

# Downside Risk Evaluation with the R Package GAS

by David Ardia, Kris Boudt and Leopoldo Catania

**Abstract** Financial risk managers routinely use non-linear time series models to predict the downside risk of the capital under management. They also need to evaluate the adequacy of their model using so-called backtesting procedures. The latter involve hypothesis testing and evaluation of loss functions. This paper shows how the R package GAS can be used for both the dynamic prediction and the evaluation of downside risk. Emphasis is given to the two key financial downside risk measures: Value-at-Risk (VaR) and Expected Shortfall (ES). High-level functions for: (i) prediction, (ii) backtesting, and (iii) model comparison are discussed, and code examples are provided. An illustration using the series of log-returns of the Dow Jones Industrial Average constituents is reported.

## Introduction

The GAS package of Catania et al. (2016) provides a complete framework for modeling, estimating and predicting time series processes for which the time-variation in the parameters is driven by the score of the conditional density function. This increasingly popular class of score-driven models has been introduced by Creal et al. (2013) and Harvey (2013). Ardia et al. (20xx) describe the general functionality implemented in the GAS package, but do not cover its functionality useful for the estimation and backtesting of Value-at-Risk (VaR) and Expected Shortfall (ES) which are the two leading risk measures used in finance. The aim of this paper is to show how the functions available in the GAS package can be used for VaR and ES evaluation, prediction, and backtesting.

The economic relevance of this topic follows partly from the Basel Accords (currently the Basel III Accords), which impose that banks and financial institutions have to meet capital requirements, and must rely on state-of-the-art risk systems. In particular, they must assess the uncertainty about the future values of their portfolios and estimate the extent and the likelihood of potential losses using a risk measure. Nowadays, VaR and ES risk measures are the standards (Jorion, 1997). For an asset (or portfolio) return, the VaR at a given time horizon equals the return such that lower returns only occur with a given probability level  $\alpha$  (referred to as the *risk level*, and which is typically set to one or five percent, that is  $\alpha \in \{0.01, 0.05\}$ ). The ES risk measure is the expectation of the asset (or portfolio) return when the return is below the VaR level.

The estimation of the VaR and ES thus requires first to accurately estimate the conditional distribution of the future portfolios' or assets' returns. Formally, assuming a continuous cumulative density function (*cdf*) with time-varying parameters  $\theta_t \in \mathbb{R}^d$  and additional static parameters  $\psi \in \mathbb{R}^q$ ,  $F(\cdot; \theta_t, \psi)$ , for the log-return at time  $t$ ,  $r_t \in \mathbb{R}$ , the  $VaR_t(\alpha)$  is given by:

$$VaR_t(\alpha) \equiv F^{-1}(\alpha; \theta_t, \psi),$$

where  $F^{-1}(\cdot)$  denotes the inverse of the *cdf*, that is, the quantile function. It follows that  $VaR_t(\alpha)$  is nothing more than the  $\alpha$ -quantile of the return distribution at time  $t$ .<sup>1</sup> The ES metric measures the expected loss after a violation of the VaR level and it is defined as:

$$ES_t(\alpha) \equiv \frac{1}{\alpha} \int_{-\infty}^{VaR_t} z dF(z, \theta_t, \psi).$$

It follows that a crucial point for correct VaR and ES assessment is the determination of  $F(\cdot)$  and its parameters  $\theta_t$  and  $\psi$ . For a general overview of existing methods, we refer the reader to Nieto and Ruiz (2016). In this paper, we illustrate how this can be achieved using the framework of Generalized Autoregressive Score (GAS) models introduced by Creal et al. (2013) and Harvey (2013). GAS models are also referred to as Score Driven (SD) models and Dynamic Conditional Score (DCS) models and have been used extensively for financial risk management purposes. For a comparison between the accuracy of VaR and ES estimates obtained by the GAS approach against alternative volatility models, we refer the reader to Bernardi and Catania (2016), Gao and Zhou (2016), Ardia et al. (2018) and Ardia et al. (20xx).

Formally, in GAS models the vector of time-varying parameters,  $\theta_t$ , is updated through a dynamic

<sup>1</sup>Sometimes VaR is defined with respect to the loss variable  $l_t = -r_t$ . All the arguments of this paper can be easily adapted to this case.

equation based on the score of the conditional<sup>2</sup> probability density function of  $r_t$ ,  $f(\cdot; \boldsymbol{\theta}_t, \boldsymbol{\psi})$ , that is:

$$\boldsymbol{\theta}_{t+1} \equiv \boldsymbol{\kappa} + \mathbf{A}\mathbf{s}_t + \mathbf{B}\boldsymbol{\theta}_t, \quad (1)$$

where  $\mathbf{s}_t$  is the (possibly) scaled score of  $f(\cdot; \boldsymbol{\theta}_t, \boldsymbol{\psi})$  with respect to  $\boldsymbol{\theta}_t$ , evaluated in  $r_t$ ; see [Creal et al. \(2013\)](#) and also the Appendix for examples. The coefficients  $\boldsymbol{\kappa}$ ,  $\mathbf{A}$  and  $\mathbf{B}$  control for the evolution of  $\boldsymbol{\theta}_t$  and need to be estimated along with  $\boldsymbol{\psi}$  from the data, usually by Maximum Likelihood.

We focus on the three major steps a practitioner involved in risk management faces during his job: (i) prediction of future downside risk, (ii) backtesting, and (iii) comparison with alternative models. The empirical part of the article deals with these three points from an applied perspective while the computational part details the GAS functionalities devoted to downside risk.

## A flexible GAS specification for modeling financial returns

Financial returns exhibit several stylized facts that need to be taken into consideration to produce reliable risk forecasts. Empirically, the distribution of returns is (left-)skewed and fat-tailed, and its variance is time-varying (*i.e.*, returns exhibit the so-called *volatility clustering*); see, for example, [McNeil et al. \(2015\)](#).

