the empirical section.

Fig. 1 shows some STW densities, and log-densities, for the range of parameter estimates found for k_1 in the real return series we analyse (i.e., $k_1 \in (1, 1.22)$), as well as $k_1 = 0.95$; the skewness was kept constant. The figure demonstrates the STW distribution's flexibility, as well as the slight thinning of the tails as $k_1 > 1$.

2.2. VaR and tail conditional expectations for a two-sided Weibull

The one-period VaR, for holding an asset, and the *conditional* one-period VaR, or ES, are defined formally via

$$\alpha = Pr(r_{t+1} < \text{VaR}_\alpha | \Omega_t);$$
$$\text{ES}_\alpha = E[r_{t+1} | r_{t+1} < \text{VaR}_\alpha, \Omega_t],$$

where r_{t+1} is the one-period return from time t to time $t + 1$, α is the quantile level, and Ω_t is the information set at time t. Thus, the VaR is simply the quantile given in Eq. (4).

In practice, $\frac{\lambda_1}{k_1}$ is estimated to be much closer to 0.5 than α, as risk management focuses only on the extreme tails of returns, particularly the cases $\alpha \leq 0.05$. Thus, only the case $\alpha < \frac{\lambda_1}{k_1}$ in Eq. (4) is relevant here. In this context, the tail expectation of an STW is:

$$\text{ES}_\alpha = \frac{-\lambda_1^2}{\alpha b_p k_1} \int_{\left(\frac{-b_p VaR_\alpha}{\lambda_1}\right)^{k_1}}^{\infty} \left[\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \right]^{\frac{1}{k_1}+1-1}$$
$$\times \exp\left[-\left(\frac{-b_p x}{\lambda_1}\right)^{k_1} \right] d\left(\frac{-b_p x}{\lambda_1}\right)^{k_1}$$

$$= \frac{-\lambda_1^2}{\alpha b_p k_1} \Gamma\left(1 + \frac{1}{k_1}, \left(\frac{-b_p VaR_\alpha}{\lambda_1}\right)^{k_1} \right);$$
$$0 \leq \alpha < \frac{\lambda_1}{k_1}, \tag{5}$$

where $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt$ is the upper incomplete gamma function.

3. Model specification

This section discusses the models considered in the empirical section. We follow the common assumption that the mean of a return series is (well approximated as) zero. The generalized model for a financial return series **y** is:

$$y_t = (\epsilon_t - \mu_\epsilon)\sqrt{h_t}, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} D(1), \tag{6}$$

where $\text{Var}(y_t | \Omega_t) = h_t$ is the conditional variance and D is the conditional distribution, with variance 1 and mean μ_ϵ (often 0). The VaR and ES in this model are:

$$\text{VaR}_{t+1} = D_\alpha^{-1}\sqrt{h_{t+1}}; \quad \text{ES}_{t+1} = \text{ES}_\alpha^D \sqrt{h_{t+1}}, \tag{7}$$

where D_α^{-1} is the inverse cdf of D, and ES_α^D is the expected shortfall of D, at the $\alpha \times 100\%$ level. The Gaussian, Student-t, skewed-t (Hansen, 1994), AL (Chen et al., 2012) and STW distributions are considered. The last two have non-zero means that are subtracted in Eq. (6). Expressions for ES_α^D in the Gaussian and Student-t cases are provided by McNeil, Frey, and Embrechts (2005, pp. 45–46), while for the AL see Chen et al. (2012). Appendix B contains a derivation of ES_α^D for the skewed-t distribution.

3.1. Volatility models

Rather than focus on one volatility specification, four are considered here: GARCH, GJR-GARCH, T-GARCH and ST-GARCH. There are many other volatility models that could be considered; we chose these four because they are popular models, are used extensively in the literature, and can capture a wide range of asymmetric volatility behaviors.

The most general volatility model considered is a two regime smooth transition nonlinear (ST-)GARCH model, similar to that used by Gerlach and Chen (2008). The specified ST-GARCH model has volatility dynamics:

$$h_t = h_t^{[1]} + G(y_{t-1}; \gamma, r) h_t^{[2]},$$
$$h_t^{[i]} = \alpha_0^{[i]} + \alpha_1^{[i]} y_{t-1}^2 + \beta_1^{[i]} h_{t-1}. \tag{8}$$

This is a continuous mixture of two volatility regimes: $h_t^{[2]}$ is the difference between the conditional variances between regimes 1 and 2. $G(x_{t-1}; \gamma, r)$ is a smooth transition function defined on $[0, 1]$, taken as a logistic:

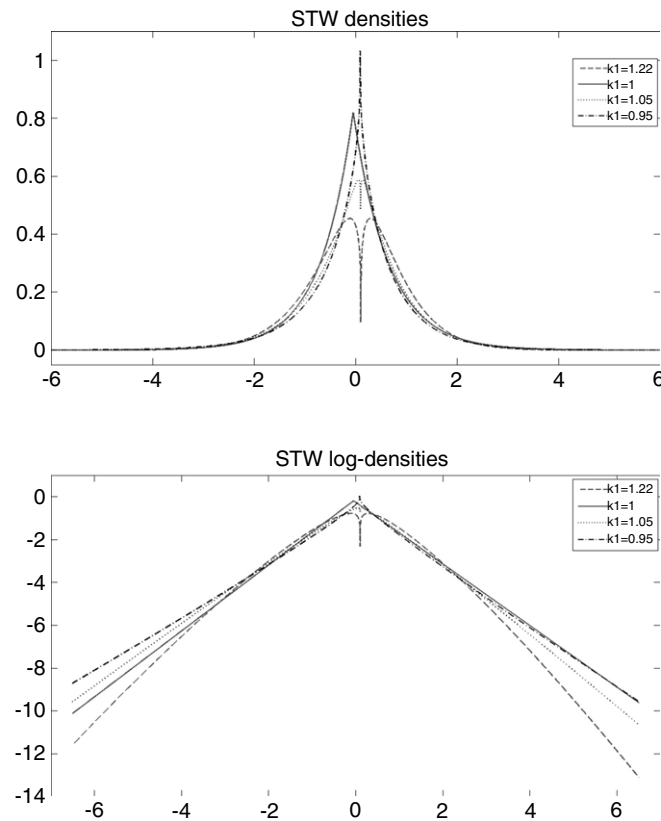$$G(y_{t-1}; \gamma, r) = \frac{1}{1 + \exp\left\{ -\gamma \left(\frac{y_{t-1}-r}{s_y} \right) \right\}},$$

Fig. 1. Some standardized two-sided Weibull densities.

Table 1
Details of volatility models.

Model (abbrev.)	Parameter restrictions	Parameter constraints:	
		Positivity	Stationarity
ST-GARCH (ST)	NA	$\alpha_0^{[1]} > 0, \alpha_1^{[1]}, \beta_1^{[1]} \geq 0$	$\alpha_1^{[1]} + \beta_1^{[1]} < 1$
		$\alpha_0^{[1]} + \alpha_0^{[2]} > 0, \alpha_1^{[1]} + \alpha_1^{[2]} \geq 0$	$\alpha_1^{[1]} + 0.5\alpha_1^{[2]} + \beta_1^{[1]} + 0.5\beta_1^{[2]} < 1$
		$\beta_1^{[1]} + \beta_1^{[2]} > 0$	$\alpha_1^{[1]} + \alpha_1^{[2]} + \beta_1^{[1]} + \beta_1^{[2]} < 1$
T-GARCH (TG)	$\gamma = \infty$	as for ST-GARCH	as for ST-GARCH
GJR-GARCH (GJR)	$\gamma = \infty, r = 0$	$\alpha_0^{[1]} > 0, \alpha_1^{[1]} > 0$	$\alpha_1^{[1]} + 0.5\alpha_1^{[2]} + \beta_1^{[1]} < 1$
	$\alpha_0^{[2]} = 0, \beta_1^{[2]} = 0$	$\alpha_1^{[1]} + \alpha_1^{[2]} > 0$	
GARCH (G)	$\gamma = \infty, r = 0$	$\alpha_0^{[1]} > 0$	$\alpha_1^{[1]} + \beta_1^{[1]} < 1$
	$\alpha_0^{[2]} = 0, \alpha_1^{[2]} = 0$	$\alpha_1^{[1]}, \beta_1^{[1]} \geq 0$	
	$\beta_1^{[2]} = 0$		

where γ is the smoothness or speed of transition param-
eter, which is assumed to be positive for identification; s_y
is the sample standard deviation of the observed threshold
variable **y**, allowing γ to be independent of the scale of **y**.

