Supplementary Material: Robust Estimation of Markov-Switching GARCH Models

1 Introduction

The supplementary materials present further information on the paper. Sections 2 and 3 present some properties of the single-regime and Markov-switching GARCH model of Haas et al. (2004), respectively. Details of the simulation are presented in Section 4 and of the application in Section 5

2 Single-regime GARCH model

The conditionally heteroskedastic autoregressive (ARCH) and its extension, the generalized ARCH (GARCH) models were introduced by Engle (1982) and Bollerslev (1986), respectively, with the aim of explaining and predicting volatility in economic and financial scenarios. Although, in the initial paper, the ARCH model was applied to explain the volatility of the UK and US inflation series, this model, together to the GARCH have been expanded to several other applications. Engle (2001) showed applications of these models in portfolio analysis, mainly using measures such as VaR. One of the reasons for its popularity is that, as cited by Francq and Zakoian (2019) (Section 2, pg. 17), the GARCH model is able to capture the main stylised facts that characterise financial series.

According to Bollerslev (1986), we say that y_t follows a GARCH(p,q) model if:

$$\begin{aligned} y_t &= \sqrt{h_t} z_t, \\ h_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \end{aligned} \tag{1}$$

where $\{z_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean 0 and variance 1; $\alpha_0 > 0$; and $\alpha_i \geq 0$, $\beta_j \geq 0, \forall i = 1, \ldots, p, j = 1, \ldots, q$ are the sufficient conditions for h_t to be positive. In this model, large values of y_{t-1}^2 or h_{t-1} imply that h_t is large. That is, large values of conditional volatility $t-1$ are often followed by large values in t, characterising volatility conglomerates, one of the stylised facts of volatility.

As quoted by Francq and Zakoian (2019) (Section 2.3.2, pg. 39), it is useful to consider the ARCH(∞) representation of a GARCH(p,q) process.

This representation allows that the conditional variance h_t to be explicitly written as a function of the infinite past of the squares of the returns. Considering the GARCH model (1,1) and satisfied the positivity conditions, $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and also considering that $\beta_1 < 1$, we get that:

$$h_t = \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{i=0}^{\infty} \beta_1^i y_{t-i-1}^2. \tag{2}$$

And, as Haas et al. (2004) pointed out, the ARCH(∞) representation reveals the role of the parameters α_1 and β_1 in the GARCH(1,1) model. Therefore, according to Campbell et al. (1997) (Section 12.2, pg. 483) and Haas et al. (2004), we have that α_1 measures the magnitude in which a volatility shock in time t feeds the volatility of the period $t+1$, β_1 is a parameter of inertia and indicates the memory in the variance, $\alpha_1 + \beta_1$ measures the rate at which this effect dies over time, that is, the persistence of the variance and the total impact of a unit shock on future variances is $\alpha_1(1-\beta_1)^{-1}$.

According to Francq and Zakoian (2019) (Section 2.2.2, pg. 27), the main tool to study the strict stationarity of the GARCH model is the Lyapunov exponent γ. Let $\{\boldsymbol{A}_t, t \in \mathbb{Z}\}$ be a sequence of random matrices $(p+q) \times (p+q)$ strictly stationary, such that $E[log^+(\boldsymbol{A}_t)]$ is finite. We have that the Lyapunov exponent is given by

$$\gamma = \lim_{t\to\infty} \frac{1}{t} E[\log ||\boldsymbol{A}_t \boldsymbol{A}_{t-1} \dots \boldsymbol{A}_1||]. \tag{3}$$

Therefore, taking A_t as:

$$A_t = \begin{bmatrix} \alpha_1 z_t^2 & & \dots & & \alpha_p z_t^2 & \beta_1 z_t^2 & & \dots & & \beta_p z_t^2 \\ 1 & 0 & \dots & & 0 & 0 & & \dots & & 0 \\ 0 & 1 & \dots & & 0 & 0 & & \dots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \dots & 1 & 0 & 0 & & \dots & 0 & 0 \\ \alpha_1 & & \dots & & \alpha_p & \beta_1 & & \dots & & \beta_p \\ 0 & & \dots & & 0 & 1 & 0 & \dots & & 0 \\ 0 & & \dots & & 0 & 0 & 1 & \dots & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & \dots & 0 & 0 & 0 & & \dots & 1 & 0 \end{bmatrix}, \tag{4}$$

one obtains that a necessary and sufficient condition for the existence of strict stationarity for the GARCH(p,q) model is that the Lyapunov exponent γ of the matrix (4) is less than 0.

However, the most common is to use the second-order stationarity. In this case, the necessary and sufficient condition for the existence of second-order stationarity for GARCH(p,q) is that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. And when $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$, the model is called an IGARCH(p,q) or integrated GARCH.

Under the stationarity condition, the mean and conditional variance of y_t, that is, $E[y_t]$ and $Var[y_t]$ are given by:

$$
\begin{aligned}
E[y_t] &= E[E(y_t|\mathcal{F}_{t-1})] = E[E(\sqrt{h}_t z_t|\mathcal{F}_{t-1})] = E[\sqrt{h}_t E(z_t|\mathcal{F}_{t-1})] = E[\sqrt{h}_t E(z_t)] = 0. \\
Var[y_t] &= E[y_t^2] = E[E(y_t^2|\mathcal{F}_{t-1})] = E[E(h_t z_t^2|\mathcal{F}_{t-1})] = E[h_t E(z_t^2|\mathcal{F}_{t-1})] = E[h_t E(z_t^2)] \\
&= E[h_t] = \alpha_0 + \sum_{i=1}^{p} \alpha_i E[h_t] + \sum_{j=1}^{q} \beta_j E[h_t] = \alpha_0 + E[h_t] \left(\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j \right) \\
&= \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j},
\end{aligned}
$$

where $\mathcal{F}_{t-1}$ denotes all information up to time $t-1$.

Analogously, we obtain that the unconditional autocovariance function of y_t is obtained by:

$$
E[y_t y_{t+\tau}] = E[E(y_t y_{t+\tau}|\mathcal{F}_{t-1})] = \begin{cases} Var[y_t], & \text{if } \tau = 0 \\ 0. & \text{if } \tau \geq 1. \end{cases}
$$

Now, for the autocorrelation function of y_t^2, writing the process y_t^2 as an ARMA(q, max(p,q)) process, that is,

$$
y_t^2 = \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) y_{t-i}^2 = \alpha_0 + (y_t^2 - h_t) - \sum_{i=1}^{q} \beta_i (y_{t-i}^2 - h_{t-i}),
$$

we can obtain the autocorrelation function of y_t^2.

As quoted by Francq and Zakoian (2019) (Section 2.4.2, pg. 45), in the case of GARCH processes, note the difference between the tails of the marginal and conditional distributions. For this, kurtosis is used. Assuming that there is a fourth moment of y_t, the excess kurtosis of the GARCH(p,q) model, when z_t follows normal distribution, is given by:

$$
\kappa_{y_t} = \frac{E[y_t^4]}{\{E[y_t^2]\}^2} - 3 = \frac{6a + \kappa_{z_t}(1 + 3a)}{1 - a(\kappa_{z_t} + 2)},
$$

where κ_{z_t} is the excess kurtosis of the normal distribution and $a = \sum_{i=1}^{\infty} \psi_i^2$, such that ψ_i is given by:

$$1+\sum_{i=1}^{\infty}\psi_i k^i=\left\{1-\sum_{i=1}^{\max(p,q)}(\alpha_i+\beta_i)k^i\right\}^{-1}\left(1-\sum_{i=1}^{p}\beta_i k^i\right).$$

