TECHNICAL TRANSACTIONS | CZASOPISMO TECHNICZNE
FUNDAMENTAL SCIENCES | NAUKI PODSTAWOWE
2-NP/2014

BARTOSZ STAWIARSKI*

DATA-DRIVEN SCORE TEST OF FIT FOR CLASS OF GARCH MODELS

ADAPTACYJNY TEST ZGODNOŚCI DLA KLASY MODELI GARCH

Abstract

A data-driven score test of fit for testing the conditional distribution within the class of stationary GARCH(p,q) models is presented. In this paper extension of the complete results obtained by Inglot and Stawiarski in [7], as well as in Stawiarski [15] for the parsimonious GARCH(1,1) case is proposed. The null (composite) hypothesis subject to testing asserts that the innovations distribution, determining the GARCH conditional distribution, belongs to the specified parametric family. Generalized Error Distribution (called also Exponential Power) seems of special practical value.

Applying the pioneer idea of Neyman [13] dating back to 1937, in combination with dimension selection device proposed by Ledwina [10] in 1994, lead to derivation uf the efficient score statistic and its data-driven version for this testing problem. In the case of GARCH(1,1) model both the asymptotic null distribution of the score statistic has been already established in [7] and [15], together with the asymptotics of the data-driven test statistic with appropriately regular estimators plugged in place of nuisance parameters. Main results are only stated herewith, while for detailed proofs inspection and power simulations, ample reference to these papers is provided. We show that the test derivation and asymptotic results carry over to stationary ARCH(q) models for any $q \in N$. Moreover, thanks to ARCH(∞) representation of the GARCH(p,q) model, the test can asymptotically encompass the full GARCH family, which as a final result provides the flexible testing tool in the GARCH(p, q) framework.

Keywords: GARCH model, Neyman smooth test, data-driven score test of fit, Generalized Error Distribution

Streszczenie

W pracy przedstawiono adaptacyjny test zgodności dla testowania warunkowego rozkładu w klasie stacjonarnych modeli GARCH(p,q). Jest to rozszerzenie kompletnych wyników uzyskanych w pracach [7] oraz [15] dla przypadku mniej rozbudowanego modelu GARCH(1,1). Podlegająca testowaniu (złożona) hipoteza zerowa postuluje, że rozkład szumu determinujący warunkowy rozkład szeregu GARCH, należy do określonej rodziny parametrycznej. Szczególne znaczenie w kontekście zastosowań ma klasa rozkładów GED.

Zastosowanie pionierskiego pomysłu Neymana z 1937 r. [13] w połączeniu z kryterium wyboru wymiaru, zaproponowanym przez Ledwinę w 1994 r. w pracy [10] pozwala wyprowadzić dla omawianego zagadnienia testowego efektywną statystykę wynikową i jej adaptacyjną wersję. W przypadku modelu GARCH(1,1) zarówno asymptotyczny rozkład statystyki przy hipotezie zerowej, jak i asymptotyka jej adaptacyjnej wersji z odpowiednio regularnymi estymatorami parametrów zakłócających zostały już uzyskane w [7] oraz [15]. Główne wyniki są tu tylko przywołane z przywołaniem licznych referencji do tych prac. Pokazano, że konstrukcja testu i wyniki asymptotyczne przenoszą się na stacjonarne modele ARCH(q) dla dowolnego $q \in N$. Ponadto dzięki reprezentacji modeli GARCH poprzez ARCH(∞) test pozwala asymptotycznie objąć całą klasę GARCH, co w ostateczności daje elastyczne narzędzie testowe dla modeli GARCH(p,q).

Słowa kluczowe: Model GARCH, gładki test Neymana, adaptacyjny test zgodności, rozkład GED

*Institute of Mathematics, Cracow University of Technology; bstawiarski@pk.edu.pl

1. Introduction

The Autoregressive Conditionally Heteroscedastic class of time series models, ARCH, was introduced by Engle in 1982 and four years later extended to GARCH by Bollerslev in [3]. Modelling financial time series was main empirical motivation standing behind introduction of such models, allowing for fluctuation of their conditional variance. Throughout more than two decades of research, vast theoretical and computational research concerning GARCH models has been done. A good deal of summaries or even books dedicated specifically to this class of time series has appeared, e.g. Francq and Zakoïan (2010), [5].

Widely exploited in the early days, conditionally normal GARCH models were soon questioned because of unsatisfactory fitting to econometric data. Therefore, the problem of checking conditional distribution assumptions in nonlinear time series has become vital. Apart from ad-hoc imposed innovations distribution (t, α-stable or Laplace), several constructive tests have been proposed only in early 2000's. Specifically, Chen in 2002 [4] proposed a characteristic function based test, while Bai in 2003 [1] presented a martingale transform approach to testing conditional distribution of some dynamic models. Inglot and Stawiarski [7] reconsidered - in the GARCH(1,1) time series context - the idea of smooth tests conceived by Neyman [13] and their data-driven versions devised by Ledwina in [10]. A data-driven score test of fit for conditional distribution for the simple hypothesis i.e. when the null density is fully specified, was derived. Just as the composite hypothesis case for *i.i.d.* random variables was considered in [6] and [8], Stawiarski [15] extended that result to the composite hypothesis case in the GARCH(1,1) framework. This was also desirable from the applicational point of view as testing conditional distribution for financial time series should naturally allow for flexible, parametric families.

In Sections 2 and 3 of this paper we rather succinctly refer the test construction and quote main theoretical results determining the asymptotic behavior of the test statistic. We also briefly report the test power simulation study in Section 4. Detailed theoretical derivations and proofs and simulation studies, due to capacity constraints, can be found in [7] and [15]. The substantially new results are reported in Section 5, which contains extension of the proposed test to the case of ARCH(q), then GARCH(p,q) models. Main theoretical results carry over from the two papers due to general properties such as imposed stationarity (strict and weak), but the present general GARCH context results in some differences of technical nature (ampler parametrization influencing the final test statistic). Therefore the derivation of complete theoretical results leading to fully-fledged data-driven score test of fit for the whole class of stationary GARCH models calls for some future refinements at several stages. Some power simulation study can be a potential subject of further interest. We conclude the paper with Appendix and discussion concerning the above-mentioned possible paths of future research.

2. Score test of fit for composite hypothesis in the GARCH(1,1) model

In [7] and [15] the following GARCH(1,1) time series model $\{X_t\}_{t\in Z}$, introduced by Bollerslev [3], was the object of research:

$$\begin{cases} X_t = \sqrt{h_t}\varepsilon_t \\ h_t = \omega + \alpha X_{t-1}^2 + \beta h_{t-1} \end{cases} \tag{2.1}$$

where $\vartheta = [\omega, \alpha, \beta]^T$ is a column vector of model parameters and $\{\varepsilon_t\}_{t\in Z}$ is a sequence of *i.i.d.* random variables, satisfying $E\varepsilon_t = 0$, $Var\varepsilon_t = 1$. Denoting by $\mathcal{F}_t = \sigma\{..., \varepsilon_{t-1}, \varepsilon_t\}$ the process filtration up to time t, it is evident that the conditional variance h_t is $\mathcal{F}_{t-1}$–measurable. It is assumed hereafter that $\vartheta \in \Theta = \{[\omega, \alpha, \beta]^T : \omega, \alpha, \beta > 0;\ \alpha+\beta < 1\}$, which ensures weak and strict stationarity as well as ergodicity of X_t (see e.g. [3]).

The null hypothesis in [7] asserted that the zero-mean and unit-variance innovations ε_t have a fully specified unknown density $f(x)$ on R. Stawiarski [15] allowed for extension to a parametric family of noise distribution densities, namely $\mathcal{G} = \{f(x, \lambda) : \lambda \in \Lambda,\ \Lambda \subset R^m\}$ (Λ - an open set), satisfying $\int_R x f(x, \lambda)dx = 0$, $\int_R x^2 f(x, \lambda)dx = 1$ for all $\lambda \in \Lambda$.