Sufficient conditions for 2nd order stationarity and
positivity for the ST-GARCH model are given in row
1 of Table 1, which also contains abbreviations for
each volatility model, and the parameter restrictions and
constraints for the other three volatility models employed.
All four volatility models are nested in Eq. (8): e.g., the
T-GARCH model is Eq. (8) when $\gamma \to \infty$.

4. Estimation and forecasting methodology

This section briefly discusses the Bayesian MCMC
procedures for estimation and forecasting. For more
details, see Chen et al. (2012) and Gerlach and Chen (2008);
or Ardia (2008) for alternative MCMC schemes for GARCH-
type models.

4.1. Bayesian estimation methods

MCMC methods are employed for all models that
require a likelihood function and a prior. Each model,

and its likelihood, follows from the choice of the error distribution D, the mean (Eq. (6)) and the volatility equation, from Eq. (8), with the relevant restrictions specified in Table 1.

The volatility model parameters are grouped, either as one group for the G or GJR models, or separated by regime for the TG and ST models: each group is generated as a block in the MCMC scheme, from its full conditional distribution. For the G and GJR models, this grouping is denoted as θ, while for the TG and ST models, the groups are denoted as $\theta_{[1]}, \theta_{[2]}$, with $\theta = (\theta_{[1]}, \theta_{[2]})$. The priors for the parameters in the most general case, the ST, are as per Gerlach and Chen (2008), for θ and r, and as per Chen, Gerlach, Choy, and Lin (2010) for γ. These simplify directly for the TG, GJR and G models. They are mostly flat and bounded, to ensure that γ is not close to 0, and also to avoid the function $G(\cdot)$ being too flat over [0, 1], in which case $\theta_{[1]}$ and $\theta_{[2]}$ are unidentified for the ST and TG models, while also not allowing $\gamma \to \infty$ in the ST model (which is the TG model).

For the STW, the parameters λ_1 and k_1 have a flat prior over their permissable region:

$$\pi(\lambda_1) \propto I(0 < \lambda_1 < k_1).$$

The AL distribution has $k_1 = 1$ and the same prior on $\lambda_1 = p$.

For the skewed t distribution, the degrees of freedom and shape parameters, ν and ζ respectively, have prior densities:

$$\pi(\nu) \propto I(4 < \nu < 30); \qquad \pi(\zeta) \propto I(-1 < \zeta < 1).$$

The same flat prior is employed for the degrees of freedom for the symmetric Student-t distribution, where $\zeta = 0$.

The parameters in each error distribution are generated one at a time from their full conditional distributions in the MCMC sample. None of the parameters or groups have a standard conditional posterior density, and, as such, Metropolis and Metropolis-Hastings methods are employed. Gerlach and Chen (2008) illustrate the efficiency gains from employing an adaptive scheme where iterates in the burn-in period are used to build a Gaussian proposal density for use in the sampling period. Chen et al. (2012) extend this method to cover a mixture of Gaussian proposals, in both the burn-in and sampling periods. This latter mixture of proposals method is adapted to all of the models here; it is a special simplified case of the more general and flexible "AdMit" mixture of Student-t proposals procedure suggested by Hoogerheide, Kaashoek, and van Dijk (2007).

All of the estimation and forecasting results are realized via these adaptive MCMC algorithms, programmed in Matlab. Convergence is assessed by running the MCMC scheme from many widely different starting points and checking the trace plots of iterates for convergence to the same posterior. The simulation results are available from the authors on request.

4.2. VaR and ES forecasts

One-step-ahead forecasting is considered. In MCMC methods, the entire parameter vector for each model,

Table 2
The quantile levels where ES at level α occurs.

α	δ_α	
	$N(0, 1)$	AL
0.01	0.0038	0.0037
0.05	0.0196	0.0184

Note: Shows the CDF of each distribution, evaluated at the ES for level α.

denoted Θ, is simulated from the posterior at each stage, combining to give a Monte Carlo sample of size N: $\Theta^{[1]}, \ldots, \Theta^{[N]}$. Each of these iterates allows the computation of a one-step-ahead forecast of h_t, via Eq. (8) and Table 1, which can be combined with the relevant formulas for VaR and ES, to give MC iterate forecasts of VaR and ES, i.e., $\text{VaR}^{[i]}, \text{ES}^{[i]}$ for $i = 1, \ldots, N$, for each model; for example, for the STW distribution, Eqs. (4) and (5) are employed. These VaR and ES iterates are simply averaged over the sampling period of the MCMC scheme, to give a one-step-ahead forecast of VaR and ES for each model. The same approach is taken for all of the different models and error distributions, using their specific expressions for volatility, VaR and ES in each case.

4.3. Back-testing VaR models

As is recommended by Basel II, VaR forecasts are obtained at the 1% risk level, while 5% is considered for illustration. Each model's forecasts are evaluated by considering their violation rate (VRate):

$$\text{VRate} = \frac{1}{m} \sum_{t=n+1}^{n+m} I(y_t < \text{VaR}_t),$$

and comparing their violation ratios VRate/α, where $\text{VRate}/\alpha \approx 1$ is preferred. The formal back-tests considered are the unconditional coverage (UC) test of Kupiec (1995), the conditional coverage (CC) test of Christoffersen (1998), and the Dynamic Quantile (DQ) test of Engle and Manganelli (2004).

4.4. Back-testing ES models

Although there are a few existing back-testing methods in the literature for ES, such as the censored Gaussian method of Berkowitz (2001), the functional delta approach of Kerkhof and Melenberg (2004) and the saddle point techniques of Wong (2008), they appear to be based on the Gaussian distribution, and also seem overly-complex and difficult to implement. Kerkhof and Melenberg (2004) made the excellent suggestion to compare ES models in the same manner as VaR models: on an equal quantile level. After all, ES does occur at a specific quantile of the return distribution. In particular, for the standard Gaussian and AL distributions, the ES quantile level at a fixed α is (different, but) constant: the ES quantile level only depends on α for the Gaussian and AL (and does not depend on the unknown AL shape parameter p). Denote by δ_α^{ES} the nominal levels for ES at the VaR level α. For the Gaussian and AL distributions, these are given in Table 2.

Chen et al. (2012) exploited this result in order to employ the standard VaR back-testing methods, discussed

Table 3
Summary descriptive statistics for the seven return series.

