Proceedings of the 63rd Annual Conference of the South African Statistical Association
© 2022 South African Statistical Association

Optimal window size detection in Value-at-Risk forecasting: A case study on conditional generalised hyperbolic models

Chun-Kai Huang[1], *Chun-Sung Huang*[2], *Jahvaid Hammujuddy*[3] *and Knowledge Chinhamu*[3]

[1]Curtin University, Bentley, Western Australia
[2]University of Cape Town, Cape Town, South Africa
[3]University of KwaZulu-Natal, Durban, South Africa

The conventional parametric approach for financial risk measure estimation involves determining an appropriate quantitative model, as well as a suitable historical sample period in which the model can be trained. While a lion's share of the existing literature entertains the identification of the most appropriate model for different types of financial assets, or across conflicting market conditions, little is known about the optimal choice of a historical sample period size (or window size) to train the model and estimate model parameters. In this paper, we propose a method to identify an optimal window size for model training when estimating risk measures, such as the widely-utilised Value-at-Risk (VaR) or Expected Shortfall (ES), under the generalised hyperbolic subclasses. We show that the accuracy of VaR estimates may increase significantly through our proposed method of optimal window size detection. In particular, our results demonstrate that, by relaxing the usual restriction of a fixed window size over time, superior VaR forecasts may be produced as a result of improved model parameter estimates.

Keywords: Hyperbolic, MSCI, Normal-inverse Gaussian, Value-at-Risk, Variance-gamma, Window size.

1. Introduction

An increasing number of studies in the ongoing literature has been dedicated to modelling the behaviour and characteristics of financial time series. Noticeably, a significant portion of these studies also includes contributions toward the estimation of financial risk measures. To adequately estimate financial risk measures, a robust methodology that can unequivocally describe the continuous movements of the time series needs to be identified at the onset. Subsequently, a procedure is implemented to accurately estimate the respective risk measures. Such a procedure typically involves specifying a sample period size (or window size) to employ the historical data for model training and the estimation of model parameters. This is usually imposed through a rule-of-thumb method instead of an adequate optimisation approach. However, errors in the estimation of model parameters may be exacerbated through an opaque choice of window sizes, leading to inferior risk measure estimates.

Corresponding author: Chun-Sung Huang (chun-sung.huang@uct.ac.za)
MSC2020 subject classifications: 62G32, 62M10, 91G70

Notwithstanding the above, there is a shortfall in the current literature on the identification of an optimal window size to effectively estimate model parameters when forecasting risk measures (such as VaR or ES). In practice, window sizes are often arbitrarily selected without any clear consensus or robust methodology. However, evidence from a number of prior research papers suggest that most estimation procedures (parametric or non-parametric) are sensitive to changes in window size (see, for example, Chen and Spokoiny, 2009; Halbleib and Pohlmeier, 2012; Sharma, 2012; Laker et al., 2017). In particular, a larger window size often results in low variance of estimates but raises the risk of modelling bias. On the contrary, small window sizes produce estimates that react efficiently to changing market conditions, but suffers from larger variations. Hence, the identification of an optimal window size becomes a critical task. Related studies on improving parameter estimation includes, but is not limited to, exponential smoothing, structural breaks, regime switching and adaptive point-wise estimation (see Čížek et al., 2009).

A wealth of models and methods for risk measure estimation have already been proposed in the existing body of knowledge. Prominent methodologies include, among others, the use of extreme value analysis (McNeil and Frey, 2000), the generalised lambda distribution (Corlu and Corlu, 2015) and quantile regression (Engle and Manganelli, 2004). In this paper, we focus on another popular class of distributions for describing financial returns, namely the generalised hyperbolic distributions (GHDs), when estimating risk measures. Such family of distributions are particularly suitable for capturing stylised facts, such as asymmetric and varying tail behaviours (including semi-heavy and heavy tails), embedded within financial data. The novel work of Eberlein and Keller (1995) was among the first to apply these extreme value distributions to financial modelling. The successes of GHDs in modelling financial data were further advocated by various subsequent studies, such as Eberlein and Prause (2002), Aas and Haff (2006), Hu and Kercheval (2007), and Huang et al. (2014), among others.

In this paper, we first deploy a GARCH(1,1) model in describing the daily returns volatility of our chosen dataset, the MSCI All Country World Index (ACWI). Specifically, we allow the distribution of the resulting GARCH(1,1) innovations to follow different subclasses of the GHDs (namely, the hyperbolic (HYP), the normal-inverse Gaussian (NIG), the generalised hyperbolic skewed-t (GHSt) and the variance-gamma (VG) subclasses). We show that the resulting VaR estimates, using the above models, can change considerably across different window sizes on the same out-of-sample set. This challenges the common practice of utilising an arbitrary fixed window size, and motivates a need for determining optimal window sizes when estimating risk measures.