Given a finite "data set" of observations $X_{(n)} = (X_1, ..., X_n)$ of the process $\{X_t\}_{t\in Z}$ obeying (2.1), the hypothesis we consider is as follows:

$$H_0\text{: } \varepsilon_t \text{ 's have a density } f(x, \lambda) \text{ belonging to } \mathcal{G}\ ;\ \vartheta \in \Theta.$$

Exploiting the concept of "smooth tests" conceived in [13], we restate the above null hypothesis in the equivalent, parametric form, subject to testing. To this end, choose a natural number k and $\Phi(x) = [\Phi_1(x), ..., \Phi_k(x)]^T$– a vector of bounded orthonormal functions in $L_2[0, 1]$ satisfying $\int_0^1 \Phi(x)dx = 0$. Denoting by$F(x, \lambda)$ the *cdf* of $f(x, \lambda)$, we immerse the hypothetical density into a $k-$parametric exponential family

$$\exp\{\tau^T \Phi(F(x, \lambda)) - C_k(\tau)\} \cdot f(x, \lambda) \tag{2.2}$$

where $\tau = [\tau_1, ..., \tau_k]^T \in R^k$ and $C_k(\tau)$ - the normalizing constant.

Let $\eta=[\vartheta^T,\lambda^T]^T \in \Theta \times \Lambda \subset R^{3+m}$ denote a vector of all nuisance parameters appearing in this testing problem, stemming from the GARCH(1,1) model and density f, respectively. Hence we get the parametric reformulation of the above hypothesis:

$$H_0^* : \tau = 0;\ \eta \in \Theta \times \Lambda. \tag{2.3}$$

For notational simplicity, we shall denote by ε the r.v. distributed as ε_t, i.e. under H_0^* with true values of all parameters. Densities (2.2) no longer have zero mean or unit variance. Still, small values of $||\tau||$ accounting for moderate departures from the null density are of major interest from the perspective of the test sensitivity.

Let us quote basic assumptions – being rather standard regularity conditions – imposed on the family $\mathcal{G}$ in our testing problem. For every $\lambda \in \Lambda$:

(A1) $\frac{\partial \log f(x,\lambda)}{\partial \lambda}$ exists and is continuous with respect to λ for almost all $x \in R$;

Fisher information matrix $I(\lambda) = E\left\{\frac{\partial \log f(\varepsilon,\lambda)}{\partial \lambda}\frac{\partial \log f(\varepsilon,\lambda)}{\partial \lambda^T}\right\}$ is well-defined, continuous w.r.t. λ and nonsingular;

(A2) $f(\bullet,\lambda)$ is absolutely continuous on R and the function $\varsigma(x,\lambda) = x\frac{\partial \log f(x,\lambda)}{\partial x}+1$ defined on the set $\{x : f(x,\lambda) > 0\}$ is not almost everywhere constant;

(A3) the following set of functions is linearly independent:

$$\left\{\Phi_1(F(x,\lambda)), ..., \Phi_k(F(x,\lambda)), \frac{\partial \log f(x,\lambda)}{\partial \lambda_1}, ..., \frac{\partial \log f(x,\lambda)}{\partial \lambda_m}\right\}.$$

(A4) $E\left|\varsigma(\varepsilon,\lambda)\right|^3 < \infty$ and $E\left\|\frac{\partial \log f(\varepsilon,\lambda)}{\partial \lambda}\right\|^3 < \infty$.

Fundamental for our derivation will be the following representation of the conditional variance obtained from (2.1) by successive iteration for $t = 1, 2, ..., n$

$$h_t = \omega \sum_{s=0}^{t-2} \beta^s + \alpha \sum_{s=0}^{t-2} \beta^s X_{t-1-s}^2 + \beta^{t-1} h_1. \tag{2.4}$$

Thus the conditional variance h_t is expressed in terms of the observed sample path $X_{(n)}$. From now on, all calculations will be carried out conditionally on $h_1 = h$ with $h > \omega(1-\beta)^{-1}$, which establishes the link between the infinite past and the present of the process. The constant h is assumed given and its influence asymptotically vanishes due to the exponentially decaying memory of the series. Let P_h be the probability on the σ-field $\sigma(X_1, X_2, ...)$ induced by the family of conditional distributions of $(X_1, ..., X_n)$, $n = 1, 2, ...$, conditionally on $h_1 = h$ and under the true value of η. Accordingly, E_h will stand for respective expectation.

Following the lines in [15] we obtain the log-likelihood $L_k(X_1, ..., X_n; \tau, \vartheta, \lambda)$ of $X_{(n)}$, which further is employed in calculating a score vector $\ell = \ell(\eta)$, defined as derivatives of L_k with respect to all parameters involved, evaluated under H_0^*:

$$\ell(\eta) = \left[\left(\frac{\partial L_k(X_{(n)}; \tau, \vartheta, \lambda)}{\partial \tau}\right)^T \left(\frac{\partial L_k(X_{(n)}; \tau, \vartheta, \lambda)}{\partial \vartheta}\right)^T \left(\frac{\partial L_k(X_{(n)}; \tau, \vartheta, \lambda)}{\partial \lambda}\right)^T\right]^T\Bigg|_{\tau=0}. \tag{2.5}$$

More explicitly, the constituent vectors – respectively: k-, 3-, and m-dimensional – are as follows

$$\begin{aligned}
\ell_\tau &= \tfrac{\partial L_k}{\partial \tau}\Big|_{\tau=0} = \sum_{t=1}^{n} \Phi(F(X_t/\sqrt{Q_t},\lambda)),\\
\ell_\vartheta &= \tfrac{\partial L_k}{\partial \vartheta}\Big|_{\tau=0} = -\tfrac{1}{2}\sum_{t=2}^{n} \tfrac{\varsigma(X_t/\sqrt{Q_t},\lambda)}{Q_t}\tfrac{\partial Q_t}{\partial \vartheta},\\
\ell_\lambda &= \tfrac{\partial L_k}{\partial \lambda}\Big|_{\tau=0} = \sum_{t=1}^{n} \tfrac{\partial \log f(X_t/\sqrt{Q_t},\lambda)}{\partial \lambda},
\end{aligned} \tag{2.6}$$

where $Q_1 = h$, and for $t \geqslant 2$

$$\begin{aligned}
Q_t &= Q_t(X_1,...,X_{t-1};h,\vartheta) = \omega\sum_{s=0}^{t-2}\beta^s + \alpha\sum_{s=0}^{t-2}\beta^s X_{t-1-s}^2 + \beta^{t-1}h\ ,\\
\tfrac{\partial Q_t}{\partial \vartheta} &= \left[\sum_{s=0}^{t-2}\beta^s \quad \sum_{s=0}^{t-2}\beta^s X_{t-1-s}^2 \quad \sum_{s=1}^{t-2} s\beta^{s-1}(\omega+\alpha X_{t-1-s}^2) + (t-1)\beta^{t-2}h\right]^T .
\end{aligned} \tag{2.7}$$

Now, in the form of brief remarks we cite three results concerning the elementary properties of $\ell(\eta)$, holding true for $h > \omega(1-\beta)^{-1}$ and $t \geqslant 1$, cf. Proposition 3.1 – 3.3 in [15].

For $\{X_t\}_{t\in Z}$ following (2.1) random variables $\tilde{\varepsilon}_t = X_t/\sqrt{Q_t}$ are *i.i.d.* under P_h and have the same distribution as ε_t's. Under former assumptions (A1) – (A4), $E_h\ell(\vartheta,\lambda) = 0$ and $E_h||\ell(\vartheta,\lambda)||^2 < \infty$, where $||\cdot||$ denotes the Euclidean norm; moreover, the components of the score vector given by (2.6) are linearly independent random variables in $L_2(P_h)$ for any $\eta \in \Theta\times\Lambda$.

