UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

LUIZ BEZERRA DE OLIVEIRA LIMA NETO

TAIL RISK FORECASTING WITH GAS MODELS

Porto Alegre
2018

LUIZ BEZERRA DE OLIVEIRA LIMA NETO

TAIL RISK FORECASTING WITH GAS MODELS

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Prof. Dr. Hudson da Silva Torrent

Porto Alegre
2018

CIP - Catalogação na Publicação

Lima Neto, Luiz Bezerra de Oliveira
TAIL RISK FORECASTING WITH GAS MODELS / Luiz Bezerra de Oliveira Lima Neto. -- 2018.
41 f.
Orientador: Hudson da Silva Torrent.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2018.

1. Generalized Autoregressive Score (GAS). 2. Semiparametric GAS. 3. Volatility Forecasting. 4. Value-At-Risk (VaR). 5. Expected Shortfall (ES). I. da Silva Torrent, Hudson, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

LUIZ BEZERRA DE OLIVEIRA LIMA NETO

TAIL RISK FORECASTING WITH GAS MODELS

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Aprovada em: Porto Alegre, 11 de Junho de 2018.

BANCA EXAMINADORA:

Prof. Dr. Hudson da Silva Torrent - Orientador
Universidade Federal do Rio Grande do Sul - UFRGS

Prof. Dr. Flávio Augusto Ziegelmann
Universidade Federal do Rio Grande do Sul - UFRGS

Prof. Dr. João Frois Caldeira
Universidade Federal do Rio Grande do Sul - UFRGS

André Alves Portela Santos
Universidade Federal de Santa Catarina - UFSC

AGRADECIMENTOS

Agradeço a todos que estiveram comigo nesta caminhada, em especial aos meus pais pelo apoio incondicional.

Ao Prof. Hudson, pela confiança, ensinamentos e tempo dedicado na orientação deste projeto. Aos demais professores do PPGE e PPGA pela excelência profissional e conhecimentos transmitidos.

RESUMO

Este artigo compara previsões de Value-at-Risk (VaR) e Expected Shortfall (ES) obtidas através do modelo GAS proposto por Creal et al. (2013) com modelos alternativos. Primeiramente é feita uma investigação dentro-da-amostra do GAS semiparamétrico de Blasques et al. (2014) e é proposta uma metodologia para a escolha da bandwidth. A parte empírica é realizada através de dois estudos distintos. O primeiro tem como foco explorar a capacidade preditiva da especificação semiparamétrica contra modelos paramétricos. A segunda compara os diversos modelos paramétricos para avaliar se o aumento da complexidade do modelo tem impacto direto na qualidade da previsão de medidas de risco. O estudo empírico é conduzido com oito indices, quatro taxas de câmbio de moedas internacionais e duas taxas de câmbio de cripto-moedas. Os modelos são comparados em termos de Quantile Loss de Gonzales-Rivera (2014) e através do procedimento MCS de Hansen et al. (2011). Os resultados indicam que os modelos semiparamétricos melhoram a qualidade de previsão dos modelos GAS(n) para o VaR e a ES. Modelos mais complexos como Beta-Skew-t-EGARCH de Surracat e Harvey (2013) produzem melhores resultados que os modelos mais básicos. Contudo, utilizando o procedimento MCS encontramos que na maioria dos casos os modelos são estatísticamente equivalentes em termos da sua habilidade em prever o VaR.

Palavras-chaves: Generalized autoregressive score (GAS). GAS semiparamétrico. Densidade kernel. Previsão de volatilidade. Value at risk (VaR). Expected shortfall (ES).

ABSTRACT

This paper compares Value-At-Risk (VaR) and Expected Shortfall (ES) forecasts delivered by the recently developed GAS models of Creal et al. (2013) with alternative models. First, we perform an in-sample investigation of the Semiparametric model of Blasques et al. (2014) and propose a methodology for bandwidth selection. The empirical part is carried out through two studies. The first aims to explore the semiparametric specification against its parametric counterparts. The second compares various parametric specifications with the objective of evaluating whether increasing the model's complexity improves its ability to forecast risk measures. The empirical study is conducted with eight index series, four foreign exchange rates and two Crypto-Currency exchange rates. The models are compared in terms of Quantile Loss of Gonzales-Rivera (2014) and with the MCS procedure of Hansen et al. (2011). We find that the semiparametric model enhance the forecasting ability of the GAS(n) model for both VaR and ES. More complex models such as the Beta-Skew-t-EGARCH two-components from Surracat and Harvey (2013), produced better results than less complex models. However by performing the MCS procedure we found that in most of cases the models are statistically equivalent in terms of the ability to forecast VaR levels.

Key-words: Generalized autoregressive score (GAS). Semiparametric GAS. Kernel density. Volatility forecasting. Value at risk (VaR). Expected shortfall (ES).

SUMÁRIO

1 INTRODUCTION **8**

2 METHODOLOGY **13**

2.1 GAS Models 13

2.1.1 Semiparametric framework for GAS models 14

2.1.2 Estimation Algorithm 16

2.1.3 Bandwidth Selection 17

2.1.4 Kernel Density Derivatives 18

2.2 Beta-t-EGARCH Models 19

2.3 Other Conditional Variance Models 20

2.3.1 GARCH Models 21

2.4 Prediction methods of VaR and ES 22

2.5 Evaluating the different approaches 23

2.6 Codes and packages 24

3 SIMULATION STUDIES **25**

3.1 Estimation accuracy 25

3.2 Impact of bandwidth b_w on parameter estimates 28

4 EMPIRICAL APPLICATION **31**

4.1 Empirical Performance - Semiparametric GAS 32

4.2 Empirical Performance - Parametric Models 35

5 CONCLUSION **38**

REFERÊNCIAS **39**

1 INTRODUCTION

The attention placed into risk management has increasingly grown throughout the last decades, in significant part due to the previous financial crises, e.g., Internet bubble (2000) and Subprime (2008), but also as a result of the financial implications of the capital requirements banks are restricted. In this context, the search for metrics that can model adequately the risk embedded into financial products is of vital role. After Basel Committee I, VaR (Value-at-Risk) has emerged as the industry standard for risk measures. The VaR provides, for a fixed time horizon, the asset's loss or return that is expected to be exceeded with a given confidence (or, probability) level. More strictly speaking the VaR of a portfolio at the level of confidence α is provided by the smallest number l such that the probability that the loss L exceeds l is no larger than $(1-\alpha)$. Formally,

$$VaR_{\alpha} = inf\{l \in \mathbb{R} : P(L>l) \leq (1-\alpha)\} = inf\{l \in \mathbb{R} : F_L(l)>\alpha\}, \tag{1}$$

In a statistical point of view, VaR is simply a quantile of the loss distribution which grants its user with a straightforward way of looking at risk, Roccioletti (2015). This intuitive character has helped to boost its popularity and turn it into the most adopted risk metric into the financial industry. Despite its broad acceptance, VaR was criticized by Artzner (1999), who proved that it is not a coherent measure, for the reason that it is against the sub-additivity axiom, which means that does not always take into account diversification benefits. Let r_1 and r_2 be the log-returns of any two assets, a risk measure $\varrho(r)$ is considered a coherent measure if respects the following axioms:

1. Monotonicity: $\varrho(r_1) \leq \varrho(r_2)$ if asset 1 has weak stochastic dominance over asset 2.
2. Sub-additivity: $\varrho(r_1 + r_2) \leq \varrho(r_1) + \varrho(r_2)$ which means that a portfolio should have less risk than the risk of two assets measured individually.
3. Homogeneity: $\varrho(kr_1) = k\varrho(r_1)$, This axiom states that risk preference has nothing to do with the risk metric.
4. Risk free condition: $\varrho(kr_1 + \gamma) = k\varrho(r_1) - \gamma$, if the return associated with asset γ is deterministic, then there is no risk associated to this asset.

Artzner (1999) paved the way for the arrival of the ES (Expected Shortfall), which circumvented the problems related to the VaR by being a coherent risk measure. The expected shortfall is defined as the expected return on the portfolio in the worst $100\alpha\%$ of the cases. Following Patton, Ziegel and Chen (2017), if X_t is the return on some asset over some horizon with conditional (on information set $\mathcal{F}_{t-1}$) distribution F_t, the α-level ES is

$$\begin{aligned} & ES_t(x) = -E[X_t|X_t \leq -VaR_t|\mathcal{F}_{t-1}], \\ where \quad & VaR_t = F_t^{-1}(\alpha), \;\; for \;\; \alpha \in (0,1), \\ and \quad & X_t|\mathcal{F}_{t-1} \sim F_t \;, \end{aligned} \tag{2}$$

Where F_t is strictly increasing with finite mean. Roccioletti (2015) shows that expected Shortfall is thus related to VaR by

$$ES_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\beta(L)d\beta, \tag{3}$$

Another drawback of using the VaR is that it is tail insensitive. As we can see in eq. (1.3), the ES integrates over the entire left tail, and thus it is sensitive to the tail shape. The ES, however, has a defect of its own, it is not elicitable, which is a mathematical property important for assessing forecasting performance. Gneiting (2011), argues that due to the lack of elicitability the ES cannot be backtested. On the other hand, a current of researchers disagree with this point of view. For instance, Acerbi and Szekely (2014) proposed three different non-parametric backtesting methods for the ES. In the sense of being a coherent risk measure, the ES is a better alternative to VaR. On this ground, the most recent Basel Committee has released a consultation paper with the recommendation to replace VaR with ES for determining the capital charge of internal based models but to kept VaR as the measure to be backtested in the usual way.

Volatility estimation is a crucial part of both VaR and ES computation. The process is usually addressed by GARCH models, that are by far the most accepted procedure for volatility modelling. More recently Creal, Koopman and Lucas (2013) and Harvey (2013) proposed to use the score of the conditional density function as the driver of time variation in the parameters of the time series process. The model is known as Generalized Autoregressive Score models (GAS) and is related to GARCH models in the sense that estimation process is carried out via maximum likelihood, supposing that the residuals have some known parametric distribution. For instance, it nests many other successful time series models, such as the gaussian GARCH model, the Autoregressive Conditional Duration (ACD) model of Engle and Russel (1998), and the Multiplicative Error model (MEM) of Engle and Gallo (2006). As noted by Creal, Koopman and Lucas (2013), since the GAS framework is based on the score, it exploits the complete density structure rather than the mean and higher moments only, as it is the case of GARCH models.

