Using Autoregressive Logit Models to Forecast the Exceedance Probability for Financial Risk Management

James W. Taylor*
Saïd Business School, University of Oxford, Park End Street, Oxford, OX1 1HP, UK.
(james.taylor@sbs.ox.ac.uk)

Keming Yu
Department of Mathematical Sciences, Brunel University, Uxbridge, UB8 3PH, UK.
(keming.yu@brunel.ac.uk)

Journal of the Royal Statistical Society, Series A, 2016, Vol. 179, pp. 2069-1092.

Summary

We present new autoregressive logit models for forecasting the probability of a time series of financial asset returns exceeding a threshold. The models can be estimated by maximizing a Bernoulli likelihood. Alternatively, to account for the extent to which an observation does or does not exceed the threshold, we propose that the likelihood is based on the asymmetric Laplace distribution, which has been found to be useful for quantile estimation. We incorporate the exceedance probability forecasts within a new time-varying extreme value approach to value at risk and expected shortfall estimation. We provide empirical illustration using daily stock index data.

Keywords: Financial risk management; Probability forecasting; Asymmetric Laplace distribution; Extreme value theory.

* Corresponding author

1. Introduction

A forecast of the time-varying probability of a financial asset return exceeding a given fixed threshold is of interest in a variety of contexts. For an extreme threshold, the forecast provides an assessment of tail risk. To estimate value at risk and expected shortfall, exceedance probability forecasts can be used within time-varying adaptations of extreme value theory (EVT) (see, for example, Chavez-Demoulin et al. 2014). The probability of a sudden fall in an exchange rate below a given threshold serves as an indicator of a currency crisis (Kumar et al. 2003). To support trading strategies, Christoffersen and Diebold (2006) and Linton and Whang (2007) consider exceedance above 0, while Chung and Hong (2007) focus on non-zero thresholds, noting that exceedance needs to be large enough to ensure a profit after allowing for transaction costs. They also note that investors may respond differently to a signal of large versus small changes leading to different dynamics in thresholds of different sizes and signs. This motivates Thomakos and Wang (2010) to consider the choice of threshold that will maximize the responsiveness of the exceedance probability to changes in the volatility.

Exceedance probability forecasts are also needed in many other applications. The probability of an inflation rate rising by more than a given percentage might prompt a central bank to increase interest rates (Granger and Pesaran 2000). In energy risk management, probability forecasts are required for price spikes, which are typically defined as exceedances over a pre-specified threshold (Kanamura and Ōhashi 2007). The occurrence of precipitation above a threshold is important for flood and drought risk management (Mason et al. 2007).

In this paper, we consider the forecasting of the probability of daily financial returns exceeding a chosen threshold. Our interest is in thresholds that are not close to 0, as this is of greater relevance for risk management. For a return y_t, we wish to estimate the time-varying probability p_t of y_t falling below a fixed chosen threshold Q. If Q lies in the lower tail of the returns distribution, then p_t is the exceedance probability, while if Q lies in the upper tail, the exceedance probability is equal to ($1-p_t$). For a time series of daily returns, the presence of heteroskedasticity causes p_t to be time-varying. One approach to estimating p_t is to fit a model to y_t, and use the estimate of the conditional distribution of y_t to make predictions of p_t. For example, a GARCH model could be fitted. However, these models require a distributional assumption, and assume the dynamics are the same for different parts of the distribution. Both of these assumptions are questionable. In view of this, it is interesting to consider the direct modelling of p_t. Direct modelling of a particular feature of the conditional distribution of y_t has also been the motivation for quantile modelling. In this paper, we focus on autoregressive modelling of p_t. We introduce a set of autoregressive logit models that can be viewed as an extension of the models of Rydberg and Shephard (2003), which they developed for the rather different problem of modelling the probability of the dichotomous event of a price change in trade-by-trade data.

A natural way to model p_t is to use a binary response variable defined as $I(y_t \leq Q)$, which takes a value of 1 if y_t is less than or equal to Q, and 0 otherwise. p_t is the conditional expectation of this binary variable. To estimate the parameters in a model for p_t, maximum likelihood can be used, based on a Bernoulli density. This is the approach taken with logistic regression. The use of the Bernoulli density is reasonable when modelling a binary response variable that was binary in its original form. However, if the binary response variable has been created, as in this paper, to indicate exceedances of an original variable that was not binary, then the use of the Bernoulli density seems inefficient, because it is affected only by whether or not the variable y_t is below the threshold Q. It would seem to be preferable to capture also the degree to which y_t is below Q. We do this by performing constrained maximum likelihood based on an asymmetric Laplace (AL) density. Maximizing an AL likelihood has been shown to be equivalent to quantile regression, which involves a time-varying quantile and a constant probability level. We adapt this for probability modelling by using a constant quantile Q and a time-varying probability p_t.

Section 2 discusses autoregressive models for exceedance probability prediction. Section 3 considers model estimation. Section 4 presents an empirical study that evaluates probability forecast accuracy using stock indices. Section 5 shows how the exceedance probability forecasts can be used within a new time-varying EVT approach for estimating value at risk (VaR) and expected shortfall (ES). Although VaR has received much attention in the research literature (see, for example, Kuester et al. 2006), future regulatory frameworks will put increased emphasis on ES (Embrechts et al. 2014).

2. Autoregressive modelling of the exceedance probability

2.1. A review of the literature

Given the common use of logistic regression for modelling a probability, it seems natural to consider some form of logistic autoregression for the time series modelling of an exceedance probability. Slud and Kedem (1994) model the exceedance probability using logistic regression, with lagged values of y_t as regressors. They apply the model to rainfall runoff data. Within a decomposition approach to modelling trade by trade data, Rydberg and Shephard (2003) model the probability of whether or not there is a change in the price of an asset, as each trade occurs. Their autoregressive logit model appears to be the first to include a lagged logit term. The dynamic binary response models of Kauppi and Saikkonen (2008) are of a similar form, with the difference that they generalise the choice of link function to be any distribution function. They developed the models for predicting the probability of a recession. The models are used by Nyberg (2011), who includes macroeconomic explanatory variables in order to forecast the direction of monthly excess stock returns. This is close to our focus, because it amounts to forecasting the probability of exceedance, albeit over a threshold of 0. Anatolyev and Gospodinov (2010) implement similar models, except that they use a logistic link function and no lagged logit term. The autoregressive models discussed in this section are estimated by maximum likelihood based on the Bernoulli density. Theoretical support for the models is presented by de Jong and Woutersen (2011).

2.2. A new set of autoregressive models

In this section, we introduce a new set of models that we term *conditional autoregressive logit* (CARL). These differ from previously developed autoregressive logit models, which have either not modelled exceedance probability or have focused on exceedance over a threshold of 0.

The standard approach to modelling a probability using a logit model is to allow the logistic function to vary between 0 and 1. However, this has little appeal for our application to financial returns with chosen threshold Q not close to 0, because the exceedance probability will be less than 0.5. In view of this, we formulate the CARL models as in expression (1), which restricts the probability p_t to vary between 0 and 0.5 for a negative threshold, and between 0.5 and 1 for a positive threshold. We found that preliminary empirical results supported the use of these restricted ranges. Our various CARL models differ in the specification of x_t, which is the logit of $(2p_t - I(Q>0))$.

$$p_t = \frac{0.5}{1+\exp(-x_t)} + 0.5I(Q>0) \tag{1}$$

The *CARL-Ind* model of expression (2) involves the lagged indicator $I(y_t<Q)$, as in the model of Anatolyev and Gospodinov (2010), and lagged logit x_t, as in the models of Rydberg and Shephard (2003).

$$x_t = \alpha_0 + \alpha_1 I(y_{t-1} < Q) + \beta_1 x_{t-1} \tag{2}$$

where α_i and β_1 are constant parameters. To allow a faster response to the change in volatility, we also consider the inclusion of $I(y_{t-1}>-Q)$, resulting in the *CARL-AsymInd* model:

$$x_t = \alpha_0 + \alpha_1 I(y_{t-1} < Q) + \alpha_2 I(y_{t-1} > -Q) + \beta_1 x_{t-1}$$

where α_i and β_1 are constant parameters. The binary variables in this model limit the model's ability to respond to changes in the volatility. This motivates the *CARL-Abs* model of expression (3). This contains the lagged absolute value of the return, which can be viewed as a proxy for the volatility.

$$x_t = \alpha_0 + \alpha_1 |y_{t-1}| + \beta_1 x_{t-1} \tag{3}$$

where α_i and β_1 are constant parameters.

The *CARL-AsymAbs* model of expression (4) allows for an asymmetric response of x_t, and hence p_t, to changes in the lagged absolute return. The similarity between the form of the GJRGARCH model of Glosten et al. (1993) and this CARL model suggests that it has the potential to capture the impact on p_t of the leverage effect, which is the tendency for volatility to be greater following a negative return than a positive return of equal size.

$$x_t = \alpha_0 + \alpha_1 |y_{t-1}| I(y_{t-1} \geq 0) + \alpha_2 |y_{t-1}| I(y_{t-1} < 0) + \beta_1 x_{t-1} \tag{4}$$

where α_i and β_1 are constant parameters.