We contribute to the existing literature by proposing a method to identify the required optimal window size, and show that such a method may effectively improve the estimation of model parameters and the resulting VaR forecasts. Furthermore, we proceed with a method that follows a daily rolling window procedure to detect an optimal size for each iteration. Our findings demonstrate the importance of relaxing the usual fixed window size restriction, and allow for time-varying window sizes when forecasting VaR. To the best of the authors' knowledge, there exists no literature relating to window size optimisation in VaR estimation under the GHD framework. In addition, although prior research exists in identifying systematic breaks (or structural breaks) and the maximum period of stability (see, for example, Spokoiny, 2009; Härdle et al., 2003), very few have been applied under the GHD framework. Hence, our study also provides further insight towards the limited research on GHDs' benefits in financial risk modelling.

The remainder of this paper proceeds as follows. In Section 2, we present the VaR methodology. Discussions around the subclasses of GHDs and optimal window size derivations are provided in Sections 3 and 4, respectively. Finally, we reveal our empirical results in Section 5 and conclude the paper in Section 6.

2. Value-at-Risk

While there are criticisms on the use of VaR, it remains a popular benchmark risk measure among banks and financial institutions for evaluating and estimating financial risks. In particular, it is directly linked to the adequate amount of market risk capital that financial entities must set aside to compensate for unprecedented large losses, as recommended by the Basel Committee on Banking Supervision. Even with the ongoing migration towards the more sophisticated Expected Shortfall as a measure of risk, in accordance with Basel III, VaR continues to be widely utilised by market participants in conjunction. Hence, further research to improve the forecast of VaR may continue to bear fruit for the fragile financial sector.

Formally, VaR is defined as a threshold amount such that the probability of the realised loss on a portfolio, over a given time horizon, exceeding this value is equal to a pre-specified confidence level. For a sequence of daily log-returns, R_t, on an existing portfolio, we assume $R_t = \mu_t + \sigma_t Z_t$, where Z_t represents the innovation characterised by some marginal distribution $F_Z(z)$. The parameters μ_t and σ_t are measurable with respect to Ω_{t-1}, all information on the process up to time $t-1$. Furthermore, if $F_R(r)$ denotes the distribution of R_t, we can deduce that

$$F_{R_{t+1}|\Omega_t}(r) = P\left(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \leq r \mid \Omega_t\right) = F_Z\left(\frac{r-\mu_{t+1}}{\sigma_{t+1}}\right). \tag{1}$$

Consequently, we can express VaR for day $t+1$, with probability of exceedance equal to $1-p$, as

$$\mathrm{VaR}_p(t+1) = \mu_{t+1} + \sigma_{t+1} z_p, \tag{2}$$

where z_p denotes the lower p^{th} quantile of Z_t. For forecasting purposes, we need to first specify a model for the dynamics of the mean, μ_{t+1}, and volatility, σ_{t+1}. We utilise the celebrated GARCH(1,1) process for the volatility and the AR(1) process for the mean, i.e.,

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta \sigma_t^2 \quad \text{and} \quad \mu_{t+1} = \phi R_t, \tag{3}$$

where $\varepsilon_t = \sigma_t Z_t, \alpha_0 > 0, \alpha_1 \geq 0, \beta \geq 0, \alpha_1 + \beta < 1$, and ϕ is the AR(1) coefficient.

Following McNeil and Frey (2000), we fit the GARCH(1,1) model using a pseudo maximum likelihood (PML) procedure, which minimises the assumptions about the distribution of innovations, and estimates μ_{t+1} and σ_{t+1} using standard one-day ahead forecasts. We further suggest this to be amalgamated with the assumption that the innovations are distributed according to a GHD subclass, and estimate the resulting z_p accordingly. This may then be implemented in a rolling window procedure to produce daily out-of-sample forecasts of VaR. Consequently, as per standard procedure, the resulting forecasts are then backtested against the realised daily returns observed. We utilised two widely-accepted backtests for VaR, namely, the Kupiec likelihood ratio test (Kupiec, 1995) and the Christoffersen conditional coverage test (Christoffersen et al., 2001). While the former tests for the unconditional coverage of the correct number of exceedances in our VaR estimates, the latter tests for the conditional coverage.