Index	Period	Mean	Std	Skewness	Kurtosis	Min	Max
Aus	98–05	0.028	0.73	−0.53	7.12	−5.85	3.39
	06–09	0.007	1.36	−0.54	7.23	−8.55	5.36
US	98–05	0.013	1.20	0.00	5.36	−7.04	5.57
	06–09	−0.011	1.65	−0.22	11.44	−9.47	10.96
UK	98–05	0.003	1.20	−0.11	5.19	−5.59	5.90
	06–09	−0.001	1.53	−0.10	10.00	−9.26	9.38
HK	98–05	0.019	1.64	0.20	8.64	−9.29	13.40
	06–09	0.033	2.13	0.09	9.26	−13.58	13.41
AU/US	98–05	0.006	0.72	−0.19	5.53	−4.45	4.82
	06–09	0.022	1.13	−0.72	15.10	−8.21	7.70
EUR/US	98–05	0.006	0.61	0.02	3.72	−2.47	2.71
	06–09	0.010	0.74	0.39	7.36	−3.00	4.62
IBM	98–05	−0.012	2.68	−9.45	253.42	−71.13	12.36
	06–10	0.044	1.62	0.18	7.68	−6.10	10.90

Note: 'Period' indicates the years over which the data were summarised; 'Std' is the standard deviation.

above, to back-test ES models. For the Student-t and skewed-t distributions, however, the quantile level of ES depends on α, as well as the degrees of freedom ν and λ for the skewed-t. Similarly, for the STW, the ES quantile level depends on the parameters λ_1 and k_1. To back-test ES models with these distributions, we approximate by considering the ES level for the average estimated parameters during the forecast sample. This works well, since these parameters do not change very much during the forecast period, and therefore neither do the ES quantile levels.

5. Empirical study

5.1. Data

The model is illustrated by applying it to daily return series from four international stock market indices: the S&P 500 (US), FTSE 100 (UK), the AORD All ordinaries index (Australia), and the HANG SENG Index (Hong Kong); as well as two exchange rate series: the AU dollar to the US dollar and the Euro to the US dollar; and a single asset series: IBM. The data are obtained from Yahoo! Finance, and cover twelve years, January 1998 to January 2010, except for the Euro to US dollar exchange rate, which starts in January 1999. The daily return series is $y_t = (ln(P_t) - ln(P_{t-1})) \times 100$, where P_t is the closing price/value on day t.

The sample is initially divided into two periods: the period from January 1998 to December 2005 (roughly the first 2000 returns) is used as an initial learning period, then the period from January 2006 to January 2010 is used as the forecasting period. The forecast sample sizes vary from 770 to 1050 days, due to different trading day holidays, etc., and this period includes the entire GFC. Table 3 shows summary statistics for the seven return series in the learning and forecast samples. Clearly, the forecast period is generally more volatile and more fat-tailed (higher kurtosis), except, notably, for IBM. The estimation results in each series (not shown to save space) are mostly as expected, and as well-documented in the literature: showing high volatility persistence ($\alpha_1 + \beta_1$),

and having fat-tailed (e.g., $\nu < 10$ in the Student-t and skewed-t error models) and mildly negatively skewed (e.g., $\lambda_1/k_1 > 0.5$ in STW, $p > 0.5$ in AL and $\lambda < 0$ in skewed-t) conditional distributions.

5.2. VaR forecast comparison

Table 4 shows the summaries of the violation ratios, i.e., the ratios of observed VRates to the true nominal levels $\alpha = 0.01, 0.05$, across all series. The summaries shown are the average ('Mean') and deviation ('Std'), which is the square root of the average squared distance of the observed ratio away from the preferred ratio of 1. In the table, a box around a number indicates the mean ratio closest to one (or the minimum deviation in ratios from one) for each series over all models. The average ratios that are furthest from 1, and the maximum deviations from 1, are indicated in bold.

First, it is clear that the differences between models are dominated by the choice of an error distribution: models with the same distribution but different volatility equations are much closer to each other in violation ratios than to models with other distributions. Thus, models with the same error distributions appear together in the table. As such, our discussion focuses on the different distributions. At $\alpha = 0.01$, models with Gaussian errors are consistently anti-conservative, under-predicting risk levels in all series: average VRates are double or more than double the nominal 1%. Alternatively, models with AL errors over-predict risk levels: the average VRates are half the nominal 1%, and are thus conservative; this agrees with the results of Chen et al. (2012). Models with skewed-t errors tend to under-predict risk, but less so than Gaussian models, with average VRates about 20%–30% too high. Models with Student-t and STW errors are clearly the best performers and the most favoured, with VRates closest to nominal. The GJR-t model ranks the highest, with an average VRate closest to 1 (1.02), followed by the GARCH-STW (1.03), which also has the minimum deviation from 1 (0.3), being equal best with the ST-GARCH-STW model. Models with STW errors have the lowest VRate ratio deviations from 1. Informally, then, models with STW

Table 4
Summary of violation ratios, VRatio $= \hat{\alpha}/\alpha$, and counts of model rejections at $\alpha = 0.01, 0.05$, over the seven return series, for each model.

Model	$\alpha = 0.01$						$\alpha = 0.05$					
	Mean	Std.	UC	CC	DQ	Total	Mean	Std.	UC	CC	DQ	Total
G-n	2.07	1.08	**7**	**6**	**7**	**7**	1.14	0.22	1	**2**	**5**	**5**
GJR-n	2.06	1.17	5	4	6	6	1.21	0.30	4	1	3	4
TG-n	2.16	1.27	5	4	5	5	1.25	**0.31**	3	**2**	3	3
ST-n	**2.24**	**1.34**	6	4	6	6	**1.26**	0.34	**4**	**2**	3	4
G-t	0.88	0.38	1	1	2	2	0.88	0.21	1	**2**	2	2
GJR-t	1.02	0.40	1	0	4	4	0.97	0.21	1	**2**	2	2
TG-t	1.06	0.52	3	0	3	3	0.96	0.24	2	**2**	2	2
ST-t	1.07	0.50	2	0	5	5	0.99	0.23	2	**2**	2	2
G-SKT	1.29	0.34	0	0	3	3	1.10	0.14	0	1	3	3
GJR-SKT	1.22	0.32	0	0	1	1	1.16	0.21	2	0	2	2
TG-SKT	1.28	0.48	2	0	2	2	1.20	0.23	2	1	1	2
ST-SKT	1.28	0.44	1	0	1	1	1.17	0.20	1	0	2	2
G-AL	0.50	0.52	1	1	3	3	0.91	0.12	0	0	2	2
GJR-AL	0.48	0.56	3	1	1	3	0.89	0.16	1	1	3	3
TG-AL	0.50	0.54	2	0	3	3	0.93	0.15	1	0	2	2
ST-AL	0.52	0.52	2	1	1	2	0.93	0.13	1	0	2	2
G-STW	1.03	0.30	0	0	2	2	1.03	0.11	0	0	3	3
GJR-STW	0.81	0.35	1	0	1	1	1.02	0.16	1	0	0	1
TG-STW	1.09	0.38	1	0	3	3	1.14	0.21	0	0	1	1
ST-STW	1.07	0.30	0	0	3	3	1.14	0.20	1	0	3	3

Note: 'Mean' is the average VRatio for each model across the seven series: the one closest to one is boxed, the furthest away is in bold; 'Std.' is the standard deviation in VRatios from one. The VaR model rejections from each test (at the 5% significance level) are counted for each model across the seven series: boxes indicate the least rejected model, bold indicates the most rejected model. 'Total' counts the number of series for which a model was rejected by at least one test.

errors are best at forecasting risk levels for $\alpha = 0.01$, very marginally ahead of models with Student-t errors.

Similar results hold for $\alpha = 0.05$, with Gaussian and skewed-t error models consistently under-predicting the risk, while models with AL, Student-t and STW errors have VRate ratios which are quite close to 1 across the seven series.

