Noname manuscript No.
(will be inserted by the editor)

Analysis of coverage and structural change tests in financial data

Wilton Bernardino [1] · **Raydonal Ospina**[2] · **Yuri Santos**[1]

Abstract Detecting structural changes in time series financial data provides important information to identify breakpoints over time, improving the risk measurement for future crash events. This paper explores structural change and coverage testing focused on Value-at-Risk (VaR) models. Quantile autoregression (QAR) is our framework for modeling the time series since it is quite flexible and robust enough to fit more general changes in the tail. As a result, we highlight how an empirical analysis of exchange-traded fund returns indicates instability in extreme returns during the subprime crisis and how methods of change-point testing are the keys to detecting and quantifying these effects. Simulations based on QAR models performed under different scenarios with and without structural breaks indicate that the structural change tests can identify change points. Also, the results indicated that coverage tests tend to be more conservative.

JEL Classification: G11, G12, G17.

Keywords Coverage test · Quantile autoregression · Subprime crisis · Structural change · Value-at-Risk

1 Introduction

Time series of financial data are complex and involve different risk characteristics depending on the asset, such as stocks, bonds, currencies, and others. Financial institutions need to measure reserve levels to cover solvency and counterparty risks of their operations once the return distributions, as measured by volatility, skewness, kurtosis, and empirical quantiles, are different from traditional assets and generate a higher

[2] Department of Statistics, CAST - Computational Agriculture Statistics Laboratory, Federal University of Pernambuco,, Cidade Universitária, Recife/PE, 50740–540, Brasil. E-mail: yurisantos.hm@gmail.com, raydonal@de.ufpe.br. *Correspondence to*: raydonal@de.ufpe.br

[1] Department of Accounting and Actuarial Sciences, Federal University of Pernambuco, Cidade Universitária, Recife/PE, 50740–540, Brasil. E-mail: wilton.silva@ufpe.br

level of uncertainty. One of the most relevant indicators of evaluation and measurement of financial data risk is given by the Value-at-Risk (VaR), which is the minimum expected loss for a given time horizon and confidence level, taking into account a set period of time that allows obtaining more precise data on the loss that a company, independent of the sector and segment it serves, may be able to support without affecting its normal operations. Given the importance of VaR, it is necessary to adopt reliable methodologies for estimating VaR. However, most of the parametric methodologies used for its calculation have some limitations imposed on the distribution of returns on financial data.

Quantile regression is an estimation method that allows dealing with such restrictions. This approach allows calculating the conditional quantile directly, without imposing assumptions about the error distribution associated with the model. In addition, their estimators are robust to outliers. Additionally, a specification of conditional quantile may include an autoregressive structure to embedding the volatility, that is very present in the financial series. From Koenker and Zhao (1996) pioneering work, which developed the quantile autoregressive model, there have been important additions to the modeling and prediction of volatility associated with financial time series (Koenker and Xiao, 2006; Xiao and Koenker, 2009). The results indicate that they have a good performance for cases in which the distributions have heavy tails and then this approach is an interesting way to explore VaR.

On the other hand, another topic connected to VaR is the structural break analysis, which seems to be an interesting approach when analyzing extremal quantiles once neglected parameter changes on the level and persistence of volatility can generate potential confusion of structure in the series evidencing for example erroneously non-stationarity of data (Hwang et al., 2004; Stărică and Granger, 2005; Andreou and Ghysels, 2002). From a practical point of view, breaks can be associated with financial crisis events, as is the case of the empirical evidence of highly volatile with possible jumps in the subprime mortgage crisis Demyanyk and Van Hemert (2011); Guidolin and Tam (2013). In this sense, seems to be useful that the structural change investigation by using statistical procedures as (Qu, 2008) can be based on quantile autoregression structure under the VaR since it is a statistical measure associated with extreme quantiles.

In this work, we propose to use quantile autoregression in estimating VaR. Here, we consider four important Exchange Traded Funds (ETFs) time series data set, namely, SPDR S&P 500 (SPY), Vanguard Information Technology (VGT), Industrial Select Sector SPDR (XLI), and Consumer Staples Select Sector SPDR (XLP) that were present in the period of the subprime mortgage crisis. By using computer simulations to evaluate the nominal level of Christoffersen (1998) and Candelon et al. (2010) coverage tests as well as the power of the Qu's test (Qu, 2008) under several structural break scenarios we analyze the target financial data set.

2 Background

In finance, a structural break happens when the analyzed time series abruptly changes its regime at a certain point over time. This behavior can involve changes in the mean or other parameters associated with the model used to describe the data-generating process. In this sense, testing hypotheses of structural break regard, identifying and monitoring adequately the variability with evidence that the parameters are or are not affected. The analysis of structural breaks can be found in Page (1954, 1955); Aue and Horváth (2013); Jeng (2015); Dette and Wied (2016); Song and Kang (2018) and Truong et al. (2019); Casini and Perron (2019) for a recent review. The literature that studies structural breaks based on financial time series is abundant. Among several studies, we can cite Granger and Hyung (2004), which analyze how occasional break modeling incorporates incremental information to the future volatility on stocks. Babikir et al. (2012) study the relevance of considering structural breaks in forecasting stock returns volatility. Jung and Maderitsch (2014) focus on understanding the volatility transmission between three stock markets by considering a period of analysis that incorporates various crisis events. Hamilton (1990) argue that many of the major exogenous economic events that influence financial series are shocks such as the doubling of oil prices experienced over the past decades. These events can be considered as episodes (breaks) with an identifiable duration in which the response of economic series might be expected to have a noteworthy difference from that seen outside these periods.

Several studies on structural changes in financial data consider the breaks only in the conditional mean. However, structural change detection in conditional quantiles seems to be more suitable (Wolters and Tillmann, 2015; Zhou et al., 2015; Bonaccolto et al., 2018). In this sense, Qu (2008); Su and Xiao (2008) investigate structural change tests on conditional quantiles. Qu (2008) adopted the methodology of quantile regression based on the CUSUM tests (Page, 1954) and proposed two types of test statistics for structural change occurring in a specific quantile. The tests are based on sequentially weighted empirical subgradient (Assaf and Ritov, 1988) because it has excellent size properties even in small samples, local power, and only requires estimating the model under the null hypothesis. The QU's test based on subgradient (Qu, 2008) is attractive by considering conditional heteroskedasticity models under the quantile regression framework. Here, the heteroskedasticity patterns (volatility presented in many financial time series) are explored via the conditional variance linked to a regression structure to model quantiles by considering explanatory variables Koenker and Bassett (1978).

Autoregressive Conditional Heteroskedasticity (ARCH) models was projected by Engle (1982) to measure the conditional variability or volatility that includes the lagged value and gives higher weights to recent past observations than others. To allow a longer memory and a more flexible lag structure Bollerslev (1986) introduced the generalized autoregressive conditional heteroskedasticity (GARCH) models. Variations of GARCH include GARCH in Mean (GARCH-M) Engle et al. (1987), Exponential

GARCH (EGARCH) (Nelson, 1991), Integrated GARCH (IGARCH) (Engle and Bollerslev, 1986), Glosten Jagnnathon Runkle GARCH (GJR-GARCH) (Glosten et al., 1993), Quadratic GARCH (QGARCH) (Engle and Ng, 1993) , Asymmetric GARCH (AGARCH) (Nelson, 1991; Xiao and Aydemir, 2007; Miron and Tudor, 2010), among others

However, all these models are deficient in describing nonlinearities in the data produced by general forms of heterogeneity when different regimes coincide in the conditional mean process, a behavior generally observed in financial time series. In this direction, quantile regression Koenker and Zhao (1996) and autoregressive (QAR) model Koenker and Xiao (2006) has attracted considerable attention by allowing to allocate in flexible form the existence of different regimes depending on the quantile of the series to be modeled (Engle and Manganelli, 2004; Xiao and Koenker, 2009; Baur et al., 2012).

Forecasting is, therefore, crucial in many areas of finance, such as option pricing, Value-at-Risk applications, and portfolio selection. The Value-at-Risk (VaR) (Jorion, 2000; Tsay, 2005) is possible to relate to quantile regression once VaR can be seen as an extremal quantile which can be estimated by using quantile regression techniques. A Value-at-Risk proposes to show the worst expected loss over a given trading horizon at a given confidence. For example, a financial institution could say that the daily VaR of some trading portfolio is \$2 million with 99% of confidence. In this case, there is a 1% probability of losing more than \$2 million under stable market conditions. For relevant related work see (Lee and Noh, 2013; Xiao et al., 2015; Haugom et al., 2016; Hagfors et al., 2016; Taylor, 2019; Christou and Grabchak, 2019). The modeling strategy of Var via QR or QAR is also justified by the simplicity of implementing the quantile regression and by the usefulness of the method to deal with asymmetric dynamics and local persistence when modeling financial time series Xiao et al. (2015); Haugom et al. (2016); Xu et al. (2016); Kuck and Maderitsch (2019); Christou and Grabchak (2019). Finally, for financial institutions and their regulators is of vital importance to evaluate the accuracy of VaR estimation. In this sense, several backtesting procedures (Philippe, 2001) are available, which include the likelihood ratio (LR) test of Kupiec (1995), the independence and conditional coverage test (Christoffersen, 1998) and the duration-based approach of (Candelon et al., 2010).

The remainder of the paper is organized as follows: In Section 2, we describe our approach to benchmarking systems. In Section 3, we briefly discuss LCDA. Section 4 reports the results of our computational study. Some conclusions follow in Section 5.