To motivate two additional CARL models, let us consider, for a GARCH model, the probability p_t of y_t falling below a fixed threshold Q. With the usual GARCH assumption of a constant distribution F for the standardized returns, we can write $p_t = F((Q-\mu)/h_t^{\frac{1}{2}})$, where h_t is the variance, and μ is the mean, which we assume to be constant. This suggests that p_t can be estimated using a logistic function with the logit term a linear function of $h_t^{-\frac{1}{2}}$. Using a GARCH(1,1) structure for the variance, we get the logit model of expressions (5) and (6), which we term *CARL-Vol*.

$$x_t = \phi_0 + \phi_1 h_t^{-\frac{1}{2}} \tag{5}$$

$$h_t = \alpha_0 + \alpha_1 (y_{t-1} - \mu)^2 + \beta_1 h_{t-1} \tag{6}$$

where ϕ_i, α_i, β_1 and μ are constant parameters. We estimate μ using the mean of the in-sample returns, and, given that the variance of y_t is stationary, we set α_0=(1-α_1-β_1)h, where h is the variance of the in-sample returns (see, for example, the analogous expression for GARCH in Section 4.1.1, Franses and Van Dijk 2000). To avoid negative variance and ensure stationarity, we impose the constraints α_1, $\beta_1 \geq 0$ and $\alpha_1+\beta_1<1$. In comparison with GARCH modelling, the appeal of the *CARL-Vol* model, and indeed all the CARL models, is that a separate exceedance probability model can be estimated for different thresholds. This avoids a distributional assumption, and allows the dynamics to differ across the distribution. For example, the left tail of the distribution may evolve at a different rate to the right tail. By contrast, standard GARCH modelling involves a distributional assumption, and assumes an autoregressive model for the variance. However, we acknowledge that GARCH models have the computational advantage that just one model is needed to deliver exceedance probabilities for multiple thresholds. Another advantage of GARCH models is that, unlike CARL models, the exceedance probability estimate cannot be larger for the more extreme of two thresholds of the same sign. This is analogous to quantile crossing in quantile estimation (see Section 2.5, Koenker 2005).

To allow for the leverage effect, we use the structure of a GJRGARCH(1,1) model to give the *CARL-AsymVol* model of expressions (7) and (8):

$$x_t = \phi_0 + \phi_1 h_t^{-\frac{1}{2}} \tag{7}$$

$$h_t = \alpha_0 + \alpha_1 I(y_{t-1} \geq 0)(y_{t-1} - \mu)^2 + \alpha_2 I(y_{t-1} < 0)(y_{t-1} - \mu)^2 + \beta_1 h_{t-1} \tag{8}$$

where ϕ_i, α_i, β_1 and μ are constant parameters. We make the reasonable assumption that P(I($y_t \geq 0$))≈0.5, so that we can estimate α_0 as (1-0.5($\alpha_1+\alpha_2$)-β_1)h (see, for example, the analogous expression for GJRGARCH in Section 4.1.2, Franses and Van Dijk 2000). To avoid negative variance estimate and ensure stationarity, we impose the constraints α_1, α_2, $\beta_1 \geq 0$ and $0.5(\alpha_1+\alpha_2)+\beta_1<1$.

3. Parameter estimation for the CARL models

3.1. Maximum likelihood based on the Bernoulli distribution

Maximizing a Bernoulli likelihood is a standard approach to estimating a probabilistic model for a binary response variable. It can be used to estimate a model for the probability p_t of y_t falling below a chosen threshold Q, with the Bernoulli density specified as in the following expression:

$$f(y_t) = p_t^{I(y_t \leq Q)} (1 - p_t)^{(1 - I(y_t < Q))}$$

However, for modelling the exceedance probability, this approach seems inefficient, as it does not capture the extent to which the variable y_t is above or below the threshold Q. We aim to capture this by constructing the likelihood function from an asymmetric Laplace (AL) distribution.

3.2. Constrained maximum likelihood based on the asymmetric Laplace distribution

Before describing how we use the AL distribution to estimate models for the exceedance probability, we first discuss quantile regression. The quantile regression minimization is presented in expression (9). It can be used to estimate the parameters in a model for the quantile Q_t of a variable y_t, corresponding to a chosen probability level p, based on n observations (Koenker 2005).

$$\min \sum_{t=1}^{n} (y_t - Q_t)(p - I(y_t \leq Q_t)) \tag{9}$$

The resulting estimates of Q_t satisfy expression (10) (see Theorem 2.2, Koenker 2005), which has the interpretation that the quantile estimates have correct in-sample unconditional coverage.

$$\frac{1}{n} \sum_{t=1}^{n} I(y_t \leq Q_t) = p \tag{10}$$

In this paper, we are interested in the time-varying probability p_t of exceedance over a constant threshold Q. This can be viewed as a constant quantile Q for which there is a time-varying probability level p_t. If the sequence of p_t is known, the quantile Q can be estimated by the following simple adaptation of the minimization in expression (9):

$$\min \sum_{t=1}^{n} (y_t - Q)(p_t - I(y_t \leq Q)) \tag{11}$$

Following a similar derivation to that used for expression (10), we find that the estimate of Q satisfies expression (12) (see Appendix A), which shows that, on average, the estimate of Q has correct in-sample unconditional coverage.

$$\frac{1}{n}\sum_{t=1}^{n} I\left(y_t \leq Q\right) = \frac{1}{n}\sum_{t=1}^{n} p_t \tag{12}$$

We will return to the issue of a time-varying probability level for a fixed quantile after discussing the connection between quantile regression and the AL distribution. A fundamental appeal of quantile regression is that an assumption is not needed for the distribution of y_t. However, Koenker and Machado (1999) point out that the quantile regression minimization in expression (9) is equivalent to maximum likelihood based on y_t specified as having the AL density function of expression (13), where p is chosen as the probability level of interest, σ is a scale parameter, and Q_t is the AL density's time-varying location, which is also the p-th quantile of the density.

$$f\left(y_t\right) = \frac{p(1-p)}{\sigma} \exp\left(-\left(y_t - Q_t\right)\left(p - I\left(y_t \leq Q_t\right)\right)/\sigma\right) \tag{13}$$

Koenker and Machado have described as "rather implausible" the idea that data would follow an AL distribution. This view seems reasonable for our application, as the AL density plots of Yu and Zhang (2005) are very different in shape to the densities typically assumed for financial returns data. For example, GARCH models with a Student-*t* distribution are often used to model such data. Although it is rather unrealistic to assume that the data follows an AL distribution, its use within a quasi-maximum likelihood framework has led to useful developments for quantile regression, including statistical inference (Koenker and Machado 1999) and Bayesian quantile regression (Yu and Stander 2007).

In this paper, we propose a quasi-maximum likelihood based on an AL distribution to estimate models for the exceedance probability. We are not motivated by an interest in statistical inference or Bayesian methods, but instead by the desire to capture, in the model estimation, the degree to which an observation exceeds the threshold. In our proposal, we rewrite the AL density, so that instead of a fixed probability and time-varying location, the density has a time-varying probability p_t and fixed location parameter Q, as in expression (14). Note that, for this density, p_t is the probability of y_t falling below Q.

$$f(y_t)=\frac{p_t(1-p_t)}{\sigma_t}\exp\left(-(y_t-Q)(p_t-I(y_t\le Q))/\sigma_t\right) \tag{14}$$

Since the asset returns y_t possess heteroskedasticity, p_t will vary over time, in which case it seems overly restrictive to constrain the scale to remain constant. Therefore, we allow the scale σ_t to be time-varying. In expression (13) of their paper, Yu and Zhang (2005) provide the maximum likelihood estimator of the scale of a static AL density. We adapt this to give the time-varying scale σ_t in expression (15), where μ is the mean of the in-sample observations.

$$\sigma_t=\frac{p_t(1-p_t)(\mu-Q)}{(1-2p_t)} \tag{15}$$

Our proposal is to estimate a model for the probability p_t of y_t falling below a chosen threshold Q by maximizing the likelihood based on the AL density of expression (14), with σ_t replaced by expression (15). However, if we do not constrain this optimisation, it will not necessarily deliver consistent estimates for the probability p_t of y_t falling below Q. Indeed, the estimates for p_t may well not satisfy expression (12), which is a necessary condition for the probability p_t of y_t falling below Q. Given the one-to-one correspondence between the quantiles and probability levels of a monotonic probability distribution function, we ensure consistent estimation for the probability p_t of y_t falling below Q by performing the likelihood maximisation with expression (12) as a constraint. We impose this constraint by subtracting, from the sum of log likelihoods, a penalty term equal to 10^5 multiplied by the square of the difference between the left and right hand sides of expression (12). The resultant penalised log likelihood is given in expression (16). It is well-known that, under some regularity conditions, penalised likelihood parameter estimators are asymptotically consistent and Gaussian (see, for example, Green, 1987).

$$\sum_{t=1}^{n}\left(\ln\left(p_t(1-p_t)\right)-\ln\left(\frac{(1-2p_t)}{p_t(1-p_t)(\mu-Q)}\right)-(y_t-Q)(p_t-I(y_t\le Q))\frac{(1-2p_t)}{p_t(1-p_t)(\mu-Q)}\right)$$
$$-10^5\left(\frac{1}{n}\sum_{t=1}^{n}I(y_t\le Q)-\frac{1}{n}\sum_{t=1}^{n}p_t\right)^2 \tag{16}$$

In proposing this AL-based estimation approach, as an alternative to a Bernoulli likelihood, we are essentially responding to Gneiting's (2008) call to give consideration to the estimation of probability models when the ultimate aim is prediction. In the next section, we compare the accuracy of prediction based on the two different approaches to estimation described in Sections 3.1 and 3.2.

4. Empirical evaluation of exceedance probability forecasts

We evaluated day-ahead forecasts using daily log returns for the FTSE 100, NIKKEI 225 and S&P 500 stock indices. Each series consisted of the 3500 daily log returns ending on 16 April 2013. We downloaded the data from Yahoo! Finance (finance.yahoo.com). In order to evaluate the models of Section 2 for a variety of low and high thresholds, of different signs, we considered the following six thresholds: -3%, -2%, -1%, 1%, 2% and 3%. To give an idea of the locations of these thresholds, for the S&P 500 series, the proportions of observations below these thresholds are 1.9%, 5.4%, 16.3%, 84.4%, 95.3% and 98.2%, respectively. For each series, we used the first 2500 observations to estimate model parameters, and then produced out-of-sample day-ahead forecasts for the next 250 periods. We then moved the estimation window of 2500 periods forward by 250 periods, re-estimated the parameters, and again produced out-of-sample day-ahead forecasts for the next 250 periods. We did this twice more to give a total of 1000 out-of-sample probability forecasts. We should emphasise that these are out-of-sample forecasts, because each forecast was generated using only data on or before the forecast origin. The structure of our analysis is that approximately 10 years of data is used for model estimation, and the models are re-estimated approximately each year.