In practice the conditional distribution is unknown and normality assumption on the innovation distribution is often assumed. Even though empirical facts about financial series show evidence of asymmetry and heavy tails, quasi-maximum-likelihood (QML) estimation under normality conditions does not show serious econometric problems as they are unbiased and consistent, even if the true distribution is far from Gaussian, Bollerslev and Wooldridge (1992). However, the misspecification of the error's conditional distribution may lead to a significant efficiency loss. Many approaches have been proposed to deal with these features since Bollerslev and Andersen (1986). For instance, the GJR model from Glosten, Jagannathan and Runkle (1993) and Nelson's EGARCH, Nelson (1991), extended the scope to allow an asymmetric reaction pattern for the volatility. The latter parameterizes the model in terms of the log-variance which allows relaxing the non-negativity assumptions about the model's parameters while the exponential link assures that the variance is always positive. A drawback in Nelson's model resides in the fact that if the conditional distribution of the observation is student's t with finite degrees of freedom the conditions needed for the existence of the unconditional moments are rarely satisfied, Harvey (2013).

The issues around the EGARCH were addressed in Harvey and Chakravarty (2008) where they proposed the Beta-t-EGARCH model. It consists of an EGARCH model in which the variance is driven by an equation that depends on the conditional score, Harvey and Sucarrat (2014). Therefore the model can be viewed as an unrestricted version of the GAS model while belonging to the class of EGARCH models. Harvey (2013) argues that recognizing the conditional score as a linear function of a beta variable enables for the moments and the autocorrelations of absolute values of the observations raised to any positive power to be derived. Therefore, it is a very convenient model that has all the advantages of the EGARCH but without the disadvantages.

Another line of research deals with the uncertainty of the DGP by fitting a nonparametric density and therefore improving the efficiency of the estimates. These models are called semiparametric, and the procedure is usually two-fold. First, the vector of parameters θ is estimated parametrically assuming normality or some known density. With the sample residuals, the model is re-estimated using a non-parametric density to construct the maximum likelihood function, which is usually performed via a kernel estimator. The second part of the procedure provides a new efficient estimate of θ. Despite the rich history of the literature concerning GARCH models, a small body of studies cover semiparametric methods. Engle and Rivera (1991) was perhaps the first paper to address a semiparametric way to estimate a GARCH model. For the density estimation of the standardized residuals calculated by

QMLE, they used the discrete maximum likelihood estimation (DMPLE). Drost and Klassen and Sun and Stengos (2006) tackled this task by estimating the error density with a kernel method. The only paper to apply the semiparametric approach to the GAS framework was Blasques, Ji and Lucas (2016). In the GAS context, the semiparametric estimation has the further advantage that the score of the nonparametric density participates in the dynamics of the time-varying volatility.

Despite only recently introduced, the GAS models have become a very active field of research during the last years and have been successfully applied in a great variety of areas, such as risk modeling. For instance, Bernardi and Catania (2015) used the model confidence set (MCS) to compare the VaR forecasting performance of several GAS model specifications with its GARCH counterparties and Engle's CaViaR models. Catania (2017) performed a large scale empirical study to evaluate the forecasting performance of Markov-switching GARCH models compared with standard single regime specifications. Bartels and Ziegelmann (2016) forecasted market risk measures, such as Value at Risk (VaR) and Expected Shortfall (ES), for large dimensional portfolios via copula modeling. Their empirical results, for assets negotiated at Brazilian BOVESPA stock market from January, 2008 to December, 2014, suggest that, compared to the other copula models, the GAS dynamic factor copula approach has a superior performance in terms of AIC (Akaike Information Criterion) and a non-inferior performance with respect to VaR and ES forecasting. In line with the aforementioned works, the aim of this paper is to contribute with the empirical literature by comparing industry standard models with GAS models and its recent specifications. Our goal is to provide risk managers with more information about the performance recent models and ultimately assess if their implementation is justified by a sufficiently higher performance.

The first contribution of the paper concerns to the semiparametric GAS framework, developed by Blasques, Ji and Lucas (2016). We deepen the discussion about the bandwidth effect on parameter's estimation and propose a methodology for bandwidth selection that is similar to likelihood cross-validation. We evaluate this methodology through a Monte Carlo study. Additionally, we provide an extensive empirical study of the semiparametric model comparing its performance with three parametric specifications. Up to our knowledge, this is the first paper to apply a semiparametric GAS model to forecast risk measures. Our second contribution is to set a performance study, considering the parametric framework of GAS models, the Beta-t-EGARCH with one and two-components and the traditional models GJR-GARCH, APARCH, EGARCH, three GARCH parametric specifications and the semiparametric GAS model. We compare these models over fourteen financial time series, composed of eight stock indices, four exchange rate series and two Crypto-Currency rates.

The remainder of the paper is structured as follows. In Section 2 we present the GAS models in their parametric and semiparametric specification and the Beta-t-EGARCH one and two components, we present the benchmark models, the risk forecasting measures and the metrics used to compare the performance of the various models. In Section 3 we perform a simulation study of the semiparametric GAS model to test its in-sample properties. In Section 4 we carry out an extensive empirical study concerning ES and VaR forecasting ability. First we study the semiparametric model's performance. Second we compare the GAS parametric models with the alternative models presented in Section 2. In Section 5 we conclude with the final remarks.

2 METHODOLOGY

This section is divided in five parts. In the first we cover the basics of GAS models. In the second and third we present the semiparametric framework for GAS models and the Beta-t-EGARCH models. In the fourth we briefly show the benchmark models. In the fourth subsection we present the estimation procedure of forecasting of risk measures. In the final subsection we discuss the metrics used to compare the evaluated models.

2.1 GAS Models

The Generalized Autoregressive Score (GAS) model, was proposed by Creal, Koopman and Lucas (2013) and by Harvey (2013) under the name of DCS consists in a time-varying parameter model where drive mechanism used to update the parameters over time is the score function. The model is remarkably versatile, for instance, by scaling the score function appropriately, we can recover standard observation-driven models such as the GARCH, ACD, and ACI.

Cox (1981) classified models as parameter driven and observation driven. In the case of parameter driven models the associated likelihood function is not available in closed form, making the estimation process more demanding. Examples are: the stochastic volatility (SV) model and the stochastic intensity models of Bauwens and Hautsch (2006). On the contrary, in an observation driven model the likelihood function is available in closed form and in most of cases forecasting is straightforward. Alike GARCH models, the GAS model is classified as an observation driven model. The key difference between both models is that the GAS framework explores all the information coming from the entire distribution, as it is the case of state space models. On the other hand, GARCH models explore only the expected value of the conditional distribution. Another important feature of the GAS framework is that it can accommodate extensions to asymmetry, long memory, and other more complicated dynamics without further complexities.

Formally, the GAS model is specified as following. Defining y_t the dependent variable of interest vector, f_t the time varying parameter vector, $\hat{\theta}$ the unknown parameter vector and conditional density of observations $p(y_t|f_t)$. The available information set at time t consists of $\{Y^{t-1}, F^{t-1}\}$ where

$$F_t = \{Y^{t-1}, F^{t-1}\}, \quad for \quad t = 1, ..., n \, , \tag{4}$$

Assuming that y_t is generated by the observation density

$$y_t \sim p(y_t|f_t; \theta) \, , \tag{5}$$

The updating scheme for f_t is given by the equation

$$f_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{i=1}^{p} B_i f_{t-j+1} \ , \tag{6}$$

Where s_t is an appropriate function of the past data and ω is a vector of constants. When a new observation y_t is realized, the time-varying parameter f_t is updated to period $t+1$ through

$$\begin{aligned} s_t &= S_t \nabla_t, \qquad S_t = S(t, f_t, \mathcal{F}, \theta) \ , \\ \nabla_t &= \frac{\partial \, log \, p(y_t | f_t; \theta)}{\partial f_t} \ , \end{aligned} \tag{7}$$

Where $S(\cdot)$ is a matrix function. The choice of an appropriate scale for this function is of paramount importance since the score is the driving mechanism that updates the model. As one can notice the score depends on the entire density and not just on the first and second moments of observation series y_t. The intuition behind the use of the score stems from the fact that it defines a steepest accent direction for improving the model's local fit in terms of the likelihood or density at time t given the current position of the parameter f_t. Hence providing a natural direction for updating the parameter, Creal, Koopman and Lucas (2013).

For the GARCH parametrization, we have that $\nabla_t = \frac{1}{2f_t^2}(y_t^2 - f_t)$ and $S_t = 2f_t^2$. Therefore the expression for the score $S_t \cdot \nabla_t$ reduces to $y_t^2 - f_t$. Plugging it in the recursive equation, the GAS update scheme reduces to a $GARCH(1,1)$

$$f_{t+1} = \omega + Bf_t + A(y_t^2 - f_t) \ , \tag{8}$$

We can clearly see that the β parameter from the original $GARCH(1,1)$ model is here represented by $(B - A)$. In this paper we use three different GAS specifications: A Gaussian GAS - GAS(n), Student's t GAS - GAS(t) and a skew Student't GAS - GAS(st). Full detail on these specifications can be found in Creal, Koopman and Lucas (2011), Creal, Koopman and Lucas (2012) and in Ardia, Boudt and Catania (2016).

2.1.1 Semiparametric framework for GAS models

The first proposal for a semiparametric model in the GARCH context was presented by Engle and Rivera (1991). The density estimation was carried out with the discrete maximum penalized likelihood estimation technique of Tapia and Thompson, (1978). Their simulation studies suggested that the semiparametric approach can improve the efficiency of the parameter estimates up to 50% over QMLE. Drost and Klassen (1997) and Yang (2006), followed the two step procedure, using a kernel estimator instead. More recently Blasques, Ji and Lucas (2016) extended the semiparametric framework to GAS models. The new semiparametric model was a step forward into the semiparametric time-varying volatility estimation as it permitted the kernel density estimator to take part not only in the likelihood estimation but also in the volatility dynamics.

The GAS parametrization pursued in this paper follows Blasques, Ji and Lucas (2016) estimation procedure. In this paper however our focus is to shed some light on questions that are not fully explored. For instance, the bandwidth's choice which is a crucial part of kernel estimation. It seems to be a consensus that values of bandwidth that lies within the range of $0.3 \leq b_w \leq 0.8$ usually work well. Di and Gagopadhyay (2011), evaluated the impact of the bandwidth selection on the performance of the estimator in the context of GARCH models. We want to extend this investigation into the semiparametric GAS framework, since the kernel density is involved in the recursion equation as well, thus making the discussion around the bandwidth even more relevant. A major problem in adaptive estimation stems from the fact that the bandwidth has to be pre-chosen based on pre-fitted residuals through initial estimates of parameters.