To account for these features, we consider a very flexible specification, in which we assume that the log-return at time  $t$ ,  $r_t$ , is distributed conditionally on past observations as follows:

$$r_t | \mathcal{I}_{t-1} \sim SKST(r_t; \mu, \sigma_t, \zeta, \nu),$$

where  $\mathcal{I}_{t-1}$  is the information set up to time  $t-1$ , and  $SKST(r_t; \cdot)$  denotes the skew Student-t distribution of [Fernández and Steel \(1998\)](#) with location  $\mu \in \mathbb{R}$ , time-varying scale (volatility)  $\sigma_t > 0$ , and skewness and shape parameters  $\zeta > 0$  and  $\nu > 2$ , respectively. We parametrize the  $SKST$  distribution as in [Bauwens and Laurent \(2005\)](#) such that  $E[r_t | \mathcal{I}_{t-1}] = \mu$  and  $V[r_t | \mathcal{I}_{t-1}] = \sigma_t^2$ . To ensure positivity of the volatility parameter, we set the time-varying GAS parameter  $\boldsymbol{\theta}_t$  in (1) to  $\boldsymbol{\theta}_t \equiv \boldsymbol{\theta}_t \equiv \log \sigma_t$ , and we define  $\boldsymbol{\psi} \equiv (\mu, \zeta, \nu)$ . In the Appendix, we show that the corresponding score  $\mathbf{s}_t$  which enters linearly in the updating equation (1) is given by:

$$\mathbf{s}_t \equiv \left( \frac{z_t(\nu+1)(z_t-m)}{(\tilde{\zeta}_t^*)^2(\nu-2) + z_t^2} - 1 \right),$$

with  $\tilde{\zeta}_t^* \equiv \zeta^{I\{z_t \geq 0\}} - I\{z_t \leq 0\}$ , where  $I\{\cdot\}$  is the indicator function equal to one if the condition holds, and zero otherwise, and with  $z_t \equiv \left( \frac{r_t - \mu}{\sigma_t} \right) k + m$ , where expressions for  $k$  and  $m$  are provided in the Appendix. For a fixed value of the asymmetry parameter  $\zeta$ , we see that the effect of shock on future values is dampened when  $\nu$  decreases. Starting from the general  $SKST$  distribution, we recover as special cases:

- (i) the Student-t distribution,  $ST$ , imposing  $\zeta = 1$ ;
- (ii) the Normal distribution,  $\mathcal{N}$ , imposing  $\nu = \infty$  and  $\zeta = 1$ ;

see the Appendix for the score in these cases. Given a series of  $T$  log-returns,  $r_1, \dots, r_T$ , the model parameters are estimated by maximizing the log-likelihood function; see [Blasques et al. \(2014\)](#). Prediction with GAS models is straightforward thanks to the recursive nature of the updating equation (1). Specifically, the one-step ahead predictive distribution  $F(\cdot; \hat{\boldsymbol{\theta}}_{T+1}, \hat{\boldsymbol{\psi}})$  is available in closed form whereas the  $h$ -step ahead distribution ( $h > 1$ ) needs to be simulated; see [Blasques et al. \(2016\)](#). VaR and ES forecasts are easily obtained from the predictive distribution.

## Evaluating downside risk forecasts

The recursive method of forecasting is usually employed to backtest the adequacy of a statistical model, as well as to perform models comparisons in terms of VaR and ES predictions ([Marcellino et al., 2006](#)). The objective of a backtesting analysis is to verify the precision of the prediction by separating the estimation window and the evaluation period. The objective of a model comparison analysis is usually to order models according to a loss function.

To this end, the full sample of  $T$  returns is divided into an in-sample period of length  $S$ , and an out-of-sample period of length  $H$ . Model parameters are first estimated over the in-sample period,

<sup>2</sup>The conditioning is intended with respect to the past observations  $r_{t-s}$  ( $s > 0$ ), however, for notational purposes, this is not always reported.

subsequently the  $h$ -step ahead prediction of the return distribution at time  $S + h$  is generated along with the corresponding VaR and ES measures. These steps are repeated augmenting the in-sample period with new observations in a recursive way until we reach the end of the series,  $T$ . If during the data augmentation step, past observations are eliminated, we are considering a rolling window, otherwise we have an expanding window. In this paper, we follow the standard approach in daily risk management of a typical trading desk and use the rolling window configuration with  $h = 1$ .

### VaR backtesting

Once a series of VaR predictions is available, forecasts adequacy is assessed through backtesting procedures. VaR backtesting procedures usually check the correct coverage of the unconditional and conditional left-tail of the log-returns distribution. Correct unconditional coverage (UC) was first considered by Kupiec (1995), while correct conditional coverage (CC) by Christoffersen (1998). The main difference between UC and CC concerns the distribution we are focusing on. For instance, UC considers correct coverage of the left-tail of the unconditional log-return distribution,  $f(r_t)$ , while CC deals with the conditional density  $f(r_t | \mathcal{I}_{t-1})$ . From an inferential perspective, UC looks at the ratio between the number of realized VaR violations observed from the data and the expected number of VaR violations implied by the chosen risk level,  $\alpha$ , during the forecast period, that is,  $\alpha H$ . In order to investigate CC, Christoffersen (1998) proposed a test on the series of VaR exceedance  $\{d_t, t = S, \dots, S + H\}$ , where  $d_t \equiv I\{r_t < VaR_t(\alpha)\}$ , usually referred to as the *hitting series*. Specifically, if correct conditional coverage is achieved by the model, VaR exceedances should be independently distributed over time.