3 Markov-switching Garch Model

Definition: The MS(K)-GARCH(1,1) model proposed by Haas et al. (2004) can be expressed as follows:

$$y_t=\sqrt{h_t^{(\Delta_t)}}z_t, \tag{5}$$
$$h_t^{(k)}=\alpha_0^{(k)}+\alpha_1^{(k)}y_{t-1}^2+\beta_1^{(k)}h_{t-1}^{(k)}, \tag{6}$$

where $h_t^{(k)}$ are the conditional variances of regime k. In this definition, the term accompanying the parameter $\beta_1^{(k)}$, that is, $h_{t-1}^{(k)}$, is what differentiates this model from the one proposed by Gray (1996). $\alpha_0^{(k)}>0$, $\alpha_1^{(k)}\geq 0$, $\beta_1^{(k)}\geq 0$ for all $k=1,\ldots,K$ are sufficient conditions for $h_t^{(k)}$ to be positive. $\{z_t\}$ is a sequence of i.i.d. random variables with mean 0 and variance 1, commonly chosen as N(0.1). And finally, Δ_t is a Markov chain with finite state space $\mathcal{S}=\{1,\ldots,K\}$ and a $K\times K$ transition matrix $\boldsymbol{P}$ denoted as:

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K}\\ p_{21} & p_{22} & \cdots & p_{2K}\\ \vdots & \vdots & \ddots & \vdots\\ p_{K1} & p_{K2} & \cdots & p_{KK}\end{bmatrix}, \tag{7}$$

where, $p_{ij}=P(\Delta_t=j|\Delta_{t-1}=i)$, and the sum of each row equal to 1 and all elements of the matrix p_{ij} are greater than zero. Assuming that the Markov chain is irreducible and aperiodic, the stationary distribution $\boldsymbol{\pi}_\infty=(\pi_\infty^{(1)},\pi_\infty^{(2)},\ldots,\pi_\infty^{(k)})'$ can be obtained by $\boldsymbol{\pi}_\infty=(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{v}$, where

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{I}_k-\boldsymbol{P}\\ \mathbf{1}_k'\end{bmatrix}\quad\text{and}\quad \boldsymbol{v}=\begin{bmatrix}\mathbf{0}_k\\ \mathbf{1}_k\end{bmatrix}, \tag{8}$$

$\boldsymbol{I}_k$ is the $k\times k$ identity matrix, and $\mathbf{0}_k$ and $\mathbf{1}_k$ are $k\times 1$ vectors of zeros and ones, respectively.

The conditional variance $h_t^{(k)}$ depends only on the respective parameters of its regime k, in contrast to the one given by Gray (1996). Therefore, if $\max\{\beta_1^{(1)},\ldots,\beta_1^{(K)}\}<1$, like to the GARCH models with a single-regime, we have that $\alpha_1^{(k)}$ represents the impact of a shock at $t-1$ on $h_t^{(k)}$; $\beta_1^{(k)}$ represents the memory in $h_t^{(k)}$; and $\alpha_1^{(k)}(1-\beta_1^{(k)})^{-1}$ reflects the total impact of a unit shock on future variations of the regime k.

To study the second-order stationarity of the model consider the $K^2 \times K^2$ matrix $\boldsymbol{M}$:

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{21} & \dots & \boldsymbol{M}_{k1} \\ \boldsymbol{M}_{12} & \boldsymbol{M}_{22} & \dots & \boldsymbol{M}_{k2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{M}_{1k} & \boldsymbol{M}_{2k} & \dots & \boldsymbol{M}_{kk} \end{bmatrix}, \tag{9}$$

where $\boldsymbol{M}_{ij} = p_{ij}(\boldsymbol{\beta}+\boldsymbol{\alpha}_1\boldsymbol{e}_k^{'})$ are $K\times K$ matrices, $\boldsymbol{\beta} = \text{diag}(\beta_1^{(1)}, \beta_1^{(2)}, \dots \beta_1^{(K)})'$, $\boldsymbol{\alpha_1} = (\alpha_1^{(1)}, \alpha_1^{(2)}, ..., \alpha_1^{(K)})'$, p_{ij} denotes the transition probabilities defined in (7) and $\boldsymbol{e}_k^{'}$ is a $K \times 1$ vector that contains zeros, except in the k-th element. According to Haas et al. (2004), the process is weakly stationary if and only if $\rho(\boldsymbol{M}) < 1$, where $\rho(.)$ is the largest eigenvalue in modulus of the matrix $\boldsymbol{M}$. Similar to the single-regime GARCH model, the strict stationarity condition is complicated. However, Liu (2006) presented the conditions for strict stationarity. Furthermore, Haas et al. (2004) stated that a necessary condition for the weak stationarity of the process is that, for all k, $\beta_1^{(k)} < 1$. However, such conditions refer to the MS(K)-GARCH(1,1) model, while for the MS(K)-GARCH(p,q) model, Abramson and Cohen (2007) presented the strict and second order stationarity conditions Under the stationarity conditions, the mean and unconditional variance of y_t are given by,

$$E[y_t] = 0;$$

$$Var[y_t] = E[y_t^2] = \sum_{j=1}^{k} \pi_j E[y_t^2|\Delta_t = j] = (\text{vec } \boldsymbol{P})^{'}(I_{k^2} - \boldsymbol{M})^{-1}(\boldsymbol{\pi}_\infty \otimes \boldsymbol{\alpha}_0),$$

where $\boldsymbol{\alpha_0} = (\alpha_0^{(1)}, \alpha_0^{(2)}, ..., \alpha_0^{(K)})'$.

To present the condition for the existence of the kurtosis of the process define the $K^3 \times K^3$ matrix $\boldsymbol{Q}$ that contains the $K^2 \times K^2$ matrices $\boldsymbol{Q}_{ij}$ such that:

$$\boldsymbol{Q}_{ji} = p_{ji}[3(\boldsymbol{\alpha}_1 \otimes \boldsymbol{\alpha_1})\text{vec}[\text{diag}(\boldsymbol{e}_i)]^{'} + (\boldsymbol{\alpha_1}\boldsymbol{e}_i^{'}) \otimes \boldsymbol{\beta} + \boldsymbol{\beta} \otimes (\boldsymbol{\alpha_1}\boldsymbol{e}_i^{'}) + \boldsymbol{\beta} \otimes \boldsymbol{\beta}].$$

The condition for the existence of the fourth unconditional moment of the process y_t is given by $\rho(\boldsymbol{Q}) < 1$, and the excess of kurtosis of the process is given by:

$$\kappa_{y_t} = \frac{E[y_t^4]}{\{E[y_t^2]\}^2} - 3 = \frac{3\boldsymbol{\omega}^{'}\boldsymbol{\eta}}{[(\text{vec } \boldsymbol{P})'(I_{k^2} - \boldsymbol{M})^{-1}(\boldsymbol{\pi}_\infty \otimes \boldsymbol{\alpha}_0)]^2} - 3,$$

where $\boldsymbol{\omega} = [\text{vec}(\text{diag}(p_1)^{'}, \ldots, \text{vec}(\text{diag}(p_k))^{'}]^{'}$, such that p_i is the i-th column of the matrix P (7); and $\boldsymbol{\eta} = (I_{k^3} - \boldsymbol{Q})^{-1}[\boldsymbol{A}_{\infty}^2 + \boldsymbol{R}(I_{k^2} - \boldsymbol{M})^{-1}\boldsymbol{A}_{\infty}^1]$, so that $\boldsymbol{R}$ is a $K^3 \times K^2$ matrix with $K^2 \times K$ matrix elements $\boldsymbol{R}_{ji} = p_{ji}[(\boldsymbol{\alpha}_1 \boldsymbol{e}_i^{'}) \otimes \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_0 \otimes (\boldsymbol{\alpha}_1 \boldsymbol{e}_i^{'}) + \boldsymbol{\alpha}_0 \otimes \boldsymbol{\beta} + \boldsymbol{\beta} \otimes \boldsymbol{\alpha}_0]$ and

$$\boldsymbol{A}_t = \begin{bmatrix} \boldsymbol{A}_t^1 \\ \boldsymbol{A}_t^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\pi}_t \otimes \boldsymbol{\alpha}_0 \\ \boldsymbol{\pi}_t \otimes (\boldsymbol{\alpha}_0 \otimes \boldsymbol{\alpha}_0) \end{bmatrix},$$

where $\boldsymbol{\pi}_t = (\pi_t^{(1)}, \ldots, \pi_t^{(K)})^{'}$ is the probability distribution over the state space $\mathcal{S}$ of Δ_t.