4.1. Probability forecasting methods

As a simple benchmark, for each threshold and each series, we produced probability forecasts for each out-of-sample period as the proportion of the previous m=2500 observations that were less than or equal to the threshold. We present this estimator in expression (17), and refer to it as historical simulation because, in using the distribution of historical observations, it is similar to the historical simulation VaR estimator. We also implemented the estimator using m=250. In addition, we produced probability forecasts using a filtered historical simulation approach, which involved an exponentially weighted moving average model for the variance, with optimised parameter, and the distribution of all 2500 standardised historical returns.

$$\hat{p}_t = \frac{1}{m}\sum_{i=1}^{m} I\left(y_{t-i} \le Q\right) \tag{17}$$

We also fitted GARCH(1,1) and GJRGARCH(1,1) models, with parameters estimated using a Student-*t* distribution. The GJRGARCH(1,1) model is written as:

$$y_t = \mu + \varepsilon_t$$

$$\varepsilon_t = h_t^{\frac{1}{2}}\eta_t$$

$$h_t \;=\; \alpha_0 + I(\varepsilon_{t-1} \ge 0)\alpha_1\varepsilon_{t-1}^2 + I(\varepsilon_{t-1} < 0)\alpha_2\varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

where y_t is the return; μ is a constant mean; ε_t is a heteroskedastic term with variance h_t; η_t is a white noise process with Student-*t* distribution; and α_i and β_1 are constant parameters. Setting α_1=α_2 delivers the GARCH(1,1) model. The probability of y_t falling below threshold Q is given by $t_v((Q-\mu)/h_t^{\frac{1}{2}})$, where t_v is the Student-t distribution function. Day-ahead probability forecasts were produced using the model's forecast for the variance. We also fitted the (asymmetric power) APARCH(1,1) model of Ding et al. (1993), which models the standard deviation σ_t as:

$$\sigma_t^{\delta} \;=\; \alpha_0 + \alpha_1\left(\left|\varepsilon_{t-1}\right| - \alpha_2\varepsilon_{t-1}\right)^{\delta} + \beta_1\sigma_{t-1}^{\delta}$$

where α_i, β_1 and δ are constant parameters. In addition, we implemented this model with a non-central *t* distribution, as proposed by Krause and Paolella (2014). We refer to this as *NCTAPARCH*.

We implemented the six CARL models, with estimation first based on a Bernoulli likelihood, as discussed in Section 3.1, and then using the constrained maximization of the AL likelihood, as described in Section 3.2. For the CARL-Ind, CARL-AsymInd, CARL-Abs and CARL-AsymAbs models, we calculated the initial probability estimate p_0 as the proportion of the first 100 returns that were less than the threshold. Using p_0 and expression (1), we obtained an initial value x_0 for use in the autoregressive logit models. This requires that $p_0 \in (0,0.5)$ for a negative threshold, $p_0 \in (0.5,1)$ for a positive threshold, and so when this was not true, we recalculated p_0 from all in-sample observations. For the CARL-Vol and CARL-AsymVol models, we calculated μ as the mean of the estimation sample of 2500 returns. These two models require an initial variance estimate h_0, and we calculated this as the variance of the first 100 returns. For the CARL models, we performed a similar optimization approach to that used by Engle and Manganelli (2004) for CAViaR models. The optimisation proceeded by sampling 10^4 vectors of parameters using a uniform random number generator between a lower and upper bound, which were set for each parameter based on initial experimentation. Of the 10^4 sampled vectors, the three that delivered the highest likelihood values were used, in turn, as the initial vector in a quasi-Newton algorithm. The resulting vector, corresponding to the highest likelihood, was chosen as the final parameter vector. As discussed in Section 3.2, when estimation was based on an asymmetric Laplace density, a penalised likelihood was used as the objective function.

In the following expressions, we present the six CARL models for the threshold Q=-2%, with parameter estimation based on the AL density, using the first 2500 S&P 500 returns.

CARL-Ind: $x_t = -0.220 + 0.662 I(y_{t-1} < Q) + 0.919 x_{t-1}$

CARL-AsymInd: $x_t = -0.211 + 0.668 I(y_{t-1} < Q) - 0.047 I(y_{t-1} > -Q) + 0.922 x_{t-1}$

CARL-Abs: $x_t = -0.224 + 8.141 |y_{t-1}| + 0.933 x_{t-1}$

CARL-AsymAbs: $x_t = -0.141 - 2.562 |y_{t-1}| I(y_{t-1} \geq 0) + 11.506 |y_{t-1}| I(y_{t-1} < 0) + 0.956 x_{t-1}$

CARL-Vol:
$$x_t = 1.423 - 0.045 h_t^{-\frac{1}{2}}$$
$$h_t = 0.000 + 0.036 (y_{t-1} - \mu)^2 + 0.940 h_{t-1}$$

CARL-AsymVol:
$$x_t = 1.695 - 0.050 h_t^{-\frac{1}{2}}$$
$$h_t = 0.000 + 0.000 I(y_{t-1} \geq 0)(y_{t-1} - \mu)^2 + 0.073 I(y_{t-1} < 0)(y_{t-1} - \mu)^2 + 0.930 h_{t-1}$$

The following expressions present the models estimated using the Bernoulli likelihood:

CARL-Ind: $x_t = -0.131 + 0.556 I(y_{t-1} < Q) + 0.958 x_{t-1}$

CARL-AsymInd: $x_t = -0.137 + 0.549 I(y_{t-1} < Q) + 0.039 I(y_{t-1} > -Q) + 0.956 x_{t-1}$

CARL-Abs: $x_t = -0.256 + 12.794 |y_{t-1}| + 0.942 x_{t-1}$

CARL-AsymAbs: $x_t = -0.170 - 2.578 |y_{t-1}| I(y_{t-1} \geq 0) + 18.431 |y_{t-1}| I(y_{t-1} < 0) + 0.961 x_{t-1}$

CARL-Vol:
$$x_t = 1.643 - 0.047 h_t^{-\frac{1}{2}}$$
$$h_t = 0.000 + 0.045 (y_{t-1} - \mu)^2 + 0.949 h_{t-1}$$

CARL-AsymVol:
$$x_t = 1.793 - 0.049 h_t^{-\frac{1}{2}}$$
$$h_t = 0.000 + 0.000 I(y_{t-1} \geq 0)(y_{t-1} - \mu)^2 + 0.077 I(y_{t-1} < 0)(y_{t-1} - \mu)^2 + 0.955 h_{t-1}$$

In the Ind and AsymInd models, the positive coefficients for $I(y_{t-1}<Q)$ are intuitive, as they imply that if the return is below the threshold, the logit rises, and hence the probability rises. The AsymAbs and AsymVol models show that, in comparison with a positive return, a negative return of equal size leads to a larger rise in the logit, and hence a larger rise in the probability.

In Fig. 1, for Q=-2%, and for the first 2750 days of the S&P 500 series, we plot the exceedance probability estimates produced by GJRGARCH, and the CARL-AsymVol models. For the last 250 days, the probability estimates are out-of-sample day-ahead forecasts. It is reassuring to see correspondence between the returns series exceeding the threshold and the magnitude of the probability forecasts. With CARL model parameters estimated using the Bernoulli likelihood, the predictions can be seen to be more responsive to changes in the returns series than when estimation was based on the AL likelihood. This tended also to be the case for the other CARL models, thresholds and series that we considered.

For the threshold of -2% for the S&P 500 returns, Table 1 shows the correlation matrix for the 1000 out-of-sample probability forecasts from the six CARL models estimated using the two different approaches. We have underlined the correlations corresponding to the cases where the same CARL model has been estimated using the two different approaches. As expected, the correlation values are reasonably high. However, many values are noticeably lower than 1, and this was also the case for the other thresholds and stock indices. In view of this, we should anticipate differences in the resulting out-of-sample accuracy when using the different CARL models and the two different estimation approaches.

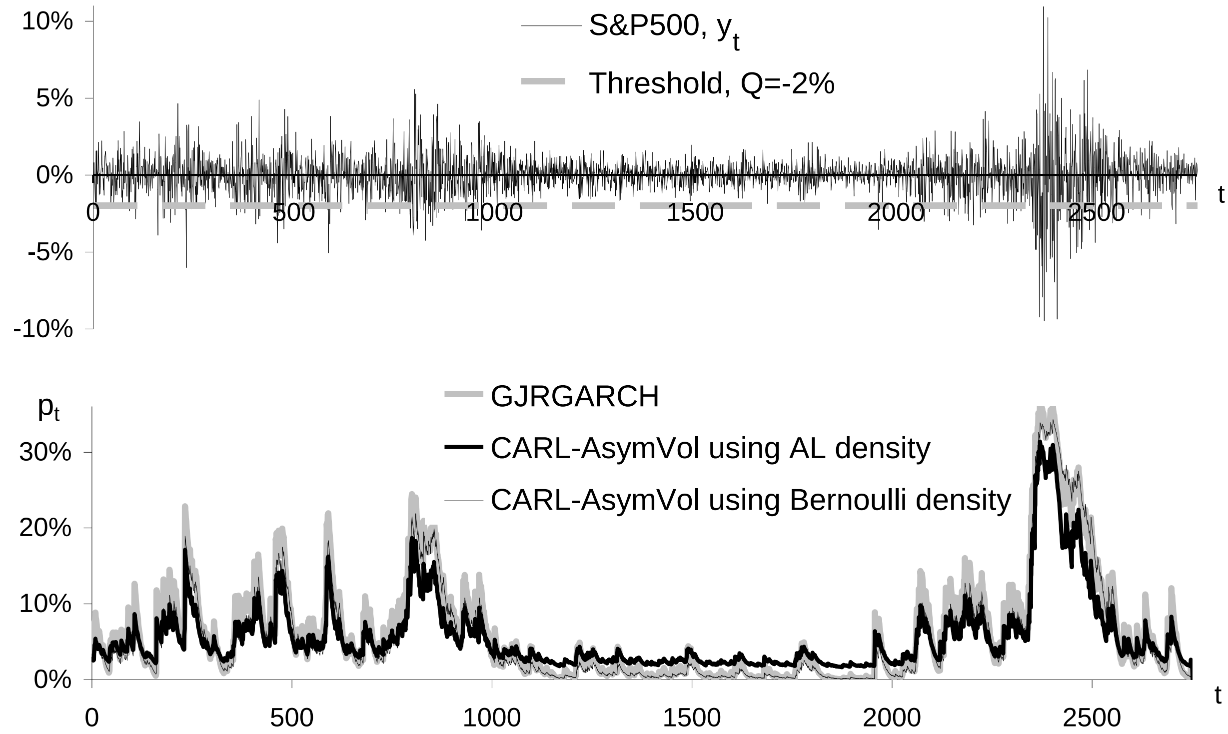

Fig. 1. S&P 500 returns (in upper panel), and probability estimates of exceedance below the threshold *Q*=-2% (in lower panel).