Determining the covariance matrix $\tilde{B}^{(n)}(\eta)$ of the normalized score vector $n^{-1/2}\ell(\eta)$ is another step. Thanks to (2.6) and by orthonormality of the system Φ, $\tilde{B}^{(n)}(\eta)$ admits the block representation

$$\tilde{B}^{(n)}(\eta) = n^{-1}E_h\{\ell(\eta)(\ell(\eta))^T\} = n^{-1}E_h\begin{bmatrix} \ell_\tau\ell_\tau^T & \ell_\tau\ell_\eta^T \\ \ell_\eta\ell_\tau^T & \ell_\eta\ell_\eta^T \end{bmatrix} = \begin{bmatrix} I_k & \tilde{B}_{12}^{(n)}(\eta) \\ \tilde{B}_{21}^{(n)}(\eta) & \tilde{B}_{22}^{(n)}(\eta)\end{bmatrix}. \tag{2.8}$$

Obviously, $\tilde{B}_{21}^{(n)}(\eta) = [\tilde{B}_{12}^{(n)}(\eta)]^T$and the block matrix $\tilde{B}_{22}^{(n)}(\eta)$ is invertible for any n and $\eta \in \Theta\times\Lambda$, by Proposition 3.3 in [15]. Explicit forms of $\tilde{B}_{12}^{(n)}(\eta)$ and $\tilde{B}_{22}^{(n)}(\eta)$ are given there in (6.7) and (6.12), respectively.

An efficient score vector $\ell^*(\eta)$ is defined as the residual of the orthogonal projection in $L_2(P_h)$ of $n^{-1/2}\ell_\tau$ upon the subspace generated by components of ℓ_η. Standard theorems, cf. [14], imply that $\ell^*(\eta) = n^{-1/2}(\ell_\tau - \tilde{B}_{12}^{(n)}(\eta)[\tilde{B}_{22}^{(n)}(\eta)]^{-1}\ell_\eta)$. This leads to the explicit formula for $\ell^*(\eta)$, which we state in the following proposition for the sake of convenient reference.

Proposition 2.1. *[Proposition 3.4 in [15]]. Suppose $\{X_t\}_{t\in Z}$ obeys (2.1) and (A1)–(A4) are satisfied. Then for any $h > \omega(1-\beta)^{-1}$ the efficient score vector $\ell^*(\eta)$ for testing H_0^* has the form*

$$\ell^*(\eta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Phi(F(\tilde{\varepsilon}_t, \lambda)) - \frac{1}{\sqrt{n}} \tilde{B}_{12}^{(n)}(\eta)[\tilde{B}_{22}^{(n)}(\eta)]^{-1} \begin{bmatrix} -\frac{1}{2} \sum_{t=2}^{n} \varsigma(\tilde{\varepsilon}_t, \lambda) \frac{1}{Q_t} \frac{\partial Q_t}{\partial \vartheta} \\ \sum_{t=1}^{n} \frac{\partial \log f(\tilde{\varepsilon}_t, \lambda)}{\partial \lambda} \end{bmatrix} \tag{2.9}$$

and its covariance matrix under P_h is given by

$$\tilde{M} = \tilde{M}^{(n)}(\eta) = I_k - \tilde{B}_{12}^{(n)}(\eta)[\tilde{B}_{22}^{(n)}(\eta)]^{-1} \tilde{B}_{21}^{(n)}(\eta). \tag{2.10}$$

The martingale-difference array structure of the vector $\ell^*(\eta)$ in (2.9) paves the way for main limit result, leading in the sequel to establishing the asymptotics of our score test statistic. We quote verbatim Theorem 3.5 from [15] and in Appendix below we provide the outline of its proof.

Theorem 2.2. *Suppose $\{X_t\}_{t \in Z}$ obeys (2.1) and (A1)–(A4) are satisfied. Then for almost every $h > \omega(1-\beta)^{-1}$ (with respect to the Lebesgue measure) it holds*

$$[\tilde{M}^{(n)}(\eta)]^{-1/2} \ell^*(\eta) \xrightarrow{D} N(0, I_k) \tag{2.11}$$

under P_h in R^k as $n \to \infty$.

Now, for fixed natural k introduce a score statistic being a quadratic form

$$W_k = W_k(\eta) = \ ||[\tilde{M}^{(n)}(\eta)]^{-1/2} \ell^*(\eta)||^2 = (\ell^*(\eta))^T [\tilde{M}^{(n)}(\eta)]^{-1} \ell^*(\eta). \tag{2.12}$$

Hence, as a direct implication of the above theorem we get, under its assumptions,

$$W_k \xrightarrow{D} \chi_k^2 \tag{2.13}$$

under P_h as $n \to \infty$, where χ_k^2 is a central chi-square random variable with k degrees of freedom.

Now, let us proceed to define a dimension selection rule. Choose fixed $K \geqslant 1$ - a maximal dimension of the exponential family (2.2) built on $\mathcal{G}$. Define a selection rule $S(\eta)$ as

$$S(\eta) = \min\{k : 1 \leqslant k \leqslant K,\ W_k(\eta) - k \log n \geqslant W_j(\eta) - j \log n\ \forall\ j = 1, ..., K\}. \tag{2.14}$$

Thus, the resulting data-driven score statistic is $W_{S(\eta)}(\eta)$. Since K is fixed, (2.13) implies that $P_h(S(\eta) = 1) \to 1$ as $n \to \infty$ (cf. Section 3.3 in [7]). Accordingly, we obtain the asymptotic behaviour of $W_{S(\eta)}(\eta)$ under H_0^*. Therefore, under the assumptions of Theorem 2.2 it holds

$$W_{S(\eta)} \xrightarrow{D} \chi_1^2 \tag{2.15}$$

under P_h as $n \to \infty$, where $S(\eta)$ is given by (2.14).

Still, $W_{S(\eta)}$ is not suitable for testing as it depends on an unknown nuisance parameter. In the following subsection we will focus on the estimated test statistic $\hat{W}_{\hat{S}}$ with a square-root consistent estimator $\hat{\eta}$ instead of η, briefly recalling main results.

3. Data-driven test statistic and its asymptotics

Let $\hat{\eta} = [\hat{\vartheta}^T,\ \hat{\lambda}^T]^T$ be a square-root consistent estimator of the nuisance parameter $\eta = [\vartheta^T,\ \lambda^T]^T$. For $t \geqslant 2$ we set $\hat{\varepsilon}_t = X_t \Big/ \sqrt{\hat{Q}_t}$, $\hat{Q}_t = Q_t|_{\vartheta=\hat{\vartheta}}$ and $\frac{\partial \hat{Q}_t}{\partial \vartheta} = \frac{\partial Q_t}{\partial \vartheta}\Big|_{\vartheta=\hat{\vartheta}}$. Suppose that $\hat{B}$is a consistent estimator of $\tilde{B}^{(n)}(\eta)$, which naturally implies consistency of the block matrices estimators $\hat{B}_{12}, \hat{B}_{22}$ for $\tilde{B}_{12}^{(n)}(\eta)$ and $\tilde{B}_{22}^{(n)}(\eta)$, respectively. Then the estimated efficient score vector $\hat{\ell}^*(\hat{\eta})$, with $\hat{\eta}$ and $\hat{B}$ plugged into it, is as follows

$$\hat{\ell}^*(\hat{\eta}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Phi(F(\hat{\varepsilon}_t, \hat{\lambda})) - \frac{1}{\sqrt{n}} \hat{B}_{12} \hat{B}_{22}^{-1} \begin{bmatrix} -\frac{1}{2} \sum\limits_{t=2}^{n} \varsigma(\hat{\varepsilon}_t, \hat{\lambda}) \frac{1}{\hat{Q}_t} \frac{\partial \hat{Q}_t}{\partial \vartheta} \\ \sum\limits_{t=1}^{n} \frac{\partial \log f(\hat{\varepsilon}_t, \hat{\lambda})}{\partial \lambda} \end{bmatrix} \tag{3.1}$$

while

$$\hat{M} = I_k - \hat{B}_{12} \hat{B}_{22}^{-1} \hat{B}_{21} \tag{3.2}$$

is a consistent estimator of the covariance matrix $\tilde{M}$.