The DQ test by Engle and Manganelli (2004) assesses the joint hypothesis that  $E[d_t] = \alpha$  and the hit variables are independently distributed. The implementation of the test involves the de-measured process  $Hit_t^\alpha \equiv d_t - \alpha$ . Under correct model specification, unconditionally and conditionally,  $Hit_t^\alpha$  has zero mean and is serially uncorrelated. The DQ test is then the traditional Wald test of the joint nullity of all coefficients in the following linear regression:

$$Hit_t^\alpha = \delta_0 + \sum_{l=1}^L \delta_l Hit_{t-l}^\alpha + \delta_{L+1} VaR_{t-1}(\alpha) + \epsilon_t.$$

Under the null hypothesis of correct unconditional and conditional coverage, we have that the Wald test statistics is asymptotically chi-square distributed with  $L + 2$  degrees of freedom. Engle and Manganelli (2004) set  $L = 4$  lags which has become the standard choice.

### VaR model comparison

Real world applications consider several models for VaR prediction. If correct unconditional/conditional coverage is achieved by more than one model, the practitioner faces the problem of not being able to choose between different alternatives. In this situation, model comparison techniques are used to choose the best performing model. Model ranking is achieved thanks to the definition of a loss function. Among several available loss functions for quantile prediction (McAleer and Da Veiga, 2008), the Quantile Loss (QL) used for quantile regressions (Koenker and Bassett, 1978) is one of the most frequent choices in the VaR context; see González-Rivera et al. (2004). Formally, given a VaR prediction at risk level  $\alpha$  for time  $t$ , the associated quantile loss,  $QL_t(\alpha)$ , is defined as:

$$QL_t(\alpha) \equiv (\alpha - d_t)(r_t - VaR_t(\alpha)). \quad (2)$$

QL is an asymmetric loss function that penalizes more heavily with weight  $(1 - \alpha)$  the observations for which we observe returns showing VaR exceedance. Quantile losses are then averaged over the forecasting period, and models with lower averages are preferred. The outperformance of model  $A$  versus model  $B$  is finally assessed looking at the ratio between the average QLs, associated with the two models, that is, if  $QL_A / QL_B < 1$  then model  $A$  outperforms model  $B$  and vice versa.

### Joint VaR and ES model comparison

The quantile loss function in (2) for VaR assessment is appropriate since quantiles are elicited by it, that is, when the conditional distribution is static over the sample, the VaR can be estimated by minimizing the average quantile loss function. Unfortunately, there is no loss function available for which the ES risk measure is elicitable; see, for instance, Bellini and Bignozzi (2015) and Ziegel (2016). However, it has been recently shown by Fissler and Ziegel (2016) (FZ) that the couple (VaR, ES) is jointly elicitable,

as the values of  $v_t$  and  $e_t$  that minimize the sample average of the following loss function:

$$FZ(r_t, v_t, e_t, \alpha, G_1, G_2) \equiv (d_t - \alpha) \left( G_1(v_t) - G_1(r_t) + \frac{1}{\alpha} G_2(e_t) v_t \right) - G_2(e_t) \left( \frac{1}{\alpha} d_t r_t - e_t \right) - \mathcal{G}_2(e_t),$$

where  $G_1$  is weakly increasing,  $G_2$  is strictly positive and strictly increasing, and  $\mathcal{G}_2' = G_2$ . The **GAS** package implements the FZ loss function for a specific choice of  $G_1$ ,  $G_2$  and  $\mathcal{G}_2$ . We set  $G_1(x) = 0$  and  $G_2(x) = -1/x$  and assume the values of VaR and ES to be strictly negative; see [Patton et al. \(2017\)](#) and [Ardia et al. \(2018\)](#) for a similar approach. For VaR and ES predictions at risk level  $\alpha$  for time  $t$ , the associated joint loss function (FZL) is then given by:

$$FZL_t^\alpha \equiv \frac{1}{\alpha ES_t^\alpha} d_t (r_t - VaR_t^\alpha) + \frac{VaR_t^\alpha}{ES_t^\alpha} + \log(-ES_t^\alpha) - 1, \quad (3)$$

for  $ES_t^\alpha \leq VaR_t^\alpha < 0$ . As for QL, FZ losses are averaged over the forecasting period and models with lower averages are preferred.

Summarizing, given a set of available models, a typical downside risk forecasting exercise consists of the following three major steps:

- (i) perform rolling forecast during the out-of-sample period,
- (ii) perform statistical backtest of VaR predictions using UC, CC and DQ tests,
- (iii) perform VaR and joint VaR and ES model comparison looking at the average QL and FZL of each model.

The next section is devoted to detailing the implementation of each of these steps with the **GAS** package.

## Empirical illustration

We first briefly review how to make predictions with GAS models using the **GAS** package. The main illustration is for the last  $T = 2,500$  observations of the daily log-returns of the General Electric stock in the dataset `dji30ret`, which is a dataframe consisting of the thirty Dow Jones Industrial Average constituents:

```
> library("GAS")
> data("dji30ret", package = "GAS")
> dji30ret <- tail(dji30ret, 2500)
```

The corresponding returns are shown as gray points in [Figure 1](#). We see the time-variation in the volatility of the daily return series, which we model next using the GAS specification with skewed Student- $t$  innovations, as previously described. Thanks to the **GAS** package, it is straightforward to estimate the GAS model on rolling windows of the available data and to make one-step ahead rolling forecasts.

Specifically, the user needs to specify the model through the `UniGASSpec()` function, and then perform rolling predictions with the `UniGASRoll()` function. In the **GAS** package, models are specified through the definition of the conditional distribution assumed for the data, `Dist`, and the list of time-varying parameters, `GASpar`.