Table 1. For the threshold of -2% for the S&P 500 time series, correlation between out-of-sample probability forecasts from CARL models.

	Ind AL	Ind Bernoulli	AsymInd AL	AsymInd Bernoulli	Abs AL	Abs Bernoulli	AsymAbs AL	AsymAbs Bernoulli	Vol AL	Vol Bernoulli	AsymVol AL	AsymVol Bernoulli
Ind AL	1											
Ind Bernoulli	0.941	1										
AsymInd AL	1.000	0.936	1									
AsymInd Bernoulli	0.936	1.000	0.930	1								
Abs AL	0.867	0.887	0.861	0.887	1							
Abs Bernoulli	0.861	0.919	0.854	0.920	0.986	1						
AsymAbs AL	0.873	0.814	0.872	0.810	0.913	0.880	1					
AsymAbs Bernoulli	0.900	0.887	0.898	0.884	0.945	0.939	0.978	1				
Vol AL	0.846	0.855	0.840	0.856	0.984	0.960	0.919	0.937	1			
Vol Bernoulli	0.838	0.882	0.830	0.885	0.976	0.970	0.899	0.936	0.989	1		
AsymVol AL	0.886	0.833	0.884	0.830	0.946	0.912	0.976	0.967	0.962	0.939	1	
AsymVol Bernoulli	0.880	0.891	0.875	0.891	0.965	0.955	0.949	0.971	0.977	0.983	0.976	1

Notes. Underlining indicates the same CARL model estimated using the two different approaches.

4.2. Out-of-sample evaluation

We evaluated the day-ahead probability forecasts from each method using the Brier score (see, for example, Gneiting et al. 2007; Wilks 2011), which is presented in expression (18).

$$\text{Brier score} = \frac{1}{N}\sum_{t=n+1}^{n+N}\left(I\left(y_t \leq Q\right) - \hat{p}_t\right)^2 \tag{18}$$

where $\hat{p}_t$ is the probability forecast for the event that y_t falls below threshold Q; n is the size of the estimation sample; and N is the number of out-of-sample periods. The Brier score can be viewed as the mean squared error for a set of probability forecasts, with $I(y_t \leq Q)$ acting as the proxy for the actual probability. The results for the S&P 500 series are presented in Table 2, with standard errors calculated using the expression of Bradley et al. (2008) with Wilks' (2010) adjustment for autocorrelation.

Table 2. For probability forecasts, Brier score (×100) for S&P 500 with standard errors in parentheses.

	Threshold					
	-3%	-2%	-1%	1%	2%	3%
Historical simulation 2500	1.20(0.33)	4.21(0.58)	11.99(0.74)	13.43(0.81)	4.02(0.57)	1.00(0.30)
Historical simulation 250	1.40(0.33)	4.57(0.55)	12.46(0.72)	13.61(0.73)	4.25(0.55)	1.13(0.30)
Filtered historical simulation	1.18(0.31)	4.13(0.54)	11.76(0.73)	12.87(0.76)	3.75(0.51)	0.95(0.27)
GARCH *with Student-t*	1.19(0.31)	4.14(0.53)	11.90(0.68)	12.90(0.71)	3.77(0.49)	0.96(0.27)
GJRGARCH *with Student-t*	1.17(0.30)	4.14(0.54)	11.77(0.69)	12.67(0.72)	3.70(0.49)	0.93(0.26)
APARCH *with Student-t*	1.17(0.31)	4.13(0.54)	11.72(0.71)	12.65(0.73)	3.68(0.49)	0.93(0.26)
NCTAPARCH	1.16(0.31)	4.11(0.56)	**11.64**(0.75)	**12.70**(0.78)	**3.69**(0.52)	**0.92**(0.28)
CARL using Asymmetric Laplace						
CARL-Ind	1.18(0.33)	4.12(0.57)	11.78(0.75)	13.43(0.80)	3.93(0.55)	0.99(0.30)
CARL-AsymInd	1.18(0.33)	4.12(0.57)	11.77(0.74)	12.88(0.80)	3.80(0.55)	0.96(0.29)
CARL-Abs	1.17(0.33)	4.11(0.57)	11.68(0.73)	12.96(0.81)	3.85(0.56)	0.95(0.29)
CARL-AsymAbs	1.17(0.33)	4.12(0.58)	11.69(0.74)	12.85(0.81)	3.86(0.57)	0.96(0.30)
CARL-Vol	1.16(0.32)	**4.09**(0.56)	11.72(0.72)	12.90(0.78)	3.76(0.52)	0.94(0.28)
CARL-AsymVol	**1.15**(0.31)	**4.09**(0.56)	11.66(0.72)	12.73(0.78)	3.70(0.51)	**0.92**(0.27)
CARL using Bernoulli						
CARL-Ind	1.18(0.33)	4.13(0.56)	11.81(0.75)	13.43(0.72)	4.03(0.53)	0.99(0.30)
CARL-AsymInd	1.18(0.33)	4.14(0.56)	11.81(0.75)	12.84(0.81)	3.88(0.57)	0.94(0.29)
CARL-Abs	1.17(0.32)	4.12(0.56)	11.85(0.71)	12.91(0.79)	3.82(0.54)	0.95(0.29)
CARL-AsymAbs	1.17(0.33)	4.12(0.57)	11.86(0.73)	12.81(0.81)	**3.68**(0.53)	0.96(0.30)
CARL-Vol	1.17(0.31)	4.12(0.55)	11.80(0.71)	12.90(0.78)	3.77(0.52)	0.94(0.27)
CARL-AsymVol	1.16(0.31)	4.11(0.54)	11.72(0.72)	12.71(0.77)	3.71(0.51)	**0.92**(0.26)

Notes. Lower values are better. Bold indicates best Brier score method in each column. Underlining indicates whether AL or Bernoulli was better for each CARL model.

Table 3. For probability forecasts, Brier skill score for the S&P 500.

	Threshold						
	-3%	-2%	-1%	1%	2%	3%	Geometric mean
Historical simulation 250	-17.0	-8.6	-3.9	-1.3	-5.6	-13.3	-8.4
Filtered historical simulation	1.4	2.0	1.9	4.2	6.8	5.4	3.6
GARCH *with Student-t*	0.2	1.6	0.7	3.9	6.3	3.6	2.7
GJRGARCH *with Student-t*	1.8	1.8	1.8	5.6	8.1	6.6	4.3
APARCH *with Student-t*	2.2	2.0	2.2	**5.8**	8.4	7.1	4.6
NCTAPARCH	3.0	2.5	**2.9**	5.5	8.3	7.9	5.0
CARL using Asymmetric Laplace							
CARL-Ind	1.4	2.2	1.7	0.0	2.2	1.0	1.4
CARL-AsymInd	1.5	2.2	1.8	4.1	5.4	3.8	3.1
CARL-Abs	2.0	2.4	2.6	3.5	4.3	4.7	3.2
CARL-AsymAbs	2.4	2.2	2.5	4.3	4.0	4.1	3.3
CARL-Vol	3.0	2.8	2.2	4.0	6.5	5.9	4.0
CARL-AsymVol	**3.7**	**3.0**	2.7	5.2	8.0	8.1	**5.1**
CARL using Bernoulli							
CARL-Ind	1.2	1.9	1.5	0.0	-0.3	1.1	0.9
CARL-AsymInd	1.2	1.8	1.5	4.4	3.6	6.2	3.1
CARL-Abs	1.8	2.3	1.1	3.9	5.0	5.1	3.2
CARL-AsymAbs	2.2	2.3	1.1	4.6	**8.5**	3.6	3.7
CARL-Vol	2.2	2.3	1.6	4.0	6.4	6.1	3.7
CARL-AsymVol	2.6	2.5	2.2	5.4	7.8	**8.4**	4.8

Notes. Higher values are better. Bold indicates best method in each column. Underlining indicates whether AL or Bernoulli was better for each CARL model.