When it comes to GAS models this issue is even more challenging as one needs to choose the bandwidth not only for the kernel's density but also for the kernel's density derivative. The Semiparametric GAS(1,1) specification is given by

$$\begin{aligned} y_t &= \mu + \xi_t = \mu + e^{\frac{f_t}{2}}\epsilon, \qquad \epsilon_t \sim q(\epsilon_t), \\ f_{t+1} &= \omega \cdot (1-\beta) + \alpha s_t + \beta f_t, \end{aligned} \tag{9}$$

The GAS model is estimated by replacing ω by $\omega/(1-\beta)$. The value of ω then corresponds to the annualized volatility. In eq. (2.6) f_t is parametrized in terms of the log-variance. Denoting h_t as the variance in a standard GARCH model, the relationship between h_t and f_t is $h_t = exp(f_t)$. This parametrization is convenient bacause it ensures a positive variance for each t without the need of imposing restrictions in the model's parameters. We can obtain $p(y_t|f_t)$ as following

$$
\begin{aligned}
F_Y(y) = P(Y_t \leq y) =& P(\mu + e^{\frac{f_t}{2}}\epsilon_t \leq y), \\
=& P(\epsilon_t \leq (y-\mu)e^{-\frac{f_t}{2}}), \\
=& F_\epsilon((y-\mu)e^{\frac{-f_t}{2}}), \\
f_Y(y) =& f_\epsilon((y-\mu)e^{\frac{-f_t}{2}})e^{-\frac{f_t}{2}},
\end{aligned} \tag{10}
$$

Supposing that the observations y_t have a conditional distribution with log-density

$$
ln\, p(y_t|f_t) = ln\, q\big((y_t-\mu)e^{-f_t/2}\big) - \frac{f_t}{2}, \tag{11}
$$

Deriving with respect to f_t

$$
\begin{aligned}
\frac{\partial\, log\, p(y_t|f_t;\theta)}{\partial f_t} &= \Big(-\frac{1}{2}\Big)\nabla q\big((y_t-\mu)e^{-\frac{f_t}{2}}\big)(y_t-\mu)e^{-\frac{f_t}{2}} - 1/2, \\
&= \Big(-\frac{1}{2}\Big)(\nabla q((y_t-\mu)e^{-\frac{f_t}{2}})(y_t-\mu)e^{-\frac{f_t}{2}} + 1),
\end{aligned} \tag{12}
$$

Where ∇_q is the log derivative of $q(\epsilon)$ and by definition eq. (2.9) equals s_t, when s_t is the identity. Parameter estimation in the context of GAS is carried out via maximum likelihood. The loglikehood function is given by

$$
L_T(\theta, q) = \frac{1}{T}\sum_{t=1}^{T} -1/2 f_t(\theta) + log\, q(e^{-\frac{f_t(\theta)}{2}}(y_t-\mu)), \tag{13}
$$

Note that forecasting the one-step-ahead f_t is straightforward. The procedure is carried out in the usual way by iterating the volatility filter forward using the observed sample $(y_1, y_2, ..., y_t)$. Hence, f_{t+1} is the last filtered value.

2.1.2 Estimation Algorithm

The estimation algorithm is two-fold. The first step consists in the estimation of the parametric model supposing some known distribution. In this step we obtain a consistent set of parameters θ, which is used as input into the model to generate the standardized residuals. The second step re-estimates the model's parameters using the standardized residuals obtained in the first step using a non-parametric density. The step-by-step procedure follows

1. Estimate the GAS parameters $\hat{\theta} = (\omega, \alpha, \beta, \mu)$ through QMLE,

2. Obtain the $f_t(\hat{\theta})$ vector,

3. Obtain the vector of standardized residuals,

$$\hat{\epsilon}_t = \frac{(y_t - \mu)}{e^{\frac{f_t(\hat{\theta})}{2}}}, \tag{14}$$

4. Density estimation using a *kernel* and its derivative, which enables to compute:

$$\nabla q = \tfrac{\partial \, log \, q}{\partial x} = \tfrac{q^{'}(\cdot)}{q(\cdot)},$$

   The calculation of the derivatives is futher explored in section (2.1.4).

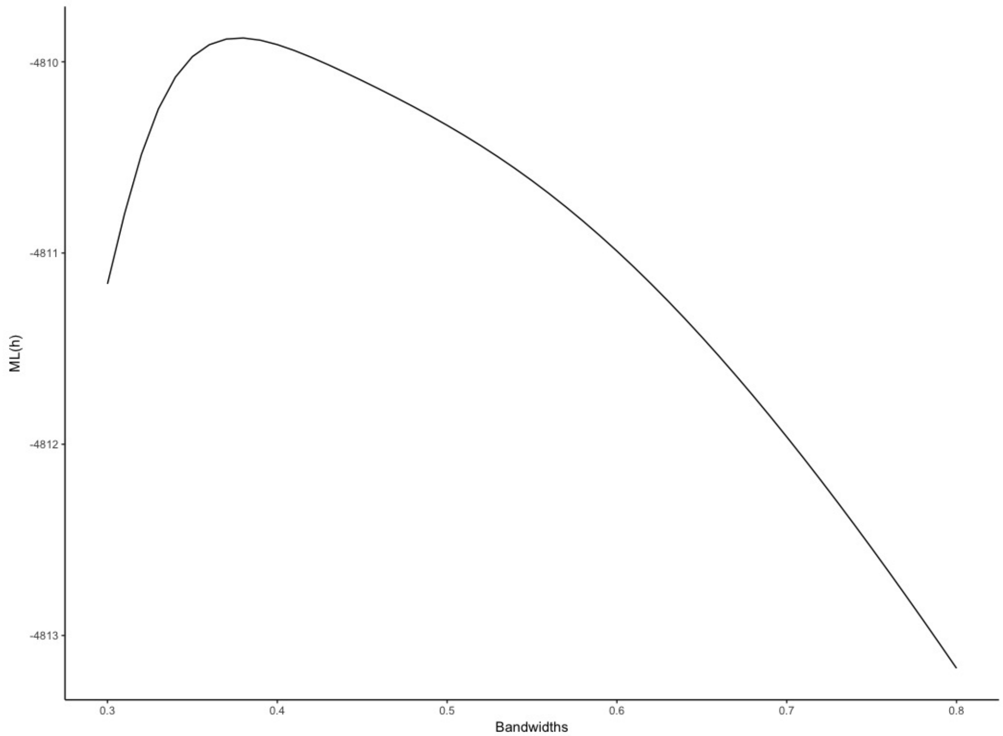
5. Re-estimate θ using the semiparametric equation's (2.6) e (2.10), obtaining a set of efficient parameters $\tilde{\theta} = (\tilde{\omega}, \tilde{\alpha}, \tilde{\beta}, \tilde{\mu})$

2.1.3 Bandwidth Selection

The bandwidth selection method proposed in this paper derives from the maximal likelihood cross-validation (MLCV) of Habbema (1974). In this paper we call it maximum likelihood bandwidth selector (*mlbs*). The procedure starts by creating a grid with all the bandwidth candidates. Once the model is parametrically estimated (step 1), we estimate the semiparametric step for each bandwidth candidate. The bandwidth candidate that generates set of parameters that yield the highest maximum likelihood value is selected. If we are interested in selecting different choices of bandwidth for the kernel density and its derivative, then the grid is bi-dimensional. The problem then scales up to n^2 iterations, so as we refine our grid (i. e., select more points in a pre-determined range) the procedure becomes increasingly more burdensome. Therefore, in this paper we opted to use the same bandwidth for both kernel and its derivative.

Some remarks about our procedure are worth mentioning. In practice, researchers choose b_w from an interval in which, robustness of the parameter estimates usually hold. The interval usually ranges from $0.3<b_w<0.8$. However in most of cases, no theoretical guidance is presented in terms of selecting the most adequate value. In the GAS framework this issue is even more relevant due to the reasons already mentioned. In this sense, we have presented an heuristic method, which is based on the idea of selecting the best fitted model without directly interfering with the optimization process, i. e., the estimation is carried out with the bandwidth held constant. We also observe that the model's generalization does not need to be vouched by, for example a leave-one-out procedure, as the parametric structure of the underlying model assures overfitting avoidance. The relative accuracy of SP-GAS *mbls* remains in the out-of-sample period, what reinforces that the accuracy gains are not caused by overfitting. The plot of the maximal likelihood function in terms of the bandwidth is shown bellow for a DGP with density student's-t with 7 degrees of freedom.

Figura 1: Plot of the maximal likelihood function - *mbls* procedure



2.1.4 Kernel Density Derivatives

In the GAS framework, calculation of the derivatives of the probability distribution function is vital, since it joins the recursion equation. Under the case of the current parametrization, we are interested in kernel derivatives. Consider f a kernel density estimator, we tackle the problem of estimating the following derivative

$$f^{(r)}(x) = \frac{d^r}{dx^r} f(x) \, , \tag{15}$$

Naturally, an estimator can be found by taking derivatives of the kernel density estimator, assuming the following form

$$\hat{f}^{(r)}(x) = \frac{d^r}{dx^r} \hat{f}(x) = \frac{1}{nb^{1+r}} \sum_{i=1}^{n} k^r(x) \Big(\frac{X_i - x}{b} \Big) \, , \tag{16}$$

The estimator $\hat{f}(x)$ makes sense if the kernel's derivative $k^{(r)}(x)$ exists and it is non-zero.