3 Models and methods applied to analyze Exchanged Traded Funds

QAR models to estimate VaR: Let $\{\varepsilon_t\}$ (a returns series) be a stochastic process such that $\varepsilon_t = \sigma_t v_t = \left(\alpha_0 + \sum_{i=1}^{p} \alpha_i |\varepsilon_{t-i}|\right) v_t$, where $\alpha_0 \in (0, \infty), \alpha_i \geq 0$, $i = 1, \ldots, p$, $\{v_t\}$ are i.i.d. random variables with distribution function F_{v_t}. Given the information set $\mathcal{F}_{t-1}$ representing observed returns, the linear

ARCH formulation for the conditional quantile of the returns can be written as (Koenker and Zhao, 1996)

$$Q_{\epsilon_t}(\tau|\mathcal{F}_{t-1}) = \left(\alpha_0 + \sum_{i=1}^{p} \alpha_i |\varepsilon_{t-i}|\right) F_{v_t}^{-1}(\tau), \tag{1}$$

where $F_{v_t}^{-1}(\tau)$ denotes the τ-th quantile of the innovations $\{v_t\}$. Koenker and Zhao (1996) showed that, $r \in (1, \infty)$, $\mu_r = \mathbb{E}(|v_t|^r)^{1/r} < \infty$, and the polynomial $\phi(x) = x^p - \mu_r(\alpha_1 x^{p-1} + \ldots + \alpha_{p-1}x + \alpha_p)$ having all roots inside of the unit circle are sufficient conditions to the stationarity of $\{\varepsilon_t\}$. Furthermore, with these conditions the process is ergodic and $\mathbb{E}(|\varepsilon_t|^r) < \infty$, where $\mathbb{E}(\cdot)$ denotes the expected value. The Equation (3) can be rewritten as

$$Q_{\epsilon_t^*}(\tau|\mathcal{F}_{t-1}) = \left(1 + \sum_{i=1}^{p} \gamma_i |\varepsilon_{t-i}|\right) F_{v_t^*}^{-1}(\tau), \tag{2}$$

where $\gamma_i = \frac{\alpha_i}{\alpha_0}$, $\sqrt{\omega_0} = \alpha_0$ are the new coefficients, $\{v_t^* = \sqrt{\omega_0} v_t\}$ represents the modified innovation having quantile function $F_{v_t^*}^{-1}$. A more general discussion about the class of QAR models can be seen in Koenker and Xiao (2006). By definition, considering $\tau \in (0, 1)$ (significance level), VaR (VaR_τ^t) corresponding to the returns $\{\varepsilon_t\}$ satisfies the equation $\Pr(\varepsilon_t \leq \mathrm{VaR}_\tau^t|\mathcal{F}_{t-1}) = \tau$, where $\Pr(A|B)$ denotes the conditional probability of A given B. As consequence, based on the information set available up to $(t-1)$-period, and using the Equation (2), we can consider the equality $Q_{\epsilon_t^*}(\tau|\mathcal{F}_{t-1}) = \mathrm{VaR}_\tau^t$.

Supose that $Q_{\epsilon^*}(\tau|\mathcal{F}) = Q_{\epsilon^*}(\tau|\mathcal{X})$ is a linear funtion of parameters, i.e., $Q_{\epsilon^*}(\tau|\mathcal{X}) = \mathcal{X}'\beta_\tau$, where $\beta_\tau = \beta F_{v^*}^{-1}(\tau) = (1, \gamma_1, \ldots, \gamma_p)' F_{v^*}^{-1}(\tau)$. The estimation of QAR models in Equation (2) can be made by solving the problem (see Koenker and Bassett (1978))

$$\hat{\beta}_\tau = \mathrm{argmin}_{b \in \mathbb{R}^p} \left\{ \sum_{t=1}^{n} \rho_\tau(\varepsilon_t - \mathcal{X}_t' b) \right\}, \tag{3}$$

where $\mathcal{X}_t = (1, |\varepsilon_{t-1}|, \ldots, |\varepsilon_{t-p}|)'$, $\{(\varepsilon_t, \mathcal{X}_t),\, t = 1, \ldots, n\}$ denotes a sample of size n, and $\rho_\tau(v) = v(\tau - I(v < 0))$ is the check function.

Qu's test for structural changes detection: The Qu's test (Qu, 2008) seeks to determine whether the coefficients of a linear quantile regression remain the same over time. For the accomplishment of this task he proposed the following inferential procedure: (1) consider the random variable $\{\varepsilon_{t_i}\}$, where the subscript indicates that process is observed in t_i-period; (2) suppose that the τ-th conditional quantile of $\{\varepsilon_{t_i}\}$ can be written as a linear function $Q_{\varepsilon_{t_i}}(\tau|\mathcal{X}_{t_i}) = \mathcal{X}_{t_i}'\beta_{t_i\tau}$, where $\mathcal{X}_{t_i} = (1, |\varepsilon_{t_i-1}|, \ldots, |\varepsilon_{t_i-p}|)'$ is the vector of explanatory variables; (3) the inferential procedure proposes to test $\mathcal{H}_0 : \beta_{t_i\tau} = \beta_{0\tau}$ for all i against $\mathcal{H}_1 : \beta_{t_i\tau} = \begin{cases} \beta_{1\tau} & \text{for } t_i \in \{t_1, t_2, \ldots, t_k\} \\ \tilde{\beta}_{1\tau} & \text{for } t_i \in \{t_{k+1}, t_{k+2}, \ldots, t_n\} \end{cases}$, where t_k denotes the breakpoint; (4) since the objective function in Equation (3) is not everywhere differentiable, the test statistic is based on the subgradient and it is computed using a sorted subsample from the beginning up to $t_k = \lfloor \lambda n \rfloor,\ \lambda \in [0, 1]$, where $\lfloor . \rfloor$ denotes a floor function. The subgradient test statistic proposed by Qu (2008) is based on the random variable

$$S_n(\lambda, \tau, \vartheta) = n^{1/2} \sum_{t=t_1}^{\lfloor \lambda n \rfloor} \mathcal{X}_t \psi_\tau(\varepsilon_t - \mathcal{X}_t' \vartheta), \tag{4}$$

where $\vartheta \in \mathbb{R}^p$ is some estimates for β_τ and $\psi_\tau(u) = 1(u \leq 0) - \tau$.

Under the null hypothesis, $\psi_\tau(\varepsilon_i - \mathcal{X}_{t_i}' \beta_{0\tau})$, $i = 1, \dots, \lfloor \lambda n \rfloor$, in Equation (4) is a sequence of independent binary random variables with mean zero and variance $\tau(1-\tau)$. In this sense, it can be considered as a pivot quantity to make decisions about the rejection/nonrejection of the null hypothesis. By considering $X = (x_{t_1}', \dots, x_{t_n}')'$ and $H_{\lambda,n}(\beta_{0\tau}) = (n^{-1} X' X)^{-1/2} S_n(\lambda, \tau, \beta_{0\tau})$, Qu (2008) concludes that, under some regularity conditions, $H_{\lambda,n}(\beta_{0\tau})$ is such that $H_{\lambda,n}(\theta_{0\tau}) \xrightarrow{d} \mathcal{N}(0, \lambda^2 \tau(1-\tau))$. Replacing $\beta_{0\tau}$ with the full sample quantile regression estimate obtained under null hypothesis $(\widehat{\beta}_{0\tau})$ we can write

$$H_{\lambda,n}(\widehat{\beta}_{0\tau}) = (X'X)^{-1/2} \sum_{t=t_1}^{\lfloor \lambda n \rfloor} \mathcal{X}_t \psi_\tau(\varepsilon_t - \mathcal{X}_t' \widehat{\beta}_{0\tau}). \tag{5}$$

Qu (2008) showed that, under the null hypothesis, $H_{\lambda,n}(\widehat{\beta}_{0\tau})$ converges to a nondegenerate distribution, whereas under the alternative hypothesis, it diverges for some λ. Since the true breakpoint is unknown, it is necessary to investigate all the possibilities. In addition, Qu (2008) concludes that, if we re-center $H_{\lambda,n}(\hat{\beta}_{0\tau})$ by $\lambda H_{1,n}(\hat{\beta}_{0\tau})$ it often has a better performance in finite samples. Thus, Qu (2008) defines the alternative test statistic

$$SQ_\tau = \sup_{\lambda \in [0,1]} \left|\left| (\tau(1-\tau))^{-1/2} \left[H_{\lambda,n}(\widehat{\beta}_{0\tau}) - \lambda H_{1,n}(\widehat{\beta}_{0\tau}) \right] \right|\right|_\infty, \tag{6}$$

where $||\cdot||_\infty$ denotes the uniform norm (see, e.g., Rudin (2006)). Under the null hypothesis, the statistic SQ_τ in Equation (6) converges to a p-vector of independent Brownian bridge processes (see, e.g., Revuz and Yor (2013)) on $[0, 1]$. Based on simulations, Qu (2008) obtained critical values for the SQ_τ test under $\mathcal{H}_0$.

Coverage tests: Coverage tests are widely used to evaluate VaR estimates. In this sense, it is not enough to know if a certain VaR model produces a plausible percentage of VaR violations[1], it is necessary to evaluate if the VaR violations are independent of each other as stated by Christoffersen (1998). Tests that simply check the expected rate of violations are called unconditional coverage tests, and the others seeking a more sophisticated approach regarding clusters and violation moments are called conditional coverage tests.

[1] Consider ε_t some return at t-period. A VaR violation occurs if the observed return exceeds the VaR's estimate, i.e., if $\varepsilon_t < \mathrm{VaR}_\tau^t$.

Let $\mathbb{I}_{1-\tau}$ be the indicator function which is equal to 1 when there is a VaR_τ^t violation. The Christoffersen's null hypothesis (independence of VaR violations) is $\mathcal{H}_0 : \Pi = \Pi_\alpha = \begin{bmatrix} \tau & 1-\tau \\ \tau & 1-\tau \end{bmatrix}$. It is tested against $\Pi = \begin{bmatrix} \pi_{01} & 1-\pi_{01} \\ \pi_{11} & 1-\pi_{11} \end{bmatrix}$, where $\pi_{ij} = \Pr(\mathbb{I}_{1-\tau}^t = j | \mathbb{I}_{1-\tau}^{t-1} = i)$. The GMM test (Candelon et al., 2010) is based on an orthonormal polynomial M_{j+1} associated with a geometric distribution with a success probability $s \in (0,1)$. This polynomial M_{j+1} is given ($\forall d \in \mathbb{N}^*$) by

$$M_{j+1}(d;s) = \frac{(1-s)(2j+1) + s(j-d+1)}{(j+1)\sqrt{1-s}} M_j(d;s) - \frac{j}{j+1} M_{j-1}(d;s), \tag{7}$$

where d is the duration between two VaR violations (supposed to be geometrically distributed under the null hypothesis), $j \in \mathbb{N}$, $M_0(d;s) = 1$, and $M_{-1}(d;s) = 0$. Under independence of the durations between VaR violations, If the geometric distribution is the true random pattern of d with a success probability $s = \tau$, then the expected value of $M_j(d;\tau)$ is zero for all j.

Let us denote $\{d_i\}_{i=1}^N$ a sequence of N durations between violations. The null hypothesis of conditional coverage GMM test is $\mathcal{H}_0 : E(M(d_i;\tau)) = 0$, where M denotes a $(p,1)$ vector of M_j for $j = 1, \ldots, p$. Under some regularity conditions, and under the null hypothesis, the GMM's test statistic (J_{cc}) converges to a chi-square distribution as described by

$$J_{cc} = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(d_i;\tau) \right)^{\top} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M(d_i;\tau) \right) \xrightarrow{d} \chi_{(\ell)}, \tag{8}$$

where ℓ is the number of orthornormal polynomials used as moment conditions (see Candelon et al. (2010)).