Table 4 also shows counts of the numbers of rejections for each model, at the 5% significance level, over the seven series, under the three formal back-tests: the UC, CC and DQ. Following Engle and Manganelli (2004), we use a lag of 4 for the DQ, while using the extended CC test of Chen et al. (2012), also with a lag of 4. At $\alpha = 0.01$, the Gaussian error models are rejected in most series, while models with Student-t errors are rejected more than the remaining models on average. The three best models are rejected in only one series: the GJR-GARCH-STW, and the ST-GARCH and GJR-GARCH with skewed-t errors. Models with AL, skewed-t and STW errors are quite comparable and perform the best on these tests across the seven series. At $\alpha = 0.05$, models with Gaussian errors are again rejected in most series. The other models are quite comparable, except for the GJR-GARCH-STW and T-GARCH-STW models, which are rejected in only one series each.

In summary, models with STW and Student-t errors tend to have average VRates which are closest to the nominal at both $\alpha = 0.01, 0.05$. In terms of the deviation in VRate ratios from 1, models with STW errors again did best overall, though models with AL errors did very well at $\alpha = 0.05$. In terms of the tests, for both $\alpha = 0.01, 0.05$, a model with STW had the minimum number of rejections: one in seven series. Models with Gaussian

errors significantly under-predicted the risk in most series at $\alpha = 0.01, 0.05$ (by over 100% at $\alpha = 0.01$); models with skewed-t errors, while doing well in the tests, under-predicted the risk levels by 10%–30% on average.

5.3. Expected shortfall forecast comparison

The ES forecasts from several parametric models, for the returns on the Australian stock market and the AU to US dollar exchange rate, are shown in Fig. 2.

The plots indicate a clear ordering in ES levels across distributions: the Gaussian is least extreme, followed by the Student-t, skewed-t, and STW, with the AL distribution giving the most extreme ES forecasts. This pattern occurred consistently across the seven series, for each volatility model.

The quantile levels at which ES occurs, for various VaR quantile levels α, are well known and can be calculated using standard software for the Gaussian and Student-t distributions, using their cdf functions; the ES quantile levels for the AL distribution, constant for a fixed α, were derived by Chen et al. (2012) and are given in Appendix B and Table 2. The closed forms for the ES and the relationship between ES and VaR for the skewed-t are derived and given in Appendix B, while for the STW this is given by Eqs. (3) and (5), which allow the ES quantile level to be evaluated for a STW at VaR level α. Table 5 shows the approximate quantile levels for ES from the Student-t, skewed-t and STW models, with the ST-GARCH volatility equation, obtained using the average of the estimates of each distribution's parameters over the forecast period in each series. The quantile levels for other volatility models
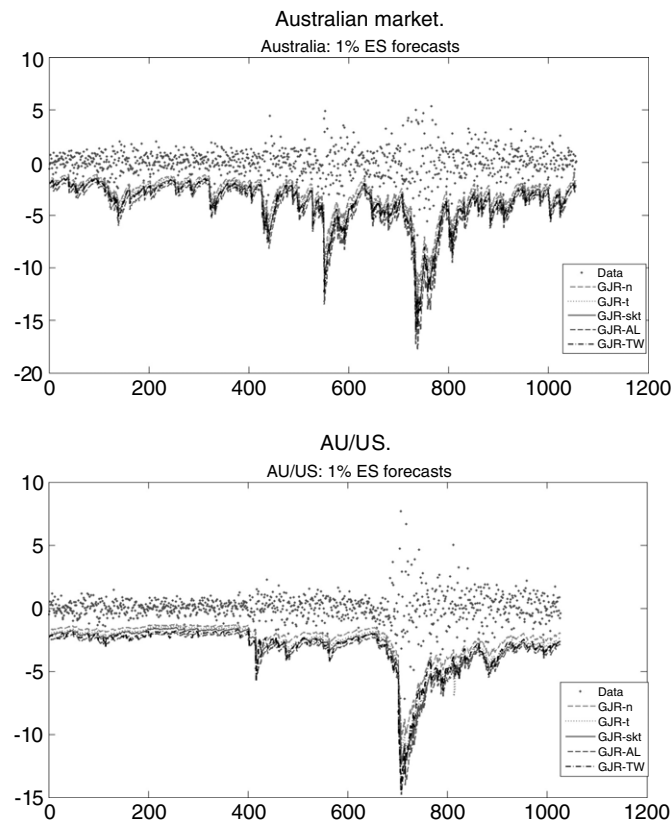
Australian market.

AU/US.

Fig. 2. 1% ES forecasts from GJR-*n*, GJR-*t*, GJR-SKT, GJR-ALCP and GJR-TW.

Table 5

The estimated actual quantile levels $\hat{\delta}_\alpha$ for the ES for the ST-*t*, ST-SKT and ST-TW models at $\alpha = 0.01, 0.05$.

	Aust	US	UK	HK	AU/US	EUR/US	IBM
$\alpha = 0.01$							
ST-*t*	0.0037	0.0037	0.0037	0.0035	0.0036	0.0036	0.0034
ST-SKT	0.0037	0.0037	0.0037	0.0035	0.0036	0.0036	0.0034
ST-TW	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
$\alpha = 0.05$							
ST-*t*	0.0186	0.0188	0.0190	0.0181	0.0182	0.0186	0.0175
ST-SKT	0.0186	0.0188	0.0190	0.0181	0.0182	0.0186	0.0175
ST-TW	0.0189	0.0188	0.0189	0.0186	0.0188	0.0188	0.0187

Note: These estimated quantile levels are obtained using the average of the estimates of each distribution's parameters over the forecast period, for each series. They estimate $\hat{\delta}_\alpha$ for the ES in each case.

are very similar, and therefore are not shown, to save space.

Using these ES quantile levels, the ES violation rate, ESRate, is defined as:

$$\text{ESRate} = \frac{1}{m} \sum_{t=n+1}^{n+m} I(y_t < \text{ES}_t),$$

and a good model should have ESRate being very close to the nominal δ_α.

Table 6 contains summaries of the ES violation rate ratios, $\hat{\delta}_\alpha / \delta_\alpha$, at $\alpha = 0.01, 0.05$ across all models and the seven series, during the forecast period. Again, the best average risk ratio (the closest to 1) is boxed, as is that with the minimum deviation from 1. At $\alpha = 0.01$, models with

Gaussian errors are again consistently anti-conservative, and significantly under-predict risk levels in all series: their ESRates are close to triple or even more than triple the nominal 1%. Furthermore, models with Student-*t* errors also under-predict the risk, sometimes significantly, with their typical ESRates being 55%–84% above the nominal. Again, models with AL errors over-predict the risk levels, but not significantly: on average, their ESRates are half the nominal 1%, and thus are conservative, which agrees with the findings of Chen et al. (2012). Models with skewed-*t* errors tend to under-predict the risk, though not significantly, with ESRates which are 16%–39% too high on average. However, the 3rd and 4th ranked models, based on average ESRate ratios, with values of 1.16

Table 6

Summary of ES violation ratios, ESRatio $= \hat{\delta}_\alpha/\delta_\alpha$, and counts of ES model rejections at $\alpha = 0.01$, 0.05.