The estimator for q' in step 4 can be found by following the procedure described above. For instance, if we have a gaussian kernel, $k^1(x)$ assumes the form

$$k^{'}(\cdot) = -\, x \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \,, \tag{17}$$

Since the Gaussian Kernel is quite easy to handle and has derivatives of all orders, it is a common choice for derivative estimation. Considering that in step 4 we are interested in the derivative of the log-density ∇q, our estimator can be written as

$$\nabla q = \frac{\partial \, log \, \hat{q}(x)}{\partial x} = \frac{\frac{1}{nb_n^2} \sum_{i=1}^{t} \left[\left(\frac{x - \hat{\epsilon}_t}{b_n} \right) \frac{1}{\sqrt{2\pi}} e^{\frac{-\left(\frac{\hat{\epsilon}_t - x}{b_n} \right)^2}{2}} \right]}{\frac{1}{nb_n} \sum_{i=1}^{t} \left[\frac{1}{\sqrt{2\pi}} e^{\frac{-\left(\frac{\hat{\epsilon}_t - x}{b_n} \right)^2}{2}} \right]} \,, \tag{18}$$

2.2 Beta-t-EGARCH Models

The Beta-t-EGARCH Model was first presented by Harvey and Chakravarty (2008) and it can be viewed as a particular case of a GAS model. The main feature of this model is to accommodate important characteristics that are present in financial time series, such as time varying volatility, fat-tailedness, leverage, conditional skewness and a decomposition of volatility into a short-term and a long-term component. Harvey and Sucarrat (2014) extended the Beta-t-EGARCH model defining that when it has a skewed-t conditional distribution, the log density is represented by

$$\begin{aligned} ln(f_t) = {} & ln(2) - ln(\gamma + \gamma^{-1}) + ln\Gamma((\nu + 1)/2) - ln\Gamma(\nu/2) \\ & -\frac{1}{2} ln(\nu) - \lambda_{t|t-1} - \frac{\nu + 1}{2} ln \left(1 + \frac{(y_t - \mu)^2}{\gamma^{2sgn(y_t - \mu)} \nu e^{2\lambda_{t|t-1}}} \right), \end{aligned} \tag{19}$$

The first order one-component Beta-Skew-t-EGARCH is given by

$$\begin{aligned} & y_t = exp(\lambda_t)\epsilon_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim st(0, \sigma_\epsilon^2, \nu, \gamma), \quad \nu > 2, \quad \gamma \in (0, \infty), \quad t = 1, 2, ..., n \,, \\ & \lambda_t = \omega + \lambda_t^{\dagger}, \\ & \lambda_t^{\dagger} = \phi_1 \lambda_{t-1}^{\dagger} + k_1 u_{t-1} + k^* sgn(-y_{t-1})(u_{t-1} + 1), \qquad |\phi_1| < 1, \end{aligned} \tag{20}$$

Where u_t is the conditional distribution score, σ_t is the scale, ϵ_t is the error and is defined by $\epsilon_t = \epsilon_t^* - \mu_\epsilon$. ϵ_t^* follows a skewed-t distribution with ν degrees of freedom, asymmetry γ and mean μ_{ϵ^*}. The forecasting equation for the conditional volatility σ_t h steps ahead is

$$E_t(\sigma_{t+h}) = exp(\omega + \phi_1^h \lambda_t^\dagger) \prod_{i=1}^{h} E_t[exp(\phi_1^{h-i} g_{t+i-1})], \tag{21}$$

Where g_t is i.i.d and equal to $k_1 u_t + k^* sgn(-\epsilon)(u_t + 1)$. When $i = 1$ the value of $E_t[exp(\phi_1^{h-i})g_{t+i-1}]$ is equal to $(\phi_1^{h-1} g_t)$, Muller (2016).

Harvey and Sucarrat (2014), extended the one component model by decomposing the volatility into a long term component and a short-term component. The reason for a two component model comes from the fact that after a large movement it is likely to be after-shocks, the DCS model by its nature does not react quickly enough to the short term movements, therefore justifying the need for the short term component. The two component Beta-Skew-t-EGARCH model is given by

$$\begin{aligned}
y_t &= exp(\lambda_t)\epsilon_t = \sigma\epsilon_t, \quad \epsilon_t \sim st(0, \sigma_\epsilon^2, \nu, \gamma),\ \nu > 2, \quad \nu, \gamma \in (0, \infty) \quad t = 1, 2, ..., n, \\
\lambda_t &= \omega + \lambda_{1,t}^\dagger + \lambda_{2,t}^\dagger, \\
\lambda_{1,t}^\dagger &= \phi_1 \lambda_{1,t-1}^\dagger + k_1 u_{t-1}, \qquad |\phi_t| < 1, \\
\lambda_{2,t}^\dagger &= \phi_2 \lambda_{2,t-1}^\dagger + k_2 u_{t-1} + k^* sgn(-y_{t-1})(u_{t-1} + 1), \qquad |\phi_2| < 1, \quad \phi_1 \neq \phi_2,
\end{aligned} \quad , \tag{22}$$

Where $\lambda_{1,t}^\dagger$ and $\lambda_{2,t}^\dagger$ can be interpreted as the long and short run components. To assure that the model is stationary and stable it is required that $|\phi_1|$ and $|\phi_2|$ to be both less than one. The forecasting equation for the two component model is

$$E_t(\sigma_{t+h}) = exp(\omega + \phi_1^h \lambda_{1,t}^\dagger + \phi_2^h \lambda_{2,t}^\dagger) \prod_{(i=1)}^{h} E_t[exp(\phi_1^{h-i} g_{1,t+i-1} + \phi_2^{h-i} g_{2,t+i-1})]\ , \tag{23}$$

Where $g_{1,t} = k_1 u_t$ and $g_{2,t} = k_2 u_t + k^* sgn(-\epsilon)(u_t + 1)$,

2.3 Other Conditional Variance Models

In this section we briefly discuss the other conditional variance models used throughout the empirical investigation section. A more complete review of these models can be found on Bauwens, Hafner and Laurent (2012).

2.3.1 GARCH Models

The seminal paper of Engle (1982) addressed the empirical findings about financial time series by introducing the ARCH(q) model. Thereafter, numerous refinements have been introduced, from which, the most known is certainly the GARCH model conceived by Bollerslev and Andersen (1986).

1. GARCH

   The process $(Y_t)_{t \in \mathbb{Z}}$ is a GARCH(p,q) process if it is strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $(\sigma_t)_{t \in \mathbb{Z}}$, the equations

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_{t-i} Y_t^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \end{aligned} \tag{24}$$

   where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, ..., p$, and $\beta_j \geq 0, j = 1, ..., q$. Heavy tails may be accommodated by using the conditional distribution t-student. The one step ahead variance forecast $\hat{\sigma}_{t+1|t}$ for the GARCH(p, q) model equals

$$\hat{\sigma}_{t+1|t} = \alpha_0^t + \sigma_{i=1}^q \alpha_i^{(t)} \epsilon_{t-i+1}^2 + \sum_{j}^{(t)} \sigma_{t-j+1}^2, \tag{25}$$

2. EGARCH

   A critique about GARCH models stems from the fact that it does not take into consideration leverage effects. The EGARCH(p,q) model from Nelson (1991), proposes to overcome this problem by specifying a exponential link function. It assumes that the conditional volatility dynamics follows

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t, \\ ln(\sigma_t^2) &= \omega + \sum_{j=1}^{q} ln(\sigma_{t-j}^2) + \sum_{i=1}^{q} k_i g(\epsilon_t - i), \end{aligned} \tag{26}$$

   And the exponential link function

$$g(\epsilon_t) = \theta \epsilon_t + \gamma[|\epsilon_t| - E(|\epsilon_t|)], \tag{27}$$

Where θ and γ are real parameters; ϵ_t and $|\epsilon_t| - E(|\epsilon_t|)$ are a sequence of random variables i.i.d with zero mean and continuous distribution.

3. GJR-GARCH

   The GJR-GARCH, proposed by Glosten, Jagannathan and Runkle (1993), introduces asymmetry into the volatility process by adding an indicator variable, Cuthbertson and Nitzsche (2005),

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 h_{t-1} + \gamma(\epsilon_{t-1}^2 D_{t-1}), \tag{28}$$

   where,

$$D_{t-1} = \begin{cases} 1 & \text{if } \epsilon_{t-1} \leq 0 \\ 0 & \text{otherwise,} \end{cases}$$

4. APARCH

   The variance equation of the APARCH(p, q) model is defined as

$$\begin{aligned} \epsilon_t &= z_t \sigma_t, \\ z_t &\sim \mathcal{N}(0, 1), \\ \sigma_t^\rho &= \omega + \alpha k(\epsilon_{t-1})^\delta + \beta \sigma_{t-1}^\delta, \\ k(\epsilon_{t-1}) &= |\epsilon_{t-1}| - \gamma \epsilon_{t-1}, \end{aligned} \tag{29}$$

   Where γ captures the leverage or asymmetry effect in return volatility and δ is the Taylor effect, regarding the difference in the sample autocorrelations of absolute and squared returns.

2.4 Prediction methods of VaR and ES

To account for the stylized empirical financial facts, the estimation of risk measures through conditional variance models has become a very popular choice. To proceed the calculation of the conditional risk measures, we make the following assumption, Mcneil, Frey and Embrechts (2015),

Assumption 1 *The process of losses $(L_t)_{t \in \mathbb{Z}}$ is adapted to the filtration $(\mathcal{F}_t)_{t \in \mathbb{Z}}$ and follows a stationary model of the form $L_t = \mu_t + \sigma_t Z_t$, where μ_t and σ_t are $\mathcal{F}_{t-1}$ measurable and the (Z_t) are SWN(0,1) innovations.*

Under the assumption, if we write G for the cdf of (Z_t), we can calculate that

$$F_{L_{t+1}|\mathcal{F}_t}(l) = P(\mu_{t+1} + \sigma_{t+1}Z_{t+1} \leq l \mid \mathcal{F}_t) = G((l - \mu_{t+1})/\sigma_{t+1}), \tag{30}$$

From the conditional mean equation of the aforementioned conditional volatility models, we obtain the one-step-ahead conditional mean forecast $\hat{\mu}_{t+1}$ and from the conditional variance equation we obtain the one-step-ahead conditional variance forecast $\hat{\sigma}^2_{t+1}$. Thus, the estimated VaR at time $t+1$ can be obtained through the following equation

$$VaR^{t+1}_{\alpha} = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} q_{\alpha}(Z), \tag{31}$$

Where Z a generic rv with df G and q_{α} is the left tail quantile. In the parametric framework q_{α} estimates are performed by inverting the underlying conditional cumulative density distribution function. For the semiparametric GAS q_{α} is the empirical quantile of the residual distribution. Therefore the VaR forecast for the SP-GAS case is fully semiparametric in the sense that both volatility and quantile are non-parametrically estimated. Our intention is to recover efficiency loss in both ends. By definition the one step-ahead Expected Shortfall (ES) is computed as

$$ES^{t+1}_{\alpha} = E(X|X{<}VaR^{t+1}_{\alpha}) = \frac{\int_{-\infty}^{VaR^{t+1}_{\alpha}} xf(x)\,dx}{p}, \tag{32}$$

2.5 Evaluating the different approaches

We select two backtesting methods to evaluate the precision of the VaR predictions. These methods are: The Loss Function of Rivera, Mishra and Lee (2004) and the Model Confidence Set procedure from Hansen (2005). A loss function QL based backtest for a given α is given by

$$QL = \frac{1}{P}\sum_{t=R}^{T}(\alpha - d^{\alpha}_{t+1})(y_{t+1} - VaR^{\alpha}_{t+1}), \tag{33}$$

Where d^{α}_{t+1} is an indicator function equal to $1(y_{t+1}{<}VaR^{\alpha}_{t+1})$. As one can observe the Q function is asymmetric, penalizing more intensively returns showing VaR exceedance with weight $(1-\alpha)$. A lower value of QL represents a better goodness of fit. The models are compared in terms of the averaged QL over the out-sample period. We normalize the performances using the GAS(n), which is the benchmark model. Therefore, if the model presents a result less than 1, the model is outperforming the GAS(n) model. The MCS procedure consists of a sequence of statistic tests which permits to construct a set of superior models, the Superior Set Models (SSM), where the null hypothesis of equal predictive ability (EPA) is not rejected at a certain confidence level α. The EPA statistic tests is calculated for an arbitrary loss function, in our case we use lost function of Gonzalez-Rivera above described.