4 An empirical motivation

4.1 Data

In the initial study, we analyzed four important ETFs, namely, SPDR S & P 500 ETF (SPY), Vanguard Information Technology ETF (VGT), Industrial Select Sector SPDR ETF (XLI), and Consumer Staples Select Sector SPDR ETF (XLP). These ones use different criteria to formulate their portfolios, focusing, respectively, on important stocks listed on NYSE and NASDAQ exchanges, Information Technology, industrial, and consumption sectors.

In the literature focusing on ETFs, empirical studies have been performed in order to assess the efficiency of this kind of investment option. The United States, for example, represents almost 70% of the ETF market in terms of assets under management, although the number of exchange-traded products in the U.S. accounts for only 30% of the world total (Abner, 2013).

4.2 Empirical analysis

The main point of this section is to use the quantile regression methods discussed in Section 3 in order to evaluate the VaR measures corresponding to the four ETFs mentioned in Subsection 4.1. The period of analysis corresponds to the daily returns from 7/03/2007 to 6/29/2009 (502 trading days). This two-year period was chosen to focus on the main extension of the subprime mortgage crisis. In this sense, we focused on analyzing the estimates of 5% and 1% VaRs corresponding to the ETFs. The Augmented Dickey–Fuller (ADF) Dickey and Fuller (1981) and McLeod-Li McLeod and Li (1983) tests were implemented, however it is important to note that the ADF test is biased towards nonrejection of nul hypothesis (unit root) (Perron, 1990). Partial autocorrelation function (PACF) and autocorrelation function (ACF) were plotted for regular and absolute returns. This procedure was done to get insights into the orders of the ARCH processes. Finally, ten fitted models were evaluated. To help on choosing the models, AIC (Akaike, 1974), pseudo R^2 (Koenker et al., 2005) were computed. In addition, GMM's, Christoffersen's, and Qu's tests were implemented.

Serial correlations in return series are not significant because they are not predictable (see, e.g., Issler (1999)). On the other hand, when considering absolute (or squared) returns, it is possible to evaluate the predictability of conditional variance by plotting Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions. The plots in Figures 1 and 2 show only a few significant values for autocorrelations and partial autocorrelations. On the other hand, the plots corresponding to the absolute returns (Figures 3 and 4) show a completely different scenario. These plots reveal some visual evidence of serial correlation in all ETFs, which justifies the ARCH modeling. In addition, after implementing the ADF test, we did not find any evidence of unity roots. As expected, McLeod-Li's test detected a serial correlation between absolute returns in all ETFs returns.

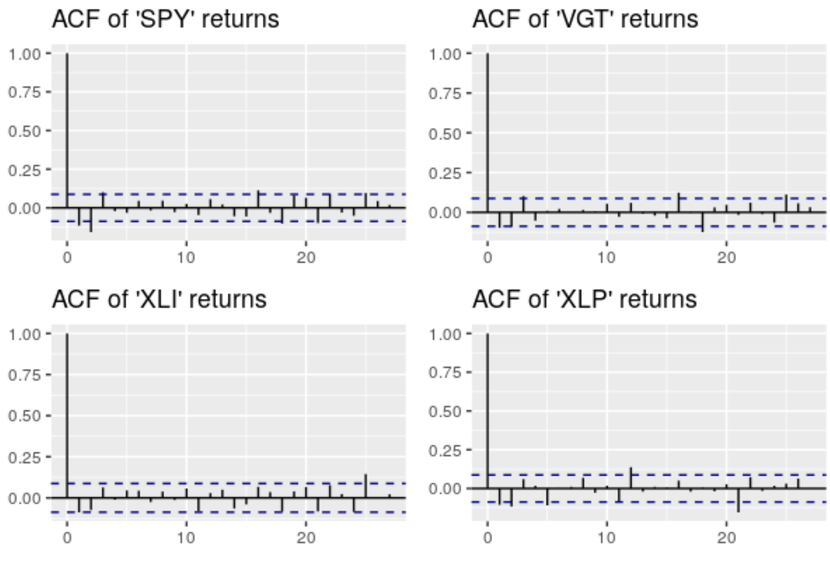

Fig. 1: ACF of returns for all ETFs.

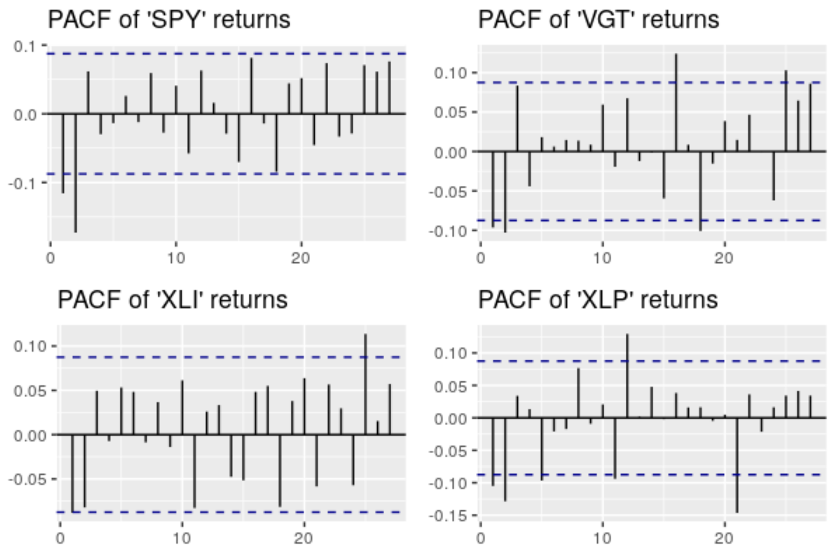

Fig. 2: PACF of returns for all ETFs.

As a second step, focusing on τ-quantiles ($\tau = 1\%, 5\%$), we fitted ten ARCH(p), $p = 1, \ldots, 10$. AIC and pseudo R^2 were computed. The results are shown in Tables 1 and 2. Tables 3 to 6 show the GMM's and Christoffersen's p-values corresponding to $\tau = 1\%$ and 5%.

The model selection started observing the AIC and pseudo R^2 measures. The AIC (Table 1) selects high order for the quantile linear ARCH models ($\tau = 5\%, 1\%$), none lower than six and some, especially when $\tau = 1\%$, reaching $p = 10$. A similar behavior occurs when using pseudo R^2 values, reaching a stable and less increasing state at higher orders. It is important to note that AIC and pseudo R^2 are not sufficient to evaluate VaR violations. This is the reason we did not focus on these measures as the

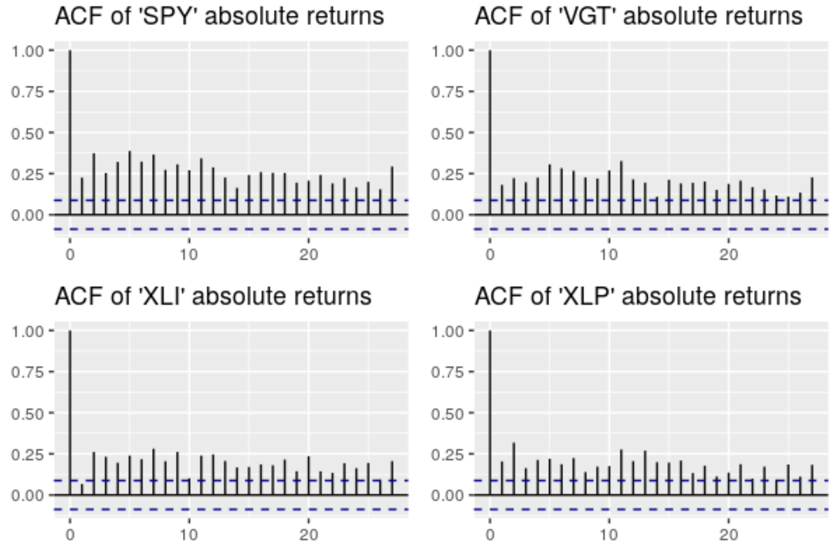

Fig. 3: ACF of absolute returns for all ETFs.

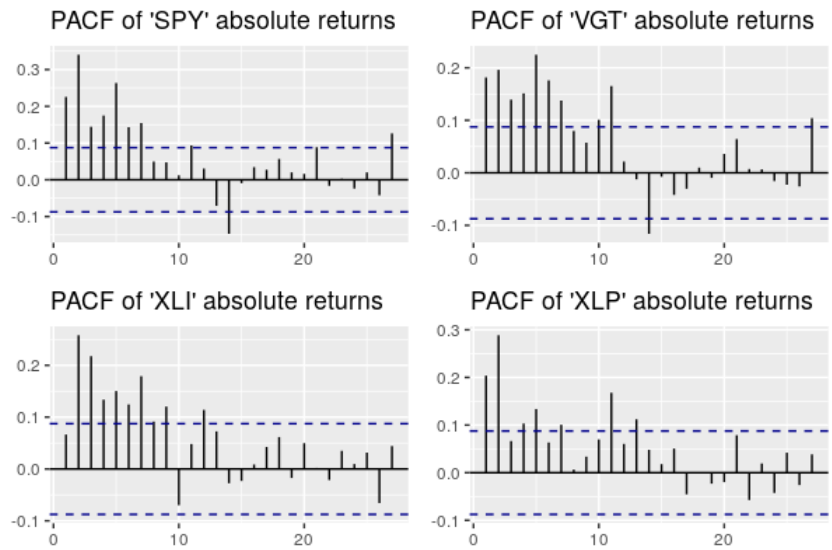

Fig. 4: PACF of absolute returns for all ETFs.

main criteria to qualify the goodness of fit of the VaR estimates. In this sense, we used the coverage tests described in Subsection 3. The results from the Christoffersen's test are shown in Tables 3 and 4. As can be seen, considering a nominal level of 5% and $\tau = 5\%, 1\%$, the null hypotheses were not rejected for almost all ETF data when considering the ARCH order $p = 1$. For $\tau = 5\%$, when $p = 2$ or 3, the null hypotheses were rejected only for SPY (CT_{uc} and CT_{cc}, when $p = 2$, and CT_{cc}, for $p = 3$) and XLI (CT_{uc}, for $p = 3$). Regarding Christoffersen's coverage test, in the next section, we show simulation evidence that Christoffersen's test is a permissive test, and we recommend considering this question in the model selection.