3. Generalised hyperbolic distributions (GHD)

GHDs, such as the HYP, NIG, VG and GHSt distributions, have the ability to cater for asymmetric, heavy and semi-heavy tailed datasets. They enable researchers to model data across a wide variety of disciplines, including finance and economics. By adequately capturing the above-mentioned stylised facts embedded in financial data, the resulting VaR estimates may also be greatly improved (see, for example, Huang et al., 2014). In this section, we shall introduce the full GHD and its range of subclasses.

3.1 The full GHD model

The probability density function (pdf) of the full GHD is given by

$$f_{GHD}(x) = \frac{\left(\alpha^2 - \beta^2\right)^{\lambda/2} \left(\delta^2 + (x-\mu)^2\right)^{(\lambda - 1/2)/2} K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) \exp(\beta(x-\mu))}{\sqrt{2\pi}\alpha^{\lambda-1/2}\delta^{\lambda} K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)}, \quad (4)$$

where K_j is the modified Bessel function of the third kind with order j (Abramowitz and Stegun, 1972), and μ is the location parameter. It should also be noted that the domain of the parameters must satisfy the following conditions

$$\begin{aligned} \delta > 0, |\beta| < \alpha, &\text{ if } \lambda = 0, \\ \delta > 0, |\beta| \leq \alpha, &\text{ if } \lambda < 0, \\ \delta \geq 0, |\beta| < \alpha, &\text{ if } \lambda > 0, \end{aligned}$$

where δ serves as a scaling factor, α determines the shape, β determines the skewness, and λ influences the kurtosis (Necula, 2009). We utilise the maximum likelihood estimation (MLE) for parameter estimates of all GHD subclasses. The various subclasses of the GHD can be obtained by considering different assumptions and asymptotic behaviours of the parameters above. We demonstrate this in the sequel.

3.2 The Hyperbolic (HYP) distribution

The HYP distribution (with $\lambda = 1$) allows us to determine the shape of the distribution by controlling both the gradient and skewness parameters. The HYP distribution is characterised by having a hyperbolic log-density function and exponential tails. A random variable follows the HYP distribution if its pdf is given by

$$f_{HYP}(x) = \frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)} e^{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)}, \quad (5)$$

where K_1 denotes the Bessel function of the third kind with order 1. The parameters α and β, with $\alpha > 0$ and $0 \leq |\beta| < \alpha$, represent the gradient and the skewness, respectively. Finally, $\delta > 0$ is the scale parameter and $\mu \in R$ is the location parameter.

3.3 The Normal-Inverse Gaussian (NIG) distribution

The NIG distribution is well-known for its ability to capture the asymmetric semi-heavy tails of financial returns (Andersson, 2001; Venter and de Jongh, 2002). In particular, the NIG distributions are most appropriate when the two extreme tails of the returns distribution to be modelled are not too heavy (Aas and Haff, 2006). The pdf of the NIG, as a subclass of GHDs with $\lambda = -1/2$, can be expressed as

$$f_{NIG}(x) = \frac{\alpha\delta}{\pi} e^{\delta\sqrt{\alpha^2-\beta^2}+\beta(x-\mu)} \frac{K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\sqrt{\delta^2+(x-\mu)^2}}, \tag{6}$$

where K_1 denotes the Bessel function of the third kind with order 1.

3.4 The Variance-Gamma (VG) distribution

The VG distribution has tails that decrease less rapidly than that of a Gaussian distribution. Such a characteristic makes the VG a suitable model for phenomena where extreme values are more probable than in the case of a Gaussian distribution, such as logarithmic returns from financial assets (Madan and Seneta, 1990). We attain the pdf of the VG distribution from the full GHD when $\lambda > 0$ and $\delta \to 0$. Hence, we have

$$f_{VG}(x) = \frac{\left(\alpha^2-\beta^2\right)^{\lambda} |x-\mu|^{\lambda-1/2} K_{\lambda-1/2}(\alpha|x-\mu|)}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-1/2}} e^{\beta(x-\mu)}, \tag{7}$$

where $K_{(\lambda-1/2)}$ denotes the Bessel function of the third kind with order $\lambda - 1/2$.

3.5 The Generalised Hyperbolic Skewed-t (GHSt) distribution

Finally, the pdf of the GHSt distribution is obtained by letting $\alpha \to |\beta|$ in the full GHD. This results in the following expression

$$f_{GHSt}(x) = \frac{2^{1/2+\lambda}\delta^{-2\lambda}|\beta|^{1/2-\lambda} K_{1/2-\lambda}\left(\sqrt{\beta^2\left(\delta^2+(x-\mu)^2\right)}\right)\exp(\beta(x-\mu))}{\Gamma(-\lambda)\sqrt{\pi}\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{1/2-\lambda}}, \tag{8}$$

for $\beta \neq 0$ and $\lambda < 0$. If $\beta = 0$, we obtain the non-central (scaled) Student's t-distribution. Notably, the GHSt distribution exhibits one heavy polynomial tail and one semi-heavy exponential tail. This unique property makes the GHSt distribution particularly dissimilar to the range of subclasses mentioned above. More importantly, it allows the GHSt distribution to uniquely model skewed data with dissimilar tail behaviours, which are commonly observed in financial data (Aas and Haff, 2006).