For each method, we calculated the Brier skill score, which is presented in expression (19) (see, for example, Wilks 2011). This measure compares the Brier score to a reference method, which we chose as the historical simulation approach based on 2500 observations.

$$\text{Brier skill score} = \left(1 - \frac{\sum_{t=n+1}^{n+N}\left(I(y_t \le Q) - \hat{p}_t\right)^2}{\sum_{t=n+1}^{n+N}\left(I(y_t \le Q) - \hat{p}_{reference,t}\right)^2}\right) \times 100 \tag{19}$$

The Brier skill score results for the S&P 500 series are presented in Table 3. Higher values indicate superior accuracy, and positive values indicate outperformance of the reference method. The final column summarizes performance, for each model, across the six thresholds. To obtain this column, we calculated the geometric mean of the ratios of the Brier score for each method to the Brier score for the reference method, then subtracted this from one, and multiplied the result by 100. A similar calculation was used to produce Table 4, which averages the Brier skill score across the three indices. In the tables, for each threshold, and for the summary column, bold indicates the best result. For each CARL model, underlining indicates whether the AL or Bernoulli likelihoods led to superior performance. Model asymmetry led to improved accuracy, with GJRGARCH, CARL-AsymInd, CARL-AsymAbs and CARL-AsymVol outperforming GARCH, CARL-Ind, CARL-Abs and CARL-Vol, respectively. Asymmetry is a feature of APARCH, and, interestingly, the two versions of this model also perform well. Table 3 shows that, for the S&P 500 series, the best results overall were produced by CARL-AsymVol and NCTAPARCH. This is also the case in Table 4 for the results averaged across the three series. The tables show that, for the CARL models, estimation based on the AL density was generally slightly better than using the Bernoulli density.

Table 4. For probability forecasts, Brier skill score averaged across the three stock indices.

	Threshold						
	-3%	-2%	-1%	1%	2%	3%	Geometric mean
Historical simulation	-11.9	-5.1	-2.8	-1.0	-3.6	-7.9	-5.5
Filtered historical simulation	-1.0	1.5	0.3	2.2	4.7	6.6	2.3
GARCH with Student-t	-1.2	1.1	-0.3	2.3	4.1	5.1	1.8
GJRGARCH with Student-t	0.0	1.4	1.0	3.4	5.3	6.5	2.9
APARCH with Student-t	1.1	1.8	1.2	**3.7**	5.5	7.0	3.4
NCTAPARCH	2.5	**2.5**	**1.7**	3.4	**6.1**	8.1	**4.0**
CARL using Asymmetric Laplace							
CARL-Ind	1.8	<u>1.9</u>	<u>0.9</u>	-0.6	<u>1.4</u>	<u>1.7</u>	<u>1.2</u>
CARL-AsymInd	<u>1.9</u>	<u>1.8</u>	<u>0.8</u>	<u>2.3</u>	<u>4.5</u>	4.1	<u>2.6</u>
CARL-Abs	<u>1.7</u>	1.7	<u>1.0</u>	2.0	3.6	4.6	2.4
CARL-AsymAbs	<u>2.6</u>	<u>2.0</u>	<u>1.6</u>	2.3	3.8	4.9	2.8
CARL-Vol	<u>1.7</u>	2.0	<u>0.6</u>	2.3	<u>4.7</u>	6.6	<u>3.0</u>
CARL-AsymVol	<u>**2.8**</u>	<u>2.4</u>	<u>1.5</u>	<u>3.1</u>	<u>5.8</u>	8.0	<u>3.9</u>
CARL using Bernoulli							
CARL-Ind	<u>1.9</u>	1.5	0.8	<u>0.0</u>	0.3	1.4	1.0
CARL-AsymInd	1.5	1.5	0.6	2.2	3.1	<u>5.4</u>	2.4
CARL-Abs	0.6	1.7	-0.3	<u>2.2</u>	<u>4.3</u>	<u>6.0</u>	2.4
CARL-AsymAbs	2.2	1.9	0.4	<u>2.4</u>	<u>**6.1**</u>	<u>6.2</u>	<u>3.2</u>
CARL-Vol	0.9	2.0	0.0	2.3	4.6	<u>7.0</u>	2.8
CARL-AsymVol	1.4	2.3	1.0	3.0	<u>5.8</u>	<u>**8.7**</u>	3.7

Notes. Higher values are better. Bold indicates best method in each column. Underlining indicates whether AL or Bernoulli was better for each CARL model.

5. An application of exceedance probability forecasts within a new EVT method

In this section, we use exceedance probability forecasts within a peaks over threshold (POT) EVT approach to estimate value at risk (VaR), which is a conditional tail quantile, and expected shortfall (ES), which is the conditional expectation of the return, given that it exceeds the VaR. After reviewing the POT method for i.i.d. observations, we present our adaptation for financial returns and an empirical study.

5.1. Peaks over threshold EVT for i.i.d. observations

The POT method considers exceedances of a variable y_t over a typically high threshold, Q. Consider first estimation of the VaR in the upper tail of the returns distribution. The essence of the POT approach is captured by expression (20), which indicates that an alternative to estimating the VaR from the returns distribution is to work with the distribution of exceedances.

$$\Pr\left(y_t > VaR_t \mid y_t > Q\right) = \frac{\Pr\left(y_t > VaR_t\right)}{\Pr\left(y_t > Q\right)} \tag{20}$$

Assuming i.i.d. observations, the number of exceedances has a Poisson distribution and, for large samples and a high threshold, the exceedances z_i obey a generalized Pareto distribution (GPD), with scale parameter s and shape parameter ξ. The GPD has the following form:

$$G(z) = \begin{cases} 1-\left(1+\xi z/s\right)^{-1/\xi} & \xi \neq 0 \\ 1-\exp\left(-z/s\right) & \xi = 0 \end{cases} \tag{21}$$

Substituting 1-$G(VaR_t\text{-}Q)$ into the left hand side of expression (20) delivers the VaR estimate:

$$\hat{VaR}_t = \begin{cases} Q + \dfrac{s}{\xi}\left(\left(\dfrac{1-\theta}{\Pr\left(y_t > Q\right)}\right)^{-\xi} - 1\right) & \xi \neq 0 \\ Q + s\ln\left(\dfrac{\Pr\left(y_t > Q\right)}{1-\theta}\right) & \xi = 0 \end{cases} \tag{22}$$

where θ is the VaR probability level. The corresponding ES estimate is given by:

$$\hat{ES}_t = \left(\frac{\hat{VaR}_t + s - \xi Q}{1-\xi}\right) \tag{23}$$

For the estimation of VaR in the lower tail of the distribution, the VaR and ES expressions are:

$$\hat{VaR}_t = \begin{cases} Q - \dfrac{s}{\xi}\left(\left(\dfrac{\theta}{\Pr(y_t < Q)}\right)^{-\xi} - 1\right) & \xi \neq 0 \\ Q - s\ln\left(\dfrac{\Pr(y_t < Q)}{\theta}\right) & \xi = 0 \end{cases} \tag{24}$$

$$\hat{ES}_t = \left(\frac{\hat{VaR}_t - s + \xi Q}{1-\xi}\right) \tag{25}$$

Assuming i.i.d. observations, the exceedance probability, which is $\Pr(y_t > Q)$ in expression (22) and $\Pr(y_t < Q)$ in expression (24), is the mean of the Poisson process, and is estimated as the proportion of observations exceeding Q. The GPD parameters, s and ξ, are estimated using maximum likelihood.

The i.i.d. assumption is inappropriate for financial returns, because they typically possess heteroskedasticity. To address this, McNeil and Frey (2000) apply the POT method to residuals standardised by GARCH conditional volatility estimates. However, this implicitly assumes that the tails have the same dynamic behaviour as the rest of the distribution (Manganelli and Engle 2004). Bali and Neftci (2003) consider autoregressive models for the location and scale of the GPD of the POT method. Our method has similarities to this, although we use different autoregressive models, and we allow the exceedance probability to vary over time. Chavez-Demoulin and Davison (2005) consider a time-varying POT approach for modelling extremes in a data set with multiple time series. Another dynamic POT method is the Bayesian approach of Chavez-Demoulin et al. (2014), which enables the POT parameters to adapt to possible non-stationarities.

5.2. A new time-varying peaks over threshold EVT method

In this section, we introduce a new time-varying POT (TVPOT) approach. It involves three steps. First, we choose a suitable threshold. Second, we estimate the time-varying exceedance probability using a CARL model. Third, we fit a GPD to the exceedances, with an autoregressive model for the scale. From these models, we get forecasts of the exceedance probability and scale, which we use in expressions (22) to (25) of Section 5.1 to deliver VaR and ES forecasts. GAUSS computer code is available on request.

Step 1: Finding a suitable threshold

The choice of the threshold Q is important for the POT method. The GPD is more appropriate for exceedances beyond a more extreme threshold, but a less extreme threshold provides more exceedances from which to estimate the GPD parameters. Chavez-Demoulin et al. (2014) note that, in practice, there is some arbitrariness in the choice of Q. They set Q to be such that 10% of the observations from the last year are exceedances. In our TVPOT approach, we have a time-varying exceedance probability, and to estimate the GPD, a necessary condition for the choice of Q is that, for VaR with probability level θ, the exceedance probability is greater than min(θ,1-θ) for each in-sample period with an exceedance. In view of this, we need to select a value for Q that satisfies this condition, rather than simply set Q to be 10%. Our approach to selecting Q involves a numerical search. We start with a value of Q for which 10% of the in-sample periods are exceedances. We then increase Q in increments of 1% until the condition is satisfied. In our empirical work, the resulting percentages of exceedances were similar for all three indices, with values of approximately 10%, 10%, 14%, 18%, 11% and 10% for θ=0.5%, 1%, 5%, 95%, 99% and 99.5%, respectively. The corresponding values of Q were also quite similar for the three indices, with averages of approximately -1.6%, -1.6%, -1.2%, 1.0%, 1.4% and 1.5%, respectively.

Step 2: Modelling the exceedance probability

The time-varying probability of exceeding Q is estimated using a CARL model. Our empirical work used the CARL-AsymVol model, estimated using constrained maximum likelihood based on the AL density. We chose this CARL model, as it was the best performing CARL model in terms of probability forecasting in Section 4.