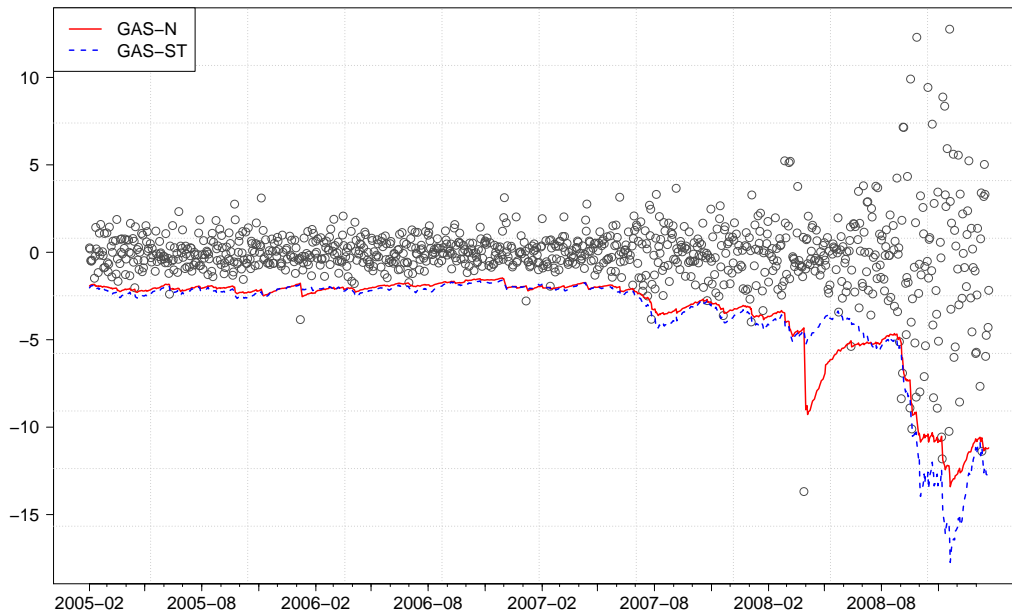
`Dist` is a character equal to the label of the distribution. For instance, *SKST* is identified as "sstd", *ST* as "std" and *N* as "norm"; see [Table 1 of Ardia et al. \(20xx\)](#) for the list of distributions and associated labels available in the **GAS** package.

`GASpar` is a list with named boolean elements. Entries name are: `location`, `scale`, `skewness` and `shape`, and indicate whether the associated distribution parameters are time-varying or not. By default we have `GASpar = list(location = FALSE, scale = TRUE, skewness = FALSE, shape = FALSE)`, that is, only volatility is time-varying. For instance, in order to specify the three GAS models: *GAS-N*, *GAS-ST* and *GAS-SKST*, we need to execute the following lines:

```
> GASSpec_N <- UniGASSpec(Dist = "norm", GASpar = list(scale = TRUE))
> GASSpec_ST <- UniGASSpec(Dist = "std", GASpar = list(scale = TRUE))
> GASSpec_SKST <- UniGASSpec(Dist = "sstd", GASpar = list(scale = TRUE))
```

`UniGASSpec()` delivers an object of the class `uGASSpec` which comes with several methods; see `help("UniGASSpec")`.

The `UniGASRoll()` function accepts an object of the class `uGASSpec`, `GASSpec`, a numeric vector for the series of returns, `data`, and other arguments such as:



**Figure 1:** One-step ahead VaR forecasts for General Electric (GE) at the  $\alpha = 1\%$  risk level for the GAS- $\mathcal{N}$  (solid) and GAS- $\mathcal{ST}$  (dotted) models. Gray points indicate realized log-returns calculated as the differences between the natural logarithm of two consecutive prices. The forecasting period ranges from February 14, 2005, to February 3, 2009, for a total of  $H = 1,000$  out-of-sample observations.

- the length of the out-of-sample period: `ForecastLength`;
- the type of the rolling window used to update the data: `RefitWindow`;
- the number of observations within each model re-estimation: `RefitEvery`,

among others; see `help("UniGASRoll")`. `ForecastLength` and `RefitEvery` are numeric elements while `RefitWindow` is a character equal to "moving" (the default) for a rolling window scheme or "recursive" for an expanding window. As previously mentioned, in this paper we consider the case `RefitWindow = "moving"`. We fix the length of the out-of-sample period to  $H = 1000$ , and re-estimate the model parameters at the weekly frequency (*i.e.*, every 5 new observations). One-step ahead rolling predictions for the first series of returns using the GAS- $\mathcal{N}$  model are then computed as:

```
> library("parallel")
> cluster <- makeCluster(2)
> H <- 1000
> Roll_N <- UniGASRoll(dji30ret[, "GE"], GASSpec_N, RefitEvery = 5,
  cluster = cluster, ForecastLength = H)
```

where we have also made use of parallel processing (with 2 cores) through the definition of a `cluster` object exploiting the **parallel** package included in R since version 2.14.0.

The output of `UniGASRoll()` is an object of the class `uGASRoll` which comes with several methods; see `help("UniGASRoll")`.

VaR and ES one-step ahead rolling forecasts at the risk level  $\alpha = 0.01$  can be computed from `Roll_N` using the `quantile` and `ES` methods, respectively:

```
> alpha <- 0.01
> VaR_N <- quantile(Roll_N, probs = alpha)
> ES_N <- ES(Roll_N, probs = alpha)
```

where `VaR_N` and `ES_N` are matrices of dimension  $1,000 \times 1$  containing the VaR and ES forecasts at the 1% risk level.<sup>3</sup> While VaR predictions are obtained by numerical inversion of the predicted cumulative density function, ES predictions are computed by numerical adaptive integration of the

<sup>3</sup>The `probs` argument in `quantile` and `ES` can also be a numeric vector of  $p$  VaR levels. In this case, `VaR_N` and `ES_N` would be a  $1000 \times p$  matrix.

predicted density relying on the `cubature` package (Narasimhan and Johnson, 2017). Multi-step ahead prediction of VaR and ES are obtained by Monte Carlo simulation and are still computed with the quantile and ES methods. To compute multi-step ahead predictions an object of the class `uGASFor` delivered by the `UniGASFor()` has to be provided; see `help("UniGASFor")`.