4. Optimal window size

The choice of an appropriate window size can affect the resulting model parameter estimates, and consequently the accuracy of the final VaR forecasts. However, identifying an optimal window size remains a difficult task. In the current literature, most analyses are conducted by utilising a fixed window size that is arbitrarily chosen according to a rule-of-thumb, or is only tested against a few alternative choices in order to determine an appropriate size. The chosen window size is then used to perform a rolling window procedure to estimate VaR at each time step of the out-of-sample data. Even though such methods of window size selection are deemed reasonable by prior studies, it can produce biased parameter estimations and inadequate VaR forecasts as a result. To remedy such drawback, a more effective procedure that can accommodate varying window sizes at each time step needs to be derived. Moreover, for robustness, the said procedure needs to optimise some criterion related to time homogeneity. In our current study, we deploy five different criteria for selecting an optimal window size at each rolling window iteration. The window sizes are chosen to either (i) minimise or maximise the standard deviation; (ii) minimise or maximise the kurtosis; or (iii) include a change-point with a fixed right-end point. While the reasons for our choice of (i) and (ii) are more apparent for risk measure focus, with the estimation of tail events, the justification for (iii) is more inline with the notion of structural breaks detection in a given dataset.

A change-point is defined as a location in the dataset in which the statistical properties of the sequence experiences a significant change. We identify change-points optimally using the at-most-one-change-point (AMOC) procedure (Silva and Teixeira, 2008) and the binary segmentation (BinSeg) procedure (Scott and Knott, 1974), with variance as the optimisation measure. The detection of a change-point can be viewed as a hypothesis test, whereby the null, H_0, corresponds to no change-point, and the alternative, H_1, advocates the existence of a change-point. The likelihood ratio method (i.e., AMOC) involves calculating the maximum likelihood under both hypotheses above. Subsequently, the ratio is maximised over all possible change-point locations. The BinSeg approach, on the other hand, is a generalisation of the AMOC, whereby the data sequence is segmented into two parts once a change-point is detected. Finally, each segment is then tested for change-points and the process continues until a pre-specified threshold is reached, or until no further change-points are identified.

5. Data and Empirical Results

In our study of optimal window size detection and the proposed varying window size approach, we use daily log-returns of the MSCI ACWI index over a 15-year period, ranging from 27 August 2001 to 25 August 2016. The MSCI ACWI is a flagship global index that aims to capture equity returns of large- and mid-cap stocks across 23 developed and 24 emerging markets. This offers investors a fully integrated view of exposure to all sources of equity returns using just a single index.

Table 1 shows the descriptive statistics of the original return series over the entire sample period, as well as the resulting innovations after fitting the GARCH(1,1) model to the same data. The large excess kurtosis of the original return series is a common characteristic found in financial data, which implies a vast tail deviation from that of the Gaussian distribution. In addition, we observe that the resulting GARCH(1,1) innovations still exhibit heavy tails, albeit to a lesser degree. These are both well-known stylised facts of financial time series (Cont, 2001).

Notably, the heavy-tails of the residuals are even more pronounced when we implement a rolling

Table 1. Summary statistics for MSCI ACWI and its GARCH innovations.

Data	**Mean**	**Std. dev.**	**Min**	**Max**	**Excess kurtosis**	**Skewness**
ACWI	−0.000130	0.010289	−0.089030	0.073713	8.172448	0.396315
Innovations	0.041017	0.999549	−3.841127	6.237750	1.311213	0.303292

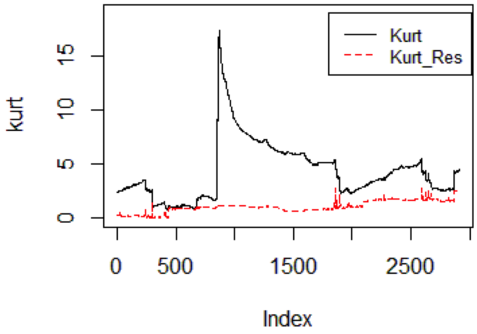

Figure 1. Rolling excess kurtosis for ACWI returns and its GARCH innovations (1000-day rolling window size).

window procedure to analyse the varying kurtosis over time. Figure 1 records the time-varying excess kurtosis of the original return series, as well as the corresponding innovations, when iterated at each time step through a 1000-day rolling window procedure. It is evident that the conditional kurtosis can deviate significantly from that of a Gaussian distribution at isolated time periods within the return series (as depicted by the sudden spikes and long periods of consistent non-zero values).