Assuming that $\rho(\boldsymbol{Q}) < 1$, when $K = 2$, we have the process autocovariance function of the squared y_t^2 for $l \geq 1$ is given by:

$$\begin{aligned} \text{cov}(y_t^2, y_{t-l}^2) =& (\text{vec}P)^{'}(\boldsymbol{M}^{l-1}\boldsymbol{\Theta}\boldsymbol{P}^{\star}\boldsymbol{\eta} + (\boldsymbol{I}_4 - M)^{-1}(\boldsymbol{I}_4 - M^l)[\boldsymbol{C}_1\tilde{\boldsymbol{\sigma}}^{(2)} \otimes \boldsymbol{\alpha}_0] + \\ & (\delta \boldsymbol{I}_4 - M)^{-1}[(\boldsymbol{C}_2\tilde{\boldsymbol{\sigma}}^{(2)}) \otimes \boldsymbol{\alpha}_0], \end{aligned}$$

where $\delta = p_{11} + p_{22} - 1$ is the degree of persistence due to Markov chain effects; $\tilde{\boldsymbol{\sigma}}^{(2)} = [\text{diag}(p_1), \ldots, [\text{diag}(p_2)](\boldsymbol{I}_{K^2} - \boldsymbol{M})^{-1}\boldsymbol{A}_{\infty}^{-1}$,

$$\boldsymbol{\Theta} = \begin{bmatrix} 3\boldsymbol{\alpha}_1 e_1^{'} + \boldsymbol{\beta} & \mathbf{0}_{K\times K} & \ldots & \mathbf{0}_{K\times K} \\ \mathbf{0}_{K\times K} & 3\boldsymbol{\alpha}_1 e_2^{'} & \ldots & \mathbf{0}_{K\times K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{K\times K} & \mathbf{0}_{K\times K} & \ldots & 3\boldsymbol{\alpha}_1 e_k^{'} \end{bmatrix}, \quad \boldsymbol{C}_1 = \begin{bmatrix} \boldsymbol{\pi}_{\infty}^1 & \boldsymbol{\pi}_{\infty}^1 \\ \boldsymbol{\pi}_{\infty}^2 & \boldsymbol{\pi}_{\infty}^2 \end{bmatrix}, \quad \boldsymbol{C}_2 = \begin{bmatrix} \boldsymbol{\pi}_{\infty}^2 & -\boldsymbol{\pi}_{\infty}^1 \\ -\boldsymbol{\pi}_{\infty}^2 & \boldsymbol{\pi}_{\infty}^1 \end{bmatrix}$$

$$\text{and} \quad \boldsymbol{P}^{\star} = \begin{bmatrix} p_{11}\boldsymbol{I}_K & \mathbf{0}_{K\times K} & \ldots & \mathbf{0}_{K\times K} & \ldots & p_{k1}\boldsymbol{I}_K & \mathbf{0}_{K\times K} & \ldots & \mathbf{0}_{K\times K} \\ \mathbf{0}_{K\times K} & p_{12}\boldsymbol{I}_K & \ldots & \mathbf{0}_{K\times K} & \ldots & \mathbf{0}_{K\times K} & p_{k1}\boldsymbol{I}_K & \ldots & \mathbf{0}_{K\times K} \\ \vdots & \vdots & \ddots & \vdots & \ldots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{K\times K} & \mathbf{0}_{K\times K} & \ldots & p_{1k}\boldsymbol{I}_K & \ldots & \mathbf{0}_{K\times K} & \mathbf{0}_{K\times K} & \ldots & p_{kk}\boldsymbol{I}_K \end{bmatrix}.$$

We observe that the autocovariance function depends on the powers of $\boldsymbol{M}$ and δ, that is, the autocovariance function is controlled by the exponential rate decay $\max\{\rho(\boldsymbol{M}), \delta\}$.

3.1 Volatility prediction

As Paolella et al. (2012) pointed out, a feature of MS models is that conditional expectations can be calculated analytically, which is in contrast to many other non-linear time series models. Using the forecast form presented in Paolella et al. (2012) for the MS(K)-GARCH(1,1) model, Equation (6) in written in matrix form:

$$\boldsymbol{X}_t = \boldsymbol{\Psi} + \boldsymbol{C}_{\Delta_{t-1},t-1}\boldsymbol{X}_{t-1}, \tag{10}$$

where $\boldsymbol{X}_t = \{h_t^{(1)}, \ldots, h_t^{(k)}, y_{t-1}^2\}'$, $\boldsymbol{C}_{\Delta_t,t} = \boldsymbol{\alpha}\boldsymbol{1}'_{\Delta_t} z_t^2 + \boldsymbol{\beta}$, such $\boldsymbol{1}^{(k)}$ is the unity vector in $\mathbb{R}^{K+1}$, for $k = 1, \ldots, K$, and

$$\boldsymbol{\Psi} = \begin{bmatrix} \alpha_0^{(1)} \\ \vdots \\ \alpha_0^{(K)} \\ 0 \end{bmatrix}, \qquad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1^{(1)} \\ \vdots \\ \alpha_1^{(K)} \\ 1 \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1^{(1)} & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & \beta_1^{(K)} & 0 \\ 0 & \ldots & 0 & 0 \end{bmatrix}. \tag{11}$$

For any function $f : S \to M_{n\times n'}(\mathbb{R})$, such that $M_{n\times n'}(\mathbb{R})$ is the space of $n \times n^{'}$ real matrices, and S is the state space of $\{\Delta_t\}$, define the matrix

$$\mathbb{P}_f^{(d)} = \begin{bmatrix} p_{11}^{(d)} f(1) & \ldots & p_{K1}^{(d)} f(1) \\ \vdots & \ldots & \vdots \\ p_{1K}^{(d)} f(k) & \ldots & p_{KK}^{(d)} f(K) \end{bmatrix}, \tag{12}$$

and also define

$$C_{mm}(j) = E(C_{jt}^{\otimes m}), \;\; j = 1, \ldots, K, \;\; m \in \mathbb{N}, \tag{13}$$

where $A^{\otimes m}$ denotes the m-th Kronecker power of A^2.

Following Paolella et al. (2012), the forecast volatilities can be calculated using a simple recursive scheme. Thus, repeated substitution in Equation (10) yields:

$$\begin{aligned} \boldsymbol{X}_{t+\tau+1} &= \boldsymbol{\Psi} + \boldsymbol{C}_{\Delta_{t+\tau},t+\tau}\boldsymbol{X}_{t+\tau} \\ &= \boldsymbol{\Psi} + \boldsymbol{C}_{\Delta_{t+\tau},t+\tau}\boldsymbol{\Psi} + \boldsymbol{C}_{\Delta_{t+\tau},t+\tau}\boldsymbol{C}_{\Delta_{t+\tau-1},t+\tau-1}\boldsymbol{X}_{t+\tau-1} \\ &\vdots \\ &= \sum_{i=1}^{\tau} \left\{ \prod_{l=1}^{i=1} \boldsymbol{C}_{\Delta_{t+\tau+1-l},t+\tau+1-l} \right\} \boldsymbol{\Psi} + \prod_{l=1}^{\tau} \left\{ \boldsymbol{C}_{\Delta_{t+\tau+1-l},t+\tau+1-l} \right\} \boldsymbol{X}_{t+1}, \end{aligned}$$

where $\prod_{l=1}^{0} \boldsymbol{C}_{\Delta_{t+\tau+1-l},t+\tau+1-l} \stackrel{\text{def}}{=} \boldsymbol{I}_{K+1}$. Let $\Delta_t = \{\Delta_s : s \leq t\}$. Then

$$E_t(\boldsymbol{X}_{t+\tau+1}|\Delta_{t+\tau}) = \sum_{i=1}^{\tau} \left\{ \prod_{l=1}^{i-1} C_{11}(\Delta_{t+\tau+1-l} \right\} \boldsymbol{\Psi} + \left\{ \prod_{l=1}^{i-1} C_{11}(\Delta_{t+\tau+1-l} \right\} \boldsymbol{X}_{t+1}, \tag{14}$$

where $E_t(.)$ denotes the expectation given $\mathcal{F}_t$. Applying some lemmas shown in Paolella et al. (2012), we obtain that:

$$E_t(\boldsymbol{X}_{t+\tau+1}) = (\mathbf{1}_K \otimes I_{k+1})M(\tau), \quad \tau \geq 0, \tag{15}$$

where $M(0) = \pi_t \otimes \boldsymbol{X}_{t+1}$ and

$$\begin{aligned}
M(\tau) &= \sum_{i=1}^{\tau} \mathbb{P}_{C_11}^{i-1}(\pi_{t+\tau+1-i} \otimes \boldsymbol{\Psi}) + \mathbb{P}_{C_11}^{d}(\pi_t \otimes \boldsymbol{X}_{t+1}) \\
&= \pi_{t+\tau} \otimes \boldsymbol{\Psi} + \mathbb{P}_{C_11} \left\{ \sum_{i=1}^{\tau-1} \mathbb{P}_{C_11}^{i-1}(\pi_{t+(\tau-1)+1-i} \otimes \boldsymbol{\Psi}) + \mathbb{P}_{C_11}^{d-1}(\pi_t \otimes \boldsymbol{X}_{t+1}) \right\} \\
&= \pi_{t+\tau} \otimes \boldsymbol{\Psi} + \mathbb{P}_{C_11} M(\tau - 1), \quad \tau \geq 1.
\end{aligned}$$