Model	$\alpha = 0.01$						$\alpha = 0.05$					
	Mean	Std.	UC	CC	DQ	Total	Mean	Std.	UC	CC	DQ	Total
G-*n*	**3.35**	**2.51**	**6**	**5**	6	6	1.53	0.56	3	**5**	**6**	**6**
GJR-*n*	2.88	2.00	**6**	4	**7**	**7**	1.59	0.71	4	3	5	5
TG-*n*	2.95	2.24	4	3	6	6	1.66	0.80	4	**4**	5	5
ST-*n*	2.84	2.16	5	3	**7**	**7**	**1.72**	**0.86**	5	**4**	5	5
G-*t*	1.55	1.01	1	1	4	4	1.37	0.46	2	3	5	5
GJR-*t*	1.77	1.12	2	1	3	3	1.47	0.55	3	3	**6**	**6**
TG-*t*	1.69	1.09	2	1	2	2	1.52	0.63	3	**4**	4	4
ST-*t*	1.84	1.24	2	1	2	2	1.54	0.67	3	3	4	4
G-SKT	1.17	0.58	1	1	3	3	1.17	0.29	1	1	5	5
GJR-SKT	1.16	0.47	0	0	1	1	1.22	0.31	2	1	2	2
TG-SKT	1.28	0.58	0	0	1	1	1.30	0.40	3	0	2	3
ST-SKT	1.39	0.73	1	0	0	1	1.30	0.39	3	0	2	3
G-AL	0.45	0.62	1	0	0	1	0.86	0.28	1	1	5	5
GJR-AL	0.37	0.67	1	0	0	1	0.86	0.27	1	0	2	2
TG-AL	0.46	0.57	0	0	0	☐0	0.81	0.30	2	0	3	3
ST-AL	0.41	0.64	1	0	0	1	0.84	0.27	1	1	3	3
G-STW	☐1.02	0.45	0	0	2	2	☐0.98	☐0.20	☐0	1	5	5
GJR-STW	0.79	☐0.35	0	0	0	☐0	0.93	0.28	☐0	0	☐0	☐0
TG-STW	1.05	0.41	0	0	1	1	1.08	☐0.20	☐0	1	2	2
ST-STW	1.20	0.49	0	1	1	1	1.07	0.26	1	1	2	2

Note: 'Mean' is the average ESRatio across the seven series, with the ratio closest to one being boxed and the furthest from one being in bold. 'Std.' is the standard deviation of these ESRatios from one. The ES model rejections are counted across the seven series, with boxes indicating the model with the fewest rejections and bold indicating the model with most rejections (at the 5% level). 'Total' counts the number of series for which a model was rejected by at least one test.

and 1.17 respectively, are the GARCH and GJR-GARCH models with skewed-*t* errors. The two top ranked models based on average ESRate ratios, with values of 1.02 and 1.05, are the GARCH and T-GARCH with STW errors. The GJR-GARCH and ST-GARCH with STW are ranked 5th and 6th, respectively, using this measure. Furthermore, based on the minimum deviation of ratios from 1, models with STW errors are ranked 1st, 2nd and 3rd, with the ST-GARCH-STW ranking 5th. The 4th ranked model is the GJR-GARCH with skewed-*t* errors. Under these criteria, it is clear that models with STW errors perform most favourably, followed by the GARCH and GJR-GARCH with skewed-*t* errors.

A similar story now holds for $\alpha = 0.05$. Models with Gaussian errors are significantly anti-conservative, but now by $\approx$50%–70% on average, while Student-*t* error models perform similarly, and are generally rejected in three of the seven series according to the UC test. Models with AL errors now only marginally over-predict the risk levels, with ESRates being 15%–20% below nominal on average, while the skewed-*t* error models again under-predict risk levels by 17%–30% on average. Here, the four top ranked models, those which have ESRates that are clearly closest to the nominal on average, are the four STW error models. Three of these, excluding the GJR-GARCH-STW, also occupy the top ranked positions based on the minimum deviation in ratios from 1.

Table 6 also shows counts of the number of rejections for each ES forecast model, at the 5% significance level, across the seven series, under the UC, CC and DQ tests, using the ES quantile levels discussed above. At $\alpha = 0.01$ and 0.05, the Gaussian error models are again rejected in all or most series by all tests, while the models with Student-*t* errors are again rejected more than the other models on average. At $\alpha = 0.05$, the Student-*t* error models are

rejected in all or most series for ES forecasting. On the other hand, the two best models, the T-GARCH-AL and the GJR-GARCH-STW, could not be rejected in any series. Models with AL, STW and skewed-*t* errors were generally rejected in only one series at $\alpha = 0.01$, and thus are quite comparable on these tests across the seven series. At $\alpha = 0.05$, only the GJR-GARCH with STW errors could not be rejected in any series; all of the other models were rejected at least twice.

Overall, for forecasting ES during this forecast period, models with STW errors performed better than any of the other models and error distributions considered, with ESRates which are generally closest to the nominal in both average and squared deviations, and ES forecasts which are mostly not rejected by the formal tests, across the seven return series. Under each criterion, a model with STW errors ranked first. The models with AL errors may also be attractive for regulatory purposes, since they have very small violation ratios, roughly half the number of violations expected. However, these smaller violation ratios do signal an over-estimation of risk, and therefore an excessive allocation of capital. Models with STW errors provided adequate and accurate risk coverages.

5.4. Pre-financial-crisis and post-financial-crisis forecast performances

The forecast sample period covers the well-known GFC. Model performances may vary between the pre-financial-crisis and post-financial-crisis periods (the latter of which contains returns both during the GFC and post-crisis). We thus present a pre-crisis and during/post-crisis comparison of risk forecasting performances.

A date for the start of the crisis must be chosen here, but this need not be exactly the same in each
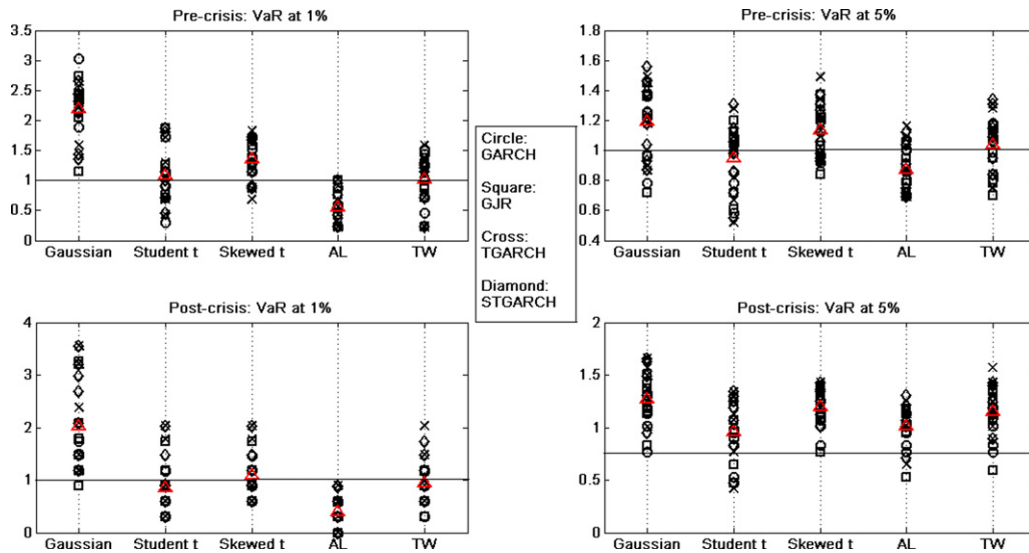
Fig. 3. Circles: GARCH; squares: GJR; crosses: TGARCH; diamonds: STGARCH; large triangles: mean of VRates for each distribution.
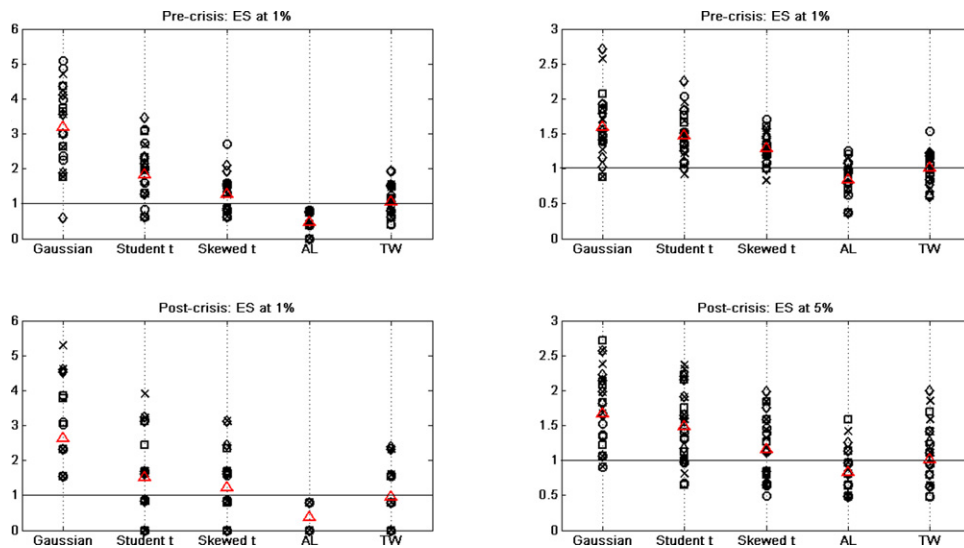