To encapsulate the effects of window size selection, we first conduct a VaR estimation procedure using a fixed window size approach. Our estimation procedure is then repeated across a range of different window sizes on the same dataset. Specifically, we will estimate the rolling window daily VaR at each time step of the out-of-sample period (from day 1501) using fixed window sizes ranging from 100 to 1500 days (at 25-day increments). For each window size, we forecast the daily VaR using the GARCH(1,1) filter with a conditional distribution following a GHD subclass. Finally, the sequence of VaR estimates are then backtested against the actual daily returns observed, and the respective p-values recorded.

Under both the Kupiec likelihood ratio and the Christoffersen conditional coverage tests, where the null hypothesis advocates for the model being 'correct' or well-specified, a higher p-value is desired. In Table 2, we present the mean, standard deviation, minimum, maximum and coefficient of variation (CV) of the different p-values obtained for both backtests across the various GHD subclasses. Interestingly, across all GHD subclasses evaluated, a range of 400-500 days appears to be the optimal choice when implementing a fixed window size. Figures 2 to 6 presents the changing p-values, for both the Kupiec and Christoffersen tests, across the range of fixed window sizes. These observations provide further empirical evidence that the performance of VaR models may depend heavily on the appropriate choice of window sizes. Apart from our explicit evidence to infer 400-500 days as an optimal range for window sizes, we observe that a larger window size tends to consistently produce inferior VaR estimates across all GHD subclasses. On the contrary, smaller window sizes, which allows more emphasis on recent market data, tends to provide more ideal VaR estimates.

To implement a varying window size selection process, within a rolling window procedure, an optimising criterion is needed to determine an adequate window size at each iteration. We shall utilise a wide range of different criteria and compare the resulting model performances through the two

Table 2. Summary statistics of VaR backtesting p-values from Kupiec and Christoffersen tests, using rolling fixed window sizes ranging from 100 to 1500 days, at the 97.5% VaR level.

GHD subclass	VaR test	Mean	Std. dev.	Min	Max	Window size for max	CV
GHD	Kupiec	0.006118	0.006207	0.000731	0.037242	400	1.014557
	Christoffersen	0.021915	0.018777	0.002515	0.107563	400	0.856805
HYP	Kupiec	0.006556	0.006174	0.000478	0.027580	400/475	0.941610
	Christoffersen	0.023240	0.019076	0.001644	0.084333	400/475	0.820814
NIG	Kupiec	0.006569	0.006794	0.000310	0.037242	400	1.034228
	Christoffersen	0.023146	0.020220	0.001061	0.107563	400	0.873588
VG	Kupiec	0.003027	0.003771	0.000198	0.027580	400	1.245834
	Christoffersen	0.011702	0.011861	0.000676	0.084333	400	1.013565
$GHSt$	Kupiec	0.005125	0.005726	0.000478	0.037242	400	1.117432
	Christoffersen	0.018682	0.017402	0.001644	0.107563	400	0.931458

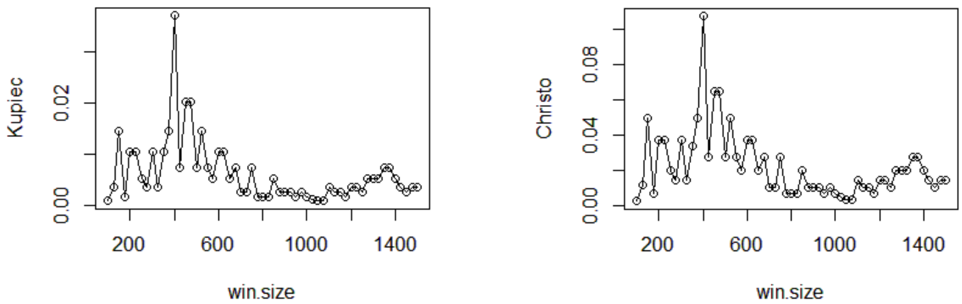

Figure 2. p-values of Kupiec and Christoffersen tests for rolling fixed window sizes ranging from 100 to 1500 days when using GHD.