The results from GMM test are shown in Tables 5 and 6. As can be seen, the null hypotheses were rejected in almost all cases (unconditional and conditional tests). An interesting exception occurred when $\tau = 1\%$ and $p = 1$ (Table 6). In this case, the p-values corresponding to XLI and XLP suggest not rejection of the null hypotheses for all coverage tests. It is important to note that the aim of coverage tests is to provide feedback on whether VaR is being well estimated or not. Thus, coverage tests precede other model criteria, such as AIC and pseud-R^2, when fitting VaR models. In our study, lower p-order models needed to be considered.

Table 1: AIC values for the fitted models

ARCH	$\tau = 5\%$				$\tau = 1\%$			
order	SPY	VGT	XLI	XLP	SPY	VGT	XLI	XLP
1	-1887	-1959	-1895	-2434	-1443	-1608	-1552	-1995
2	-1977	-2010	-1950	-2451	-1614	-1691	-1692	-2168
3	-1972	-2012	-1971	-2449	-1631	-1711	-1733	-2169
4	-2063	-2055	-2016	-2458	-1802	-1750	-1797	-2164
5	-2077	-2053	-2027	-2483	-1880	-1787	-1792	-2224
6	-2096	-2088	-2028	-2475	-1918	-1915	-1801	-2226
7	-2120	-2095	-2027	-2479	-1973	-1978	-1863	-2316
8	-2119	-2103	-2021	-2490	-2015	-2020	-1862	-2330
9	-2112	-2096	-2017	-2482	-2011	-2014	-1917	-2336
10	-2111	-2093	-2011	-2483	-2012	-2008	-1924	-2341

Table 2: pseudo R^2 values for the fitted models

ARCH	$\tau = 5\%$				$\tau = 1\%$			
order	SPY	VGT	XLI	XLP	SPY	VGT	XLI	XLP
ARCH	$\tau = 5\%$				$\tau = 1\%$			
order	SPY	VGT	XLI	XLP	SPY	VGT	XLI	XLP
1	0.668	0.696	0.698	0.696	0.893	0.910	0.912	0.902
2	0.699	0.713	0.717	0.704	0.910	0.918	0.924	0.918
3	0.700	0.716	0.725	0.706	0.913	0.920	0.928	0.919
4	0.728	0.730	0.739	0.711	0.927	0.924	0.933	0.919
5	0.734	0.732	0.744	0.721	0.933	0.927	0.933	0.925
6	0.742	0.743	0.746	0.721	0.936	0.937	0.934	0.926
7	0.750	0.747	0.748	0.725	0.940	0.941	0.938	0.933
8	0.752	0.751	0.749	0.730	0.943	0.944	0.939	0.934
9	0.752	0.752	0.750	0.731	0.943	0.944	0.943	0.935
10	0.754	0.753	0.750	0.733	0.944	0.944	0.943	0.936

4.2.1 Graphical analysis using in-sample VaR estimates

In an in-sample estimation strategy, 1% and 5% VaRs were estimated based on quantile linear ARCH(p), $p = 1, \ldots, 3$. The plots corresponding to each ETF are shown in Figures 5 to 8. In each graph, the black line is the returns, the read line represents 1%-quantiles ($\mathrm{VaR}_{1\%}$) and the orange line is the 5%-quantiles

Table 3: p-values of Christoffersen's test ($\tau = 5\%$). CT_{uc} and CT_{cc} denotes unconditional and conditional tests, respectively.

ARCH	SPY		VGT		XLI		XLP	
order	CT_{uc}	CT_{cc}	CT_{uc}	CT_{cc}	CT_{uc}	CT_{cc}	CT_{uc}	CT_{cc}
1	0.081	0.053	0.324	0.062	0.553	0.788	0.429	0.013
2	0.034	0.036	0.235	0.064	0.235	0.493	0.052	0.016
3	0.114	0.058	0.830	0.135	0.033	0.099	0.230	0.064
4	0.001	0.003	0.012	0.020	0.001	0.005	0.226	0.063
5	0.000	0.000	0.019	0.027	0.001	0.003	0.048	0.016
6	0.001	0.002	0.030	0.066	0.011	0.041	0.004	0.002
7	0.001	0.000	0.004	0.005	0.002	0.006	0.018	0.004
8	0.004	0.005	0.001	0.002	0.002	0.006	0.001	0.001
9	0.003	0.001	0.006	0.006	0.010	0.037	0.003	0.004
10	0.003	0.008	0.017	0.003	0.002	0.008	0.000	0.000

Table 4: p-values of Christoffersen's test ($\tau = 1\%$). CT_{uc} and CT_{cc} denotes unconditional and conditional tests, respectively.

ARCH	SPY		VGT		XLI		XLP	
order	CT_{uc}	CT_{cc}	CT_{uc}	CT_{cc}	CT_{uc}	CT_{cc}	CT_{uc}	CT_{cc}
1	0.666	0.142	0.048	0.059	0.329	0.610	0.996	0.951
2	0.048	0.059	0.048	0.059	0.397	0.632	0.008	0.016
3	0.020	0.032	0.008	0.002	0.003	0.008	0.008	0.016
4	0.000	0.000	0.000	0.000	0.000	0.000	0.047	0.058
5	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.001
6	0.000	0.000	0.000	0.000	0.003	0.007	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.001	0.019	0.031
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 5: p-values for GMM test ($\tau = 5\%$). GMM_{uc} denotes the unconditional test. GMM_{cc_3} and GMM_{cc_5} correspond to the conditional tests with $\ell = 3$ and 5, respectively.

ARCH	SPY			VGT		
order	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}
1	0.000	0.000	0.000	0.021	0.000	0.000
2	0.056	0.000	0.000	0.014	0.000	0.000
3	0.090	0.000	0.000	0.094	0.000	0.001
4	0.004	0.000	0.000	0.005	0.000	0.000
5	0.001	0.000	0.000	0.003	0.000	0.000
6	0.006	0.000	0.000	0.004	0.000	0.000
7	0.004	0.000	0.000	0.001	0.000	0.000
8	0.014	0.000	0.000	0.006	0.000	0.000
9	0.014	0.000	0.000	0.001	0.000	0.000
10	0.012	0.000	0.000	0.038	0.000	0.000
ARCH	**XLI**			**XLP**		
order	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}
1	0.009	0.004	0.005	0.027	0.009	0.002
2	0.000	0.000	0.000	0.006	0.002	0.001
3	0.003	0.000	0.000	0.001	0.001	0.005
4	0.000	0.000	0.000	0.001	0.000	0.001
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
7	0.009	0.000	0.000	0.000	0.000	0.000
8	0.009	0.000	0.000	0.000	0.000	0.000
9	0.007	0.000	0.000	0.001	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000

($\mathrm{VaR}_{5\%}$). In addition, a dashed horizontal blue line at 5% represents a daily hypothetical loss limit in an ETF investment. According to Figures 5 to 8, in terms of VaR estimates, SPY seems to be the most aggressive index. In the opposite direction, we can highlight the XLP index, which is, in fact a conservative ETF.

Table 6: p-values for GMM test when $\tau = 1\%$. GMM_{uc} denotes the unconditional test. GMM_{cc_3} and GMM_{cc_5} correspond to the conditional tests with $\ell = 3$ and 5, respectively.

ARCH	SPY			VGT		
order	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}
1	0.040	0.021	0.010	0.006	0.001	0.000
2	0.007	0.002	0.000	0.004	0.000	0.000
3	0.006	0.001	0.000	0.008	0.002	0.000
4	0.000	0.000	0.000	0.002	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.002	0.000	0.000
7	0.000	0.000	0.000	0.003	0.001	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000
ARCH	XLI			XLP		
order	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}	GMM_{uc}	GMM_{cc_3}	GMM_{cc_5}
1	0.447	0.722	0.944	0.080	0.068	0.081
2	0.057	0.057	0.116	0.009	0.004	0.005
3	0.002	0.000	0.000	0.005	0.001	0.000
4	0.000	0.000	0.000	0.009	0.002	0.001
5	0.000	0.000	0.000	0.005	0.002	0.001
6	0.002	0.000	0.000	0.001	0.000	0.000
7	0.001	0.000	0.000	0.016	0.010	0.011
8	0.000	0.000	0.000	0.013	0.005	0.001
9	0.000	0.000	0.000	0.001	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000

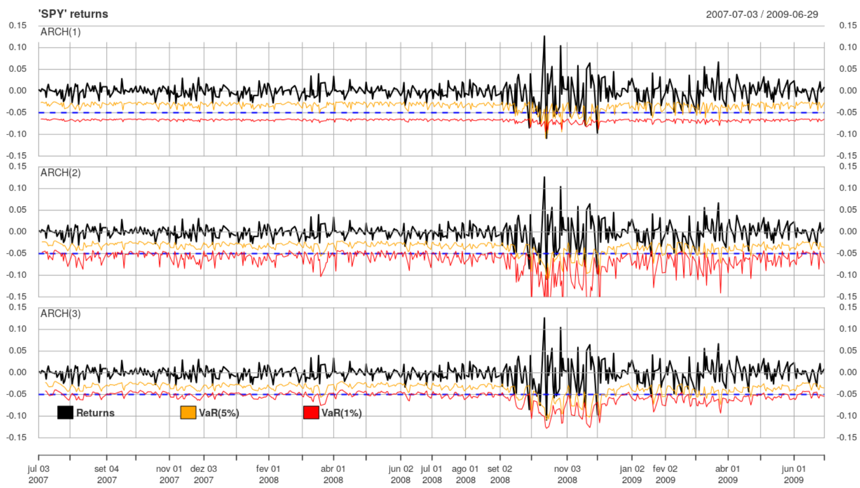

Fig. 5: SPY returns and its $VaR_{5\%}$ and $VaR_{1\%}$ fitted as quantile linear ARCH of order 1 to 3, top-bottom.