Fig. 4. Circles: GARCH; squares: GJR; crosses: TGARCH; diamonds: STGARCH; large triangles: mean of ESRates for each distribution.

market. Based on news media accounts and Wikipedia, it is generally agreed that the effects of the crisis were initially apparent in international markets in September and/or October, 2008. We choose appropriate dates for each market based on a maximization of the sample return variance in the post-crisis period from among possible days in September/October 2008. The dates chosen thus for each market were: Australia, 22nd September; US and IBM, 19th September; UK, 10th September; HK, 18th September; AU/US, 23rd September; and EUR/US, 23rd September, all in 2008. The forecast sample up to the day before these dates is the pre-crisis period, while the post-crisis period is from these dates up to January, 2010. For each market, there are approximately 700 days in the pre-crisis period and approximately 350 days in the post-crisis sample.

Figs. 3 and 4 show the ratios of VRate/α and ESRate/δ_α at $\alpha = 0.01, 0.05$ for the pre-crisis and post-crisis periods

for the VaR and ES forecast models, as labelled. The results for the pre-crisis sample are very similar to those for the forecast sample as a whole, since it makes up the majority of the sample: models with STW and Student-t errors forecast VaR most accurately at the 1% and 5% risk levels, with the VRate averaging close to one, though the STW error models have VRate ratios with slightly lower variations around one. Furthermore, only models with STW errors have ESRate ratios which are consistently, and averaging, close to one. Models with AL errors are again the only consistently conservative risk forecasters for both VaR and ES.

The results for the during/post-crisis period tell a different story. For VaR forecasting, models with Student-t, skewed-t and STW errors perform well at $\alpha = 0.01$, all with average ratios which are close to one and similar deviations about one, across the seven series. For ES

forecasting, models with STW errors are clearly the best post-crisis, with average ratios which are closest to one and the smallest deviations about one. At the 5% risk level, however, the models with AL and Student-t errors perform best for VaR forecasting, with the STW models under-predicting the risk levels slightly on average. For ES forecasting at $\alpha = 0.05$, the STW has the closest average ratio to one post-crisis, but the AL also does well and has the smallest deviation in ratios from one.

5.5. Loss function

Loss functions can also be applied to the assessment of quantile forecasts. The applicable loss function is the criterion function, minimized via quantile regression estimation, as was done by Koenker and Bassett (1978), for example; it can be written as:

$$LF = \sum_{t=n+1}^{n+m} (y_t - R_t)(\alpha - I_t),$$

where I_t is the indicator variable of a violation (i.e., $y_t < R_t$), R_t is the risk forecast (we use VaR_t for each model/method here), and α is the quantile at which the VaR is evaluated. ES forecasts can also be assessed at their approximate quantile levels, with δ_α being substituted for α above. The best risk forecasts in terms of accuracy should minimize this loss function.

Fig. 5 shows the mean of the loss function for the VaR and ES forecasts via various models, taken over the seven series and the entire forecast period. Two things are apparent: the GJR model (shown as squares) usually has the lowest average loss for each error distribution; for VaR forecasting at $\alpha = 0.01$, models with Student-t, skewed-t and STW errors have the lowest, and comparable, average losses. For VaR forecasting at $\alpha = 0.05$, however, the skewed-t, AL and STW-error models have comparable, and the lowest, average losses. For ES, the losses among the distributions (except for the Gaussian, which has the largest average loss in each case) seem quite close and comparable.

A relevant question is whether any of the models' loss function values are significantly different to any of the others. The model confidence set (MCS) approach of Hansen, Lunde, and Nason (2011) is employed, using the loss function values for each of the competing models in each return series separately. A significance level of 10% is taken, so that the final model set includes the best model with 90% confidence. A block bootstrap procedure, as per Politis and Romano (1994), is employed to estimate the standard errors for pairwise mean loss differences, and also to estimate the sampling distributions of the two t-statistics used in the tests employed. One of these t-statistics is the maximum of the absolute values of all of the t-statistics for the average pairwise loss differences between models remaining in the MCS. If there are m models, then there are $m \times (m-1)/2$ of these t-statistics, being the average loss difference divided by a bootstrap estimate of the standard error of the average loss difference. We call this the 'max loss' test below. The second t-statistic is that for the maximum of the t-statistics for the

Table 7
Summary of counts of MCS test model rejections, using a significance level of 0.1 over the two tests.

Model	Max loss				Mean loss			
	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.05$	
	VaR	ES	VaR	ES	VaR	ES	VaR	ES
G-n	1	0	0	0	0	0	0	1
GJR-n	0	0	1	0	0	0	0	0
TG-n	0	0	0	1	0	0	0	0
ST-n	0	1	0	1	0	0	0	0
G-t	4	0	2	1	0	0	2	0
GJR-t	3	1	1	0	0	0	1	0
TG-t	3	0	1	0	0	0	1	0
ST-t	1	0	1	0	0	0	0	0
G-SKT	3	2	1	1	0	0	0	1
GJR-SKT	0	2	0	0	0	0	0	0
TG-SKT	0	1	0	0	0	0	0	0
ST-SKT	0	1	0	0	0	0	0	0
G-AL	3	4	2	2	1	1	1	1
GJR-AL	3	3	0	1	1	1	0	0
TG-AL	2	4	0	1	1	1	0	0
ST-AL	2	3	0	0	1	1	0	0
G-STW	2	2	1	2	1	1	1	1
GJR-STW	1	1	0	0	0	1	0	0
TG-STW	0	1	0	0	0	0	0	0
ST-STW	0	1	0	0	0	0	0	0

Note: 'max loss' is the test using the maximum of the absolute values of all t-statistics measuring the pairwise average loss differences between models; and 'mean loss' is the test using the mean of all average pairwise loss differences of one model with each of the other models. A model is removed from the MCS if it has the maximum t-statistic and is significant at the 10% level.

mean of the pairwise average loss differences, where the mean is taken with respect to one model. See Hansen et al. (2011) for details. Each of these t-statistics is used to remove models from the MCS in each market, separately and iteratively, until a final 90% MCS is obtained. Loss functions for both the VaR and ES forecast series, at both the 5% and 1% risk levels, are employed. The results are summarised in Table 7, which contains the number of return series for which each model was not included in the final 90% MCS, for each of the two tests, and separately for each model's VaR and ES loss series.