Predictions for the  $GAS-ST$  and  $GAS-SST$  specifications using the `GASpec_ST` and `GASpec_SKST` model definitions are computed analogously. Figure 1 reports 1% VaR predictions delivered by the  $GAS-N$  (solid line) and the  $GAS-ST$  (dotted line). We see the impact of the recent Global Financial Crisis on the volatility of the series. Indeed, the 2007–2008 returns of GE present much more variability than over the period 2005–2006, which is translated into lower values for VaR. What is also clearly evident from Figure 1, is the robustness of the  $GAS-ST$  model to extreme observations compared with  $GAS-N$ . Indeed, on April 11, 2008, General Electric reported an unexpected net income drop of 6%, which in turn translated to a fall of about 12% of its market value. The signal captured by  $GAS-N$  was that of an abrupt increase in volatility, with the consequence of large VaR level predictions. In contrast, the  $GAS-ST$  model slightly increased the volatility level and continued to predict reasonable VaR levels. What happened is that,  $GAS-ST$  treated the 12% negative return as a realization from the fat-tailed Student- $t$  distribution, hence tapering its impact on the conditional volatility level. On the contrary,  $GAS-N$  treated the negative return as a realization from the Normal distribution, which is a clear signal of increase of volatility. The same behavior of the  $GAS-N$  specification is shown in the ES predictions (not reported to save space).

## VaR backtesting

Let us now show how the accuracy of the VaR forecasts can be evaluated using the `BacktestVaR()` function in the `GAS` package. This function accepts the following arguments:

- `data`, numeric containing the out-of-sample data;
- `VaR`, numeric containing the series of VaR forecasts;
- `alpha`, the VaR risk level  $\alpha$ ;
- `Lags`, the number of lags used in the DQ test, by default `Lags = 4`; see Engle and Manganelli (2004).

The function returns a list with named entries:

- `LRuc`, the test statistic and associated  $p$ -value for the UC test of Kupiec (1995);
- `LRcc`, the test statistic and associated  $p$ -value for the CC test of Christoffersen (1998);
- `DQ`, the test statistic and associated  $p$ -value for the DQ test of Engle and Manganelli (2004);
- `Loss`, the quantile loss (QL), as defined in (2), together with the average QL used by González-Rivera et al. (2004);
- `AD`, the mean and max VaR Absolute Deviation (AD) used by McAleer and Da Veiga (2008);
- `AE`, the Actual over Expected ratio.

For instance, in order to compute the VaR backtest measures defined above on the forecast series `VaR_N`, we use:

```
> VaRBacktest_N <- BacktestVaR(data = tail(dji30ret[, "GE"], H), VaR = VaR_N, alpha = alpha)
```

Then, the DQ test statistic and its associated  $p$ -value can be extracted as:

```
> VaRBacktest_N$DQ
$stat
      [,1]
[1,] 52.47578

$pvalue
      [,1]
[1,] 4.7043e-09
```

which, in this case, is against the null of correct model specification for the 1% VaR level.

Now, if we evaluate VaR forecasts using the  $GAS-ST$  model, and save the results as the `VaRBacktest_ST` object, then the DQ test reports:

```

> VaRBacktest_ST$DQ
$stat
      [,1]
[1,] 8.763418

$pvalue
      [,1]
[1,] 0.270091

```

Asset	$\alpha = 1\%$			$\alpha = 5\%$		
	GAS- $\mathcal{N}$	GAS- $\mathcal{ST}$	GAS- $\mathcal{SKST}$	GAS- $\mathcal{N}$	GAS- $\mathcal{ST}$	GAS- $\mathcal{SKST}$
AA	0.00	0.17	0.16	0.00	0.00	0.00
AIG	0.00	0.00	0.00	0.00	0.00	0.00
AXP	0.09	0.99	0.25	0.19	0.13	0.15
BA	0.29	0.98	0.30	0.13	0.55	0.54
BAC	0.00	0.00	0.06	0.00	0.03	0.06
C	0.00	0.00	0.00	0.01	0.00	0.00
CAT	0.00	0.22	0.22	0.00	0.38	0.51
CVX	0.00	0.02	0.00	0.02	0.03	0.01
DD	0.03	0.41	0.34	0.35	0.31	0.13
DIS	0.00	0.05	0.08	0.00	0.35	0.15
GE	0.00	0.04	0.00	0.16	0.30	0.03
GM	0.00	0.06	0.02	0.02	0.29	0.17
HD	0.01	0.01	0.10	0.09	0.63	0.57
HPQ	0.00	0.01	0.03	0.00	0.11	0.35
IBM	0.06	0.03	0.03	0.00	0.04	0.05
INTC	0.04	0.07	0.33	0.06	0.10	0.06
JNJ	0.95	1.00	0.35	0.00	0.00	0.00
JPM	0.18	0.95	0.21	0.27	0.42	0.17
KO	0.00	0.00	0.00	0.00	0.00	0.00
MCD	0.11	0.68	0.73	0.10	0.17	0.40
MMM	0.00	0.34	0.00	0.00	0.18	0.33
MRK	0.00	0.05	0.06	0.00	0.11	0.05
MSFT	0.11	0.28	0.42	0.05	0.43	0.52
PFE	0.15	0.37	0.38	0.13	0.81	0.75
PG	0.00	0.18	0.07	0.13	0.11	0.06
T	0.35	0.37	0.42	0.02	0.01	0.06
UTX	0.38	0.41	0.07	0.09	0.90	0.93
VZ	0.04	0.99	0.96	0.88	0.70	0.83
WMT	0.00	0.00	0.01	0.93	0.54	0.26
XOM	0.00	0.00	0.00	0.07	0.04	0.17

**Table 1:** DQ test statistic  $p$ -values for the Dow Jones Industrial Average constituents one-step ahead VaR forecasts at the two downside risk levels  $\alpha = 1\%$  and  $\alpha = 5\%$ . Under the null hypothesis, we have correct model specification for the chosen risk level. Light gray cells indicate  $p$ -values lower than 1%. The forecasting period ranges from February 14, 2005, to February 3, 2009, for a total of  $H = 1,000$  out-of-sample observations. The model parameters are re-estimated at the monthly frequency.