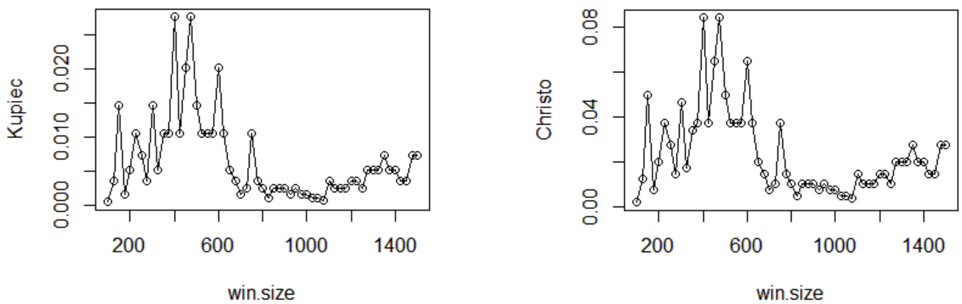

Figure 3. p-values of Kupiec and Christoffersen tests for rolling fixed window sizes ranging from 100 to 1500 days when using HYP.

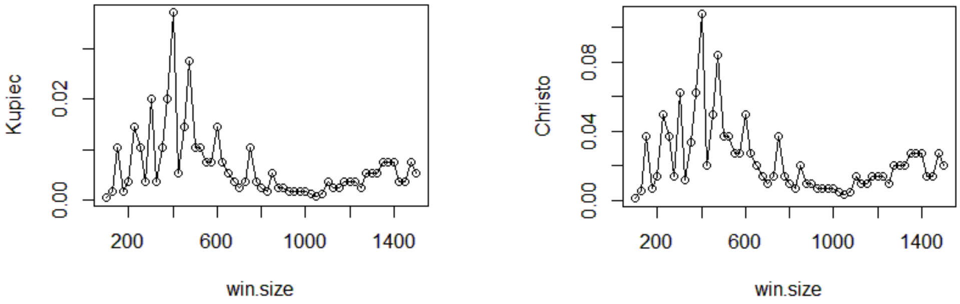

Figure 4. p-values of Kupiec and Christoffersen tests for rolling fixed window sizes ranging from 100 to 1500 days when using NIG.

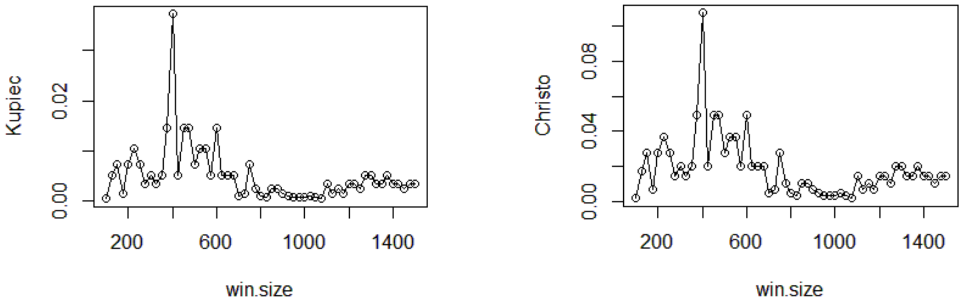

Figure 5. p-values of Kupiec and Christoffersen tests for rolling fixed window sizes ranging from 100 to 1500 days when using VG.

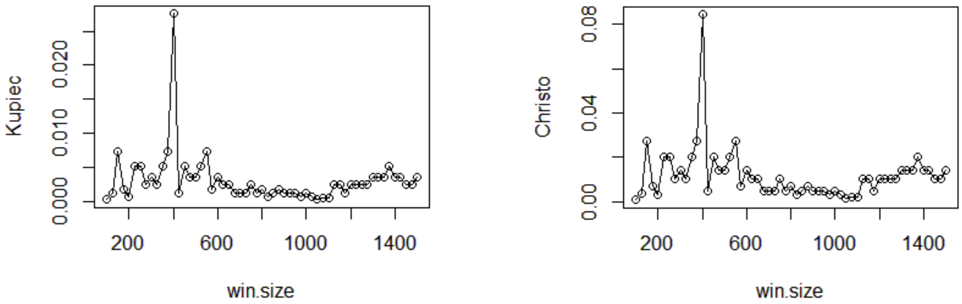

Figure 6. p-values of Kupiec and Christoffersen tests for rolling fixed window sizes ranging from 100 to 1500 days when using GHSt.

Table 3. VaR backtesting p-values from Kupiec and Christoffersen tests under varying window sizes across different optimising criteria.