4.2.2 Structural change analysis

The Qu's test application results are shown in Tables 7 and 8. At 5% nominal level, critical values, test statistics and the dates of the structural breaks are shown according to the corresponding p-order. As we can see, except for the XLP index, QU's test shows pieces of evidence of breaks at 5%- quantiles ($\mathrm{VaR}_{5\%}$) when considering $p < 5$. For $\tau = 1\%$, all the indices presented breakpoints by considering $p = 1$ (small order) and $p = 9, 10$ (large orders). It is important to note that, for small orders, the significant breakpoints were all close to September 15th, 2008, which represents the date of the Lehman Brothers bankruptcy. Furthermore, for small p orders, the breakpoints suggested to the XLP index were at this

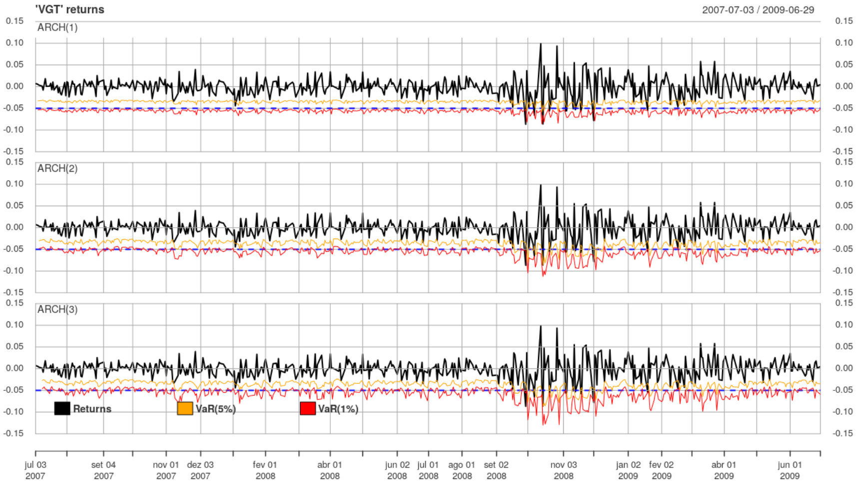

Fig. 6: VGT returns and its $VaR_{5\%}$ and $VaR_{1\%}$ fitted as quantile linear ARCH of order 1 to 3, top-bottom.

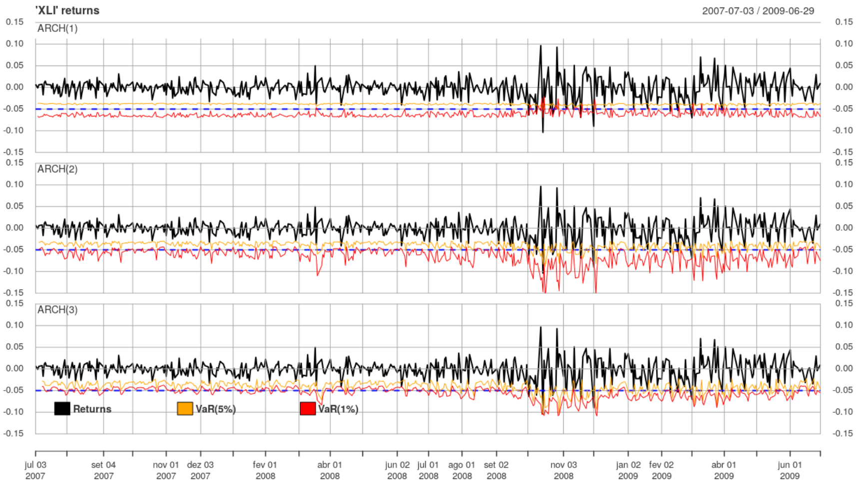

Fig. 7: XLI returns and its $VaR_{5\%}$ and $VaR_{1\%}$ fitted as quantile linear ARCH of order 1 to 3, top-bottom.

Fig. 8: XLP returns and its $VaR_{5\%}$ and $VaR_{1\%}$ fitted as quantile linear ARCH of order 1 to 3, top-bottom.

date or one day after it. We highlight that the XLP is an ETF that delivers a conservative basket of consumer-staples firms. In this sense, the investors rapidly made their decisions after the Lehman Brothers' bankruptcy, which can justify the breakpoints appointed by QU's test in the case of this ETF.

Table 7: Qu's test values for the fitted models at quantile $\tau = 5\%$

ARCH	Crit.	SPY		VGT		XLI		XLP	
order	value	SQ_τ	Break at	SQ_τ	Break at	SQ_τ	Break at	SQ_τ	Break at
1	1.329	2.947	03-09-08	3.140	12-09-08	2.810	12-09-08	2.218	16-09-08
2	1.453	2.239	03-09-08	2.339	25-08-08	2.677	03-09-08	2.136	16-09-08
3	1.517	2.206	03-09-08	2.326	12-09-08	2.134	03-09-08	1.436	16-09-08
4	1.569	2.154	03-09-08	1.796	03-09-08	1.621	23-07-08	1.521	16-09-08
5	1.601	1.763	05-06-08	1.524	03-09-08	1.512	06-08-08	1.484	29-04-08
6	1.628	1.304	03-09-08	1.083	08-07-08	1.623	06-08-08	1.485	29-04-08
7	1.650	0.619	24-12-08	1.264	15-10-08	1.952	03-09-08	1.594	16-09-08
8	1.655	1.296	16-09-08	1.380	26-09-08	2.081	03-09-08	0.889	19-05-08
9	1.684	1.167	08-09-08	1.078	05-06-08	1.967	03-09-08	1.627	02-10-08
10	1.695	0.926	08-09-08	1.312	06-11-08	2.191	03-09-08	1.335	16-09-08

Table 8: Qu's test values for the fitted models at quantile $\tau = 1\%$

ARCH	Crit.	SPY		VGT		XLI		XLP	
order	value	SQ_τ	Break at	SQ_τ	Break at	SQ_τ	Break at	SQ_τ	Break at
1	1.329	1.672	26-09-08	1.681	26-09-08	1.421	01-10-08	1.366	16-09-08
2	1.453	1.399	26-09-08	1.693	26-09-08	0.857	01-10-08	1.067	15-09-08
3	1.517	0.856	26-09-08	1.964	26-09-08	1.796	01-10-08	1.195	16-09-08
4	1.569	1.509	26-09-08	0.740	07-10-08	1.791	01-10-08	1.677	05-06-08
5	1.601	1.892	05-06-08	1.669	26-09-08	1.512	14-10-08	1.469	05-06-08
6	1.628	1.692	05-06-08	0.874	07-01-09	1.722	09-02-09	1.859	02-10-08
7	1.650	1.881	05-06-08	3.320	02-10-08	1.824	13-10-08	1.890	05-06-08
8	1.655	2.741	12-09-08	1.863	07-11-08	1.729	14-10-08	0.992	08-10-08
9	1.684	2.201	14-10-08	2.042	27-01-09	2.563	08-10-08	2.318	18-11-08
10	1.695	2.682	16-09-08	2.999	12-02-09	2.661	08-10-08	2.276	17-11-08

As a conclusion of our empirical analysis we can highlight that the model selection was harsh and unfruitful when considering models without structural break in financial crisis. For this reason, a structural change analysis was proposed. The Qu's test showed good performance when trying to identify the dates of the breakpoints, which happened close to the Lehman Brothers' bankruptcy for small p-orders. Furthermore, the analysis suggested the XLP as a conservative index, which is in accordance with the proposal of this ETF. After making the structural change analysis, we can see how it could influence the coverage test results. In fact, we found evidence of structural changes in all ETFs we analyzed, and after comparing this insight with the results from Cristofersen's and GMM's coverage tests, we were motivated to make a simulation study involving these statistical procedures and the quantile ARCH(p) approach. The study is shown in Section 5.

5 Simulation study

In the simulation study, ten different break scenarios were considered. The analysis seeks to get insights into the behavior of null rejection rates of the two coverage tests described in Subsection 3 and also to analyze the QU's test by considering a model without breaks and under several structural change scenarios. To simplify the simulations, and according to the results suggesting small p-order, the simulations were made by considering ε_t following a linear ARCH(2) model, as described below

$$\varepsilon_t = \begin{cases} (1+\gamma_{1,0}|\varepsilon_{t_i-1}|+\gamma_{2,0}|\varepsilon_{t_i-2}|)v_{t_i}\sqrt{\omega_0}, & t_i \in [0, \lfloor\lambda_{1n}\rfloor) \\ (1+\gamma_{1,1}|\varepsilon_{t_i-1}|+\gamma_{2,1}|\varepsilon_{t_i-2}|)v_{t_i}\sqrt{\omega_1}, & t_i \in [\lfloor\lambda_{1n}\rfloor, \lfloor\lambda_{2n}\rfloor) \\ \vdots & \vdots \\ (1+\gamma_{1,k}|\varepsilon_{t_i-1}|+\gamma_{2,k}|\varepsilon_{t_i-2}|)v_{t_i}\sqrt{\omega_k}, & t_i \in [\lfloor\lambda_{kn}\rfloor, n] \end{cases}, \tag{9}$$

where $0 < \lambda_1 < \ldots < \lambda_k < 1$ and $v_t\sqrt{\omega_i} \sim \mathcal{N}(0, \omega_j)$, $j \in \{0, 1, \ldots, k\}$. The random process in Equation (9) describes a reparametrized linear ARCH(2) process with k breaks in parameters $(\gamma_{1,j}, \gamma_{1,j}, \omega_j)$, $j \in \{0, 1, \ldots, k\}$. It is importante to note that each λ_j, $j \in \{1, \ldots, k\}$, represents the relative position of the j-th break across the time window $([t_0, n])$. In the simulations, the order $p = 2$ is equal to the average value between the orders we highlighted in Subsection 4.2 ($p = 1, 2, 3$).

To evaluate the tests, two thousand Monte Carlo replications from 10 different linear ARCH(2) processes were generated. In the analysis, the programming language and environment for statistical computing R (R Core Team, 2018) was used by setting the code *set.seed(42)* in the Mersenne-Twister pseudorandom number generator. Each simulated model was planned to achieve different perspectives of the tests used in the empirical analysis. In addition, we intended to simulate a return series as seen in the previous Sections, but also under different break scenarios. Figure 9 shows a particular case when simulating a model behaving like the SPY stock index.

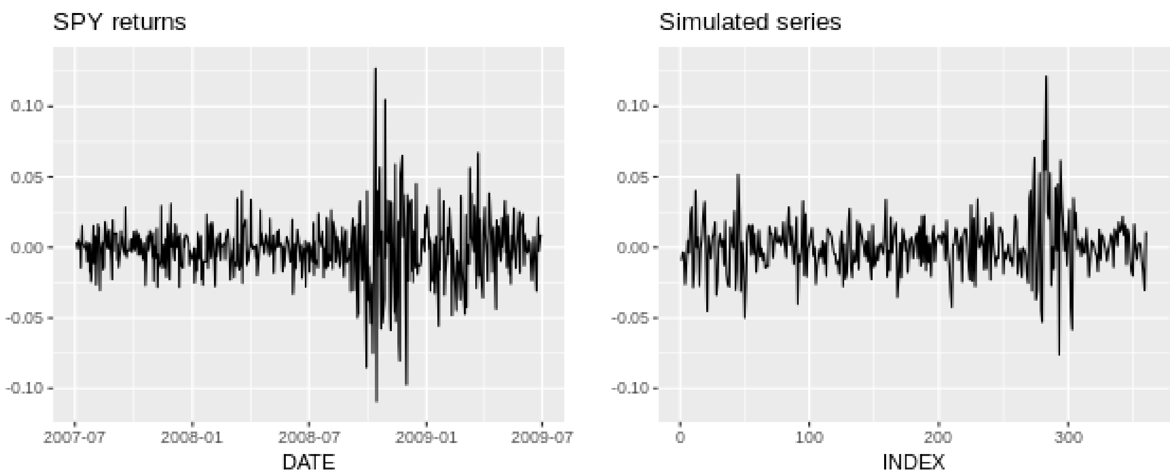

Fig. 9: SPY returns(left) and simulation process(right).