The large  $p$ -value indicates that the null of correct model specification for the 1% VaR cannot be rejected at the usual levels of significance for the GAS- $\mathcal{ST}$  model.

We reproduce this analysis for each of the thirty daily stock return series in the `dji30ret` data set previously detailed. The out-of-sample period starts on February 14, 2005, and includes the recent Global Financial Crisis of 2007–2008. We consider two VaR risk levels:  $\alpha = 1\%$  and  $\alpha = 5\%$ . The code used for this application is available in the GitHub GAS repository: <https://github.com/LeopoldoCatania/GAS/wiki>.

Table 1 reports the  $p$ -values of the DQ test for the three model specifications and the two VaR risk levels. Under the null hypothesis, we have correct model specification for the  $\alpha$ -quantile level. Clearly, our results indicate that GAS- $\mathcal{N}$  is suboptimal with respect to GAS- $\mathcal{ST}$  and GAS- $\mathcal{SKST}$  in terms of correct unconditional and conditional coverage. This result is somehow expected since GAS- $\mathcal{ST}$  and GAS- $\mathcal{SKST}$  exhibit excess kurtosis and deliver more robust updates for the volatility parameter than GAS- $\mathcal{N}$ . Moreover, we see that GAS- $\mathcal{ST}$  and GAS- $\mathcal{SKST}$  perform similarly in terms of correct unconditional and conditional coverage. Hence, the inclusion of skewness does not seem to increase the performance of VaR predictions for the considered series. Indeed, sometimes results

are even worse for GAS-SKST compared with GAS-ST, indicating that the estimation error for the additional skewness parameter could worsen VaR predictions. This is the case for example for 3M Company (MMM) when GAS-SKST rejects the null for  $\alpha = 1\%$ , whereas GAS-ST does not.

### VaR model comparison

The goal of VaR model comparison is to rank the models in terms of accuracy of their VaR forecasts. This can be done using the average value of the quantile loss in (2), as available in the output from the BacktestVaR() function.

In our application on the daily stock returns of General Electric, we find that the model comparison in terms of average QL also favors GAS-ST compared with GAS-N. Indeed the ratio between the QL of GAS-ST and GAS-N is below unity indicating a lower average quantile loss when using the GAS-ST:

```
> round(VaRBacktest_ST$Loss$Loss / VaRBacktest_N$Loss$Loss, 2)
[1] 0.94
```

This indicates that GAS-ST outperforms GAS-N by 6% in terms of average QL.

The left part of Table 2 (QL ratios) reports the results of repeating this model comparison analysis for all thirty daily stock return series in the dji30ret dataset.<sup>4</sup> It shows, for both  $\alpha = 1\%$  and  $\alpha = 5\%$ , the ratios between the average QL of the considered models over the one delivered by GAS-N. Values greater than one indicate outperformance of GAS-N, and vice versa. Consistently with the DQ test results, we find that GAS-ST and GAS-SKST perform similarly and are preferred to GAS-N.

Asset	QL ratios				FZL ratios			
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 1\%$		$\alpha = 5\%$	
	GAS-ST	GAS-SKST	GAS-ST	GAS-SKST	GAS-ST	GAS-SKST	GAS-ST	GAS-SKST
AA	0.93	0.93	0.98	0.98	0.95	0.95	0.98	0.98
AIG	1.00	1.01	1.02	1.03	0.86	0.87	0.93	0.94
AXP	0.93	0.94	0.98	0.99	0.94	0.96	0.98	1.00
BA	0.99	1.00	0.99	0.99	0.98	0.99	0.99	0.99
BAC	0.83	0.85	1.01	1.01	0.88	0.88	0.98	0.98
C	0.97	1.00	1.02	1.03	0.94	0.95	0.98	0.99
CAT	0.82	0.83	0.91	0.91	0.83	0.84	0.91	0.91
CVX	1.00	1.18	1.01	1.06	0.99	1.20	1.01	1.09
DD	0.94	0.94	0.98	0.98	0.95	0.97	0.98	0.98
DIS	0.98	0.99	1.00	1.00	0.95	0.96	0.99	1.00
GE	0.95	0.96	0.97	0.98	0.96	0.99	0.98	0.98
GM	0.87	0.88	0.93	0.94	0.91	0.93	0.96	0.97
HD	1.02	1.00	1.00	1.00	1.01	1.00	1.00	1.00
HPQ	0.95	0.95	0.94	0.94	0.95	0.95	0.95	0.95
IBM	1.00	1.00	0.94	0.94	0.92	0.93	0.93	0.93
INTC	0.97	0.97	0.96	0.96	0.94	0.94	0.96	0.96
JNJ	1.03	1.01	1.02	1.00	1.00	0.99	1.01	0.99
JPM	0.92	0.92	0.98	0.99	0.94	0.95	0.99	0.99
KO	0.97	0.96	0.96	0.96	0.97	0.96	0.93	0.92
MCD	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99
MMM	0.82	0.83	0.90	0.90	0.80	0.82	0.89	0.89
MRK	0.83	0.84	0.89	0.89	0.81	0.82	0.88	0.89
MSFT	0.86	0.87	0.90	0.90	0.87	0.90	0.91	0.91
PFE	0.83	0.84	0.91	0.91	0.83	0.83	0.90	0.91
PG	0.86	0.87	0.95	0.95	0.82	0.85	0.92	0.93
T	0.95	0.97	1.00	1.00	0.97	0.98	0.99	0.99
UTX	0.93	0.94	0.94	0.94	0.93	0.94	0.92	0.92
VZ	0.91	0.91	0.98	0.98	0.93	0.94	0.97	0.97
WMT	0.97	0.97	0.99	1.00	0.96	0.97	0.99	1.00
XOM	1.02	0.96	1.02	1.01	0.97	0.92	1.00	0.99

**Table 2:** QL ratios (left part) and FZL ratios (right part) for GAS-ST and GAS-SKST with respect to GAS-N for the two risk levels  $\alpha = 1\%$  and  $\alpha = 5\%$ . Values greater than one indicate outperformance of GAS-N and vice versa. Light gray cells indicate ratios higher than one (note that values are rounded). The forecasting period ranges from February 14, 2005, to February 3, 2009, for a total of  $H = 1,000$  out-of-sample observations. The model parameters are re-estimated at the monthly frequency.