GHD subclass	VaR test	Max. std. dev.	Min. std. dev.	Max Kurtosis	Min Kurtosis	Change-point AMOC	Change-point BinSeg
GHD	Kupiec	0.176607	< 0.000001	0.002444	0.005169	0.020189	0.027580
	Christoffersen	0.332875	0.000004	0.010141	0.019968	0.065124	0.084333
HYP	Kupiec	0.176607	< 0.000001	0.002444	0.005169	0.020189	0.010451
	Christoffersen	0.332875	0.000001	0.010140	0.019968	0.065124	0.037132
NIG	Kupiec	0.140315	< 0.000001	0.003574	0.003574	0.010451	0.014609
	Christoffersen	0.287102	0.000001	0.014331	0.014331	0.037132	0.049541
VG	Kupiec	0.176607	< 0.000001	0.001652	0.005169	0.020189	0.014609
	Christoffersen	0.332875	< 0.000001	0.007076	0.019968	0.065124	0.049541
$GHSt$	Kupiec	0.140315	< 0.000001	0.001652	0.002444	0.010451	0.014609
	Christoffersen	0.287102	< 0.000001	0.007076	0.010141	0.037132	0.049541

standard backtests. Firstly, we select an optimal window size for each iteration of the rolling window according to a maximum standard deviation, minimum standard deviation, maximum kurtosis and minimum kurtosis, over the different window sizes ranging from 100 to 1500 days (at 25-day increments). Secondly, we utilise two change-points procedures, namely, AMOC and BinSeg, for identifying the change-point(s) within each rolling window (with the base window size set to 1500 days). For AMOC, the period between the change-point and the most right-end point in a given rolling window is selected. For BinSeg, the period between the largest change-point and the most right-end point of the rolling window is selected instead. Finally, the different VaR estimates are obtained through the various optimal window sizes detected per criteria and backtested accordingly.

Table 3 presents the Kupiec test and Christoffersen test p-values for the varying window size procedure using the different optimisation methods mentioned above. The minimum standard deviation, maximum kurtosis and minimum kurtosis appears to be inadequate as optimising criteria, each exhibiting poorer results in comparison to the average performance of the alternative fixed window size approach (as shown in Table 2). Surprisingly, while producing superior results to that of the above-mentioned trio, the two change-point procedures seem to be marginally better or on par with the average performance of using fixed window sizes (at a 5% confidence level). However, both AMOC and BinSeg are still less robust than using an optimal fixed window of 400 days (when such a window size may be determined a priori). The selection of varying window sizes through maximum standard deviation overwhelmingly outperforms the alternative criteria, as well as the fixed window approach. It also consistently produces the highest p-values among all criteria across the various GHD subclasses. Overall, our results also demonstrate that a GARCH(1,1) with a conditional distribution of either the GHD, HYP or VG is the most robust model for forecasting VaR in MSCI ACWI returns.

Figure 7 shows the changing window sizes at each rolling window iteration for the maximum standard deviation, AMOC and BinSeg optimising criteria. Notably, the changes in optimal varying

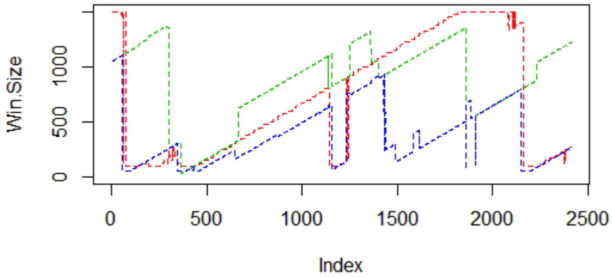

Figure 7. Varying window sizes over time using various selection criteria (red = max standard deviation, green=AMOC, blue=BinSeg).

window sizes over time also reflect the market conditions that ensued. For instance, the recommended optimal window size contracts sharply during periods of market distress, which adequately allows rolling windows to capture more recent data that better represents the prevailing market downturn. This is clearly exemplified by the infamous 2008 Global Financial Crisis, the subsequent Eurozone crisis, as well as the 2015-2016 selloff triggered by the Chinese stock market turbulence (as depicted in Figure 7). While both the maximum standard deviation and BinSeg criteria are efficient in following market trends, the AMOC, with the restriction of at most one change-point detection, tend to suffer from excessive lags in its response.