The evaluation of the accuracy of the ES forecasts is assessed using the Shortfall Test of Mcneil and Frey (2000). The test consists on calculating the difference between the ES forecasts and the realized losses for observations where the loss is greater than the VaR forecast. If the ES forecasts are appropriate, the standardized residuals should be i.i.d and have zero mean. Bootstrap is used to mitigate any bias with respect to the ES underlying distribution. We set the bootstrap samples as 1.000. The main steps of the empirical study can be summarized as

1. Implement a rolling forecast for the out-of-sample horizon;
2. Compute VaR and ES measures with eqs. (2.29) and (2.30);
3. Perform the statistic backtest and obtain the values of Q for all models;
4. Comparison between the models using Q as the main metric and perform MCS procedure;

2.6 Codes and packages

The empirical application was conducted using R. The main packages used were: *fGarch* and *rugarch* to estimate GARCH models. For the estimation of Beta-Skew-t-EGARCH models we used *betaegarch* package. The codes for the GAS and Semiparametric GAS estimation are our own and are available on github www.github.com/luizbezerraneto. Some functions from the GAS package of Ardia, Boudt and Catania (2016) were used. All the VaR backtesting was carried out using the MCS package from Bernardi and Catania (2015). The package *rugarch* was used to perform ES backtesting.

3 SIMULATION STUDIES

In this section, we study the performance of the semiparametric estimator proposed by Blasques, Ji and Lucas (2016), using simulation studies. First, we compare the performance of the SP-GAS with parametric GAS and GARCH models. Secondly, we evaluate the role of bandwidth selection on the performance of the estimator and propose a new method to select the bandwidth based on the log likelihood of the estimated model. The data generating process (DGP) is a GAS(1, 1) model with parameter values $\theta(\mu, \omega, \alpha, \beta) = (0, 2, 0.3, 0.9)$. Three innovation distributions being considered are: Gaussian, t-student and skewed t-student. The purpose of such choices is to get a grasp of the efficiency issue by adding more complexity to the DGP such as heavy tail and asymmetry. We study the efficiency of the estimators by calculating the parameters RMSE. Some options of density specification for the first step are available. For example, we could consider using a density such as a Students-t . However we find that great part of the model's novelty comes from the fact that no assumption about the density has to be made beforehand (i.e: simply considering the normal), therefore all studies conducted in this paper consider the Gaussian QMLE for the first step of the semiparametric estimation.

3.1 Estimation accuracy

Table 1 compares the performance of estimated parameters in terms of RMSE. It is possible to notice that the performance of the candidate models vary according to the choice of conditional density for the DGP. First we evaluate the in-sample estimates under Gaussianity. In this case all models performed very closely and no improvements were noticed with the semiparametric (SP) approach. The results for ω calls attention, as the GARCH models performed very poorly and the SP models performed significantly worse than the parametric models. With respect to the t-student as conditional density for DGP, we have tested three values for the degrees of freedom $\nu = 5$, $\nu = 7$, $\nu = 10$. In all cases the GAS(t) model is the best performing model. Here, however, we notice that the SP performs better than GAS(n) for all degrees of freedom and in most cases the loss of efficiency recovered is significant. For the skewed t-student density, the semiparametric models were the best performing models for all parameters, except for ω, for which GAS-t performed better. Boxplots provided in figure 1 exhibit the sampling distributions of the estimated parameters. For the sake of brevity only the results associated with t(5) distribution are shown.

The results are expected and in line with the literature, however there are some points worth mentioning. First, we can observe that the estimation of ω clearly improves from the GAS-N model to the SP-GAS models. This fact contradicts the literature, for instance Drost and Klaassen (1997) restrict their analysis only to α and β, arguing that these are the only adaptive estimable parameters. Overall, we can conclude that as the model's underlying density departure from the true DGP, the semiparametric estimation is more attractive. The semiparametric approach outperforms GAS(n) model in all density specifications, recovering up to 60% of the efficiency loss. Finally, we compare our methodology for bandwidth selection with the literature's optimal bandwidth 0.5. We notice that both perform very closely, however, it seems that fixed choice of 0.5 dominates the *mbls* approach in terms of RMSE.

Figura 2: Sampling distribution of the estimates of β. Data generated from a $t(5)$.

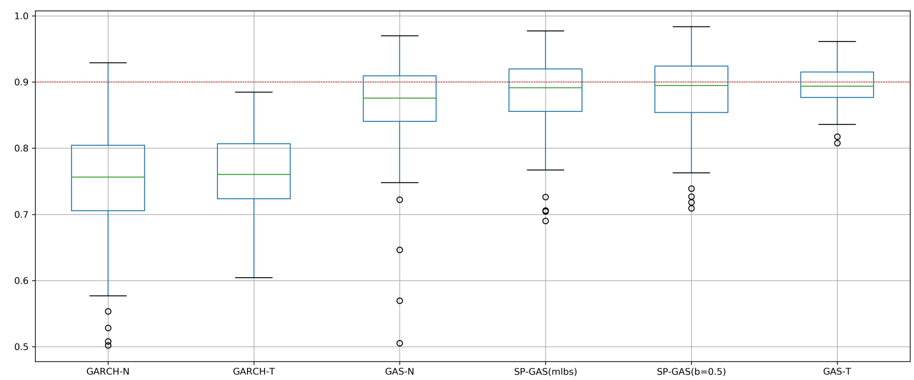


Figura 3: Sampling distribution of the estimates of α. Data generated from a $t(5)$.

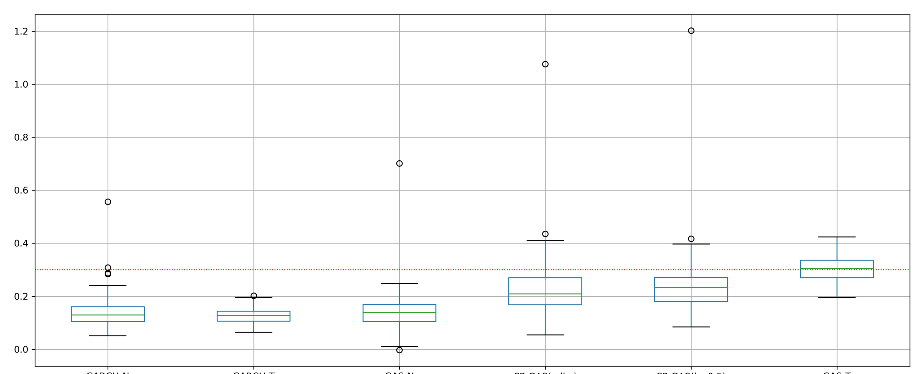


Figura 4: Sampling distribution of the estimates of ω. Data generated from a $t(5)$.

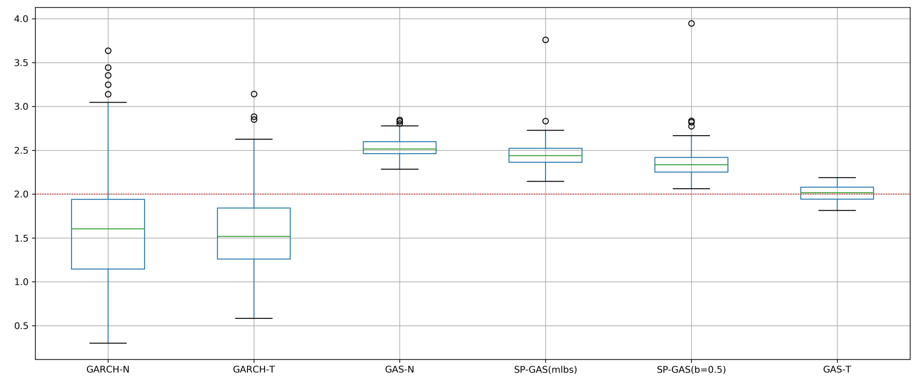


Figura 5: Sampling distribution of the estimates of μ. Data generated from a $t(5)$.

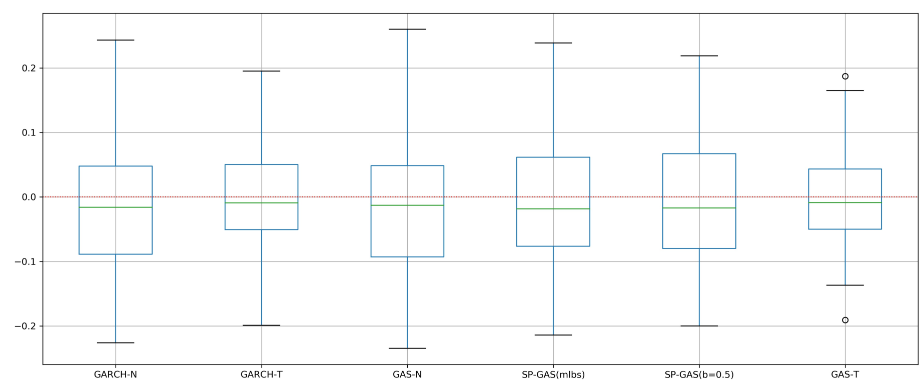


Tabela 1: Simulation results in terms of relative Root Mean Square Error (RMSE)