Step 3: Fitting a GPD with autoregressive scale

The final step of our approach involves fitting a GPD to the exceedances beyond Q, with an autoregressive model for the scale. Following Bali and Neftci (2003) and Chavez-Demoulin et al. (2014), we use a constant shape parameter. We note that, for the GPD of expression (21), the mean is $s/(1-\xi)$ and the variance is $s^2/\left((1-\xi)^2(1-2\xi)\right)$. In view of this, we initialize the scale to be

$\sqrt{(1-\xi)^2(1-2\xi)}$ multiplied by the standard deviation of the exceedances in the first 100 periods. For a period i with a non-zero exceedance z_i, we update the scale estimate using:

$$\hat{s}_i^{\,2} \;=\; a_0 + a_1\left(z_i - \hat{s}_{i-1}/(1-\xi)\right)^2 + b_1 s_{i-1}^2$$

where a_i and b_1 are constant parameters. We assume that the scale is stationary, and set $a_0=(1-a_1-b_1)S^2$, where S^2 is the square of the unconditional scale of the in-sample exceedances, which we estimate as the product of $\left((1-\xi)^2(1-2\xi)\right)$ and the variance of the exceedances. To avoid this becoming negative, we impose the constraint $\xi<0.5$. To avoid a negative value and to ensure stationarity for estimates of the scale squared, we impose the constraints a_1, $b_1 \geq 0$ and $a_1+b_1<1$.

An alternative model for the scale is the following asymmetric formulation, which responds to exceedances in both tails of the returns distribution:

$$\hat{s}_i^{\,2} \;=\; a_0 + a_1 I(z_i>0)\left(z_i - \hat{s}_{i-1}/(1-\xi)\right)^2 + a_2 I(w_i>0)\left(w_i - \hat{s}_{i-1}/(1-\xi)\right)^2 + b_1 s_{i-1}^2$$

where a_i and b_1 are constant parameters, w_i is an exceedance beyond $-Q$, and i is a period with non-zero z_i or w_i. We make the assumption that, for these periods, $P(I(z_i>0))\approx 0.5$ and $P(I(w_i>0))\approx 0.5$, so that we can estimate a_0 as $(1-0.5(a_1+a_2)-b_1)S^2$ (see, for example, the analogous expression for GJRGARCH in Section 4.1.2, Franses and Van Dijk 2000). For this model, we impose the constraints $\xi<0.5$; a_1, a_2, $b_1\geq 0$; and $0.5(a_1+a_2)+b_1<1$.

We estimated the parameters a_i, b_1 and ξ by maximum likelihood based on the GPD. The GPD density and sum of log likelihoods are given in expressions (26) and (27), respectively, where n_z is the number of exceedances beyond Q in the estimation sample.

$$g(z_i) = \begin{cases} \dfrac{1}{s_i}\left(1+\xi z_i/s_i\right)^{-(1+\xi)/\xi} & \xi \neq 0 \\ \dfrac{1}{s_i}\exp\left(-z_i/s_i\right) & \xi = 0 \end{cases} \tag{26}$$

$$LL = \begin{cases} -\sum_{i=1}^{n_z} \ln(s_i) + (1+1/\xi)\ln\left(1+\xi z_i/s_i\right) & \xi \neq 0 \\ -\sum_{i=1}^{n_z} \ln(s_i) + z_i/s_i & \xi = 0 \end{cases} \tag{27}$$

In expressions (28) and (29), we present the symmetric and asymmetric scale models, used to derive the 99% VaR, with estimation based on the first 2500 S&P 500 returns.

$$\hat{s}_i^{\,2} \;=\; 0.000001 + 0.177\left(z_i - \hat{s}_{i-1}\right)^2 + 0.821 s_{i-1}^2 \tag{28}$$

$$\hat{s}_i^{\,2} = 0.000001 + 0.095 I\left(z_i > 0\right)\left(z_i - \hat{s}_{i-1}\right)^2 + 0.250 I\left(w_i > 0\right)\left(w_i - \hat{s}_{i-1}\right)^2 + 0.826 s_{i-1}^2 \tag{29}$$

For the two models, the estimated shape parameter ξ was 0.0504 and -0.0088, respectively. Fig. 2 presents the scale estimates produced by the asymmetric model of expression (29), along with the exceedances. The scale can be seen to adjust in response to the magnitude of the exceedances,. Fig. 3 presents the resulting 99% VaR and ES estimates for the first 2750 returns. These estimates can be seen to vary with the volatility in the returns. In both figures, the final 250 are out-of-sample day-ahead forecasts.

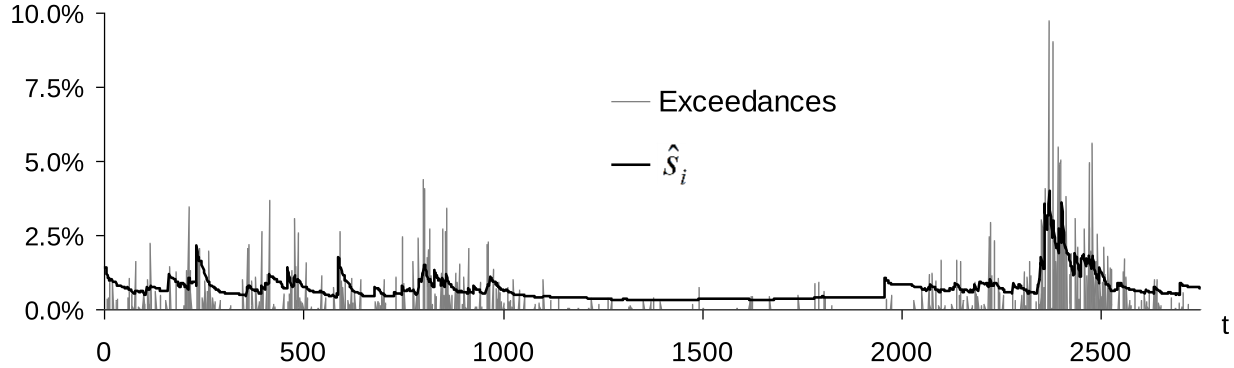

Fig. 2. For TVPOT estimation of the 99% VaR for S&P 500 returns, the plot shows exceedances beyond the threshold Q=1.21%, and time-varying scale from the asymmetric model of expression (29).

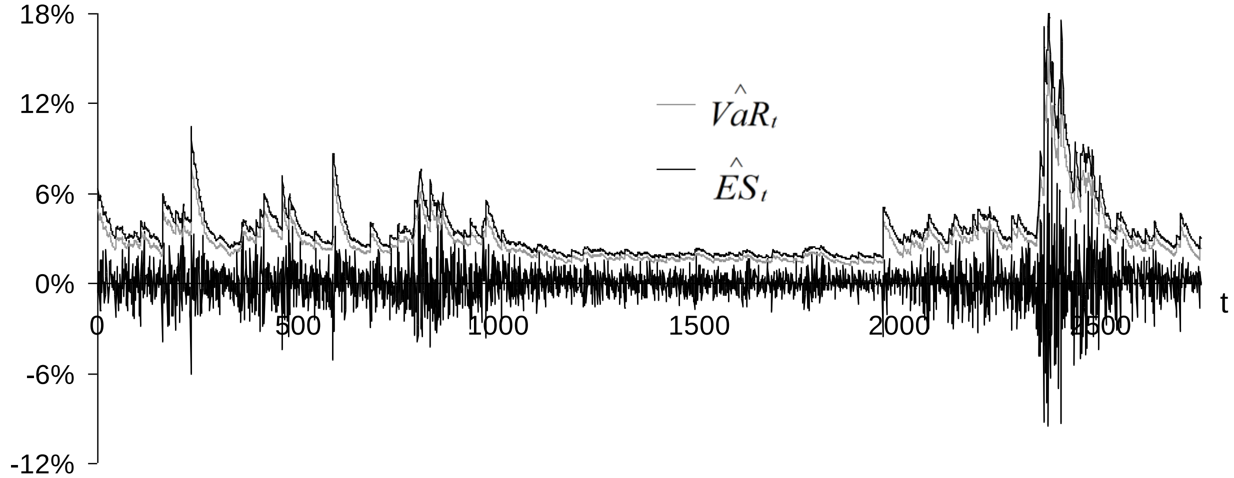

Fig. 3. S&P 500 returns with 99% VaR and ES estimated using TVPOT.

5.3. Out-of-sample VaR forecast evaluation

We evaluated day-ahead VaR forecasts for the three series of stock indices, and the following six VaR probability levels: 0.5%, 1%, 5%, 95%, 99% and 99.5%. As in our empirical study of exceedance probability forecasting, we estimated model parameters using four samples of 2500, and evaluated day-ahead forecasts for each of the next 250 periods, leading to 1000 out-of-sample forecasts.

In addition to the TVPOT approach, we generated VaR forecasts from a set of benchmark methods. We used two versions of historical simulation; the first used 2500 observations in the moving window, and the second used 250. We implemented filtered historical simulation, which involved an exponentially weighted moving average model for the variance, with optimised parameter, and historical simulation applied to all 2500 standardised in-sample returns. We also implemented GARCH(1,1), GJRGARCH(1,1) and APARCH(1,1) models, estimated using the Student-*t* distribution, and also the APARCH(1,1) model estimated using a non-central *t* distribution. We produced VaR and ES forecasts using the same distributions, and also using the method of McNeil and Frey (2000), which involves applying EVT to the standardised residuals. We also implemented the four CAViaR models of Engle and Manganelli (Section 3, 2004).

We evaluated the VaR forecasts using a test for unconditional coverage and a test for conditional coverage. For estimation of the VaR with probability level θ, we define the hit percentage as the percentage of observations falling below the estimator. We tested for unconditional coverage using a test based on the binomial distribution to examine significant difference of the hit percentage from the ideal value of p. We tested for conditional coverage using Engle and Manganelli's (2004) dynamic quantile (DQ) test. This tests whether the hit variable, defined as $Hit_t = I\left(y_t \leq V\hat{a}R_t\right) - p$, is distributed i.i.d. Bernoulli with probability p, and is independent of the VaR estimator, $V\hat{a}R_t$. Ideally, Hit_t will have zero unconditional and conditional expectations. We included four lags of Hit_t in the test's regression to give a test statistic, which, under the null hypothesis of perfect unconditional and conditional coverage, is distributed $\chi^2(6)$.