Neither test was able to distinguish clearly between the loss function outcomes across the models, in most cases. When models were excluded from the MCS, it was not obvious that they should be. For example, models with Student-t errors were not included in the 90% MCS for forecasting the 1% VaR in either three or four series out of seven (depending on the type of volatility model). However, these models had loss values below the median loss in most cases, as is shown in Fig. 5. Furthermore, models with Gaussian errors, which typically had the largest loss values across the series, were almost always included in the final MCS. This result can be explained by the inherent large and small variability, respectively, in the loss series for models with Gaussian tails compared with those with fatter tails. Models with Gaussian tails were usually anti-conservative, and thus had more large residuals, which contributed to the loss function (LF) values, and thus had an average loss difference with a larger sampling variation. In a pairwise t-test, small mean differences can be significant if both series have small variabilities, while the same size differences can
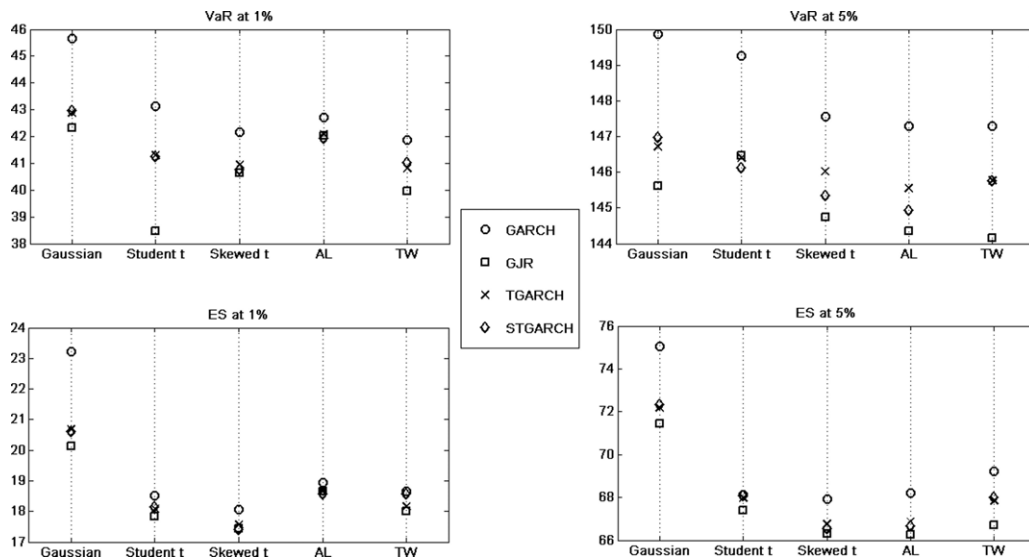
Fig. 5. Loss functions of VaR ES forecasts from various distributions across various volatility models.

be insignificant if one series has a larger variability. Thus, when models with fat tails were compared, the smaller differences in their loss functions were sometimes significant, while the larger loss differences found between models with Gaussian errors and those without were not significant, due to the large variabilities in the loss function series for the Gaussian models. This seems to be an issue which is inherent in the use of paired t-statistics for loss functions in the MCS procedure.

Overall, the STW model performs the most favourably as a risk forecaster for this forecast data period across the seven series for both VaR and ES forecasting at both $\alpha = 0.01, 0.05$ levels. Using almost all criteria, models with STW errors ranked either best or equal best, with violation rates closest to one based on the average and squared deviation, and having the minimum number of model rejections by formal tests, including almost never being excluded from the MCS, whether for the entire period or for either the pre- or post-GFC periods. Models with Student-t errors consistently did well at VaR forecasting for $\alpha = 0.01$, even though they were excluded from the MCS, while the models with AL errors were consistently conservative and exhibited violation rates which were usually below the nominal, with comparatively small variation in violation rate ratios, but were also regularly excluded from the MCS.

6. Conclusion

The recent global financial crisis has challenged market participants' abilities to provide a reasonable level of coverage for dynamic changing risk levels. As a coherent risk measurement method, expected shortfall is able to measure the size of losses in extreme cases, unlike VaR. Despite the benefits of this alternative method, expected shortfall is absent from regulations such as Basel II, perhaps mostly because the back-testing of ES models is less straightforward than that for VaR. Calculating a benchmark for allocating regulatory capital, and thus

protecting financial institutions from risks during extreme market movements, is the ultimate goal of VaR and ES models. However, as another essential function of these financial entities is to make profit, the allocation of capital is very important. In this paper, we argue that, rather than using an extremely conservative model, a more appropriate approach should be able to both relieve the burden of over-allocation of regulatory capital, and protect against the risky under-allocation of capital, by forecasting dynamic risk levels more accurately, and thus carefully and properly increasing the investment opportunities in more profitable assets. For this purpose, we propose a two-sided Weibull conditional return distribution, coupled with a volatility model. The properties of this distribution have been developed and presented, including the VaR and expected shortfall functions. An adaptive Markov chain Monte Carlo method was employed for estimation and forecasting. An empirical study of seven asset return series found that models with conditional two-sided Weibull errors were highly accurate at forecasting both VaR and ES levels, and could not be rejected or bettered consistently across several criteria, relative to the Gaussian, Student-t, skewed-t and asymmetric Laplace conditional return distributions. This accurate performance was found to hold both before the GFC hit markets, and during and after the GFC period. Hopefully, the models introduced in this paper will offer both the regulators and the financial institutions a new option or compromise between suffering from either excess violations or unnecessarily reduced profit. It is clear that the two-sided Weibull has improved the modeling of the the tails of the conditional return distribution. An extension of our model could involve allowing a distribution that had Weibull tails while preserving a single mode, perhaps through the use of a partitioned distribution; e.g., Student-t in the centre and Weibull in the tails. Future work could consider alternative conditional volatility specifications.

Appendix A. Some properties of STW

Let $X \sim \text{STW}(\lambda_1, k_1)$. Then,

$$E(X) = \frac{-\lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2} \Gamma\left(1 + \frac{1}{k_2}\right)$$
$$= \mu_X. \tag{9}$$

The skewness of a standardized two-sided Weibull random variable is

$$S(X) = -\frac{\lambda_1^4}{b_p^3 k_1} \Gamma\left(1 + \frac{3}{k_1}\right) + \frac{\lambda_2^4}{b_p^3 k_2} \Gamma\left(1 + \frac{3}{k_2}\right)$$
$$- 3\left[\frac{\lambda_1^3}{b_p^2 k_1} \Gamma\left(1 + \frac{2}{k_1}\right) + \frac{\lambda_2^3}{b_p^2 k_2} \Gamma\left(1 + \frac{2}{k_2}\right)\right]$$
$$\times \left[-\frac{\lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2} \Gamma\left(1 + \frac{1}{k_2}\right)\right]$$
$$+ 2\left[-\frac{\lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2} \Gamma\left(1 + \frac{1}{k_2}\right)\right]^3. \tag{10}$$

The kurtosis of a standardized two-sided Weibull random variable is

$$K(X) = \frac{\lambda_1^5}{b_p^4 k_1} \Gamma\left(1 + \frac{2}{k_1}\right) + \frac{\lambda_2^5}{b_p^4 k_2} \Gamma\left(1 + \frac{2}{k_2}\right)$$
$$- 4\left[-\frac{\lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2} \Gamma\left(1 + \frac{1}{k_2}\right)\right]$$
$$\times \left[\frac{-\lambda_1^4}{b_p^3 k_1} \Gamma\left(1 + \frac{3}{k_1}\right) + \frac{\lambda_2^4}{b_p^3 k_2} \Gamma\left(1 + \frac{3}{k_2}\right)\right]$$
$$+ 6\left[-\frac{\lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2} \Gamma\left(1 + \frac{1}{k_2}\right)\right]^2$$
$$\times \left[\frac{\lambda_1^3}{b_p^2 k_1} \Gamma\left(1 + \frac{2}{k_1}\right) + \frac{\lambda_2^3}{b_p^2 k_2} \Gamma\left(1 + \frac{2}{k_2}\right)\right]$$
$$- 3\left[-\frac{\lambda_1^2}{b_p k_1} \Gamma\left(1 + \frac{1}{k_1}\right) + \frac{\lambda_2^2}{b_p k_2} \Gamma\left(1 + \frac{1}{k_2}\right)\right]^4. \tag{11}$$

These formulas can be used to verify the ranges of skewness and kurtosis given in Section 2 for the STW.