In order to proceed, some further theoretical assumptions in addition to (A1)–(A4) on our model have to be imposed. These are listed precisely as (A6)–(A13) in [15] and, besides the consistency of nuisance estimators, they concern the distribution of ε - r.v. with the density $f(x, \lambda)$ and appropriate regularity conditions for f itself, as well as the system Φ. Specifically it is worth mentioning that (A5) requires $E|\varepsilon|^{2\kappa} < \infty$ for some $\kappa > 2$.

The main limit theorem being a counterpart of Theorem 2.2 is stated below, but for its lengthy proof we refer to Section 8 in [15].

Theorem 3.1. *[Theorem 4.1 in [15]] Under assumptions (A1)–(A13) in [15] for almost every $h > \omega(1-\beta)^{-1}$ the following asymptotic result holds*

$$\hat{M}^{-1/2} \hat{\ell}^*(\hat{\eta}_*) \xrightarrow{D} N(0, I_k) \tag{3.3}$$

under P_h in R^k as $n \to \infty$, where $\hat{\ell}^*(\hat{\eta}_*)$ is parallel to $\hat{\ell}^*(\hat{\eta})$ from (3.1) with discretized version $\hat{\eta}_*$ of the estimator $\hat{\eta}$ and $\hat{M}$ as in (3.2).

Consequently, the estimated score statistic corresponding to (2.12) takes the form

$$\hat{W}_k = \hat{W}_k(\hat{\eta}_*) = \ ||\hat{M}^{-1/2}\hat{\ell}^*(\hat{\eta}_*)||^2 = (\hat{\ell}^*(\hat{\eta}_*))^T\hat{M}^{-1}\hat{\ell}^*(\hat{\eta}_*), \tag{3.4}$$

while the ready-to-perform in practice data-driven score test statistic is just $\hat{W}_{\hat{S}}$ with

$$\hat{S} = \hat{S}(\hat{\eta}_*) = \min\{k : 1 \leqslant k \leqslant K,\ \hat{W}_k - k\log n \geqslant \hat{W}_j - j\log n\ \forall\ j = 1, ..., K\}.$$

Hence, as far as the testing is concerned, we directly obtain the result of major importance:

Proposition 3.2. *Under the assumptions of Theorem 3.1 it holds*

$$\hat{W}_k(\hat{\eta}_*) \xrightarrow{D} \chi^2_k \text{ and } \hat{W}_{\hat{S}} \xrightarrow{D} \chi^2_1 \tag{3.5}$$

under P_h as $n \to \infty$.

Detailed remarks concerning the LeCam's discretization method can be found in [15]. Here we just mention its basic concept. Suppose the parameter space $\Theta \times \Lambda$ is partitioned into cubes with edges of length $O(n^{-1/2})$. The estimator $\hat{\eta}_* = [\hat{\vartheta}^T_*,\ \hat{\lambda}^T_*]^T$ is defined to be a discretized version of $\hat{\eta}$ as the center of the cube to which $\hat{\eta}$ belongs (see e.g. [2], p. 44). The origination of the vector $\hat{\ell}^*(\hat{\eta}_*)$ in (3.1) and the test statistic (3.5) follow accordingly.

4. Simulation study

The test performance in practice was also examined, both for simple and composite hypothesis in [7] and [15], respectively. In the simple case, conditional normality and standard Laplace distribution were considered as null hypotheses, whereas in the composite framework the class $\mathcal{G}$ of standardized Generalized Error Distributions (GED) was taken as a hypothetical family. Recall that the *p.d.f.* of this (standardized, therefore one-parameter) class describing the random behavior of our innovation sequence $\{\varepsilon_t\}$ has the form

$$f(x, \lambda) = \frac{\lambda C_\lambda}{2\Gamma(\lambda^{-1})}\exp(-|C_\lambda x|^\lambda), \tag{4.1}$$

where $x \in R$, $\lambda \in \Lambda = R_+$ is a parameter indexing the family, $C_\lambda = \sqrt{\Gamma(3\lambda^{-1})/\Gamma(\lambda^{-1})}$ is a normalizing constant and Γ denotes the Euler gamma function. This symmetric class is flexible enough to encompass the Laplace ($\lambda = 1$) and normal ($\lambda = 2$) distributions as special cases, see e.g. [12]. Validity of the formerly mentioned assumptions from among (A1)–(A13) concerning the GED family has been checked in Section 9 of Stawiarski [15].

The issue of estimating the nuisance parameter η is vital. As for its GARCH(1,1) part ϑ, the Quasi Maximum Likelihood Estimator $\hat{\vartheta}$, QMLE, well-described in literature, was computed by iterative methods. Under mild conditions (cf. [11]) the QMLE $\hat{\vartheta}$ is square-root consistent and asymptotically normal with the asymptotic covariance matrix $\Sigma = Var(\varepsilon^2)U_\infty^{-1}$, where U_∞ appears in (6.11) below. Hence, for fixed $h > \omega\ (1-\beta)^{-1}$ we get $\hat{Q}_t$ and the estimated innovations $\hat{\varepsilon}_t = X_t \Big/ \sqrt{\hat{Q}_t}$, $t = 1, .., n$ appearing in (3.1). This estimated sequence, $\{\hat{\varepsilon}_t\}_{t=1,\ldots,n}$, serves in turn to obtain the $\sqrt{n}$-consistent estimator $\hat{\lambda}$ of λ stemming from the hypothetical family $\mathcal{G}$, given by (4.1), subject to our testing procedure. Specifically, $\hat{\lambda}$ is calculated numerically by common method of moments. Its $\sqrt{n}$-consistency is not trivial and was proved in Section 9 of [15].

Accordingly, all other matrices and quantities appearing in the data-driven score statistic $W_S$had to be estimated, yielding finally the data-driven test statistic $\hat{W}_{\hat{S}}(\hat{\eta})$, see (3.4). Details can be found in Section 5 of [15], together with versatile discussion concerning numerical issues emerging during estimating critical values and performing the test power study against vast scope of alternatives. Especially, using the asymptotic, nominal χ_1 quantiles is largely disputable due to present time series framework implying slower asymptotic convergence for moderate path lengths.

We set the quadratic scale coefficient $\omega = 0.001$ and the starting value $h = 0.1$ throughout our simulations. The maximal embedding dimension K was fixed at 10, and the cosine orthonormal system $\Phi_j(x) = \sqrt{2}\cos(j\pi x)$, $j = 1, 2, \ldots$, on [0,1] was chosen. The significance level is fixed at 0.05. To show the influence of nuisance parameters on the empirical critical values, we considered various combinations of ϑ and λ. Theoretical results imply the stability of c.v.'s with respect to changing η, hence the critical value was proposed as a common average.

Focusing upon $1 \leqslant \lambda \leqslant 2$ is motivated by empirical research in modelling real stock and commodities returns, see e.g. [4], [16]. Changing λ from 2 down to 1 results in heavier tails of the distribution. Various values of λ were paired with several combinations of the GARCH(1,1) parameters (α, β). Again, in accordance with results found in papers concerning econometric time series modelling, we deliberately focused on ϑ's such that $\beta > \alpha$ and $\alpha + \beta \geqslant 0.8$. In such cases, the influence of past conditional variances h_s on the present h_t, is stronger, providing more pronounced autocorrelation structure ("memory") within the squares of the time series. The results are reported in Table 1 as empirical .95-quantiles from 5000 Monte Carlo runs for sample size $n = 1000$.

The observed stability of estimated critical values justifies using the global average 4.501 as the critical value of our test for $n = 1000$. Under the null hypothesis, the Schwarz selection rule picks out $\hat{S} = 1$ with frequency about 95-97%. Slow convergence rate of the estimators, especially $\hat{\lambda}$, suggests caution with time series of length e.g. $n = 500$ in the composite hypothesis case, while for simple hypotheses considered in

Tab. 1: Simulated critical values for $\hat{W}_{\hat{S}}$ in case of GED family of null distributions with $\lambda \in (0.99; 2.01)$. Significance level 0.05, $\omega = 0.001$, $h = 0.1$, $K = 10$, $n = 1000$, $MC = 5000$ Monte Carlo loops

$(\alpha\ ,\ \beta)$	λ				
	1	1.25	1.5	1.75	2
(0.3; 0.5)	4.543	4.334	4.525	4.464	4.306
(0.2; 0.7)	5.025	4.770	4.871	4.280	4.195
(0.4; 0.5)	4.378	4.492	4.770	4.324	4.311
(0.25; 0.65)	4.860	4.474	4.769	4.255	4.069
Column-wise averages	4.701	4.517	4.734	4.331	4.220
Average critical value : 4.501					

[7] such length was large enough.