### Joint VaR and ES model comparison

The FZLoss function implements the FZ loss and accepts the following arguments:

<sup>4</sup>Here we use re-estimate the model parameters at the monthly frequency (*i.e.*, every 21 new observations).



- data: vector of observations;
- VaR: vector of VaR predictions;
- ES: vector of ES predictions;
- alpha: the  $\alpha$  risk level,

and returns a numeric vector of the same size of data with the FZ losses computed at each point in time. For instance, using the VaR\_N and ES\_N vectors previously computed we can evaluate the associated FZ loss as:

```
> FZL <- FZLoss(data = tail(dji30ret[, "GE"], H), VaR = VaR_N, ES = ES_N, alpha = alpha)
```

where FZL is a numeric vector of length H.

The right part of Table 2 (FZL ratios) reports the results of performing model comparison analysis for all thirty daily stock return series in the dji30ret dataset according to the FZ loss reported in (3). It shows, for both  $\alpha = 1\%$  and  $\alpha = 5\%$ , the ratios between the average FZ loss of the considered models over the one delivered by GAS- $\mathcal{N}$ . Values greater than one indicate outperformance of GAS- $\mathcal{N}$ , and vice versa. Consistently with the comparison in terms of only VaR predictions, we find that GAS- $ST$  and GAS- $SKST$  perform similarly and are preferred to GAS- $\mathcal{N}$ .

## Conclusion

Under the regulation of the Basel Accords, risk managers of financial institutions need to rely on state-of-the-art methodologies for predicting and evaluating their downside risk (Board of Governors of the Federal Reserve Systems, 2012). This article illustrates the usefulness of the R package GAS in putting the theory of modern downside risk management into practice. The recommended strategy consists of four steps: (i) model specification, (ii) downside risk predictions, (iii) backtesting and (iv) model comparison. We illustrate the proposed implementation in R using the package GAS applied to the Value-at-Risk and Expected Shortfall estimation for the daily returns of the thirty Dow Jones Industrial Average constituents.

## Computational details

The results in this paper were obtained using R 3.5.0 with the package GAS version 0.2.8 available on CRAN at <https://cran.r-project.org/package=GAS>. Computations were performed on Windows 7 x64 (build 7601) Service Pack 1, x86\_64-w64-mingw32/x64 (64-bit) with Intel(R) Xeon(R) CPU E5-2560 v3 2.30 GHz.

## Acknowledgments

The authors thank Alexios Ghalanos, participants at the R/Finance conference 2017 in Chicago and participants at the session "Financial econometrics with R" at the 11th International Conference on Computational and Financial Econometrics 2017 in London. The authors acknowledge Google for financial support via the Google Summer of Code 2016 and 2017 project "GAS".

## Appendix

To obtain the score of the  $SKST$  distribution, we first write its log-density evaluated in  $r_t$ :

$$\log f_{SKST}(r_t; \mu, \sigma_t, \xi, \nu) = \log g + \log k + c - \log \sigma_t - \frac{\nu + 1}{2} \log \left[ 1 + \frac{\left[ \left( \frac{r_t - \mu}{\sigma_t} \right) k + m \right]^2}{(\nu - 2) (\xi^*)^2} \right],$$

where:

$$\begin{aligned} m &\equiv \mu_1 \left( \zeta - \frac{1}{\zeta} \right) \\ k &\equiv \sqrt{(1 - \mu_1^2) \left( \zeta^2 + \frac{1}{\zeta^2} \right) + 2\mu_1^2 - 1} \\ g &\equiv \frac{2}{\zeta + \frac{1}{\zeta}} \\ c &\equiv \frac{1}{2} [-\log(\nu - 2) - \log \pi] + \log \Gamma \left( \frac{\nu + 1}{2} \right) - \log \Gamma \left( \frac{\nu}{2} \right), \end{aligned}$$

with:

$$\mu_1 \equiv \frac{2\sqrt{\nu - 2}}{(\nu - 1)} \frac{\Gamma(\frac{\nu + 1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})},$$

and  $\zeta_t^* \equiv \zeta^{I\{z_t \geq 0\} - I\{z_t \leq 0\}}$ , where  $I\{\cdot\}$  is the indicator function, and  $z_t \equiv \left( \frac{r_t - \mu}{\sigma_t} \right) k + m$ . Taking the partial derivative of the log-density with respect to  $\theta_t$ , yields the score  $s_t$ , given by:

$$s_t \equiv \frac{\partial \log f_{SKST}}{\partial \theta_t} = \left( \frac{z_t (\nu + 1) (z_t - m)}{(\zeta_t^*)^2 (\nu - 2) + z_t^2} - 1 \right).$$

In the case where  $\zeta = 1$ ,  $m = 0$  and the score becomes:

$$s_t = \left( \frac{z_t^2 (\nu + 1)}{(\nu - 2) + z_t^2} - 1 \right),$$

which is the score of a Student-t distribution. Finally, when  $\nu \rightarrow \infty$ , we obtain:

$$s_t = z_t^2 - 1,$$

which is the score of the Normal distribution.