Appendix B. The VaR and ES of the skewed Student-t and AL

The skewed Student-t distribution of Hansen (1994) is considered, and has the following density function:

$$g(z|\nu, \zeta) = \begin{cases} bc\left(1 + \frac{1}{\nu - 2}\left(\frac{bz + a}{1 - \zeta}\right)^2\right)^{-(\nu+1)/2}; \\ z < -a/b, \\ bc\left(1 + \frac{1}{\nu - 2}\left(\frac{bz + a}{1 + \zeta}\right)^2\right)^{-(\nu+1)/2}; \\ z \geq -a/b, \end{cases}$$

where $2 < \nu < \infty$, and $-1 < \zeta < 1$. The constants a, b and c are given by

$$a = 4\zeta c\left(\frac{\nu - 2}{\nu - 1}\right); \qquad b^2 = 1 + 3\zeta^2 - a^2;$$
$$c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu - 2)}\Gamma\left(\frac{\nu}{2}\right)}. \tag{12}$$

The inverse CDF of the skewed Student-t distribution $Skewt(\nu, \lambda)$ is:

$$F^{-1}(\alpha|\nu, \zeta) = \begin{cases} \frac{1 - \zeta}{b}\sqrt{\frac{\nu - 2}{\nu}} F_s^{-1}\left(\frac{\alpha}{1 - \zeta}, \nu\right) - \frac{a}{b}; \\ \alpha < \frac{1 - \zeta}{2}, \\ \frac{1 + \zeta}{b}\sqrt{\frac{\nu - 2}{\nu}} F_s^{-1}\left(0.5 + \frac{1}{1 + \zeta}\right. \\ \left. \times \left(\alpha - \frac{1 - \zeta}{2}\right), \nu\right) - \frac{a}{b}; \\ \alpha \geq \frac{1 - \zeta}{2}. \end{cases}$$

Here, F_s^{-1} is the inverse CDF of the Student-t distribution.

The Expected Shortfall of a $Skewt(\nu, \zeta)$ at level α for a long position can be calibrated via:

$$ES_\alpha = \frac{-\frac{c(1 - \zeta)^2}{b}\frac{\nu - 2}{\nu - 1}d^{\frac{\nu-1}{2}} - \frac{a(1 - \zeta)}{b}F_s\left(\frac{bVaR_\alpha + a}{1 - \zeta}, \nu\right)}{\alpha};$$
$$VaR_\alpha < -\frac{a}{b}, \tag{13}$$

where $d = \cos^2\left(\arctan\left(-\frac{bVaR_\alpha + a}{(1 - \zeta)\sqrt{\nu - 2}}\right)\right)$.

References

Ardia, D. (2008). *Lecture notes in economics and mathematical systems*: vol. 612. *Financial risk management with Bayesian estimation of GARCH models*. Springer.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). Thinking coherently. *Risk*, *10*(11), 68–71.

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

Bakshi, G., & Panayotov, G. (2007). *The capital adequacy puzzle*. Working Paper, Smith Business School, University of Maryland.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Brooks, C. (2001). A double-threshold GARCH model for the French Franc/Deutschmark exchange rate. *Journal of Forecasting*, *20*, 135–145.

Chen, C. W. S., Gerlach, R., Choy, B., & Lin, C. (2010). Estimation and inference for exponential smooth transition nonlinear volatility models. *Journal of Statistical Planning and Inference*, *140*, 719–733.

Chen, C. W. S., Gerlach, R., & Lin, E. M. H. (2008). Volatility forecasting using threshold heteroskedastic models of the intra-day range. *Computational Statistics and Data Analysis*, *52*(6), 2990–3010.

Chen, Q., Gerlach, R., & Lu, Z. (2012). Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics and Data Analysis*, *56*, 3498–3516.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, *39*, 841–862.

Committee on the Global Financial System (1999). *A review of financial market events in Autumn 1998*. Basel, Switzerland: Bank for International Settlements. http://www.bis.org/publ/cgfs12.pdf.

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

Engle, R. F., & Russell, J. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, *66*, 1127–1162.

Ewerhart, C. (2002). *Banks, internal models and the problem of adverse selection*. Working Paper, University of Bonn.

Gerlach, R., & Chen, C. W. S. (2008). Bayesian inference and model comparison for asymmetric smooth transition heteroskedastic models. *Statistics and Computing*, *18*(4), 391–408.

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 stock. *Journal of Finance*, *48*, 1779–1801.

Gonzalez-Rivera, G. (1998). Smooth transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, *3*(2), 61–78.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*(3), 705–730.

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

Hinkley, D. V., & Revankar, N. S. (1977). Estimation of the Pareto law from underreported data. *Journal of Econometrics*, *5*, 1–11.

Hoogerheide, L. F., Kaashoek, J. F., & van Dijk, H. K. (2007). On the shape of posterior densities and credible sets in instrumental variable regression models with reduced rank: an application of flexible sampling methods using neural networks. *Journal of Econometrics*, *139*, 154–180.

Kerkhof, F. L. J., & Melenberg, B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking and Finance*, *28*, 1845–1865.

Koenker, R. W., & Bassett, G., Jr. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*, 73–84.

Malevergne, Y., & Sornette, D. (2004). VaR-efficient portfolios for a class of super and sub-exponentially decaying assets return distributions. *Quantitative Finance*, *4*, 17–36.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management*. New Jersey: Princeton University Press.

Mittnik, M., & Ratchev, S. T. (1989). Stable distributions for asset returns. *Applied Mathematics Letters*, *2*(3), 301–304.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*, 1303–1313.

Sornette, D., Simonetti, P., & Andersen, J. V. (2000). Φ^q-field theory for portfolio optimization: fat-tails and non-linear correlations. *Physics Reports*, *335*(2), 19–92.

Wong, K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking and Finance*, *32*, 1404–1415.

Qian Chen is an Assistant Professor at the Peking University HSBC Business School, Shenzhen campus. She received her Ph.D. degree in Econometrics from the University of Sydney, Australia. She received her M.Sc. in Economics from the Hong Kong University of Science and Technology, and a BSc. in Mathematics from the Wuhan University, China. Her main research interests focus on financial risk forecasting and management, time series econometrics and the application of Bayesian statistics. She has been working on the development of Bayesian methods for the estimation, inference and forecasting of flexible models for Value-at-Risk and tail conditional expectations.

Richard Gerlach is Associate Professor at the University of Sydney Business School, Australia. His research interests lie mainly in financial econometrics, time series and quantitative risk forecasting. His methodological work has concerned the development of computationally intensive Bayesian methods for inference, forecasting, diagnosis and model comparison for time series models, with a recent focus on nonlinear volatility models, as well as Value-at-Risk and expected shortfall forecasting.