Therefore, volatility forecasts can be calculated as

$$E_t(y_{t+\tau}^2) = (\mathbf{1}_K \otimes e_{K+1})^{'} M(\tau), \quad \text{where} \tag{16}$$

$$M(\tau) = \begin{cases} \pi_t \otimes \boldsymbol{X}_{t+1}, & \text{if } \tau = 0 \\ \pi_{t+\tau} \otimes \boldsymbol{\Psi} + \mathbb{P}_{C_11} M(\tau - 1), & \text{if } \tau \geq 1. \end{cases} \tag{17}$$

In practice, we estimate π_t in Equation (17) via $\boldsymbol{\xi}_t$ in Equation (23) in the paper (estimation via Hamilton filter).

4 Simulation

The results of the estimation of the MS(2)-GARCH(1,1) model based on 500 replications are show in Table 1. The results are presented for three estimators: Two nonrobust estimators using, the quasi-maximum likelihood estimators with normal (QML-n) and t-Student(QML-t) distributions and the proposed robust estimator (Rob).

5 Application

This section presents the application to the exchange rates of euro, pound, Swiss franc and yen against the US dollar. First, the descriptive statistics of the series and graphs of the return series and the correlation function of the return series and their square are presented in Section 5.1. The estimation of the first and last windows of each series in Section 5.2 Finally, in Section 5.3 we present the performance tests of the VaR an ES estimates.

5.1 The series

In this subsection we present some descriptive and information of the return series. Table 2 presents some descriptive measures of these returns.

The evolution of the exchange rate series and their returns are given in Figure 1 and of the sample autocorrelation function of the return and squared returns of the exchange rate series in Figure 2.

5.2 Estimation of the first and last windows

The estimations of the first as last windows for euro, pound, Swiss franc and yen are presented in Tables 3-6. Table 3 presents the estimates for the euro series. There are great differences in the estimates in the two windows. The QML-n estimate the probability of staying at regime 1 in two consecutive observations almost as equal to zero, while the robust estimators estimates the stationarity probability of regime 2 as approximately 0.95.

Table 4 presents the estimation for the pound. In this series there are also differences in the estimates of the parameters between the windows. The QML-n estimate the probability of staying at the same regime in two consecutive days as small and the robust estimator, in the last window, estimate the unconditional probability of one regime almost near one. Furthermore, the QML-n estimator has a significant change in $\alpha_1^{(1)}$ estimates, which vary from 0.1095 to 0.2379, from the first to the last window.

Table 5 presents the results for the yen. There is a reasonable change in the estimation of the parameters between the windows in some parameters, mainly in the parameters of the transition matrix. The QML-n estimator does not show much difference in the estimates of the variance equation parameters, but there is a difference in p_{11} that varies from 0.1061 to 0.2778. However, the probability of staying in state 1 is still very low. Regarding the QML-t estimator, there is a very large difference in the parameters of regime 2. Furthermore, in the last window, the estimate of $\alpha_1^{(1)}$ it's very high. Again, in the last window, the estimates of the parameters in the regime with the highest probability in the stationary distribution are very similar.

Finally, Table 6 presents the parameters referring to the Swiss franc currency. For this currency, there are extreme changes in parameter estimates, mainly in relation to $\alpha_0^{(1)}$. From the first to the last window, the estimates range from 0.0614, 0.0000 and 0.0057 to 0.0603, 0.5053 and 0.1845, for the robust estimators, QML-n and QML-t, respectively. Furthermore, the robust estimator tends to estimate a regime with greater predominance, with stationary probabilities equal to 0.9633 and 0.9905 for the first and last window, respectively. For the QML-t estimator these probabilities drop to 0.7019 and

0.7297, respectively. In the case of the QML-n estimator, there is a large variation, with a value equal to 0.7631, close to the QML-t estimate, in the first window and equal to 0.9883 in the last window, close to the robust estimate. Again, the estimates of the parameters in the regime with the highest probability in the stationary distribution are very similar.

5.3 Forecasting performance tests

This section presents tests and measures to evaluate the performance of the estimated VaR and ES at α level. Consider that the VaR and ES are estimated for the t-th observation, $t = T+1, ..., T+N$, T being the size of the rolling window to estimate them. Denote by $\hat{v}_t$ and $\hat{s}_t$ the estimates of VaR_t and ES_t, respectively, where we are dropping the index α and use the same notation for estimators and estimates to simplify the notation. First, we present some tests, called calibration tests, which are based on test of hypotheses to evaluate the VaR and ES forecast accuracy, which are presented in Sections 5.3.1. Roughly, we can say that calibration tests test the null hypothesis that the risk measurement procedure is considered as adequate. When there are more than one test which pass the calibration tests, the calibration tests cannot be used to sort the methods. One method to sort the methods is to use an scoring measure. Section 5.4 presents some of scoring functions.

5.3.1 VaR calibration tests

Define the variable called *hit*, denoted by H^{α}, as the number of times that the loss is greater than the VaR. First, define the indicator variable $\mathbb{I}_t$ as

$$\mathbb{I}_t = \mathbb{I}(y_t) = \begin{cases} 1, & \text{if } y_t < \hat{v}_t \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbb{I}(.)$ is the indicator functions and y_t is the return of the series.

Kupiec (1995) proposed a coverage test based on the likelihood ratio called the unconditional coverage test. When the VaR is correctly estimated, $\mathbb{I}_t$ has independent Bernoulli(α) distributions; thus, $H = \sum_{t=T+1}^{T+N} \mathbb{I}_t \sim$ Binomial(N, α). To test the null hypothesis $H_0 : E[H] = n\alpha$, the test statistic UC is defined as

$$\text{UC} = -2\ln\left[(1-\alpha)^{N-H}\alpha^{H}\right] + 2\ln\left[\left(1-\frac{H}{N}\right)^{N-H}\left(\frac{H}{N}\right)^{H}\right],$$

and under the null hypothesis, $\text{UC} \sim \chi^2(1)$.

However, as cited by Gaglianone et al. (2011), the unconditional coverage property does not provide information about the time dependence of VaR violations, and the Kupiec (1995) test ignores conditional coverage. Christoffersen (1998) extended the likelihood ratio statistic UC in such a way that the sequence of hits must also be independent over time. For this, Christoffersen (1998) assumed that the process $\mathbb{I}_t$ follows a Markov chain with two states and transition probability matrix

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix} = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix},$$

such that $\pi_{ij} = P[\mathbb{I}_t = j | \mathbb{I}_{t-1} = i]$, $i, j = 0, 1$. Thus, under the null hypothesis,

$$H_0 : \Pi = \begin{pmatrix} 1-\alpha & \alpha \\ 1-\alpha & \alpha \end{pmatrix},$$

the test statistics for conditional coverage and for independence are given respectively by:

$$\begin{aligned}
\text{CC} &= -2\ln\left[(1-\alpha)^{N-H}\alpha^{H}\right] + 2\ln\left[(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}\right] \xrightarrow{d} \chi^2(2) \\
\text{Ind} &= -2\ln\left[(1-H/N)^{N-H}(H/N)^{H}\right] + 2\ln\left[(1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}\right] \\
&\qquad \xrightarrow{d} \chi^2(1),
\end{aligned}$$

where n_{ij} is the total of observations going from i to j, and under the null hypothesis of independence, we have that $\hat{\pi}_{11} = \hat{\pi}_{01} = (n_{01} + n_{11}) = H/N$. Also, we have CC = UC + Ind.