Gaussian	GARCH(n)	GARCH(t)	GAS(n)	GAS(t)	SP-GAS($b = 0.5$)	SP-GAS($mlbs$)
β	0.117862	0.110384	0.036232	0.037830	0.038650	0.040177
α	0.174711	0.173623	0.019516	0.019874	0.020436	0.021884
ω	1.260358	1.227295	0.078829	0.079277	0.238948	0.190514
μ	0.055228	0.057516	0.055380	0.055141	0.055712	0.057046
$t(5)$	GARCH(n)	GARCH(t)	GAS(n)	GAS(t)	SP-GAS($b = 0.5$)	SP-GAS($mlbs$)
β	0.171572	0.175711	0.171257	0.050719	0.130864	0.132402
α	0.172688	0.148343	0.082766	0.029947	0.058439	0.061224
ω	0.796043	0.657438	0.545444	0.088641	0.421639	0.490112
μ	0.085698	0.072869	0.092481	0.072372	0.086068	0.089016
$t(7)$	GARCH(n)	GARCH(t)	GAS(n)	GAS(t)	SP-GAS($b = 0.5$)	SP-GAS($mlbs$)
β	0.161150	0.162534	0.119599	0.044586	0.067788	0.071921
α	0.146821	0.147770	0.039467	0.025703	0.031350	0.035957
ω	0.897542	0.882844	0.352599	0.082150	0.170595	0.282764
μ	0.075257	0.069719	0.078229	0.068727	0.073701	0.084665
$t(10)$	GARCH(n)	GARCH(t)	GAS(n)	GAS(t)	SP-GAS($b = 0.5$)	SP-GAS($mlbs$)
β	0.157480	0.156529	0.098077	0.043979	0.057484	0.061331
α	0.147189	0.148009	0.032976	0.029863	0.030187	0.035236
ω	1.076311	1.058162	0.227184	0.086115	0.090933	0.143428
μ	0.059394	0.057138	0.060539	0.057398	0.059529	0.061542
$skt(\nu = 3, \delta = 0.5)$	GARCH(n)	GARCH(t)	GAS(n)	GAS(t)	SP-GAS($b = 0.5$)	SP-GAS($mlbs$)
β	0.104124	0.090295	0.121728	0.173491	0.110012	0.116341
α	0.282428	0.226753	0.100593	0.023782	0.076146	0.063513
ω	1.263529	1.118616	1.079046	0.405470	0.787918	0.891711
μ	0.203095	0.248333	0.212141	0.371742	0.157967	0.164462

3.2 Impact of bandwidth b_w on parameter estimates

In this subsection we analyse the stability of the estimates with regard to several bandwidth values. In practice such procedure is useful to analyse the estimates robustness. Since the GAS framework have the error density participating in the volatility dynamics as well, testing for parameter stability becomes even more relevant. In Figure 03 we can observe the variation of sample estimators with respect to different values of bandwidth. The plots indicate that a wider bandwidth is associated smaller variability, however at the cost of larger mean deviation from the true parameter. This pattern is clear in both parameters α and β. This fact clearly shows impact of the bandwidth b_n with regard to the bias-variance trade-off. Table 2 presents the variance, the deviation from the mean and Root Mean Square Error (RMSE) of the parameter estimates for certain values of bandwidth. The underlying process is a student's t with 7 degrees of freedom, based on 100 Monte Carlo replications. We can observe that the differences of RMSE between values of bandwidth varying in the range of $0.3 \leq b \leq 0.8$ are minimal, corroborating with the literature, e. g. Sun and Stengos (2006) and Blasques, Ji and Lucas (2016).

Figura 6: Distribution of the estimates of β for bandwidth values over the range $0.3 \leq b \leq 0.8$.

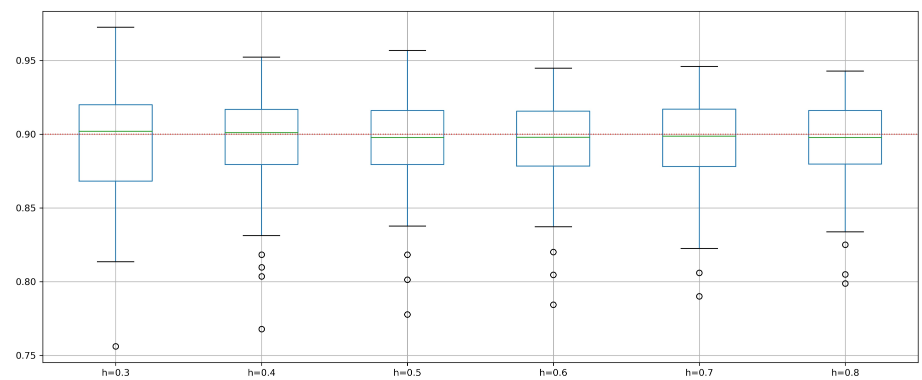


Figura 7: Distribution of the estimates of α for bandwidth values over the range $0.3 \leq b \leq 0.8$.

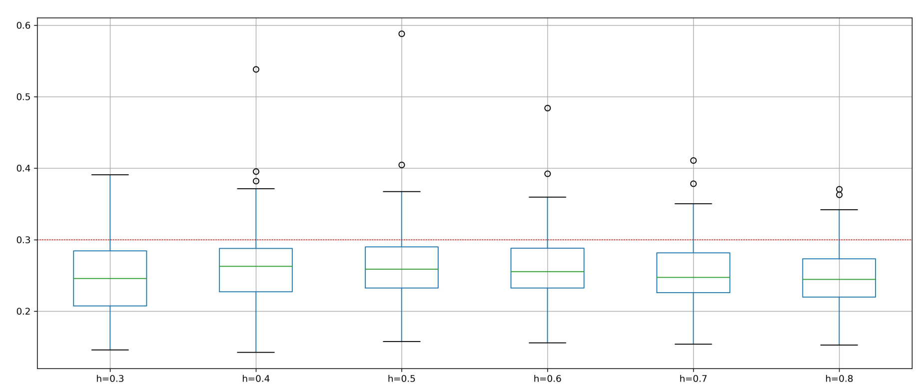


Tabela 2: Comparison of the estimates variance, mean deviation and estimation accuracy for different values of bandwidth. Data generated from a $t(7)$

	h=0.3	h=0.4	h=0.5	h=0.6	h=0.7	h=0.8
Variance	0.001410	0.001125	0.000969	0.000904	0.000899	0.000905
β - $\tilde{\beta}$	0.005373	0.005356	0.004841	0.005285	0.005787	0.006018
RMSE	0.074231	0.066775	0.067788	0.065147	0.066620	0.069602
	h=0.3	h=0.4	h=0.5	h=0.6	h=0.7	h=0.8
Variance	0.002985	0.003228	0.003305	0.002635	0.002203	0.001930
α - $\tilde{\alpha}$	0.050553	0.035534	0.036375	0.040447	0.047514	0.054164
RMSE	0.037751	0.033804	0.031350	0.030385	0.030384	0.030537

4 EMPIRICAL APPLICATION

The empirical application is performed considering three distinct sets of data, adding up to 17 time series. The first set consists on the log returns of four foreign exchange rates: USD against BRL, EUR, JPY and GBP against EUR. The data was retrieved from the website https://br.investing.com/. The second set consists on the log returns of eight stock indices: CAC 40 (France), DAX 30 (Germany), Dow Jones Industrial Average (US,DJI), Euro Stoxx 50 (Europe), FTSE 100 (UK), Bovespa (Brazil, IBOV), Nasdaq (US), Nasdaq (US) and S&P500 (US). All indices were retrived from https://finance.yahoo.com/. We also run a study with two Crypto-Currency Bitcoin (BTC) and Litecoin (LTC). The criteria used to choose the series was first based on the market capitalization and the number of data observations available. Both are among the top 5 crycurrencies with largest capitalization and also with the most data available. Yet the series are still too short. Therefore we analyse these series only over semiparametric experiment and in terms of QL. The data was collected in https://coinmarketcap.com.

Tabela 3: Descriptive statistics of return series

Exchange Rates	N. Obs	Min	Max	Mean	Stdev	Skewness	Kurtosis	E.Quantile	J-B
USD/BRL	4732.00	-9.97	8.65	0.02	1.07	0.30	7.73	-1.52	11877.81
USD/EUR	2142.00	-2.62	3.07	-0.01	0.59	0.01	1.62	-0.93	235.26
GBP/EUR	4701.00	-8.06	3.52	-0.00	0.60	-0.65	9.87	-0.93	19558.65
USD/JPY	4733.00	-3.70	5.35	0.00	0.65	-0.04	3.67	-1.00	2667.72
Indices	N. Obs	Min	Max	Mean	Stdev	Skewness	Kurtosis	E.Quantile	J-B
CAC40	4125.00	-9.47	10.59	0.00	1.44	-0.00	5.61	-2.30	5408.93
DAX	4097.00	-7.43	10.80	0.02	1.48	-0.01	4.92	-2.38	4145.12
DJI	4060.00	-8.20	10.51	0.02	1.11	-0.04	9.58	-1.68	15554.52
STOXX50	4054.00	-9.01	10.44	-0.00	1.48	-0.04	5.06	-2.35	4338.07
FTSE	4075.00	-9.26	9.38	0.01	1.18	-0.14	7.17	-1.83	8753.50
IBOV	3994.00	-12.10	13.68	0.05	1.75	-0.10	4.47	-2.80	3331.50
Nasdaq	4060.00	-11.11	11.85	0.04	1.44	-0.00	5.90	-2.38	5903.11
S&P 500	4060.00	-9.47	10.96	0.02	1.19	-0.26	10.05	-1.83	17166.17
Crypto-Currency	N. Obs	Min	Max	Mean	Stdev	Skewness	Kurtosis	E.Quantile	J-B
BTC	1756.00	-26.62	35.75	0.25	4.50	-0.19	8.11	-6.68	4837.19
LTC	1756.00	-51.39	82.90	0.23	6.98	1.81	25.13	-8.43	47288.09

The indices data set was collected over a 16-year period, from January 01, 2002 to February 18, 2018 and the currency data over a 18-year period, from January 01, 2000 to February 18, 2018, except only for the USD/EUR that covers a period from January 01, 2010 to February 18, 2018. The Crypto-Currency data was collected over 5 year period, from April 28, 2013 to February 18, 2018. All returns are calculated as the logarithmic difference of the daily index level multiplied by 100.

$$y_t = (log(p_t) - log(p_{t-1})) \times 100, \tag{34}$$

Where p_t is the closing price of each class of asset. The descriptive statistics are presented in table 2. Clearly all series display high degree of leptokurtosis. The presence of skewness is found for most of series. Furthermore, Jarque and Bera statistic strongly rejects the null hypothesis of normality for all series.

4.1 Empirical Performance - Semiparametric GAS

In order to evaluate the usefulness of the proposed semiparametric GAS model, we compare its performance with its parametric counterparts. For this empirical study we use the last 2.500 returns from the series described above. The data sets are divided into two samples: an in sample period from approximately Jul, 2008, to May, 2012 and an out-sample period from approximately May, 2012, to Feb 18, 2018. The exact date is not specified as it may vary due to local holidays. A rolling window approach is used to produce one-day ahead VaR and ES forecasts with $\alpha = 1\%$, $\alpha = 2.5\%$ and $\alpha = 5\%$ thresholds.