Proceeding now to checking the test performance, we consider $n = 1000$ and take the critical value $\overset{\wedge}{CV} = 4.501$ at 0.05 significance level (the asymptotic CV being 3.84). Following alternative distributions (centered and scaled when necessary) have been considered:

-t-Student with a degrees of freedom, referred to as $t(a)$;

-chi-square with 5 degrees of freedom, χ^2_5;

-normal mixture (bimodal) with $\mu > 0$ and $\sigma = 1$, $SN(\mu)$, with the density $f_1(x, \mu) = \frac{1}{2\sqrt{2\pi}} \left(\exp\{-(x-\mu)^2/2\} + \exp\{-(x+\mu)^2/2\}\right), x \in R$;

-first-type beta with $a, b > 0$ including uniform $(a, b = 1)$, $B(a, b)$;

-symmetric Pareto-type with shape $b > 2$, $PR(b)$, given by the density

$f_1(x, b) = \frac{b}{\sqrt{2(b-1)(b-2)}} \left(1 + \frac{|x|\sqrt{2}}{\sqrt{(b-1)(b-2)}}\right)^{-b-1}, x \in R$;

-GED(λ) with $\lambda > 2$.

To check the test performance and sensitivity when necessary, the "contaminated" alternatives f_ρ were considered, namely

$f_\rho(x, \lambda) = (1-\rho)f(x, \lambda) + \rho\ f_1(x)$,where $0 \leqslant \rho \leqslant 1$, $f(x, \lambda)$ is the null GED density (4.1) and $f_1(x)$ is taken from the above-listed alternatives. The simulations were run for the specific GARCH(1,1) model, namely with $\vartheta = (0.001;\ 0.3;\ 0.5)$, and the power was estimated as percentage of H_0^* rejections out of 2000 Monte Carlo loops. The results are collected in Tables 2 and 4.

Naturally, the power is generally weaker than that reported in [7] for the simple hypothesis. However, such deviations from GED as bimodality, skewness, excess kurtosis are detected satisfactorily well. Some light-tailed distributions are hardly distinguished from the GED null family, while ultra-thin tails, as those of GED(λ) with $\lambda > 6$ are detected quite well. Heavy Pareto-type tails are easily distinguished,

Tab. 2: Simulated powers of $\hat{W}_{\hat{S}}(\hat{\eta})$ under symmetric, unimodal alternatives. Significance level 0.05, $h = 0.1$, $K = 10$, $n = 1000$, $MC = 2000$ Monte Carlo loops; contaminations of conditionally GED GARCH(1,1) model with $\vartheta = (0.001;\ 0.3;\ 0.5)$ and three values of λ. H_0: f - GED(λ); $\lambda \in [1, 2]$

f_1	ρ	Simulated power (% of rejections)		
		$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$
$PR(2.5)$	1	41		
	0.8	19	19	20
$PR(5.5)$	1	19		
$PR(5)$	1	11		
$t(3)$	1	59		
	0.8	39	43	44
	0.6	30	33	34
$t(5)$	1	22		
GED(4)	1	5		
GED(6)	1	7		
GED(8)	1	7		
B(1,1)	0.8	97	81	62
	0.6	90	72	43
	0.4	34	30	28
B(1.2; 1.2)	1	25		
B(2,2)	1	18		

Tab. 3: Simulated powers of $\hat{W}_{\hat{S}}(\hat{\eta})$ under skew and bimodal alternatives. Significance level 0.05, $h = 0.1$, $K = 10$, $n = 1000$, $MC = 2000$ Monte Carlo loops; contaminations of conditionally GED GARCH(1,1) model with $\vartheta = (0.001;\ 0.3;\ 0.5)$ and three values of λ. $H_0 : f-$ GED(λ); $\lambda \in [1, 2]$

f_1	ρ	Simulated power (% of rejections)		
		$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$
$B(1;\ 1.5)$	1	100		
	0.8	98	93	87
	0.6	72	57	46
χ^2_5	0.6	99	100	100
	0.4	68	74	80
	0.2	16	21	23
$SN(1.6)$	0.8	99	98	98
	0.6	86	87	82
	0.4	50	43	41
$SN(1.2)$	1	19		

provided that $2 < b < 3$. Overall, the sensitivity of the test against wide range of alternatives is satisfactory.

The proposed data-driven methodology works well and can be used as an omnibus testing tool against various type alternatives. Other null hypotheses also can be considered upon checking the distributional assumptions imposed in our paper. More accurate and reliable conditional distribution testing procedure can substantially improve the quality of empirical time series modeling and resulting inference, providing fundamentals to better deal with financial engineering, including risk management, hedging, option pricing issues, especially in the face of lingering economic instability, market anomalies like asset bubbles, crashes translating into non-gaussian, heavy-tail and often skew, asymmetric indices, stocks or commodities returns.

5. Score test of fit in general GARCH(p,q) case

Now, we aim at extension of the previously obtained data-driven score test of fit, valid for GARCH(1,1) case to the general, finite dimensional GARCH class. This will provide us with useful, more flexible testing tool within the whole GARCH family, but at the price of some more complex derivations. Specifically, the technical lemmas stated in [7] and [15] carry over to the present situation under appropriate reformulations handling the higher dimensionality of the problem. Here we provide main outline of the extension construction, while some minor work might be an object of future research.

According to the pioneer definition in Bollerslev (1986), the class of symmetric GARCH(p,q) models allows q-step backward dependency upon squared series values, as well as p-step backward dependency upon its past conditional variances, namely:

$$\begin{cases} X_t = \sqrt{h_t}\varepsilon_t \\ h_t = \omega + \sum\limits_{i=1}^{q} \alpha_i X_{t-i}^2 + \sum\limits_{j=1}^{p} \beta_j h_{t-j} \end{cases} \tag{5.1}$$

with zero-mean, unit variance white noise $\{\varepsilon_t\}$, and $\omega > 0$, α_i, $\beta_j \geqslant 0$ but α_q, $\beta_p > 0$.

As we employ the data-driven score test methodology to strictly stationary series with finite second unconditional moment, let us quote some already known theorems establishing necessary and sufficient conditions for stationarity. To this end, it is convenient to express the GARCH(p,q) model as a vector Markov process $Z_t = B_t + A_t Z_{t-1}$, where

$$B_t = \left(\omega\varepsilon_t^2, 0, ..., \omega, 0, ..., 0\right)^T \in R^{p+q},\ Z_t = \left(X_t^2, ..., X_{t-q+1}^2, h_t, ..., h_{t-p+1}\right)^T \in R^{p+q},$$

$$A_t = \begin{bmatrix} \alpha_1\varepsilon_t^2 & & ... & & \alpha_q\varepsilon_t^2 & \beta_1\varepsilon_t^2 & & ... & & \beta_p\varepsilon_t^2 \\ 1 & 0 & ... & 0 & 0 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 & 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & & ... & 1 & 0 & 0 & 0 & ... & 0 & 0 \\ \alpha_1 & & ... & & \alpha_q & \beta_1 & & ... & & \beta_p \\ 0 & 0 & ... & 0 & 0 & 1 & 0 & ... & 0 & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & 1 & ... & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & ... & 0 & 0 & 0 & 0 & ... & 1 & 0 \end{bmatrix} \tag{5.2}$$

Thus A_t is a square, sparse $(p+q) \times (p+q)$ matrix with stochastic first row. Now, successively iterating the formula for Z_t we get the strictly stationary solution

$$Z_t = B_t + \sum_{i=1}^{\infty} \left(\prod_{j=0}^{i-1} A_{t-j} \right) B_{t-i} \tag{5.3}$$

provided the almost sure existence of the series. The examination of strict stationarity can be carried out by means of top Lyapunov exponent γ associated with the sequence of (strictly stationary and ergodic) matrices $A_t, t \in Z$. Assuming that $E\ \log^+ \|A_t\| < \infty$ holds, (here $x^+ = \max\{x; 0\}$) the top Lyapunov exponent can be derived as [cf. [5]]

$$\gamma = \lim_{t\to\infty} t^{-1}E(\log||A_tA_{t-1}...A_1||) = a.s. \lim_{t\to\infty} t^{-1}\log||A_tA_{t-1}...A_1||.$$

Now we cite two theorems concerning strict and weak (second-order) stationarity of the GARCH process, which are the objects of our interest.