Also, Engle and Manganelli (2004) proposed a test based on a linear regression. Let $\text{Hit}_t \equiv \mathbb{I}_t - \alpha$ and consider the following regression:

$$\text{Hit}_t = \delta_0 + \sum_{j=1}^{L} \delta_j \text{Hit}_{t-j} + \delta_{L+1}\hat{v}_{t-1} + \epsilon, \tag{18}$$

where $\epsilon \sim N(0, \sigma^2)$. Now denote $\boldsymbol{\delta} = (\delta_0 ...., \delta_{L+1})'$ the vector of all parameters and $\hat{\boldsymbol{\delta}}$ your estimate by ordinary least squares. Then, under the null hypothesis $H_0 : \boldsymbol{\delta} = \mathbf{0}$, the test statistic, denoted by DQ is given by:

$$\text{DQ} = \frac{\hat{\boldsymbol{\delta}}' \boldsymbol{Z}' \boldsymbol{Z} \hat{\boldsymbol{\delta}}}{\alpha(1-\alpha)},$$

where $\boldsymbol{Z}$ is the matrix of explanatory variables. Also, under the null hypothesis, DQ converges in distribution to a $\chi^2(L+2)$. Note that this test represents the Wald test in which we tested all the parameters simultaneously. Generally, $L = 4$ is taken. Not necessarily, the only explanatory variable in (18) needs to be $\hat{v}_{t-1}$. Other variables such as returns, for example, can enter the model. However, for practicality, and following the notation presented by Ardia et al. (2018), we chose this definition. There are other methods to test the validity of VaR and it is not within the scope of this work to list all of them. Among them, we can mention the loss function presented by González-Rivera et al. (2004) and the VaR quantile regression test of Gaglianone et al. (2011), but they will not be used in this work.

5.3.2 ES calibration tests

Most of the calibration tests for ES also involves VaR estimates due to non-elicitability and non-identifiability of the ES; see, for instance Fissler et al. (2016). Here we present five tests.

McNeil and Frey (2000) presented an ES calibration test based on the standardized residual exceedances. Define the standardized residuals as

$$e_t = \frac{y_t - \hat{s}_t}{\sqrt{h_t}}, \quad t = T+1, ..., T+N,$$

McNeil and Frey (2000) define the exceedance residuals as the set of standardized residuals where $y_t < -\hat{v}_t$, i.e.

$$\{r_t : t = T+1, ..., T+N, y_t < -\hat{v}_t\}$$

Under the assumptions that the VaR is correctly specified the the mean of the exceedance residuals is zero. So, they proposed to test the null and alternative hypotheses defined as $H_o : E(e_t|y_t < \mathrm{VaR}_t = 0$ and $H_a : E(e_t|y_t < \mathrm{VaR}_t < 0$, respectively. The underlying distribution of the exceedance residuals under the null hypothesis is obtained by bootstrap.

Nolde and Ziegel (2017) proposed the conditional calibration test (CoC). Defining

$$\boldsymbol{W}(\hat{v}_t, \hat{s}_t, y_t) = \begin{pmatrix} \alpha - \mathbb{I}(y_t < \hat{v}_t) \\ \hat{s}_t - \hat{v}_t + \mathbb{I}(y_t < \hat{v})(\hat{v}_t - y_t)/\alpha \end{pmatrix},$$

we expect that $H_O : E\left(\boldsymbol{W}(\mathrm{VaR}_t, \mathrm{ES}_t, y_t)|\mathcal{F}_{t-1}\right) = \mathbf{0}$. Nolde and Ziegel (2017) to test this hypothesis by a Wald test given by:

$$\mathrm{CoC} = N\left(\frac{1}{N}\sum_{t=T+1}^{T+N} \boldsymbol{W}(\hat{v}_t, \hat{s}_t, y_t)\right)^{\top} \hat{\boldsymbol{\Omega}}^{-1} \left(\frac{1}{N}\sum_{t=T+1}^{T+N} \boldsymbol{W}(\hat{v}_t, \hat{s}_t, y_t)\right),$$

where

$$\hat{\boldsymbol{\Omega}} = \frac{1}{N}\left(\sum_{t=T+1}^{T+N} \boldsymbol{W}(\hat{v}_t, \hat{s}_t, y_t)\right)\left(\sum_{t=T+1}^{T+N} \boldsymbol{W}(\hat{v}_t, \hat{s}_t, y_t)\right)^{\top},$$

Under the null hypothesis the test statistics distribution converge to a chi-square distribution with two degrees of freedom.

Bayer and Dimitriadis (2022) proposed three tests called Expected Shortfall Regression (ESR). They consider a double regression model

$$y_t = \boldsymbol{V}_t^\top \boldsymbol{\beta} + u_t^q \quad \text{and} y_t = \boldsymbol{W}_t^\top \boldsymbol{\gamma} + u_t^e, \tag{19}$$

where $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$ are covariate vectors with the usual regression model assumptions. Bayer and Dimitriadis (2022) proposed three sets of covariate vectors.

The first two tests are two-sided alternative. The first one, called *Auxiliary ESR Backtest* considers the regression system

$$y_t = \beta_1 + \beta_2 \hat{v}_t + u_t^q \quad \text{and} \quad y_t = \gamma_1 + \gamma_2 \hat{s}_t + u_t^e. \tag{20}$$

The second one, called *Strict ESR Backtest* considers the regression system

$$y_t = \beta_1 + \beta_2 \hat{s}_t + u_t^q \quad \text{and} \quad y_t = \gamma_1 + \gamma_2 \hat{s}_t + u_t^e. \tag{21}$$

In both tests they considered the hypotheses $H_0 : (\gamma, \gamma_2) = (0,1)$ and $H_0 : (\gamma, \gamma_2) \neq (0,1)$ and the Wald-type test statistics

$$N \left(\hat{\boldsymbol{\gamma}} - (0,1)\right) \hat{\Omega}_\gamma^{-1} \left(\hat{\boldsymbol{\gamma}} - (0,1)\right)^\top, \tag{22}$$

where $\hat{\Omega}_\gamma$ are a consistent estimator of the covariance of the estimator of $\hat{\boldsymbol{\gamma}}$; see Section 1.5 of Bayer and Dimitriadis (2022).

The third test can employ one or two side hypotheses with regression system

$$y_t - \hat{s}_t = \beta_1 + \beta_2 \hat{s}_t + u_t^q \quad \text{and} \quad y_t - \hat{s}_t = \gamma_1 + u_t^e.. \tag{23}$$

Because this is equivalent to the strict ESR backtest with the slope parameter $\gamma_2 = 1$ and test only the intercept γ_1, it is called *Intercept ESR Backtest.* In the two-sided alternatives the hypotheses are $H_0 : \gamma_1 = 0$ and $H_0 : \gamma_1 \neq 0$, while the one-sided alternatives are $H_0 : \gamma_1 \geq 0$ and $H_0 : \gamma_1 < 0$. The test statistics is the t-test. In a certain way, Bayer and Dimitriadis (2022) suggest to use the one-sided hypotheses because they comply to the Basel Accords requirements with less capital, and it is the test used in the paper.

5.4 Scoring functions

The comparison of the goodness of forecasting from several methods can be done by defining scoring or loss functions. Following Hallin and Trucíos (2021) one scoring function for VaR and three for ES.