Table 9 shows the λ-values that indicate the relative position of the break (when it exists) and the values of parameters $(\gamma_{1,j}, \gamma_{2,j}, \omega_j)$ used before and after the break. Note that there is no break in model A which implies that λ is not defined in this case. Models B, C, and D were conceived to check a single break at different points over time. Model E aimed to check the impact of only one brake in the conditional volatility. Models F and G were proposed to evaluate the behavior of the tests when a break with clusters of volatility happens. Model H stands for a more progressive break. Model 'I' has a long-length cluster. Model J has three breaks and forms interchangeable clusters. In all cases, the γ values were chosen according to Lee and Noh (2013). The simulation results are summarized in Tables 10 to 21 and in Figures 10 and 11.

Table 9: Simulated processes

Model	λ	γ-parameters	ω	Model	λ	α-parameters	ω
A	–	(0.15, 0.06)	1	G	0	(0.15, 0.06)	1
B	0	(0.15, 0.0375)	1		.75	(0.35, 0.14)	2
	.25	(0.35, 0.1400)	2		.85	(0.15, 0.06)	1
C	0	(0.15, 0.0375)	1	H	0	(0.15, 0.0375)	1
	.5	(0.35, 0.1400)	2		.4	(0.35, 0.1400)	2
D	0	(0.15, 0.0375)	1		.6	(0.15, 0.0375)	3
	.75	(0.35, 0.1400)	2	I	0	(0.15, 0.0375)	1
E	0	(0.15, 0.0375)	1		.25	(0.35, 0.1400)	2
	.5	(0.35, 0.1400)	1		.75	(0.15, 0.0375)	1
F	0	(0.15, 0.06)	1	J	0	(0.15, 0.0375)	1
	.25	(0.35, 0.14)	2		.25	(0.35, 0.1400)	2
	.35	(0.15, 0.06)	1		.5	(0.15, 0.0375)	1
					.75	(0.35, 0.1400)	2

After conducting a statistical investigation of all simulated data, we concluded that the presence of structural changes did not cause a violation of the null hypotheses of the coverage tests. Because of that, for these two tests, our analysis focused on an investigation considering the test statistics defined under the null hypothesis. The results when considering the Christoffersen (1998) tests (Tables 10 to 13) display null rejection rates much smaller than the nominal levels (5% and 1%). It suggests very conservative tests (with or without structural breaks).

Table 10: Null rejection rates of UC Christoffersen's test statistics at 5% nominal level.

	No breaks					
Models	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	One break					
B	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
C	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
E	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Two breaks					
F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
G	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
H	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
I	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Three breaks					
J	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 11: Null rejection rates of UC Christoffersen's test statistics at 1% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	One break					
B	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
C	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
E	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Two breaks					
F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
G	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
H	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
I	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Three breaks					
J	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 12: Null rejection rates of CC Christoffersen's test statistics at 5% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0020	0.0055	0.0015	0.0010	0.0015	0.0015
	One break					
B	0.0015	0.0065	0.0010	0.0020	0.0045	0.0015
C	0.0015	0.0065	0.0030	0.0000	0.0030	0.0015
D	0.0015	0.0085	0.0025	0.0005	0.0045	0.0010
E	0.0015	0.0050	0.0000	0.0015	0.0035	0.0025
	Two breaks					
F	0.0030	0.0120	0.0025	0.0020	0.0030	0.0055
G	0.0040	0.0070	0.0015	0.0020	0.0045	0.0045
H	0.0015	0.0035	0.0020	0.0005	0.0055	0.0010
I	0.0050	0.0070	0.0020	0.0015	0.0050	0.0025
	Three breaks					
J	0.0025	0.0060	0.0025	0.0020	0.0035	0.0020

Table 13: Null rejection rates of CC Christoffersen's test statistics at 1% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0000	0.0005	0.0005	0.0005	0.0000	0.0000
	One break					
B	0.0000	0.0010	0.0000	0.0010	0.0000	0.0000
C	0.0000	0.0005	0.0000	0.0000	0.0005	0.0000
D	0.0000	0.0005	0.0000	0.0005	0.0000	0.0000
E	0.0005	0.0000	0.0000	0.0005	0.0000	0.0005
	Two breaks					
F	0.0005	0.0000	0.0005	0.0005	0.0005	0.0005
G	0.0000	0.0000	0.0000	0.0020	0.0000	0.0005
H	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000
I	0.0000	0.0010	0.0000	0.0010	0.0000	0.0005
	Three breaks					
J	0.0005	0.0010	0.0005	0.0010	0.0005	0.0000

Tables 14 through 17 show the results corresponding to the GMM statistics. The unconditional test (Tables 14 and 15) is very conservative in almost all cases. However, there were some scenarios in which the null rejection rates were much larger than the nominal levels: (1) in Table 14, models H ($n = 360, 720, 1080$), I ($n = 360, 720$) and B ($n = 1080$), for $\tau = 5\%$, and models D ($n = 720, 1080$), C, H and I ($n = 1080$), for $\tau = 1\%$; (2) in Table 15, model H ($n = 720, 1080$), for $\tau = 5\%$.

The conditional tests were realized by considering the order ℓ equal to 3 and 5 (Tables 16 and 17, and, 18 and 19, respectively). In general, the test seems to be conservative. However, in some scenarios it presented a very liberal behavior: (1) in Table 16, models C, H, I, J ($n = 360, 720, 1080$), and B ($n = 1080$), for $\tau = 5\%$, and, D ($n = 720, 1080$), for $\tau = 1\%$; (2) in Table 17, models C, H, I, J ($n = 720, 1080$), and B ($n = 1080$), for $\tau = 5\%$; (3) in Table 18, models C, H, J ($n = 360, 720, 1080$), and B, D, I ($n = 720, 1080$), for $\tau = 5\%$, and, D, F, G ($n = 720, 1080$), for $\tau = 1\%$; (4) in Table 19, models C ($n = 360, 720, 1080$), and B, D, H, J ($n = 720, 1080$), for $\tau = 5\%$, and, D, G ($n = 1080$), for $\tau = 1\%$.

Table 14: Rejection rates in UC GMM test at 5% nominal level.

Models	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
	No breaks					
A	0.0005	0.0015	0.0000	0.0005	0.0000	0.0020
	One break					
B	0.0090	0.0010	0.0625	0.0035	0.2665	0.0065
C	0.0790	0.0015	0.0460	0.0540	0.0190	0.2285
D	0.0005	0.0080	0.0000	0.1665	0.0000	0.1920
E	0.0010	0.0020	0.0005	0.0065	0.0005	0.0085
	Two breaks					
F	0.0005	0.0045	0.0000	0.0305	0.0000	0.0260
G	0.0005	0.0030	0.0005	0.0185	0.0000	0.0150
H	0.1765	0.0010	0.3980	0.0320	0.3540	0.0955
I	0.1020	0.0020	0.1080	0.0545	0.0800	0.2340
	Three breaks					
J	0.0010	0.0015	0.0005	0.0035	0.0095	0.0000

Table 15: Rejection rates in UC GMM test at 1% nominal level.

Models	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
	No breaks					
A	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	One break					
B	0.0005	0.0000	0.0000	0.0000	0.0010	0.0000
C	0.0035	0.0000	0.0185	0.0000	0.0045	0.0005
D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0360
E	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Two breaks					
F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030
G	0.0000	0.0000	0.0000	0.0005	0.0000	0.0015
H	0.0035	0.0000	0.1420	0.0000	0.2500	0.0015
I	0.0015	0.0000	0.0220	0.0000	0.0215	0.0015
	Three breaks					
J	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

5.1 Qu's structure change test

For Qu's test, we analyzed the null rejection rates (model A - no breaks) and the power of the test under some structural break scenarios (models B to J). The results are shown in Tables 20 (5% nominal level) and 21 (1% nominal level). The rates of null hypothesis rejection presented greater than the nominal levels for

Table 16: Rejection rates in CC GMM test at 5% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0000	0.0015	0.0040	0.0000	0.0040	0.0000
	One break					
B	0.0010	0.0010	0.0595	0.0010	0.2230	0.0005
C	0.0675	0.0025	0.3070	0.0140	0.4685	0.0375
D	0.0125	0.0120	0.0650	0.1030	0.0885	0.1865
E	0.0005	0.0015	0.0200	0.0030	0.0320	0.0025
	Two breaks					
F	0.0015	0.0095	0.0125	0.0310	0.0165	0.0375
G	0.0015	0.0095	0.0075	0.0325	0.0180	0.0395
H	0.0180	0.0015	0.5090	0.0070	0.7890	0.0145
I	0.0155	0.0020	0.1130	0.0100	0.3260	0.0360
	Three breaks					
J	0.0005	0.0030	0.1875	0.0060	0.3625	0.0120

Table 17: Rejection rates in UC GMM test at 1% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0000	0.0000	0.0030	0.0000	0.0030	0.0000
	One break					
B	0.0000	0.0000	0.0485	0.0000	0.1915	0.0000
C	0.0410	0.0000	0.2595	0.0005	0.4290	0.0015
D	0.0090	0.0000	0.0485	0.0050	0.0640	0.0415
E	0.0005	0.0000	0.0170	0.0000	0.0260	0.0000
	Two breaks					
F	0.0000	0.0000	0.0075	0.0010	0.0110	0.0075
G	0.0010	0.0000	0.0035	0.0015	0.0130	0.0055
H	0.0060	0.0000	0.2480	0.0000	0.6280	0.0005
I	0.0005	0.0000	0.0575	0.0005	0.2630	0.0010
	Three breaks					
J	0.0005	0.0000	0.1585	0.0000	0.3210	0.0000

Table 18: Null rejection rates in conditional GMM test with $\ell = 5$ and 5% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0050	0.0040	0.0170	0.0040	0.0155	0.0030
	One break					
B	0.0080	0.0035	0.1095	0.0015	0.2765	0.0045
C	0.1930	0.0121	0.5815	0.0225	0.7980	0.0515
D	0.0885	0.0541	0.2270	0.2280	0.3230	0.4595
E	0.0190	0.0060	0.0565	0.0065	0.0800	0.0205
	Two breaks					
F	0.0260	0.0256	0.0480	0.0850	0.0655	0.1585
G	0.0235	0.0286	0.0460	0.1010	0.0685	0.1795
H	0.0865	0.0121	0.4685	0.0085	0.8715	0.0160
I	0.0185	0.0101	0.3295	0.0215	0.6535	0.0320
	Three breaks					
J	0.0980	0.0110	0.4865	0.0300	0.7545	0.0675

$\tau = 1\%$ and $n = 360, 720$. In the other cases, the rates were close to nominal levels. The power of the test was high in many scenarios. However, at 1% nominal level (Table 21), the simulations show very small powers for some scenarios in models B ($\tau = 1\%, n = 360, 720, 1080$), E ($\tau = 5\%, n = 360, 720$; $\tau = 1\%, n = 1080$), F ($\tau = 5\%, n = 360, 720$), G ($\tau = 5\%, n = 360$), H ($\tau = 1\%, n = 360$), I ($\tau = 5\%, n = 360, 720$; $\tau = 1\%, n = 360, 720, 1080$), and J ($\tau = 5\%, n = 360$; $\tau = 1\%, n = 720, 1080$).