Theorem 5.1. *[Theorem 2.4 in [5]]*
Let γ be the top Lyapunov exponent of the sequence $\{A_t\}$ given by (5.2). The process (5.1) admits strictly stationary solution if and only if $\gamma < 0$. Such a solution is also nonanticipative (with respect to process filtration) and ergodic.

Theorem 5.2. *[Theorem 2.5 in [5]]*
If the process obeying (5.1) is weakly stationary and nonanticipative, then

$$\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j < 1. \tag{5.4}$$

If, conversely, (5.4) holds, the unique strictly stationary solution of (5.1) is weakly stationary.

Remark. Under the conditions of Theorem 5.1 the unconditional variance of GARCH(p,q) model equals

$$EX_t^2 = v^2 = \frac{\omega}{1-\sum\limits_{i=1}^{q}\alpha_i - \sum\limits_{j=1}^{p}\beta_j}. \tag{5.5}$$

Moreover, (5.4) implies that $\gamma < 0$ as the weakly stationary solution stated in Theorem 5.1 is also strictly stationary. In the case of GARCH(1,1) checking the Lyapunov exponent in Theorem 5.1 simplifies greatly: the stationarity condition reads as $\gamma = E\log(\alpha\varepsilon_t^2 + \beta) < 0$.

Further vast and versatile theoretical results concerning not only the GARCH models but their various modifications (xARCH), together with broad empirical applications, simulations etc. can be found in [5] and in numerous preceding papers. Now, we proceed to extending our former results derived in [7] and [15] to finite dimensional ARCH family, and then finally for the whole GARCH(p,q) class.

5.1. ARCH($\boldsymbol{q}$) case

For ARCH(q) submodel with no dependence on past conditional variances,

$$\begin{cases} X_t = \sqrt{h_t}\varepsilon_t \\ h_t = \omega + \sum\limits_{j=1}^{q}\alpha_j X_{t-j}^2 \end{cases} \tag{5.6}$$

with $\omega > 0$ and non-negative α_j's but $\alpha_q > 0$, we shall test the conditional distribution of X_t by addressing the distribution of innovations ε_t. With the observed strictly (and weakly) stationary series $(X_1, \ldots, X_n)$ obeying (5.6), and its induced joint distribution taken, as before, conditionally on $h_1 = h > \omega(1 - \alpha_1 - \ldots - \alpha_q)^{-1}$, we aim at deriving our data-driven test statistic. This can be done by repeating vast part of calculus done for GARCH(1,1) model along the lines from Section 2 and 3, but there are indeed some differences in the score vector and the efficient score vector as now the dynamics of h_t is somewhat different than that of GARCH(1,1). The nuisance parameter in our testing problem becomes now $(q+1+m)$-dimensional, namely $\eta = (\omega, \alpha_1, \ldots, \alpha_q, \lambda_1, \ldots, \lambda_m)^T \in \Theta \times \Lambda$, where the set Θ of ARCH parameter values ensures finite variance of X_t.

In this case, too, we have to derive successively: score vector, efficient score vector and score-statistic with its data-driven version, culminating in the test statistic with nuisance parameters, estimated regularly enough. We put into closer inspection the stages in which derivation differs significantly from the previously considered GARCH(1,1) model. Such a first notable difference is obviously the conditional variance h_t in (5.6), compare (2.1). Successive iteration in a manner leading to formula (2.4) does not carry over which is not cumbersome until some details in proofs of the limit theorem paralleling Theorems 2.2 and 3.1.

The second part of the score vector $\ell(\eta)$ given in (2.6), namely ℓ_ϑ, changes accordingly to difference between dynamics of conditional variance in GARCH(1,1) and ARCH(q) models, compare (2.1) and (5.6). Now, instead of (2.7) we have, conditionally on h_1 for $t \geqslant 2$

$$Q_t = Q_t(X_1, ..., X_{t-1}; h, \vartheta) = \omega + \sum_{j=1}^{\min\{t-1;q\}} \alpha_j X_{t-j}^2 \tag{5.7}$$

with $Q_1 = h$, while

$$\frac{\partial Q_t}{\partial \vartheta} = \left[\frac{\partial Q_t}{\partial \omega}, \frac{\partial Q_t}{\partial \alpha_1}, \ ..., \ \frac{\partial Q_t}{\partial \alpha_q}\right]^T = \begin{bmatrix} 1 \\ X_{t-1}^2 \left(\mathrm{I}\{t > q\} + \mathrm{I}\{1 < t \leqslant q\}\right) \\ \vdots \\ X_{t-j}^2 \left(\mathrm{I}\{t > q\} + \mathrm{I}\{j < t \leqslant q\}\right) \\ \vdots \\ X_{t-q}^2 \left(\mathrm{I}\{t > q\} + \mathrm{I}\{q < t \leqslant q\}\right) \end{bmatrix}^T \tag{5.8}$$

except for the first row always being 1, contains X_{t-j}^2's or just zeroes depending on whether $t > q$ or whether t falls between $j + 1$ and q; $j = 1, \ldots, q$. The formula (2.6) retains its shape but it must be underlined that ℓ_ϑ is now $(q + 1)$ dimensional in accordance with (5.8), therefore the whole $\ell(\eta)$ is now $(k + q + 1 + m)$-dimensional.

The present ARCH(q) setup implies generally larger (for $q > 2$) dimensions of the block matrices constituting $\tilde{B}^{(n)}(\eta)$ in (2.8). The whole score statistic construction goes along the same lines as in Section 2. The main proofs, stating the asymptotic result concerning W_k and W_S apply, too, but with one important remark. Namely, some stages of the proofs derivation call for a bit modified re-writing whenever we deal with blocks containing the "subvector" ℓ_ϑ. Main obstacle to overcome is the fact that it is by a long chalk harder to descend down to h_1 in (5.7), than it was possible to do outright in (2.7) by successive iterations. Obviously, it is attainable via vector representation like (5.2), but in that case the crucial Lemmas 5.2 and 5.3 in [7] providing the limit behaviors of $\bar{V}_n$, $E_h\bar{V}_n$, $\bar{U}_n$ and $E_h\bar{U}_n$ (cf. (6.1) and (6.2)) under P_h as $n \to \infty$ deal with some other expressions. Specifically, in the present context, it remains to show that

$$E_h\left\{\frac{1}{n}\sum_{t=2}^{n}\frac{1}{Q_t}\right\} \stackrel{n\to\infty}{\longrightarrow} \Psi_0 \quad , \quad E_h\left\{\frac{1}{n}\sum_{t=2}^{n}\frac{X_{t-j}^2}{Q_t}\right\} \stackrel{n\to\infty}{\longrightarrow} \Psi_j \tag{5.9}$$

under P_h for $j = 1, \ldots, q$. Obviously, the same argument as in Appendix of [7] can be used here, namely the Birkhoff's ergodic theorem applied for appropriately defined L_1-integrable functional operating on the sequence of squared innovations $\{\varepsilon_t^2\}$ driving our model (5.6). In the GARCH(1,1) context, conditionally on $h_1 = h$, Q_t could be expressed outright as

$$Q_t = \omega\left(1 + \sum_{i=1}^{t-2}\prod_{j=0}^{i-1}(\beta + \alpha\tilde{\varepsilon}_t^2)\right) + h\prod_{j=0}^{t-1}(\beta + \alpha\tilde{\varepsilon}_t^2) \tag{5.10}$$