González-Rivera et al. (2004) proposed a goodness of fit measure for the VaR, which is based on the fact that VaR is a quantile. It is defined as

$$\mathrm{QL}(\hat{v}_t, \hat{s}_t, y_t) = \frac{1}{N} \sum_{t=T+1}^{T+N} \left(\alpha - \mathbb{I}(y_t < \hat{v}_t)\right)(y_t - \hat{v}_t)$$

Although ES is not elicitable, Fissler et al. (2016) showed that the pair (VaR, ES) is jointly elicitable, and suggest as possible scoring functions of the form

$$\begin{aligned} S(\hat{v}_t, \hat{s}_t, y_t) &= \left(\mathbb{I}(y_t \leq \hat{v}_t) - \alpha\right) G_1(\hat{v}_t) - \mathbb{I}(y_t \leq \hat{v}_t) G_1(y_t) \\ &\quad + G_2(\hat{s}_t)\left(\hat{s}_t - \hat{v}_t + \mathbb{I}(y_t \leq \hat{v}_t)\right)(\hat{v}_t - y_t)/\alpha + G_3(\hat{s}_t) + G_4(y_t), \end{aligned} \tag{24}$$

where the functions $G_1 - G_4$ must satisfy some restrictions (Fissler and Ziegel, 2016; Fissler et al., 2016; Taylor, 2020). We use three suggestions. The first one, proposed by Fissler et al. (2016) with

$$G_1(x) = x, \;\; G_2(x) = e^x/(1+e^x), \;\; G_3(x) = \log(1+e^x), \;\text{ and }\; G_4(x) = \log(2).$$

The second one, proposed by Nolde and Ziegel (2017) with

$$G_1(x) = 0, \;\; G_2(x) = \sqrt{-x}/2, \;\; G_3(x) = -\sqrt{-x} \;\text{ and }\; G_4(x) = 0,$$

and the third one, proposed by Taylor (2019) with

$$G_1(x) = 0, \;\; G_2(x) = -1/x, \;\; G_3(x) = -\log(-x) \;\text{ and }\; G_4(x) - = 1 - \log(1-\alpha).$$

References

Abramson, A. and Cohen, I. (2007). On the stationarity of Markov-switching GARCH processes. *Econometric Theory*, 23(3):485–500.

Ardia, D., Bluteau, K., Boudt, K., and Catania, L. (2018). Forecasting risk with Markov-switching GARCH models: A large-scale performance study. *International Journal of Forecasting*, 34(4):733–747.

Bayer, S. and Dimitriadis, T. (2022). Regression-based expected shortfall backtesting. *Journal of Financial Econometrics*, 20(3):437–471.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Campbell, J. Y., Lo, A., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 2nd edition.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and osband's principle. 44(4):1680–1707.

Fissler, T., Ziegel, J. F., and Gneiting, T. (2016). Expected shortfall is jointly elicitable with value at risk-implications for backtesting. *Risk*, 5(1):8–16.

Francq, C. and Zakoian, J.-M. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, Hoboken, NJ, 2nd edition.

Gaglianone, W. P., Lima, L. R., Linton, O., and Smith, D. R. (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*, 29(1):150–160.

González-Rivera, G., Lee, T.-H., and Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, 20(4):629–645.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1):27–62.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4):493–530.

Hallin, M. and Trucíos, C. (2021). Forecasting value-at-risk and expected shortfall in large portfolios: A general dynamic factor model approach. *Econometrics and Statistics*.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84.

Liu, J.-C. (2006). Stationarity of a Markov-switching GARCH model. *Journal of Financial Econometrics*, 4(4):573–593.

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(3-4):271–300.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, 11(4):1833–1874.

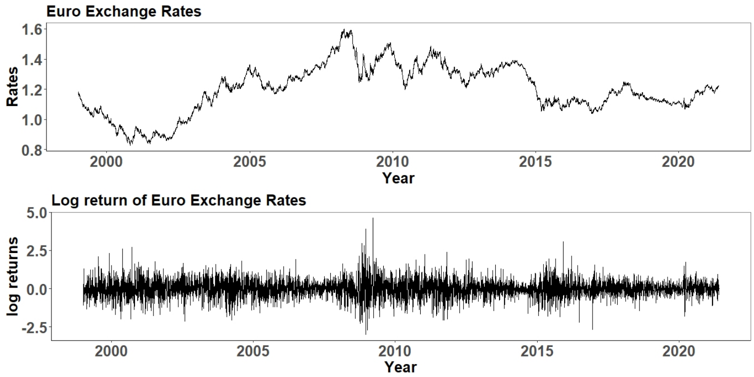
Paolella, M. S., Haas, M., Bauwens, L., Hafner, C. M., and Laurent, S. (2012). Mixture and regime-switching GARCH models. *Wiley Handbooks in Financial Engineering and Econometrics*, (3):71–102.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133.

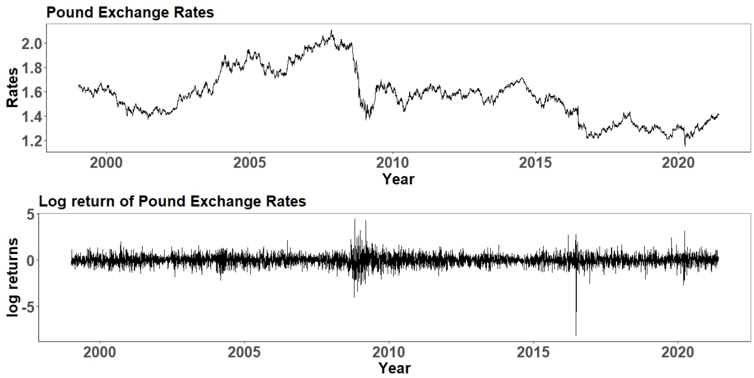
Taylor, J. W. (2020). Forecast combinations for value at risk and expected shortfall. *International Journal of Forecasting*, 36(2):428–441.

Table 1: Bias and RMSE, for the estimators of the MS(2)-GARCH(1,1) model with parameters $\alpha_0^{(1)} = 2$, $\alpha_1^{(1)} = 0.1$, $\beta_1^{(1)} = 0.6$, $\alpha_0^{(2)} = 0.3$, $\alpha_1^{(2)} = 0.35$, $\beta_1^{(2)} = 0.2$, $p_{11} = 0.96$ and $p_{22} = 0.98$. per$^{(i)}$ is the persistence of the i regime, $^{(1)}$ represents the high volatility regime and $^{(2)}$ represents the bearish regime volatility and $T = 3000$. In bold are those that presented the best results.