Table 19: Null rejection rates in conditional GMM test with $\ell = 5$ and 1% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0020	0.0005	0.0095	0.0015	0.0090	0.0000
	One break					
B	0.0025	0.0010	0.0850	0.0010	0.2410	0.0000
C	0.1410	0.0000	0.4715	0.0015	0.7165	0.0100
D	0.0485	0.0045	0.1535	0.0675	0.2320	0.2800
E	0.0085	0.0010	0.0350	0.0015	0.0510	0.0080
	Two breaks					
F	0.0105	0.0080	0.0300	0.0285	0.0350	0.0670
G	0.0135	0.0075	0.0210	0.0350	0.0375	0.0955
H	0.0530	0.0010	0.3535	0.0010	0.6320	0.0025
I	0.0035	0.0000	0.2100	0.0015	0.5360	0.0050
	Three breaks					
J	0.0385	0.0020	0.3865	0.0075	0.6605	0.0155

Table 20: Null rejection rates (model A) and power (models B to J) for Qu's test at 5% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0525	0.1710	0.0510	0.1075	0.0620	0.0785
	One break					
B	0.4070	0.1300	0.9085	0.1085	0.9875	0.1630
C	0.7570	0.3000	0.9760	0.7125	0.9995	0.9585
D	0.4170	0.4590	0.7475	0.7005	0.9105	0.8440
E	0.1325	0.2135	0.2000	0.1700	0.2960	0.1875
	Two breaks					
F	0.1640	0.2885	0.2270	0.2625	0.3520	0.2825
G	0.2060	0.3645	0.3200	0.3400	0.5065	0.3925
H	0.9485	0.2480	1.0000	0.6185	1.0000	0.9830
I	0.0510	0.1025	0.4055	0.0405	0.7015	0.0310
	Three breaks					
J	0.1480	0.1865	0.3995	0.1685	0.6570	0.1845

Table 21: Null rejection rates (model A) and power (models B to J) for Qu's test at15% nominal level.

Models	No breaks					
	n = 360		n = 720		n = 1080	
	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$	$\tau = 5\%$	$\tau = 1\%$
A	0.0125	0.0805	0.0075	0.0325	0.0115	0.0175
	One break					
B	0.1050	0.0675	0.7630	0.0360	0.9625	0.0420
C	0.5715	0.1140	0.9285	0.3255	0.9975	0.8270
D	0.1970	0.2910	0.4950	0.4825	0.7360	0.7065
E	0.0470	0.1165	0.0605	0.0605	0.1155	0.0670
	Two breaks					
F	0.0530	0.1835	0.0830	0.1345	0.1455	0.1255
G	0.0800	0.2390	0.1345	0.1865	0.2555	0.1965
H	0.8560	0.0975	0.9960	0.2400	1.0000	0.6985
I	0.0060	0.0475	0.1260	0.0110	0.4130	0.0080
	Three breaks					
J	0.0530	0.1030	0.1665	0.0670	0.4020	0.0705

Figures 10 (models B to E) and 11 (models F to J) show the simulated process (left charts) and the corresponding λ-histograms (right charts). In Figure 10, model E presented a more volatile process and the corresponding λ-histogram identified the true brake point very well. This fact also happened when considering models B, C, and D. In Figure 11, models F and G, λ-histogram identified the second brake point. The λ-histogram of model H identified the first breakpoint. The two breaks were well identified in Model I, and the λ-histogram identified only two points in Model J.

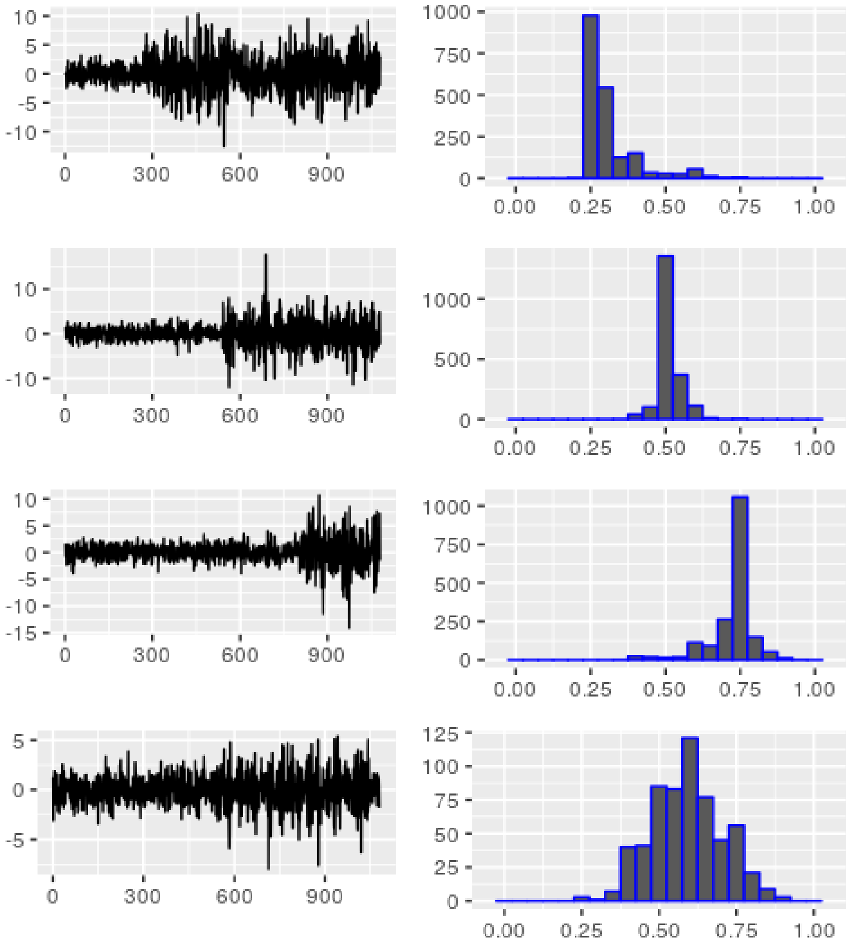

Fig. 10: Single break models (B to E), top-down. A single model simulation(left) and the histogram of λ relative position of its break points when it is rejected Qu's test hypothesis given $\tau = 5\%$ and $n = 1080$(right).

6 Conclusion

Quantile autoregression models provide a framework to cope with Value-at-Risk estimation. In this sense, in an initial study, this paper proposed to use QAR modeling to investigate 5% and 1% VaR associated with the returns from four important ETFs (SPY, VGT, XLI, and XLP), which represent firms belonging to the strategic stock market, such as the Technology Information, industrial and consumption. The VaRs were analized during the Subprime mortgage crisis in an in-sample estimation strategy. Our investigation concludes in favor to Qu's test capacity to detect structural changes in real data. However, this empirical study did not conclude positively when referring to the widely used coverage tests Christoffersen (1998); Candelon et al. (2010). This motivated our simulation study in the second part of the paper.

In the simulations, we considered two thousand Monte Carlo replications from ten different linear ARCH(2) processes, which represent several simulation scenarios focusing on designing simulated data behaving similarly to the ETFs and presenting structural breakpoints. We estimated VaR at 1% and

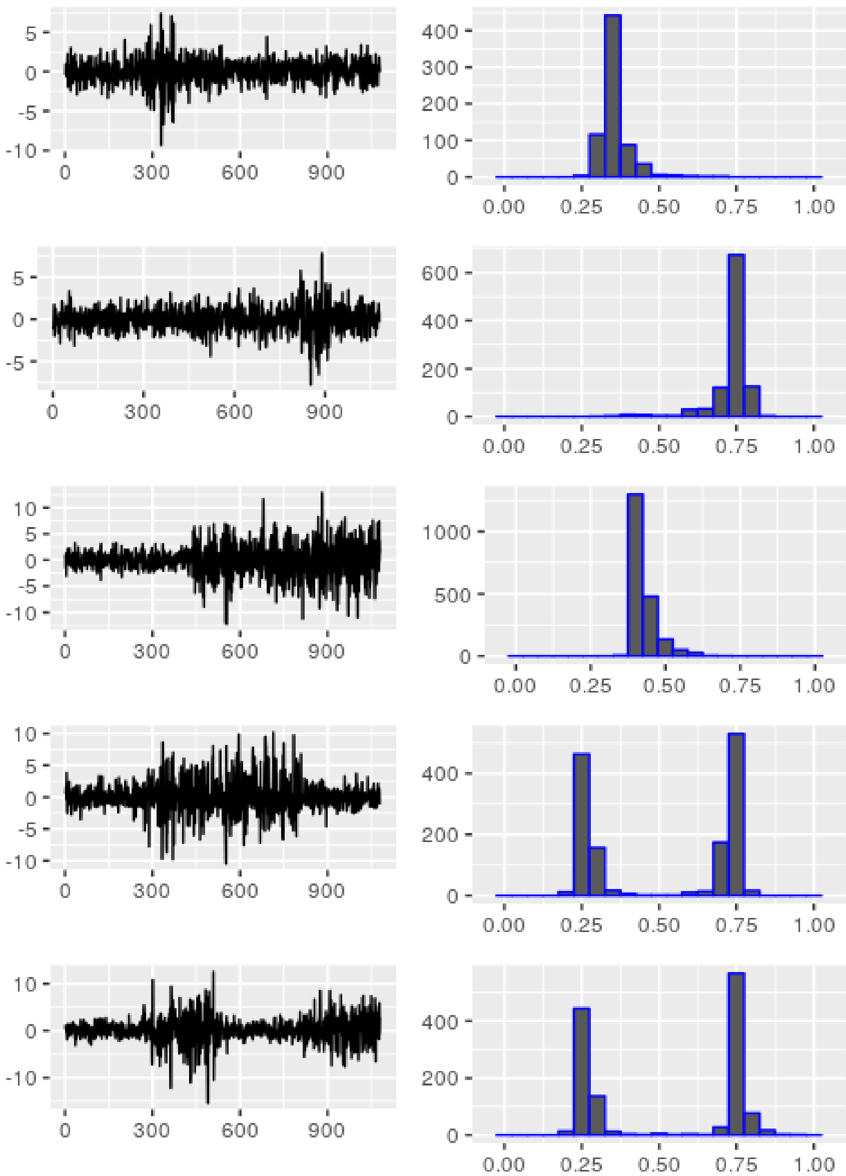

Fig. 11: Multi break models (F to J), top-down. A single model simulation(left) and the histogram of λ relative position of its break points when it is rejected Qu's test hypothesis given $\tau = 5\%$ and $n = 1080$(right).