which in Section 6 of [7] served to prove the crucial Lemmas 5.2 and 5.3, with the aid of the fact that $E\tilde{e}_t^2 = E(\alpha\tilde{\varepsilon}_t^2 + \beta) < 1$. Now, to prove (5.9) for the ARCH(q) case, one has to express Q_t tracking back down to h similarly as in (5.10) but taking the model difference into account. The analogue of (5.10) in this context will contain cumulative products of variables being now $\tilde{e}_{t,j}^2 = \alpha_j\tilde{\varepsilon}_{t-j}^2$, with $E\tilde{e}_{t,j}^2 < 1$ for $j = 1, \ldots, q$ and any t. Formal proof of (5.9) calls for readjustment in calculus and is the object of current research. Thank to the general properties of ARCH model resulting from our assumptions, the validity of (5.9) can be ascertained with next to zero risk of failure. This will imply the validity of the whole Section 4, as the martingale-difference-array structure of the efficient score vector $\ell^*(\eta)$ in (2.9) remains intact and the Kundu et al. [9] limit theorem applies accordingly.

Repeating the proof of Theorem 3.1 for the data-driven score statistic with QMLE $\hat{\vartheta}$ of the ARCH nuisance parameter, and the estimator $\hat{\lambda}$ as before follows directly. Obviously, detailed estimations and derivations in the auxiliary lemmas in Section 7 of [15] will change accordingly, but with no detriment to the validity of the main limit theorem.

5.2. GARCH($\boldsymbol{p,q}$) case

For the whole GARCH(p,q) model (5.1) with $r = \max\{p,\, q\} \geqslant 2$ we obviously have

$$h_t = \omega + \sum_{j=1}^{r} (\alpha_j \varepsilon_{t-j}^2 + \beta_j) h_{t-j}, \tag{5.11}$$

but successive iteration in (5.11) for each of h_{t-j} down to h_1 and conditionally on $h_1 = h$ exceeding the right-hand side of (5.5) will not yield any constructive form of Q_t from our test derivation standpoint. Indeed, the products of expressions $(\alpha_j \varepsilon_{t-j}^2 + \beta_j)$ will proliferate so amply that the derivation of the score vector $\ell(\eta)$ defined in (2.5), (2.6) in the closed form is unattainable due to the cumbersome derivative $\frac{\partial Q_t}{\partial \vartheta}$ in (2.7). The same remark applies to the block covariance matrix $\tilde{B}^{(n)}(\eta)$ in (2.8) and the efficient score vector $\ell^*(\eta)$ in (2.9), due to the indisposable component ℓ_ϑ. The technically intricate and computationally hardly tractable matrix representation (5.2), using higher-order Kronecker products, does not facilitate the task, either.

However, we can propose a circumvention of these major technical obstacles, transforming the GARCH(p,q) model into the reparametrized ARCH(∞) model, then consequently truncating it at some fixed lag, say Q. This comes at a minor expense of disposing the exponentially vanishing process history tracking back beyond $(t-Q)$ time scale, but allows us to employ the results from the above subsection, dealing again with ARCH(Q) model. Thus our data-driven score test of fit (asymptotically) encompasses the whole GARCH(p,q) family. Such a shortcut comes with no harm to practical applications, since it is common to consider lower or moderate model sizes, so that the model can successfully serve for applicational purposes imposed by financial time modeling objectives.

Formalizing our concept, let us quote a theorem providing the needed transformation. Such an "inversion" of GARCH model to ARCH(∞) goes in a similar spirit to the classical case of causal and invertible (linear) ARMA models. The result has already been stated in the pioneer paper of Bollerslev [3], pp. 309-310, refined later on by other authors.

Theorem 5.3. *For a strictly stationary GARCH(p,q) model $\{X_t\}$ following (5.1), with $E\varepsilon_t^2 = \sigma^2 < \infty$ and $\sigma^2 \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$, there exists a non-negative, summable sequence $\{\psi_j\}_{j\geqslant 0}$ such that*

$$h_t = \psi_0 + \sum_{j=1}^{\infty} \psi_j X_{t-j}^2 \tag{5.12}$$

with expansion coefficients given as

$$\begin{aligned} &\psi_0 = \frac{\omega}{1-\sum\limits_{r=1}^{p} \beta_r}, \\ &\sum_{j=1}^{\infty} \psi_j z^j = \frac{\sum\limits_{i=1}^{q} \alpha_i z^i}{1-\sum\limits_{r=1}^{p} \beta_r z^r},\ j \geqslant 1,\ z \in C; |\ z| \ < 1 \end{aligned} \tag{5.13}$$

With this result, having the GARCH(p,q) model with $\sigma^2 = 1$ and original parameters ω, α_i, β_j we can reparametrize it into ARCH(∞), solving the system (5.13) to find the new coefficients ψ_0, ψ_1,.... Next, the resulting conditional variance h_t in (5.12) can be truncated to ARCH(Q) representation with Q picked so that an arbitrarily chosen accuracy measured e.g. by L_1 norm can be obtained. Define

$$h_t^{(Q)} = \psi_0 + \sum_{j=1}^{Q} \psi_j X_{t-j}^2 \tag{5.14}$$

and for any fixed accuracy $\delta > 0$ choose Q so that (cf. (5.5))

$$E\left\{h_t - h_t^{(Q)}\right\} = \sum_{r=Q+1}^{\infty} \psi_r E X_{t-r}^2 = \nu^2 \sum_{r=Q+1}^{\infty} \psi_r < \delta. \tag{5.15}$$

Finally, we can employ the former derivation of our data-driven score test of fit for the ARCH(Q) model parallel to (5.1):

$$\begin{cases} X_t = \sqrt{h_t^{(Q)}}\varepsilon_t \\ h_t^{(Q)} = \psi_0 + \sum\limits_{i=1}^{Q} \psi_i X_{t-i}^2 \end{cases} \tag{5.16}$$

where after reparametrization now $\vartheta = [\psi_0, \psi_1, ..., \psi_Q]^T$. Thus, our data-driven score test of fit, designed primarily to verify the hypothetical conditional distribution in the GARCH(1,1) model has been "asymptotically" extended over the whole GARCH(p,q) class of stationary time series, and its performance also can be checked by means of computer simulations.

6. Appendix

We provide a brief proof of Theorem 2.2, in which Theorem 1.3 from Kundu et al. [9] for martingale-difference arrays is very helpful. The present arguments mimic the ones used in the proof of Theorem 3.6 in [7]. Reasoning goes along the lines of Subsection 6.2 in [15].

Let us first examine limiting behaviour of the matrices $\tilde{B}_{12}^{(n)}(\eta)$ and $\tilde{B}_{22}^{(n)}(\eta)$ appearing in (2.8), (2.9), (2.10) – originally in (6.7) and (6.12) in [15]. For $t \geqslant 2$ introduce random matrices

$$V_t = \frac{1}{Q_t}\frac{\partial Q_t}{\partial \vartheta}, \quad U_t = V_t V_t^T, \tag{6.1}$$

for notational convenience, set $V_1 = 0$, $U_1 = 0$ and define two corresponding random averages

$$\bar{V}_n = \frac{1}{n}\sum_{t=1}^{n} V_t, \quad \bar{U}_n = \frac{1}{n}\sum_{t=1}^{n} U_t. \tag{6.2}$$

By mutual independence of $\tilde{\varepsilon}_t$'s under P_h, straightforward calculation gives

$$n^{-1}E_h \ell_\tau \ell_\vartheta^T = -\Delta_1(E_h \bar{V}_n^T)/2, \tag{6.3}$$

where

$$\Delta_1 = \Delta_1(\lambda) = E\Phi(F(\varepsilon,\lambda))\varsigma(\varepsilon,\lambda) \tag{6.4}$$

with ς defined in (A2). Similarly we get

$$n^{-1}E_h \ell_\tau \ell_\lambda^T = E\Phi(F(\varepsilon,\lambda))\frac{\partial \log f(\varepsilon,\lambda)}{\partial \lambda^T} = \Delta_2(\lambda) = \Delta_2. \tag{6.5}$$