Table 4 reports the average QL ratios for the $SPGAS$, $GAS(st)$, $GAS(t)$ with respect to $GAS(n)$ for the indices data set. We report the results for the three VaR confidence levels. The best performing model was $GAS(t)$ model, closely followed by GAS(st). The $SPGAS$ outperformed $GAS(n)$ for the confidence levels of $\alpha = 1\%$ and $\alpha = 5\%$ for indices series. For the currency series the Semiparametric GAS had better average performance for all thresholds.

Tabela 4: QL Ratios for Semiparametic GAS and Parametic GAS Specifications

Model (Indices)	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
$GAS(n)$	1.0000	1.0000	1.0000
$GAS(t)$	0.9517	0.9655	0.9648
$GAS(st)$	0.9563	0.9685	0.9665
$SPGAS$	0.9896	1.0016	0.9946
Model (Currencies)	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
$GAS(n)$	1.0000	1.0000	1.0000
$GAS(t)$	0.9683	0.9610	0.9612
$GAS(st)$	0.9701	0.9646	0.9625
$SPGAS$	0.9717	0.9761	0.9705

Table 5 reports the MCS ranking in relation to the index series. For the sake of brevity the MCS procedure was applied only to VaR forecasts at 5%. We can observe that SPGAS outperforms the GAS(st) for most of the cases. This is surprising since the average QL results for $GAS(st)$ were in average 3.15% better. However a closer look at the results reveals that GAS(t) and GAS(st) models had a striking performance in DOW30, outperforming SPGAS in up to 20%. These results pushed their average QL down. Not surprisingly, the MCS model eliminated GAS(n) and SPGAS models in that series. When the ranking position is not displayed ($-$) indicates that the model was eliminated by the MCS process. Overall, we had a low number of eliminated models, indicating a high homogeneity between the models. Therefore, for greatest part of the series the models are statistically equivalent in terms of their ability to forecast future VaR levels.

Table 6, shows the MCS procedure applied to the currencies series. Once more we can observe a low number of eliminated models and a prevailing out-performance of GAS(t) model. The Semiparametric model improved GAS(n) model for most of cases, with the exception of USD/JPY series.

Tabela 5: MCS procedure for Semiparametic GAS and Parametic GAS Specifications - Indices

Asset	CAC40	DAX	DJI	STOXX50	FTSE	IBOV	Nasdaq	SP&500
GAS(n)	3	4	-	4	2	3	3	4
GAS(t)	1	1	1	2	1	4	1	1
GAS(st)	2	3	2	3	3	1	4	3
SPGAS	4	2	-	1	4	2	2	2

Tabela 6: MCS procedure for Semiparametic GAS and Parametic GAS Specifications - Foreign Exchange Rates

Model	GBP/EUR	USD/BRL	USD/EUR	USD/JPY
GAS(n)	4	4	-	3
GAS(t)	1	1	2	1
GAS(st)	2	2	3	2
SPGAS	3	3	1	4

Table 7 shows the p-values obtained performing the ES backtesting. The p-values that are greater than 0.05 indicates that models are appropriate for the forecasting of ES at the confidence levels 95%, 97.5% and 99%. As we can see, $GAS(n)$ model performs very poorly in all series with the only exception of FTSE, which had p-value greater than 0.05. On contrary, the remaining models perform rather well in all series, the only exception was IBOV, which only $GAS(t)$ model had p-value greater than 0.05. The results show that the Semiparametric GAS model consistently outperform the GAS model with gaussian innovations. For the currency series all models had p-values greater than 5%.

The QL values for the Crypto-Currency series are displayed in table 8. Here we notice

Tabela 7: The backtesting results of ES forecasting

Asset	$GAS(n)$	$GAS(t)$	$GAS(st)$	$SPGAS$
CAC40	0.01	0.57	0.19	0.49
DAX	0.01	0.95	0.66	0.51
DJI30	0.04	0.99	0.96	0.80
STOXX50	0.00	0.32	0.10	0.61
FTSE	0.10	0.95	0.80	0.66
IBOV	0.01	0.09	0.02	0.00
NASDAQ	0.00	0.77	0.00	0.62
S&P500	0.03	0.90	0.87	0.85
Asset	$GAS(n)$	$GAS(t)$	$GAS(st)$	$SPGAS$
GBP/EUR	1.00	1.00	1.00	1.00
USD/BRL	0.13	0.97	0.79	0.99
USD/EUR	1.00	1.00	1.00	1.00
USD/JPY	0.85	1.00	1.00	1.00

that the GAS(st) had the best performance among the competitors. Hence, the inclusion of skewness seems to improve the performance of the model. To some extent this was expected, since the Crypto-Currency series are more leptokurtic and asymmetric than usual financial series. As a matter of fact we could report to table 4 to verify that the inclusion of skewness did not enhance the GAS models performance under the currency and index time series. We also note that the Semiparametric GAS improves performance of the GAS(n) model for all confidence levels, with the only exception of $\alpha = 5\%$ for the Bitcoin series.

Tabela 8: QL Ratios for Semiparametic GAS and Parametic GAS Specifications - Crypto-Currencies

Model (Bitcoin)	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
GAS(n)	1.0000	1.0000	1.0000
GAS(t)	0.9622	0.9420	0.9547
GAS(st)	0.8819	0.9448	0.9837
SPGAS	0.9305	0.9636	1.0217
Model (Litecoin)	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
GAS(n)	1.0000	1.0000	1.0000
GAS(t)	1.1859	1.0805	1.0063
GAS(st)	0.9680	0.9550	0.9710
SPGAS	0.9446	0.9818	1.0143

4.2 Empirical Performance - Parametric Models

In this subsection we evaluate the performance of the parametric models. Our main goal is to verify if GAS models can generate more accurate VaR forecasts. Moreover, we investigate whether the addition of further complexities in the model can lead to more accurate out-of-sample VaR and ES prediction. In this empirical study, we use the full returns series. The data sets are divided into two samples. For the indices data set we have an in-sample period from approximately (Jan, 2002) to (May, 2012) and an out-sample period from approximately (May, 2012) to (Feb, 2018). For the currencies data set we have an in sample period from (Jan, 2002) to (May, 2012) and an out-sample period from approximately (May, 2012) to (Feb , 2018). A rolling window approach is used to produce one-day ahead forecasts of $\alpha = 1\%$, $\alpha = 2.5\%$ and $\alpha = 5\%$ thresholds.

Table 9 displays results related to the QL ratios for the 11 models. We can note that the Beta-Skew-t-EGARCH two-components model consistently outperformed all other models for the indices data. The average improvement was 7.6% comparing to the GAS(n). These results were closely followed by the Beta-Skew-t-EGARCH one-component model. The results suggest that even though the two-component model has improved the VaR forecasts, accounting for the second component did not bring expressive enhancements to VaR forecast over the series considered. The improvements were, on average, only 0.30%. Among the GARCH models, the GJR-GARCH specification had the best performance, followed closely by the APARCH model. For the currencies data set the results are similar. The best performing models are the Beta-Skew-t-EGARCH two-components model, the Beta-Skew-t-EGARCH one-component model and the GAS(st) model for the confidence levels of 1%, 2.5% and 5%, respectively. Beta-Skew-t-EGARCH two-components improved over GAS(n) model in about 7.3%, on average. Among the GARCH models the APARCH specification showed the best results followed closely by the GJR-GARCH.

Table 10 shows the MCS ranking for the 11 parametric models at $\alpha = 5\%$. First we observe a clear superiority of the Beta-Skew-t-EGARCH with two-components model over all competitors. This finding is connected to its high performance in terms of QL, since it is the main metric under MCS procedure. We can observe that only in three out of eleven series we had some kind of elimination. The models eliminated were the GAS(n) and GARCH(n). This is expected, since these are the the simplest models in our study. However, the paucity eliminations calls attention. We report to Bernardi and Catania (2015), where similar results were found. It general terms, this fact highlights the forecast equivalence of the various GAS(n) and GARCH(n) specifications, from simpler to more complex models such as Beta-Skew-t-EGARCH.

Table 11 shows the MCS procedure results for the foreign exchange rates. We can observe

Tabela 9: QL Ratios for Parametic Models

Models (Indices Data Set)	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
GAS(n)	1.0000	1.0000	1.0000
GAS(t)	0.9495	0.9817	0.9942
GAS(st)	0.9238	0.9564	0.9760
Beta-Skew-t-EGARCH (1comp.)	0.9010	0.9272	0.9522
Beta-Skew-t-EGARCH (2comp.)	0.8958	0.9263	0.9491
GARCH(n)	1.1953	1.1119	1.0653
GARCH(t)	0.9264	0.9526	0.9761
GARCH(st)	0.9564	0.9810	1.0005
APARCH	0.9357	0.9493	0.9619
GJR	0.9090	0.9391	0.9597
EGARCH	0.9352	0.9530	0.9665
Models (Currencies Data Set)	$\alpha = 1\%$	$\alpha = 2.5\%$	$\alpha = 5\%$
GAS(n)	1.0000	1.0000	1.0000
GAS(t)	1.0443	1.0291	1.0108
GAS(st)	0.9128	0.9255	0.9453
Beta-Skew-t-EGARCH (1comp.)	0.9139	0.9250	0.9458
Beta-Skew-t-EGARCH (2comp.)	0.9104	0.9281	0.9462
GARCH(n)	0.9829	0.9890	0.9931
GARCH(t)	0.9537	0.9552	0.9577
GARCH(st)	0.9695	0.9622	0.9637
APARCH	0.9128	0.9381	0.9504
GJR	0.9159	0.9410	0.9568
EGARCH	0.9114	0.9367	0.9544

that for each exchange rate we have a different number of eliminated models. Once more it is clear that the when eliminations occur, they tend to reject less complex models. For instance the most eliminated models were again the GARCH(n) and GAS(n) followed by the EGARCH model.

The p-values for the ES backtests are reported in table 12. From the first part of the table, i.e, index series, we find that the only models capable of forecasting adequately the ES were the Beta-Skew-t-EGARCH models. They reported p-values higher than 0.05 for all series, with the exceptions of STOXX50 and IBOV, for which no model had satisfactory performance. For the second part of the table, we find that all models perform well on forecasting the ES for the usual confidence levels.