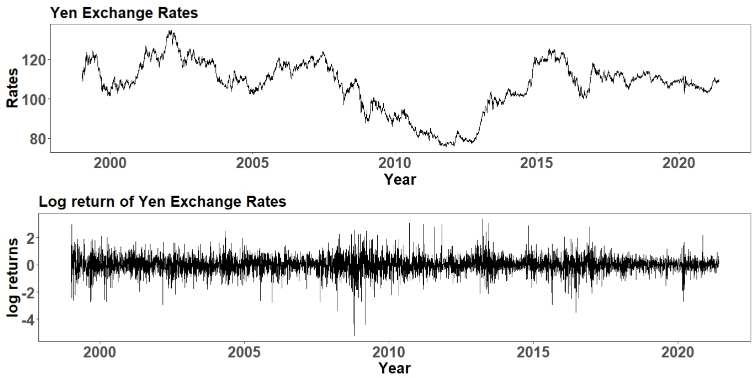
		Estim.	$\alpha_0^{(1)}$	$\alpha_1^{(1)}$	$\beta_1^{(1)}$	$\alpha_0^{(2)}$	$\alpha_1^{(2)}$	$\beta_1^{(2)}$	p_{11}	p_{22}	per.$^{(1)}$	per.$^{(2)}$
$\varepsilon = 0\%$	Bias	QML-N	0.247	**0.002**	-0.045	**0.000**	**-0.002**	**0.003**	**-0.001**	**0.000**	-0.042	**0.001**
		QML-t	0.789	0.051	-0.032	0.001	0.197	0.172	0.010	0.007	**0.019**	0.369
		Rob	**0.018**	0.043	**-0.012**	-0.041	0.146	0.082	-0.017	0.012	0.030	0.228
	RMSE	QML-N	1.231	**0.049**	0.213	**0.041**	**0.056**	**0.096**	**0.008**	**0.005**	0.190	**0.078**
		QML-t	2.110	0.132	0.244	0.067	0.228	0.220	0.013	0.030	0.217	0.376
		Rob	**0.960**	0.082	**0.172**	0.064	0.163	0.118	0.020	0.013	**0.144**	0.241
$\varepsilon = 1\%$ d = 3	Bias	QML-N	0.197	**0.002**	-0.021	0.036	**-0.050**	**0.015**	-0.014	-0.011	-0.019	**-0.036**
		QML-t	1.111	0.013	-0.028	0.077	0.163	0.114	**0.008**	**0.006**	**-0.015**	0.277
		Rob	**0.072**	0.040	**-0.003**	**-0.026**	0.143	0.083	-0.018	0.012	0.037	0.226
	RMSE	QML-N	1.737	**0.081**	0.272	0.074	**0.110**	0.219	0.019	0.015	0.229	**0.156**
		QML-t	2.511	0.112	0.262	0.108	0.204	0.193	**0.012**	**0.008**	0.232	0.293
		Rob	**1.060**	0.083	**0.178**	**0.058**	0.163	**0.125**	0.021	0.013	**0.149**	0.241
$\varepsilon = 1\%$ d = 5	Bias	QML-N	0.347	**0.001**	0.037	**-0.012**	**-0.059**	0.100	-0.075	-0.024	**0.038**	**0.042**
		QML-t	1.796	-0.014	-0.101	0.146	0.130	**0.053**	**0.006**	**0.002**	-0.115	0.183
		Rob	**-0.020**	0.045	**0.007**	-0.023	0.140	0.080	-0.019	0.012	0.051	0.220
	RMSE	QML-N	2.724	0.166	0.293	0.094	**0.142**	0.313	0.184	0.029	0.208	**0.203**
		QML-t	3.270	0.139	0.314	0.177	0.204	0.229	**0.011**	**0.007**	0.297	0.223
		Rob	**1.008**	**0.088**	**0.172**	**0.057**	0.160	**0.121**	0.022	0.013	**0.149**	0.235
$\varepsilon = 5\%$ d = 3	Bias	QML-N	-0.863	0.069	0.222	-0.206	-0.285	0.518	-0.309	-0.122	0.291	0.233
		QML-t	1.482	**-0.008**	0.066	0.309	**-0.112**	0.100	**0.001**	**-0.005**	-0.073	**-0.012**
		Rob	**0.450**	0.037	**0.010**	**0.044**	0.122	**0.091**	-0.022	0.012	**0.048**	0.213
	RMSE	QML-N	**1.058**	0.102	0.231	0.211	0.288	0.528	0.340	0.124	0.295	0.246
		QML-t	3.316	0.104	0.309	0.364	0.185	0.332	**0.011**	**0.012**	0.270	**0.230**
		Rob	1.463	**0.098**	**0.193**	**0.085**	**0.159**	**0.166**	0.026	0.014	**0.159**	0.234
$\varepsilon = 5\%$ d = 5	Bias	QML-N	2.868	0.099	0.191	-0.191	-0.279	0.513	-0.559	-0.116	0.290	0.234
		QML-t	**-0.099**	-0.080	**0.132**	0.029	-0.340	0.516	**-0.035**	-0.036	**0.052**	**0.176**
		Rob	-0.583	**0.024**	0.138	**0.010**	**0.056**	**0.168**	-0.093	**-0.007**	0.162	0.224
	RMSE	QML-N	5.481	0.169	**0.234**	0.199	0.289	0.535	0.585	0.116	0.290	**0.248**
		QML-t	2.497	**0.086**	0.318	0.244	0.342	0.589	**0.044**	0.039	0.287	0.327
		Rob	**1.398**	0.126	0.264	**0.124**	**0.149**	**0.283**	0.180	**0.027**	**0.215**	0.271
$\varepsilon = 10\%$ d = 3	Bias	QML-N	-0.508	0.099	0.195	-0.227	-0.312	0.552	-0.413	-0.221	0.295	0.240
		QML-t	**0.440**	-0.028	0.097	**0.155**	-0.319	0.422	-0.033	-0.043	0.069	**0.103**
		Rob	0.884	**0.023**	**0.041**	0.160	**0.051**	**0.118**	**-0.028**	**0.008**	**0.063**	0.169
	RMSE	QML-N	**0.658**	0.107	**0.199**	0.229	0.312	0.555	0.419	0.222	0.295	0.245
		QML-t	3.224	**0.055**	0.301	0.418	0.322	0.534	0.049	0.061	0.298	0.321
		Rob	2.304	0.112	0.242	**0.206**	**0.137**	**0.236**	**0.036**	**0.016**	**0.189**	**0.214**
$\varepsilon = 10\%$ d = 5	Bias	QML-N	2.761	0.081	0.211	-0.227	-0.333	0.606	-0.576	-0.188	0.292	0.273
		QML-t	**0.166**	**0.013**	0.275	-0.205	-0.333	0.625	**-0.367**	-0.172	**0.288**	**0.292**
		Rob	1.946	0.085	**0.205**	**-0.184**	**0.112**	**0.314**	-0.630	**-0.148**	0.290	0.426
	RMSE	QML-N	3.023	0.090	**0.215**	0.228	0.333	0.607	0.578	0.188	0.292	**0.275**
		QML-t	**0.950**	**0.036**	0.277	0.207	0.333	0.626	**0.379**	0.173	**0.289**	0.294
		Rob	2.974	0.136	0.231	**0.186**	**0.129**	**0.321**	0.634	**0.149**	0.290	0.427



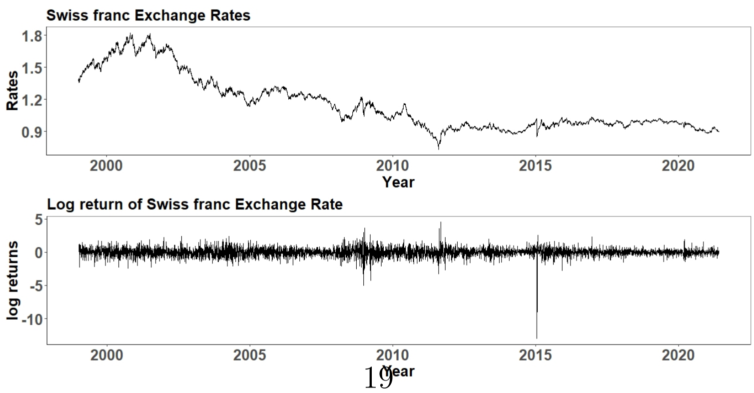
(a)



(b)

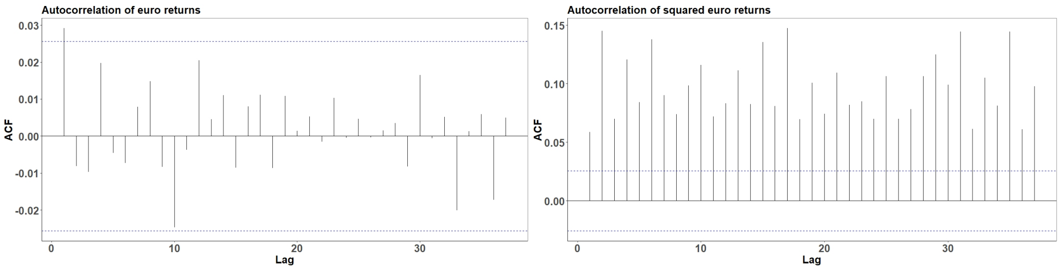


(c)

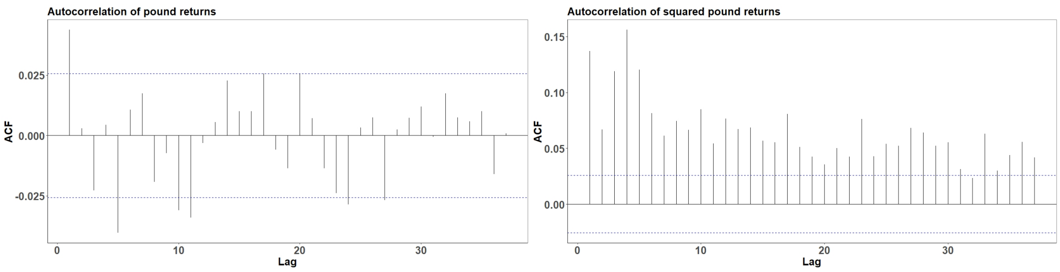


(d)

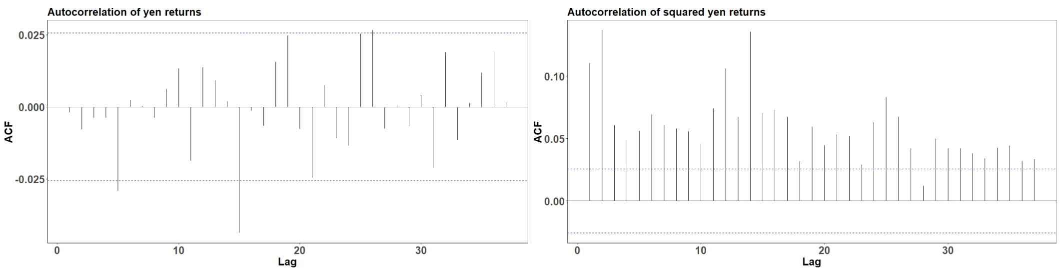
Figure 1: Daily evolution of exchange rates against the US dollar in the upper and lower charts the respective log-returns (a) euro, (b) pound, (c) yen, and (d) Swiss franc.