6. Limitations and concluding remarks

In this paper, we examine how the accuracy of VaR estimates, under conditional GHD subclasses, may vary depending on the choice of an appropriate window size when estimating model parameters. Forecasting performances were measured according to the widely-accepted Kupiec likelihood ratio and the Christoffersen conditional coverage tests. Evidently, our analyses showed that the robustness of VaR models rely heavily on the appropriate selection of window sizes for parameter estimation. In order to identify an optimal window size (for each rolling window iteration), we investigated several possible optimisation methods to enable a time-varying window size procedure, and compared our results to that of the classical fixed window implementation. The optimising criteria employed to select a suitable varying window size were given by either maximising or minimising the standard deviation, maximising or minimising the kurtosis, using an AMOC procedure, or using a BinSeg procedure. It is worthwhile noting that the AMOC and BinSeg procedures appeared to be only as good as the average performance of the fixed window size approach, and worse off when an optimal fixed window size is utilised. Maximising the standard deviation under the varying window size approach seemed to produce the best risk forecasting results under the GHD framework. Our findings advocate the critical need to optimise window sizes prior to parameter estimation when forecasting VaR. Moreover, it is necessary to relax the usual restriction of a fixed window size, and allow for time-varying window sizes instead. Lastly, it is necessary to evaluate a range of optimising criteria in order to identify the most appropriate criterion to deploy.

An important caveat to our study is the limited number of criteria investigated for optimal window size selection. Further research may include implementing other attractive methods, such as the adaptive pointwise estimation (see Čížek et al., 2009) or segment neighbourhood procedure for change-point identification (see Auger and Lawrence, 1989), and analysing the accuracy of resulting VaR forecasts. Additionally, it may be worthwhile to explore whether the suitability of selection criteria, or procedure, may change significantly under different market conditions (when certain stylised facts may become extreme), or when different distributional assumptions for the data series are implemented. Finally, with the recommended migration towards Expected Shortfall (as per the latest Basel Accords), further studies of optimal window sizes detection to improve ES estimation is paramount.

References

Aas, K. and Haff, I. H. (2006). The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, **4**, 275–309.

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series 55. Tenth Printing*. ERIC.

Andersson, J. (2001). On the normal inverse Gaussian stochastic volatility model. *Journal of Business & Economic Statistics*, **19**, 44–54.

Auger, I. E. and Lawrence, C. E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Biology*, **51**, 39–54.

Chen, Y. and Spokoiny, V. (2009). Modeling and estimation for nonstationary time series with applications to robust risk management.
URL: *https://www.academia.edu/22109590*

Christoffersen, P., Hahn, J., and Inoue, A. (2001). Testing and comparing value-at-risk measures. *Journal of Empirical Finance*, **8**, 325–342.

Čížek, P., Härdle, W., and Spokoiny, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models. *The Econometrics Journal*, **12**, 248–271.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, **1**, 223.

Corlu, C. G. and Corlu, A. (2015). Modelling exchange rate returns: which flexible distribution to use? *Quantitative Finance*, **15**, 1851–1864.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, **1**, 281–299.

Eberlein, E. and Prause, K. (2002). The generalized hyperbolic model: financial derivatives and risk measures. *In Mathematical Finance—Bachelier Congress 2000*. Springer, 245–267.

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

Halbleib, R. and Pohlmeier, W. (2012). Improving the value at risk forecasts: Theory and evidence from the financial crisis. *Journal of Economic Dynamics and Control*, **36**, 1212–1228.

Härdle, W., Herwartz, H., and Spokoiny, V. (2003). Time inhomogeneous multiple volatility modeling. *Journal of Financial Econometrics*, **1**, 55–95.

Hu, W. and Kercheval, A. (2007). Risk management with generalized hyperbolic distributions. *In Proceedings of the fourth IASTED international conference on financial engineering and applications*. ACTA Press, 19–24.

Huang, C.-S., Huang, C.-K., and Chinhamu, K. (2014). Assessing the relative performance of heavy-tailed distributions: Empirical evidence from the Johannesburg Stock Exchange. *Journal of Applied Business Research*, **30**, 1263–1286.

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

Laker, I., Huang, C.-K., and Clark, A. E. (2017). Dependent bootstrapping for value-at-risk and expected shortfall. *Risk Management*, **19**, 301–322.

Madan, D. B. and Seneta, E. (1990). The variance gamma (VG) model for share market returns. *Journal of Business*, **63**, 511–524.

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.

Necula, C. (2009). Modeling heavy-tailed stock index returns using the generalized hyperbolic distribution. *Romanian Journal of Economic Forecasting*, **10**, 118–131.

Scott, A. J. and Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, **30**, 507–512.

Sharma, M. (2012). Evaluation of Basel III revision of quantitative standards for implementation of internal models for market risk. *IIMB Management Review*, **24**, 234–244.

Silva, E. G. and Teixeira, A. A. (2008). Surveying structural change: Seminal contributions and a bibliometric account. *Structural Change and Economic Dynamics*, **19**, 273–300.

Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk. *The Annals of Statistics*, **37**, 1405–1436.

Venter, J. H. and de Jongh, P. J. (2002). Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk*, **4**, 1–24.