5% levels and analyzed the null rejection rates of the coverage tests as well as the power of QU's test under the presence of breakpoints and without breaks. Among the main conclusions, we highlight that the coverage tests behaved as very conservative. Qu's structural change test performed a notable power in almost all models proposed during the simulation. However, its power might be severely downgraded in some kinds of series structures like those showing clusters, so common in financial time series. In our simulations, Qu's test retained yet a significant power in these cluster situations. It was also noted in the multiple break models with similar traits a tendency to detect a break in the middlemost break.

In future work, we suggest developing Christorfesen's and GMM's null statistics theoretically to incorporate adjustments to deal with structural breaks, improving the null rejection rates of these tests.

References

1. Koenker, R., Zhao, Q.. Conditional quantile estimation and inference for arch models. *Econometric Theory* 1996;12(5):793–813.
2. Koenker, R., Xiao, Z.. Quantile autoregression. *Journal of the American Statistical Association* 2006;101(475):980–990.
3. Xiao, Z., Koenker, R.. Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association* 2009;104(488):1696–1712.
4. Hwang, S., Satchell, S.E., Pereira, P.L.V., et al. How persistent is volatility? an answer with stochastic volatility models with markov regime switching state equations. In: *Proceedings of The Econometric Society 2004 Latin American Meetings*; vol. 198. 2004:.
5. Stărică, C., Granger, C.. Nonstationarities in stock returns. *Review of economics and statistics* 2005;87(3):503–522.
6. Andreou, E., Ghysels, E.. Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 2002;17(5):579–600.
7. Demyanyk, Y., Van Hemert, O.. Understanding the subprime mortgage crisis. *The review of financial studies* 2011;24(6):1848–1880.
8. Guidolin, M., Tam, Y.M.. A yield spread perspective on the great financial crisis: Break-point test evidence. *International Review of Financial Analysis* 2013;26:18–39.
9. Qu, Z.. Testing for structural change in regression quantiles. *Journal of Econometrics* 2008;146(1):170–184. URL: `https://www.sciencedirect.com/science/article/pii/S0304407608000948`. doi:`https://doi.org/10.1016/j.jeconom.2008.08.006`.
10. Christoffersen, P.F.. Evaluating interval forecasts. *International economic review* 1998;:841–862.
11. Candelon, B., Colletaz, G., Hurlin, C., Tokpavi, S.. Backtesting value-at-risk: a gmm duration-based test. *Journal of Financial Econometrics* 2010;9(2):314–343.
12. Page, E.S.. Continuous inspection schemes. *Biometrika* 1954;41(1/2):100–115.
13. Page, E.. Control charts with warning lines. *Biometrika* 1955;42(1-2):243–257.
14. Aue, A., Horváth, L.. Structural breaks in time series. *Journal of Time Series Analysis* 2013;34(1):1–16.
15. Jeng, J.L.. Analyzing Event statistics in corporate finance: methodologies, evidences, and critiques. Springer; 2015.
16. Dette, H., Wied, D.. Detecting relevant changes in time series models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2016;78(2):371–394.
17. Song, J., Kang, J.. Parameter change tests for arma–garch models. *Computational Statistics & Data Analysis* 2018;121:41–56.

18. Truong, C., Oudre, L., Vayatis, N.. Selective review of offline change point detection methods. *Signal Processing* 2019;:107299.

19. Casini, A., Perron, P.. Structural breaks in time series. In: *Oxford Research Encyclopedia of Economics and Finance*. 2019:.

20. Granger, C.W., Hyung, N.. Occasional structural breaks and long memory with an application to the s&p 500 absolute stock returns. *Journal of empirical finance* 2004;11(3):399–421.

21. Babikir, A., Gupta, R., Mwabutwa, C., Owusu-Sekyere, E.. Structural breaks and garch models of stock return volatility: The case of south africa. *Economic Modelling* 2012;29(6):2435–2443.

22. Jung, R.C., Maderitsch, R.. Structural breaks in volatility spillovers between international financial markets: Contagion or mere interdependence? *Journal of Banking & Finance* 2014;47:331–342.

23. Hamilton, J.D.. Analysis of time series subject to changes in regime. *Journal of econometrics* 1990;45(1-2):39–70.

24. Wolters, M.H., Tillmann, P.. The changing dynamics of us inflation persistence: A quantile regression approach. *Studies in Nonlinear Dynamics & Econometrics* 2015;19(2):161–182.

25. Zhou, M., Wang, H.J., Tang, Y.. Sequential change point detection in linear quantile regression models. *Statistics & Probability Letters* 2015;100:98–103.

26. Bonaccolto, G., Caporin, M., Gupta, R.. The dynamic impact of uncertainty in causing and forecasting the distribution of oil returns and risk. *Physica A: Statistical Mechanics and its Applications* 2018;507:446–469.

27. Su, L., Xiao, Z.. Testing for parameter stability in quantile regression models. *Statistics & Probability Letters* 2008;78(16):2768–2775.

28. Assaf, D., Ritov, Y.. A double sequential procedure for detecting a change in distribution. *Biometrika* 1988;75(4):715–722.

29. Koenker, R., Bassett, G.. Regression quantiles. *Econometrica: journal of the Econometric Society* 1978;:33–50.

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

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

32. Engle, R.F., Lilien, D.M., Robins, R.P.. Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica: journal of the Econometric Society* 1987;:391–407.

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

34. Engle, R.F., Bollerslev, T.. Modelling the persistence of conditional variances. *Econometric reviews* 1986;5(1):1–50.
35. 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* 1993;48(5):1779–1801.
36. Engle, R.F., Ng, V.K.. Measuring and testing the impact of news on volatility. *The journal of finance* 1993;48(5):1749–1778.
37. Xiao, L., Aydemir, A.. Volatility modelling and forecasting in finance. In: *Forecasting volatility in the financial markets*. Elsevier; 2007:1–45.
38. Miron, D., Tudor, C.. Asymmetric conditional volatility models: Empirical estimation and comparison of forecasting accuracy. *Romanian Journal of Economic Forecasting* 2010;13(3):74–92.
39. Engle, R.F., Manganelli, S.. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 2004;22(4):367–381.
40. Baur, D.G., Dimpfl, T., Jung, R.C.. Stock return autocorrelations revisited: A quantile regression approach. *Journal of Empirical Finance* 2012;19(2):254–265.
41. Jorion, P.. Value At Risk The New Benchmark For Managing Financial. McGraw-Hill; 2000. ISBN 0-70-135502-2.
42. Tsay, R.S.. Analysis of financial time series; vol. 543. John wiley & sons; 2005.
43. Lee, S., Noh, J.. Quantile regression estimator for garch models. *Scandinavian Journal of Statistics* 2013;40(1):2–20.
44. Xiao, Z., Guo, H., Lam, M.S.. Quantile regression and value at risk. *Handbook of Financial Econometrics and Statistics* 2015;:1143–1167.
45. Haugom, E., Ray, R., Ullrich, C.J., Veka, S., Westgaard, S.. A parsimonious quantile regression model to forecast day-ahead value-at-risk. *Finance Research Letters* 2016;16:196–207.
46. Hagfors, L.I., Bunn, D., Kristoffersen, E., Staver, T.T., Westgaard, S.. Modeling the uk electricity price distributions using quantile regression. *Energy* 2016;102:231–243.
47. Taylor, J.W.. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics* 2019;37(1):121–133.
48. Christou, E., Grabchak, M.. Estimation of value-at-risk using single index quantile regression. *Journal of Applied Statistics* 2019;:1–16.
49. Xu, Q., Liu, X., Jiang, C., Yu, K.. Quantile autoregression neural network model with applications to evaluating value at risk. *Applied Soft Computing* 2016;49:1–12.
50. Kuck, K., Maderitsch, R.. Intra-day dynamics of exchange rates: New evidence from quantile regression. *The Quarterly Review of Economics and Finance* 2019;71:247–257.

51. Philippe, J.. Value at risk: the new benchmark for managing financial risk. *NY: McGraw-Hill Professional* 2001;.

52. Kupiec, P.. Techniques for verifying the accuracy of risk measurement models. *The J of Derivatives* 1995;3(2).

53. Rudin, W.. Real and complex analysis. Tata McGraw-Hill Education; 2006.

54. Revuz, D., Yor, M.. Continuous martingales and Brownian motion; vol. 293. Springer Science & Business Media; 2013.

55. Abner, D.J.. Visual Guide to ETFs. Bloomberg financial series; 1 ed.; Bloomberg Press; 2013. ISBN 978-1-118-20465-8.

56. Dickey, D.A., Fuller, W.A.. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society* 1981;:1057–1072.

57. McLeod, A.I., Li, W.K.. Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* 1983;4(4):269–273.

58. Perron, P.. Testing for a unit root in a time series with a changing mean. *Journal of Business & Economic Statistics* 1990;8(2):153–162.

59. Akaike, H.. A new look at the statistical model identification. *IEEE transactions on automatic control* 1974;19(6):716–723.

60. Koenker, R., Chesher, A., Jackson, M.. Quantile Regression. Econometric Society Monographs; Cambridge University Press; 2005. ISBN 9780521608275.

61. Issler, J.V.. Estimating and forecasting the volatility of brazilian finance series using arch models. *Brazilian review of econometrics* 1999;19(1):5–56.

62. R Core Team, . R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria; 2018. URL: `https://www.R-project.org/`.