Table 5 presents the values of the hit percentage for each method applied to the S&P 500 returns for each VaR probability level. The asterisks indicate significance at the 5% level, and bold indicates the best performing method for each probability level. The final column presents the number of probability levels for which the hit percentage is significantly different from the ideal. This final column shows that, for the GARCH models, using the Student-t distribution to construct VaR estimates was not as accurate as applying EVT to the standardized residuals. The CAViaR models and the two TVPOT methods performed well. Table 6 summarises the hit percentage results for the three stock indices. The TVPOT methods again can be seen to perform well, along with three of the CAViaR models, GARCH with EVT and the historical simulation approach based on 250 observations.

The DQ test results are presented in Table 7 for the S&P 500 returns, and summarised for all three indices in Table 8. The results are poor for many of the methods. The GARCH models were notably improved by using EVT instead of the Student-*t* distribution. The best results correspond to symmetric absolute value and asymmetric slope CAViaR, and the two TVPOT methods.

Table 5. VaR hit percentages for the S&P 500.

	Probability level θ						No. sig. at 5% level
	0.5%	1%	5%	95%	99%	99.5%	
Historical simulation 2500	0.1	0.5	3.9	95.6	99.6	99.9	**0**
Historical simulation 250	0.7	1.1	3.6*	96.0	98.9	**99.5**	1
Filtered historical simulation	1.0*	1.7*	5.2	94.5	99.3	99.4	2
GARCH with Student-t	**0.6**	**1.0**	4.6	96.7*	99.8*	100.0*	2
GJRGARCH with Student-t	**0.6**	1.2	4.7	96.7*	99.7*	99.9	2
APARCH with Student-t	0.7	1.2	**5.0**	96.3	99.7*	99.8	1
NCTAPARCH	0.8	1.4	5.8	98.1*	98.6	99.5	1
GARCH with EVT	0.7	1.3	5.2	94.2	**99.1**	99.6	**0**
GJRGARCH with EVT	0.7	1.5	5.5	94.1	**98.9**	99.6	**0**
APARCH with EVT	0.8	1.5	5.8	93.7	98.8	**99.5**	**0**
NCTAPARCH with EVT	0.8	1.5	5.9	93.5*	98.7	**99.5**	1
CAViaR - Adaptive	0.3	0.8	4.5	95.6	99.4	99.7	**0**
CAViaR - Symmetric Absolute Value	0.8	1.8*	5.6	94.4	98.8	99.1	1
CAViaR - Asymmetric Slope	0.7	1.5	6.0	94.1	98.2*	99.1	1
CAViaR - Indirect GARCH	0.9	1.6	5.1	**94.7**	99.3	99.4	**0**
TVPOT using Sym Scale Model	0.3	0.7	4.7	**94.7**	99.4	99.6	**0**
TVPOT using Asym Scale Model	**0.4**	1.1	5.2	94.1	**99.1**	**99.5**	**0**

Notes. Bold indicates best method in each column. * indicates significance at 5% level.

Table 6. Number of VaR hit percentages significant at the 5% level for the three stock indices.

	Probability level θ						No. sig. at 5% level
	0.5%	1%	5%	95%	99%	99.5%	
Historical simulation 2500	1	2	1	**0**	1	**0**	5
Historical simulation 250	**0**	**0**	1	1	**0**	**0**	2
Filtered historical simulation	1	1	**0**	**0**	**0**	**0**	2
GARCH with Student-t	**0**	**0**	1	2	2	2	7
GJRGARCH with Student-t	**0**	**0**	1	2	3	1	7
APARCH with Student-t	**0**	**0**	1	1	3	1	6
NCTAPARCH	**0**	**0**	1	3	**0**	**0**	4
GARCH with EVT	**0**	**0**	**0**	1	**0**	**0**	1
GJRGARCH with EVT	**0**	**0**	**0**	2	1	**0**	3
APARCH with EVT	**0**	**0**	**0**	2	1	**0**	3
NCTAPARCH with EVT	**0**	**0**	**0**	3	1	**0**	4
CAViaR - Adaptive	**0**	1	**0**	**0**	1	**0**	2
CAViaR - Symmetric Absolute Value	**0**	1	**0**	**0**	**0**	**0**	1
CAViaR - Asymmetric Slope	**0**	**0**	**0**	1	2	1	4
CAViaR - Indirect GARCH	**0**	**0**	**0**	**0**	**0**	**0**	**0**
TVPOT using Sym Scale Model	**0**	**0**	1	1	**0**	**0**	2
TVPOT using Asym Scale Model	**0**	**0**	**0**	1	**0**	**0**	1

Notes. Bold indicates best method in each column.

Table 7. VaR DQ test p-values for the S&P 500.

	Probability level θ						No. sig. at 5% level
	0.5%	1%	5%	95%	99%	99.5%	
Historical simulation 2500	0.007	0.000	0.007	0.051	0.000	0.013	5
Historical simulation 250	0.000	0.000	0.000	0.000	0.055	0.000	5
Filtered historical simulation	0.018	0.002	0.141	0.193	0.952	0.975	2
GARCH with Student-t	0.000	0.060	0.140	0.046	0.000	0.000	4
GJRGARCH with Student-t	0.000	0.164	0.215	0.032	0.006	0.010	4
APARCH with Student-t	0.000	0.184	0.125	0.099	0.006	0.384	2
NCTAPARCH	0.003	0.251	0.191	0.000	0.719	0.701	2
GARCH with EVT	0.000	0.000	0.012	0.693	0.994	0.983	3
GJRGARCH with EVT	0.000	0.229	0.049	0.245	0.980	0.749	2
APARCH with EVT	0.003	0.294	0.187	0.123	0.865	0.708	**1**
NCTAPARCH with EVT	0.003	0.300	0.148	0.112	0.737	0.709	**1**
CAViaR - Adaptive	0.000	0.001	0.018	0.006	0.000	0.305	5
CAViaR - Symmetric Absolute Value	0.003	0.016	0.189	0.456	0.876	0.428	2
CAViaR - Asymmetric Slope	0.000	0.370	0.239	0.317	0.473	0.466	**1**
CAViaR - Indirect GARCH	0.017	0.005	0.009	0.279	0.951	0.981	3
TVPOT using Sym Scale Model	0.000	0.000	0.282	0.135	0.833	0.999	2
TVPOT using Asym Scale Model	0.000	0.000	0.664	0.517	0.998	1.000	2

Notes. Bold indicates best method in the final column.

Table 8. Number of VaR DQ tests significant at the 5% level for the three stock indices.

	Probability level θ						No. sig. at 5% level
	0.5%	1%	5%	95%	99%	99.5%	
Historical simulation 2500	3	3	3	2	3	3	17
Historical simulation 250	3	3	3	3	1	3	16
Filtered historical simulation	2	2	0	1	1	1	7
GARCH with Student-t	2	2	**0**	2	2	2	10
GJRGARCH with Student-t	2	**1**	1	3	3	3	13
APARCH with Student-t	2	**1**	1	1	3	2	10
NCTAPARCH	2	1	1	3	0	0	7
GARCH with EVT	2	3	1	1	**0**	**0**	7
GJRGARCH with EVT	2	**1**	1	1	**0**	**0**	5
APARCH with EVT	2	**1**	**0**	1	**0**	**0**	4
NCTAPARCH with EVT	2	**1**	**0**	1	**0**	**0**	4
CAViaR - Adaptive	2	3	3	3	3	**0**	14
CAViaR - Symmetric Absolute Value	2	**1**	**0**	**0**	**0**	**0**	**3**
CAViaR - Asymmetric Slope	2	**1**	**0**	**0**	**0**	**0**	**3**
CAViaR - Indirect GARCH	2	2	1	**0**	**0**	**0**	5
TVPOT using Sym Scale Model	2	**1**	1	**0**	**0**	**0**	4
TVPOT using Asym Scale Model	**1**	2	**0**	1	**0**	**0**	**3**

Notes. Bold indicates best method in each column.

5.4. *Out-of-sample ES forecast evaluation*

We produced forecasts of the conditional ES from each of the methods for which we had generated VaR forecasts, with the exception of the CAViaR models, because they cannot be used to produce ES estimates. To evaluate the ES estimates, we followed the approach of McNeil and Frey (2000), which involves the discrepancy between an observation and the conditional ES estimate for periods in which the observation exceeds the corresponding VaR estimate. When standardised, these discrepancies should be i.i.d. with a mean of zero. We standardised the discrepancies by dividing each by the corresponding VaR estimate. McNeil and Frey test for zero mean with a bootstrap test to avoid distributional assumptions (see page 224 of Efron and Tibshirani 1993). Table 9 presents p-values for this test for ES estimation for the S&P 500. The final column in this table is a count for the number of probability levels for which the null is rejected at the 5% level. The N/A entries in the table indicate that the test could not be performed due to there being no exceedances beyond the VaR estimate. Table 10 summarises the test results for the three stock indices. In Tables 9 and 10, the TVPOT method using an asymmetrical scale model performs very well. The results are also very impressive for historical simulation based on 250 observations, which is consistent with this method's hit percentage results in Tables 5 and 6. However, the DQ results of Tables 7 and 8 indicate that the dynamic properties of this method are poor. As in many other empirical studies of ES, we did not perform a test for the standardised discrepancies being i.i.d., as the number of discrepancies was low.

Table 9. For the S&P 500, p-values for ES bootstrap test for zero mean in standardised discrepancies.