## Bibliography

- D. Ardia, K. Bluteau, K. Boudt, and L. Catania. Forecasting performance of Markov-switching GARCH models: A large-scale empirical study. *Econometrics and Operations Research*, 34(4):733–747, 2018. URL <https://doi.org/10.1016/j.ijforecast.2018.05.004>. [p1, 4]
- D. Ardia, K. Boudt, and L. Catania. Generalized autoregressive score models in R: The GAS package. *Journal of Statistical Software*, 20xx. URL <https://doi.org/10.2139/ssrn.2825380>. Forthcoming. [p1, 4]
- L. Bauwens and S. Laurent. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business & Economic Statistics*, 2005. URL <https://doi.org/10.1198/073500104000000523>. [p2]
- F. Bellini and V. Bignozzi. On elicitable risk measures. *Quantitative Finance*, 15(5):725–733, 2015. URL <https://doi.org/10.1080/14697688.2014.946955>. [p3]
- M. Bernardi and L. Catania. Comparison of Value-at-Risk models using the MCS approach. *Computational Statistics*, 31(2):579–608, 2016. URL <https://doi.org/10.1007/s00180-016-0646-6>. [p1]
- F. Blasques, S. J. Koopman, and A. Lucas. Maximum likelihood estimation for correctly specified generalized autoregressive score models: Feedback effects, contraction conditions and asymptotic properties. techreport TI 14-074/III, Tinbergen Institute, 2014. URL <http://www.tinbergen.nl/discussionpaper/?paper=2332>. [p2]
- F. Blasques, S. J. Koopman, K. Łasak, and A. Lucas. In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting*, 32(3):875–887, 2016. URL <https://doi.org/10.1016/j.ijforecast.2015.11.018>. [p2]
- Board of Governors of the Federal Reserve Systems. 99th annual report. Technical report, Board of Governors of the Federal Reserve Systems, 2012. URL <https://www.federalreserve.gov/publications/annual-report/files/2012-annual-report.pdf>. [p9]

- L. Catania, K. Boudt, and D. Ardia. *GAS: Generalized Autoregressive Score Models*, 2016. URL <https://github.com/LeopoldoCatania/GAS>. R package version 0.2.7. [p1]
- P. F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4):841–862, 1998. URL <http://www.jstor.org/stable/2527341>. [p3, 6]
- D. Creal, S. J. Koopman, and A. Lucas. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795, 2013. URL <https://doi.org/10.1002/jae.1279>. [p1, 2]
- R. F. Engle and S. Manganelli. CAViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381, 2004. URL <https://doi.org/10.1198/073500104000000370>. [p3, 6]
- C. Fernández and M. F. Steel. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371, 1998. URL <https://doi.org/10.1080/01621459.1998.10474117>. [p2]
- T. Fissler and J. F. Ziegel. Higher order elicibility and Osband’s principle. *The Annals of Statistics*, 44(4):1680–1707, 2016. URL <https://doi.org/10.1214/16-AOS1439>. [p3]
- C.-T. Gao and X.-H. Zhou. Forecasting VaR and ES using dynamic conditional score models and skew Student distribution. *Economic Modelling*, 53:216–223, 2016. URL <https://doi.org/10.1016/j.econmod.2015.12.004>. [p1]
- G. González-Rivera, T.-H. Lee, and S. Mishra. Forecasting volatility: A reality check based on option pricing, utility function, Value-at-Risk, and predictive likelihood. *International Journal of Forecasting*, 20(4):629–645, 2004. URL <https://doi.org/10.1016/j.ijforecast.2003.10.003>. [p3, 6]
- A. C. Harvey. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press, 2013. URL <https://doi.org/10.1017/CB09781139540933>. [p1]
- P. Jorion. *Value at Risk*. McGraw–Hill, New York, 1997. URL <https://doi.org/10.1036/0071464956>. [p1]
- R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. URL <https://doi.org/10.2307/1913643>. [p3]
- P. H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2):73–84, 1995. URL <https://doi.org/10.3905/jod.1995.407942>. [p3, 6]
- M. Marcellino, J. H. Stock, and M. W. Watson. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1):499–526, 2006. URL <https://doi.org/10.1016/j.jeconom.2005.07.020>. [p2]
- M. McAleer and B. Da Veiga. Single-index and portfolio models for forecasting Value-at-Risk thresholds. *Journal of Forecasting*, 27(3):217–235, 2008. URL <https://doi.org/10.1002/for.1054>. [p3, 6]
- A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton university press, Princeton, 2015. URL <https://doi.org/10.1111/jtsa.12177>. [p2]
- B. Narasimhan and S. G. Johnson. *cubature: Adaptive Multivariate Integration over Hypercubes*, 2017. URL <https://CRAN.R-project.org/package=cubature>. R package version 1.3-11. [p6]
- M. R. Nieto and E. Ruiz. Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, 32(2):475–501, 2016. URL <https://doi.org/10.1016/j.ijforecast.2015.08.003>. [p1]
- A. J. Patton, J. F. Ziegel, and R. Chen. Dynamic semiparametric models for expected shortfall, 2017. URL <https://doi.org/10.2139/ssrn.3000465>. Working paper. [p4]
- J. F. Ziegel. Coherence and elicibility. *Mathematical Finance*, 26(4):901–918, 2016. URL <https://doi.org/10.1111/mafi.12080>. [p3]

*David Ardia*  
*Institute of Financial Analysis*  
*University of Neuchâtel, Switzerland*  
*and*  
*Department of Finance, Insurance and Real Estate*  
*Laval University, Canada*  
ORCID 0000-0003-2823-782X  
[david.ardia@unine.ch](mailto:david.ardia@unine.ch)

*Kris Boudt*  
*Vrije Universiteit Brussel, Belgium*  
*and*  
*Vrije Universiteit Amsterdam, The Netherlands*  
ORCID 0000-0002-1000-5142  
[kris.boudt@vub.be](mailto:kris.boudt@vub.be)

*Leopoldo Catania (corresponding author)*  
*Department of Economics and Business Economics*  
*School of Business and Social Sciences*  
*Aarhus University, Denmark*  
ORCID 0000-0002-0981-1921  
[leopoldo.catania@econ.au.dk](mailto:leopoldo.catania@econ.au.dk)