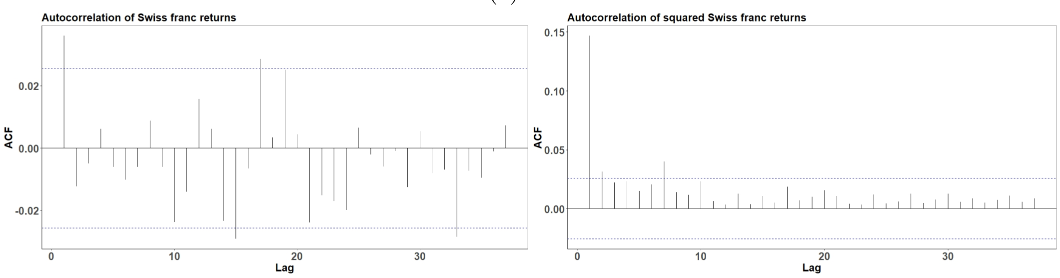
(a) Euro



(b) Pound



(c) Yen



(d) Swiss Franc

Figure 2: On the left side, the sample auto-correlation function of log-returns and on the right side, the sample auto-correlation function of thr squared log-returns.

Table 2: Summary of the exchange rates series.

Currency	Min	q_{25}	q_{50}	Mean	q_{75}	Max	Variance	Asymmetry	Kurtose
Euro	-3.00	-0.31	0.00	0.00	0.31	4.62	0.33	0.15	2.69
Pound	-8.17	-0.31	0.00	-0.00	0.32	4.43	0.32	-0.56	10.79
Yen	-5.22	-0.31	0.00	-0.00	0.33	3.34	0.37	-0.31	4.50
Swiss Franc	-13.02	-0.33	0.00	-0.01	0.32	4.54	0.41	-1.57	33.01

Table 3: Estimates of the parameters of the MS(2)-GARCH(1,1) model for the log-return series of the exchange rate of the euro against the US dollar using the normal and t-Student quasi-maximum likelihood estimators and the robust estimator,

	First window			Last window		
	Rob	QML-n	QML-t	Rob	QML-n	QML-t
$\alpha_0^{(1)}$	0.0014	0.0001	0.0047	0.0133	0.0278	0.0096
$\alpha_1^{(1)}$	0.0459	0.0018	0.0344	0.0774	0.0432	0.0503
$\beta_1^{(1)}$	0.9511	0.9911	0.9608	0.8939	0.9485	0.9377
p_{11}	0.8687	0.0792	0.6738	0.9567	0.0059	0.9926
$\pi_{1\infty}$	0.3490	0.1977	0.4118	0.0546	0.0854	0.3729
$\alpha_0^{(2)}$	0.0033	0.0032	0.0047	0.0001	0.0006	0.0001
$\alpha_1^{(2)}$	0.0513	0.0328	0.0344	0.0308	0.0323	0.0503
$\beta_1^{(2)}$	0.9416	0.9636	0.9608	0.9689	0.9575	0.9377
p_{22}	0.9296	0.7731	0.7716	0.9975	0.9072	0.9956
$\pi_{2\infty}$	0.6510	0.8023	0.5882	0.9454	0.9146	0.6271

Table 4: Parameter estimates of the MS(2)-GARCH(1,1) model for the log-return series of the exchange rate of the pound against the US dollar using the normal and t-Student quasi-maximum likelihood estimators and the robust estimator,

	First window			Last window		
	Rob	QML-n	QML-t	Rob	QML-n	QML-t
$\alpha_0^{(1)}$	0.0173	0.0269	0.0058	0.0167	0.0650	0.0634
$\alpha_1^{(1)}$	0.0564	0.1095	0.0484	0.3000	0.2379	0.0825
$\beta_1^{(1)}$	0.8794	0.8772	0.9458	0.6948	0.7524	0.8098
p_{11}	0.9625	0.2603	0.6340	0.6165	0.1544	0.9939
$\pi_{1\infty}$	0.4993	0.2677	0.0193	0.0420	0.2612	0.2824
$\alpha_0^{(2)}$	0.0014	0.0004	0.0036	0.0013	0.0010	0.0019
$\alpha_1^{(2)}$	0.0438	0.0232	0.0359	0.0396	0.0219	0.0290
$\beta_1^{(2)}$	0.9507	0.9680	0.9593	0.9550	0.9614	0.9687
p_{22}	0.9626	0.7296	0.9928	0.9953	0.7010	0.9976
$\pi_{2\infty}$	0.5007	0.7323	0.9807	0.9580	0.7388	0.7176

Table 5: Parameter estimates of the MS(2)-GARCH(1,1) model for the log-return series of the yen exchange rate against the US dollar using the normal and t-Student quasi-maximum likelihood estimators and the robust estimator,

	First window			Last First window		
	Rob	QML-n	QML-t	Rob	QML-n	QML-t
$\alpha_0^{(1)}$	0.1199	0.0482	0.1029	0.0767	0.0533	0.0008
$\alpha_1^{(1)}$	0.0014	0.0658	0.0980	0.1096	0.0747	0.9990
$\beta_1^{(1)}$	0.6241	0.9284	0.6875	0.7811	0.9149	0.0086
p_{11}	0.9783	0.1061	0.3105	0.9132	0.2778	0.3290
$\pi_{1\infty}$	0.2386	0.1743	0.1249	0.0345	0.1531	0.0722
$\alpha_0^{(2)}$	0.0012	0.0030	0.0037	0.0004	0.0027	0.0031
$\alpha_1^{(2)}$	0.0576	0.0222	0.0297	0.0431	0.0327	0.0498
$\beta_1^{(2)}$	0.9415	0.9555	0.9675	0.9561	0.9305	0.9478
p_{22}	0.9932	0.8257	0.9016	0.9969	0.8694	0.9478
$\pi_{2\infty}$	0.7614	0.9190	0.8751	0.9655	0.8469	0.9278

Table 6: Parameter estimates of the MS(2)-GARCH(1,1) model for the log-return series of the Swiss franc exchange rate against the US dollar using the normal and t-Student quasi-maximum likelihood estimators and the robust estimator,

	First window			Last window		
	Rob	QML-n	QML-t	Rob	QML-n	QML-t
$\alpha_0^{(1)}$	0.0614	0.0000	0.0057	0.0603	0.5053	0.1845
$\alpha_1^{(1)}$	0.0679	0.0025	0.0304	0.0440	0.0000	0.0977
$\beta_1^{(1)}$	0.7529	0.9888	0.9639	0.9494	0.9389	0.6412
p_{11}	0.6276	0.1539	0.5201	0.7920	0.0675	0.9946
$\pi_{1\infty}$	0.0367	0.2369	0.2981	0.0095	0.0117	0.2703
$\alpha_0^{(2)}$	0.0019	0.0050	0.0057	0.0009	0.0015	0.0014
$\alpha_1^{(2)}$	0.0330	0.0310	0.0304	0.0491	0.0291	0.0293
$\beta_1^{(2)}$	0.9628	0.9647	0.9640	0.9486	0.9602	0.0086
p_{22}	0.9858	0.7374	0.7962	0.9980	0.9890	0.9980
$\pi_{2\infty}$	0.9633	0.7631	0.7019	0.9905	0.9883	0.7297