Tabela 10: MCS procedure for Parametic Models - Index Series

Asset	CAC40	DAX	DOW	EURO	FTSE	IBOV	NASDAQ	SP500
GAS(n)	-	9	-	8	6	7	8	9
GAS(t)	8	8	6	7	7	4	7	8
GAS(st)	6	6	8	6	9	3	6	7
Beta-Skew-t-EGARCH (1comp.)	4	2	1	4	2	5	3	2
Beta-Skew-t-EGARCH (2comp.)	1	1	2	1	1	1	1	1
GARCH(n)	-	-	7	9	8	9	11	6
GARCH(t)	7	7	9	10	10	2	10	11
GARCH(st)	9	10	10	11	11	11	9	10
APARCH	2	5	4	3	4	8	4	4
GJR	5	4	3	2	3	6	2	3
EGARCH	3	3	5	5	5	10	5	5

Tabela 11: MCS procedure for Parametic Models - Foreign Exchange Series

Model	USD/BRL	USD/EUR	GBP/EUR	USD/JPY
GAS(n)	11	4	-	-
GAS(t)	10	9	-	-
GAS(st)	1	5	4	5
Beta-Skew-t-EGARCH (1comp.)	2	6	1	6
Beta-Skew-t-EGARCH (2comp.)	3	8	2	2
GARCH(n)	8	3	-	-
GARCH(t)	7	2	3	-
GARCH(st)	9	1	5	-
APARCH	4	7	6	4
GJR	6	10	7	3
EGARCH	5	-	-	1

Tabela 12: The backtesting results of ES Forecasting - Parametric Models

Currency	Beta-Skew-t-EGARCH (2comp.)	Beta-Skew-t-EGARCH (1comp.)	GJR-GARCH	APARCH
CAC40	0.03	0.02	0.00	0.00
DAX	0.18	0.19	0.01	0.00
DJI30	0.94	0.93	0.00	0.08
STOXX50	0.00	0.01	0.00	0.00
FTSE	0.11	0.48	0.00	0.09
IBOV	0.00	0.00	0.00	0.00
NASDAQ	0.21	0.43	0.00	0.00
S&P500	0.79	0.59	0.00	0.03
Currency	Beta-Skew-t-EGARCH (2comp.)	Beta-Skew-t-EGARCH (1comp.)	GJR-GARCH	APARCH
USD/BRL	1.00	1.00	1.00	1.00
USD/EUR	1.00	1.00	1.00	1.00
USD/GBP	0.81	0.86	0.01	0.88
USD/JPY	1.00	1.00	1.00	1.00

5 CONCLUSION

In this paper, we compare various volatility models in terms of their ability to forecast VaR and ES. Our objective resides in evaluating the performance of the recently developed GAS models, focusing on the semiparametric GAS model proposed by Blasques, Ji and Lucas (2016), and on other parametric specifications, such as the Beta-Skew-t-EGARCH model from Harvey and Sucarrat (2014). First, we perform an in-sample investigation of the semiparametric model. We study the impact of bandwidth selection on parameter's efficiency and propose a methodology for bandwidth selection in the volatility estimation context. Secondly, we run two distinct empirical experiments. The first aims to explore the semiparametric specification against its parametric counterparts. The second study compares various parametric specifications and our objective is to evaluate whether increasing the complexity, i. e., incorporating more features into the model, can help to improve its ability to forecast risk measures. The models are compared in terms of Quantile Loss of Rivera, Lee and Mishra (2004) and with the MCS procedure of Hansen (2005). Our main findings can be summarized as follows.

First, the semiparametric GAS showed good in-sample performance recovering partially the efficiency loss from the QMLE estimation. Our proposed procedure for the selection of bandwidth has not improved the in-sample properties of the semiparametric estimator. The dearth of theoretical guidance for bandwidth selection in the context of semiparametric estimation evoke a necessary discussion on the topic and yet the method has not improved estimation performance, it seems to be more a reasonable criteria compared to the pre-fixed bandwidth method. Further investigation points out that values of bandwidth ranging from $0.3 \leq b_w \leq 0.8$ do not significantly alter the parameter's in-sample performance in terms of RMSE. This finding is consistent to Blasques, Ji and Lucas (2016). Our emprical results, suggest that the semiparametric GAS specification improves the forecasting ability of the Gaussian GAS model in terms of both VaR and ES. Moreover, in several cases, the improvements are comparable to the GAS(st) model, exhibiting potential to be used in a forecasting combination exercises, for example, in the VaR dynamic technique of Bernardi and Catania (2015).

Second, more complex models such as the Beta-Skew-t-EGARCH with one and two components produced better results than simpler models. Although the improvements were substantial in some cases, the MCS results hardly ever had more than two models eliminated. This fact indicates a high homogeneity between the models, leading to the conclusion that in most of cases the models are statistically equivalent in terms of their ability to forecast VaR.

REFERÊNCIAS

ACERBI, C.; SZEKELY, B. Back-testing expected shortfall. *Risk*, Incisive Media Limited, p. 76, 2014.

ARDIA, D.; BOUDT, K.; CATANIA, L. Generalized autoregressive score models in r: The gas package. 2016.

ARDIA, D.; BOUDT, K.; CATANIA, L. Value–at–risk prediction in r with the gas package. 2016.

ARTZNER, P. et al. Coherent measures of risk. *Mathematical finance*, Wiley Online Library, v. 9, n. 3, p. 203–228, 1999.

BARTELS, M.; ZIEGELMANN, F. A. Market risk forecasting for high dimensional portfolios via factor copulas with gas dynamics. *Insurance: Mathematics and Economics*, Elsevier, v. 70, p. 66–79, 2016.

BAUWENS, L.; HAFNER, C. M.; LAURENT, S. *Handbook of volatility models and their applications*. [S.l.]: John Wiley & Sons, 2012. v. 3.

BAUWENS, L.; HAUTSCH, N. Stochastic conditional intensity processes. *Journal of Financial Econometrics*, Oxford University Press England, v. 4, n. 3, p. 450–493, 2006.

BERNARDI, M.; CATANIA, L. Comparison of value-at-risk models: the mcs package. *arXiv preprint arXiv:1502.04472*, 2015.

BERNARDI, M.; CATANIA, L. The model confidence set package for r. 2015.

BERNARDI, M.; CATANIA, L.; PETRELLA, L. Are news important to predict large losses? *arXiv preprint arXiv:1410.6898*, 2014.

BLASQUES, F.; JI, J.; LUCAS, A. Semiparametric score driven volatility models. *Computational Statistics & Data Analysis*, Elsevier, v. 100, p. 58–69, 2016.

BOLLERSLEV, T. Andersen, g. bollerslev, t. answering the skeptics: Yes, standard volatility model do provide accurate forecasts. international economic review, v. 39, p. 885-905, 1998. bawens, l. laurent, s. rombouts, j. multivariate garch models: A survey. journal of applied econometrics, v. 21, p. 79-109, 2006. *Journal of Econometrics*, v. 31, p. 307–327, 1986.

BOLLERSLEV, T.; WOOLDRIDGE, J. M. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, Taylor & Francis, v. 11, n. 2, p. 143–172, 1992.

CATANIA, L. et al. Forecasting risk with markov-switching garch models. 2017.

COX, D. R. et al. Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, JSTOR, p. 93–115, 1981.

CREAL, D.; KOOPMAN, S. J.; LUCAS, A. Univariate generalized autoregressive score volatility models. p. 1–14, year=2012, publisher=.

CREAL, D.; KOOPMAN, S. J.; LUCAS, A. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 29, n. 4, p. 552–563, 2011.

CREAL, D.; KOOPMAN, S. J.; LUCAS, A. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, Wiley Online Library, v. 28, n. 5, p. 777–795, 2013.

CUTHBERTSON, K.; NITZSCHE, D. *Quantitative financial economics: stocks, bonds and foreign exchange*. [S.l.]: John Wiley & Sons, 2005.

DI, J.; GANGOPADHYAY, A. On the efficiency of a semi-parametric garch model. *The Econometrics Journal*, Wiley Online Library, v. 14, n. 2, p. 257–277, 2011.

DROST, F. C.; KLAASSEN, C. A. Efficient estimation in semiparametric garch models. *Journal of Econometrics*, Elsevier, v. 81, n. 1, p. 193–221, 1997.

ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 987–1007, 1982.

ENGLE, R. F.; GALLO, G. M. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, Elsevier, v. 131, n. 1-2, p. 3–27, 2006.

ENGLE, R. F.; GONZALEZ-RIVERA, G. Semiparametric arch models. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 9, n. 4, p. 345–359, 1991.

ENGLE, R. F.; RUSSELL, J. R. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, JSTOR, p. 1127–1162, 1998.

GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, Wiley Online Library, v. 48, n. 5, p. 1779–1801, 1993.

GNEITING, T. Making and evaluating point forecasts. *Journal of the American Statistical Association*, Taylor & Francis, v. 106, n. 494, p. 746–762, 2011.

GONZÁLEZ-RIVERA, G.; LEE, T.-H.; MISHRA, S. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting*, Elsevier, v. 20, n. 4, p. 629–645, 2004.

HABBEMA, J. A stepwise discriminant analysis program using density estimetion. In: PHYSICA-VERLAG. *Compstat*. [S.l.], 1974. p. 101–110.

HANSEN, P. R. A test for superior predictive ability. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 23, n. 4, p. 365–380, 2005.

HARVEY, A.; SUCARRAT, G. Egarch models with fat tails, skewness and leverage. *Computational Statistics & Data Analysis*, Elsevier, v. 76, p. 320–338, 2014.

HARVEY, A. C. *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. [S.l.]: Cambridge University Press, 2013. v. 52.

HARVEY, A. C.; CHAKRAVARTY, T. Beta-t-(e) garch. Faculty of Economics, University of Cambridge, UK, 2008.

MCNEIL, A. J.; FREY, R. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, Elsevier, v. 7, n. 3-4, p. 271–300, 2000.

MCNEIL, A. J.; FREY, R.; EMBRECHTS, P. *Quantitative risk management: Concepts, techniques and tools*. [S.l.]: Princeton university press, 2015.

MULLER, F. M. et al. Melhoramentos inferenciais no modelo beta-skew-t-egarch. Universidade Federal de Santa Maria, 2016.

NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 347–370, 1991.

PATTON, A. J.; ZIEGEL, J. F.; CHEN, R. Dynamic semiparametric models for expected shortfall (and value-at-risk). *arXiv preprint arXiv:1707.05108*, 2017.

ROCCIOLETTI, S. *Backtesting Value at Risk and Expected Shortfall*. [S.l.]: Springer, 2015.

SUN, Y.; STENGOS, T. Semiparametric efficient adaptive estimation of asymmetric garch models. *Journal of Econometrics*, Elsevier, v. 133, n. 1, p. 373–386, 2006.

YANG, L. A semiparametric garch model for foreign exchange volatility. *Journal of Econometrics*, Elsevier, v. 130, n. 2, p. 365–384, 2006.