	Probability level θ						No. sig. at 5% level
	0.5%	1%	5%	95%	99%	99.5%	
Historical simulation 2500	0.000	0.647	0.111	0.004	0.047	0.000	4
Historical simulation 250	0.135	0.159	0.148	0.587	0.312	0.079	**0**
Filtered historical simulation	0.007	0.159	0.250	0.663	0.150	0.554	1
GARCH with Student-t	0.274	0.852	0.594	0.006	0.511	N/A	N/A
GJRGARCH with Student-t	0.012	0.351	0.060	0.050	0.609	0.000	2
APARCH with Student-t	0.895	0.546	0.029	0.049	0.292	0.507	2
NCTAPARCH	0.737	0.808	0.059	0.000	0.933	0.438	1
GARCH with EVT	0.382	0.484	0.781	0.556	0.665	0.628	**0**
GJRGARCH with EVT	0.004	0.061	0.563	0.719	0.699	0.640	1
APARCH with EVT	0.005	0.055	0.955	0.890	0.656	0.916	1
NCTAPARCH with EVT	0.527	0.688	0.264	0.959	0.562	0.795	**0**
TVPOT using Sym Scale Model	0.510	0.645	0.672	0.138	0.314	0.868	**0**
TVPOT using Asym Scale Model	0.758	0.706	0.587	0.118	0.136	0.119	**0**

Notes. Bold indicates best method in final column. N/A indicates not available due to no exceedances beyond VaR.

Table 10. For the three stock indices, number of ES bootstrap tests significant at the 5% level. Test is for zero mean in standardised discrepancies.

	Probability level θ						No. sig. at 5% level
	0.5%	1%	5%	95%	99%	99.5%	
Historical simulation 2500	1	**0**	1	3	1	3	9
Historical simulation 250	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Filtered historical simulation	1	**0**	1	**0**	**0**	**0**	2
GARCH with Student-t	**0**	**0**	**0**	2	**0**	N/A	N/A
GJRGARCH with Student-t	1	**0**	**0**	1	**0**	N/A	N/A
APARCH with Student-t	1	**0**	1	3	**0**	N/A	N/A
NCTAPARCH	1	**0**	**0**	3	**0**	**0**	4
GARCH with EVT	**0**	1	1	**0**	1	**0**	3
GJRGARCH with EVT	1	**0**	**0**	**0**	**0**	**0**	1
APARCH with EVT	1	**0**	**0**	**0**	**0**	**0**	1
NCTAPARCH with EVT	1	**0**	**0**	**0**	**0**	**0**	1
TVPOT using Sym Scale Model	1	**0**	1	**0**	**0**	**0**	2
TVPOT using Asym Scale Model	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Notes. Bold indicates best method in each column. N/A indicates not available due to no exceedances beyond VaR.

6. Concluding comments

Our empirical study of Section 4 provides encouraging results for the use of the CARL models for day-ahead prediction of exceedance probabilities for stock index returns. The best performing CARL model, CARL-AsymVol, relates exceedance probability to a proxy for the volatility, and incorporates asymmetry by allowing a different response to negative and positive shocks of equal size. We obtained slightly better accuracy when the parameters of the CARL models were estimated using our proposal of performing a constrained maximum likelihood based on the AL likelihood, rather than the standard Bernoulli likelihood. Our overall results for three series of daily stock index returns showed that the CARL-AsymVol model was very competitive in terms of the Brier score when compared with GARCH and historical simulation benchmark methods.

We applied the exceedance probability forecasts to a new time-varying POT EVT approach to VaR and ES estimation. The approach uses CARL model probability forecasts, and an autoregressive model for the scale of the GPD. We evaluated VaR and ES forecast accuracy for six probability levels for three stock indices, and found that the new method performed well in comparison with historical simulation, filtered historical simulation, CAViaR and GARCH-based approaches.

Acknowledgments

We are very grateful to an associate editor and two referees for providing useful comments that helped greatly to improve the paper.

Appendix

In this appendix, we derive expression (12), drawing heavily on Section 2.2 of Koenker (2005). The objective function $R(Q)$ of expression (11) is the following:

$$R(Q)=\sum_{t=1}^{n}(y_t-Q)(p_t-I(y_t\leq Q))$$

The function $R(Q)$ is not differentiable at the points at which any of the residuals, (y_t-Q), are equal to zero. For this reason, when considering the minimisation of $R(Q)$, we consider directional derivatives. The directional derivative of R in direction w is given by

$$\nabla R(Q,w) \equiv \frac{d}{dq} R(Q+qw) \,\Big|_{q=0}$$

$$= \frac{d}{dq} \sum_{t=1}^{n} (y_t - Q - qw)(p_t - I(y_t \le Q + qw)) \,\Big|_{q=0}$$

$$= -\sum_{\substack{t=1 \\ y_t \ne Q}}^{n} (p_t - I(y_t \le Q))w \; - \sum_{\substack{t=1 \\ y_t = Q}}^{n} (p_t - I(0 \le w))w$$

The parameter Q minimises $R(Q)$ if and only if the directional derivatives, $\nabla R(Q,w)$, are nonnegative for all directions w. We present this condition in the following expression:

$$-\sum_{\substack{t=1 \\ y_t \ne Q}}^{n} (p_t - I(y_t \le Q))w \; - \sum_{\substack{t=1 \\ y_t = Q}}^{n} (p_t - I(0 \le w))w \; \ge 0 \tag{30}$$

If we let w=-1 in expression (30), we get

$$\sum_{t=1}^{n} I(y_t < Q) \le \sum_{t=1}^{n} p_t \tag{31}$$

and if we let w=1 in expression (30), we get

$$\sum_{t=1}^{n} p_t \le \sum_{t=1}^{n} I(y_t \le Q). \tag{32}$$

If we make the reasonable assumption that y_t is not equal for Q for all t, expressions (31) and (32) constitute expression (12).

References

Anatolyev, S., and Gospodinov, N. (2010) Modeling financial return dynamics via decomposition. *Journal of Business and Economic Statistics*, **28**, 232-245.

Bali, T.G., and Neftci, S.N. (2003) Disturbing extremal behavior of spot rate dynamics. *Journal of Empirical Finance*, **10**, 455-477.

Bollerslev, T., and Wooldridge, J.M. (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11**, 143-172.

Bradley, A.A., Schwartz, S.S., and Hashino, T. (2008) Sampling uncertainty and confidence intervals for the Brier score and Brier skill score. *Weather and Forecasting*, **23**, 992-1006.

Chavez-Demoulin, V., and Davison, A.C. (2005) Generalized additive modelling of sample extremes. *Applied Statistics*, **54**, 207-222.

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014) Extreme-quantile tracking for financial time series. *Journal of Econometrics,* to appear.

Christoffersen, P.F., and Diebold, F.X. (2006) Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, **52**, 1273-1287.

Chung, J., and Hong, Y. (2007) Model-free evaluation of directional predictability in foreign exchange markets. *Journal of Applied Econometrics*, **22**, 855-889.

de Jong, R.M., and Woutersen, T. (2011) Dynamic time series binary choice. *Econometric Theory*, **27**, 673-702.

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

Efron, B., J. Tibshirani. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Embrechts, P., Puccetti, G., Ruschendorf, L., Wang, R., and Beleraj, A. (2014) An academic response to Basel 3.5. *Risks*, to appear.

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

Franses, P.H., D. Van Dijk. (2000). *Non-linear Time Series Models in Empirical Finance*. Cambridge University Press.

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

Gneiting, T. (2008) Editorial: probabilistic forecasting. *Journal of the Royal Statistical Society. Series A*, **171**, 319-321.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007) Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B*, **69**, 243-268.

Granger, C.W.J., and Pesaran, M.H. (2000) Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, **8**, 426-459.

Kanamura, T., and Ōhashi, K. (2007) A structural model for electricity prices with spikes: measurement of spike risk and optimal policies for hydropower plant operation. *Energy Economics*, **29**, 1010-1032.

Kauppi, H., and Saikkonen, P. (2008) Predicting US recessions with dynamic binary response Models. *Review of Economics and Statistics*, **90**, 777-791.

Koenker, R.W. (2005) *Quantile Regression*. Cambridge, UK: Cambridge University Press.

Koenker, R., and Machado, J.A.F. (1999) Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**, 1296-1310.

Krause, J. and Paolella, M. S. (2014) A fast, accurate method for value at risk and expected shortfall. *Econometrics*, **2**, 98-122.

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

Kumar, M., Moorthy, U., and Perraudin, W. (2003) Predicting emerging market currency Crashes. *Journal of Empirical Finance*, **10**, 427-454.

Linton, O., and Whang, Y.J. (2007) The quantilogram: with an application to evaluating directional predictability. *Journal of Econometrics*, **141**, 250-282.

Manganelli S., and Engle, R.F. (2004) A comparison of value-at-risk models in finance. In *Risk Measures for the 21st Century*, ed. G. Szegö, Chichester: Wiley, pp. 123-144.

Mason, S.J., Galpin, J.S., Goddard, L., Graham, N.E., and Rajartnam, B. (2007) Conditional exceedance probabilities. *Monthly Weather Review*, **135**, 363-372.

McNeil, A.J., and Frey, R. (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, **7**, 271-300.

Nyberg, H. (2011) Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting*, **27**, 561-578.

Rydberg, T.H., and Shephard, N. (2003) Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, **1**, 2-25.

Slud, E.V., and Kedem, B. (1994) Partial likelihood analysis of logistic regression and autoregression. *Statistica Sinica*, **4**, 89-106.

Thomakos, D.D., and Wang, T. (2010) ‘Optimal’ probabilistic and directional predictions of financial returns. *Journal of Empirical Finance*, **17**, 102-119.

Wilks, D.S. (2010) Sampling distributions of the Brier score and Brier skill score under serial dependence. *Quarterly Journal of the Royal Meteorological Society*, **136**, 2109-2118.

Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*. Oxford, UK: Academic Press.

Yu, K., and Stander, J. (2007) Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics*, **137**, 260-276.

Yu, K., and Zhang, J. (2005) A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics – Theory and Methods*, **34**, 1867-1879.