Now, (6.3) together with (6.5) yield the following block representation of $\tilde{B}_{12}^{(n)}(\eta)$:

$$\tilde{B}_{12}^{(n)}(\eta) = [-\Delta_1(E_h \bar{V}_n^T)/2 \ \Delta_2]. \tag{6.6}$$

By similar derivation we obtain

$$n^{-1}E_h \ell_\vartheta \ell_\vartheta^T = J_\varsigma(E_h \bar{U}_n)/4 \tag{6.7}$$

with $J_\varsigma = J_\varsigma(\lambda) = E\varsigma^2(\varepsilon,\lambda)$. We further have

$$n^{-1}E_h \ell_\vartheta \ell_\lambda^T = -(E_h \bar{V}_n)\Delta_3^T/2, \tag{6.8}$$

$$\Delta_3 = \Delta_3(\lambda) = E\left\{\varsigma(\varepsilon,\lambda)\frac{\partial \log f(\varepsilon,\lambda)}{\partial \lambda}\right\}. \tag{6.9}$$

Finally, $n^{-1}E_h \ell_\lambda \ell_\lambda^T = I(\lambda)$ - the Fisher information matrix defined in (A1). This and (6.7)-(6.9) lead to the following block form of $\tilde{B}_{22}^{(n)}(\eta)$

$$\tilde{B}_{22}^{(n)}(\eta) = \begin{bmatrix} J_\varsigma(E_h \bar{U}_n)/4 & -(E_h \bar{V}_n)\Delta_3^T/2 \\ -\Delta_3(E_h \bar{V}_n^T)/2 & I(\lambda) \end{bmatrix}. \tag{6.10}$$

By Lemma 5.2 and 5.3 from [7] it follows that

$$\bar{V}_n \xrightarrow{P_h} V_\infty, \ E_h \bar{V}_n \to V_\infty,$$

$$\bar{U}_n \xrightarrow{P_h} U_\infty, E_h\bar{U}_n \to U_\infty \tag{6.11}$$

as $n \to \infty$, where V_∞ and U_∞ are deterministic matrices given explicitly by (A.12) and (A.13) in Appendix of [7]. Hence

$$\tilde{B}_{12}^{(n)}(\eta) \underset{n\to\infty}{\longrightarrow} \tilde{B}_{12}^{\infty}(\eta) = [\ -\Delta_1 V_\infty^T/2 \quad \Delta_2\ ] \tag{6.12}$$

and

$$\tilde{B}_{22}^{(n)}(\eta) \underset{n\to\infty}{\longrightarrow} \tilde{B}_{22}^{\infty}(\eta) = \begin{bmatrix} J_\varsigma U_\infty/4 & -V_\infty \Delta_3^T/2 \\ -\Delta_3 V_\infty^T/2 & I(\lambda) \end{bmatrix}. \tag{6.13}$$

For further convenience, let us introduce the abbreviated notation for the projection matrix in (2.9),

$$\tilde{A} = \tilde{A}^{(n)}(\eta) = \tilde{B}_{12}^{(n)}(\eta)[\tilde{B}_{22}^{(n)}(\eta)]^{-1}. \tag{6.14}$$

Thus, (6.12) and (6.13) directly imply that for $n \to \infty$
$\tilde{A}^{(n)}(\eta) \to \tilde{A}^{\infty}(\eta) = \tilde{B}_{12}^{\infty}(\eta)[\tilde{B}_{22}^{\infty}(\eta)]^{-1}$and $\tilde{M}^{(n)}(\eta) \to \tilde{M}^{\infty}(\eta) = I_k - \tilde{B}_{12}^{\infty}(\eta)[\tilde{B}_{22}^{\infty}(\eta)]^{-1}\tilde{B}_{21}^{\infty}(\eta)$.

Now, we can view $\ell^*(\eta)$ as a sum composed of martingale difference array summands. To this end, define for $t \geqslant 1$ σ-fields $\sigma_t = \sigma(X_1, ..., X_t)$ and

$$X_{tn} = \tilde{M}^{-1/2}\frac{1}{\sqrt{n}}\left(\Phi(F(\tilde{\varepsilon}_t,\lambda)) - \tilde{A}\begin{bmatrix} -\varsigma(\tilde{\varepsilon}_t,\lambda)V_t/2 \\ \frac{\partial \log f(\tilde{\varepsilon}_t,\lambda)}{\partial\lambda} \end{bmatrix}\right), \tag{6.15}$$

with $\tilde{M}$ given by (2.10).

It is straightforward to show that the sequence $\{X_{1n}, ..., X_{nn}\}$ is, under P_h, a martingale difference array adapted to $\{\sigma_1, ..., \sigma_n\}$ (cf. Proposition 3.5 in [7]) and $\sum_{t=1}^{n} X_{tn} = \tilde{M}^{-1/2}\ell^*(\eta)$. This observation allows us to apply Theorem 1.3 from [9].

Since (A4) is satisfied, checking the Lindeberg-type condition (ii) of that theorem can be replaced with the following stronger, Lyapunov-type one
$\sum_{t=1}^{n} E_h\{||X_{tn}||^3|\sigma_{t-1}\} \xrightarrow{P_h} 0$ as $n \to \infty$.

However, this is carried out along exactly the same lines as in the proof of Theorem 3.6 in [7], therefore we omit the details citing just this reference.

Turning now to condition (i) in the aforementioned theorem of Kundu et al. it suffices to show that

$$\sum_{t=1}^{n} E_h\{X_{tn}X_{tn}^T|\sigma_{t-1}\} \xrightarrow{P_h} I_k \tag{6.16}$$

in R^k as $n \to \infty$, with X_{tn} as in (6.15). According to the form of X_{tn} the left-hand side of (6.16) can be decomposed into $S_{1n} + S_{2n} + S_{3n}$. By Proposition 3.1 in [15] and orthonormality of Φ,

$$S_{1n} = \tilde{M}^{-1/2} E\Phi(F(\varepsilon,\lambda))\Phi^T(F(\varepsilon,\lambda))\tilde{M}^{-1/2} = \tilde{M}^{-1} \to (\tilde{M}^\infty)^{-1} \text{as } n \to \infty. \quad (6.17)$$

Recalling that $\tilde{\varepsilon}_t$ is independent of σ_{t-1} and V_t is $\sigma_{t-1}-$measurable, from (6.2), (6.4) and (6.5) we immediately get

$$S_{2n} = \tilde{M}^{-1/2}\{[\ \Delta_1 \bar{V}_n^T/2 \quad -\Delta_2\ ]\tilde{A}^T + \tilde{A}[\ \bar{V}_n \Delta_1^T/2 \quad -\Delta_2^T\ ]^T\}\tilde{M}^{-1/2} \quad (6.18)$$

and consequently, by (6.11)-(6.14) and the forms of $\tilde{A}^\infty(\eta)$ and $\tilde{M}^\infty(\eta)$,

$$S_{2n} \xrightarrow{P_h} -2(\tilde{M}^\infty)^{-1/2}\tilde{B}_{12}^\infty(\tilde{B}_{22}^\infty)^{-1}\tilde{B}_{21}^\infty(\tilde{M}^\infty)^{-1/2} = 2I_k - 2(\tilde{M}^\infty)^{-1} \text{ as } n \to \infty. \quad (6.19)$$

Exploiting similar arguments and from (6.2), (6.7)-(6.14) we obtain

$$S_{3n} = \tilde{M}^{-1/2}\tilde{A}\begin{bmatrix} J_\varsigma \bar{U}_n/4 & -\bar{V}_n\Delta_3^T/2 \\ -\Delta_3\bar{V}_n^T/2 & I(\lambda) \end{bmatrix}\tilde{A}^T\tilde{M}^{-1/2} \xrightarrow{P_h} (\tilde{M}^\infty)^{-1} - I_k \text{ as } n \to \infty. \quad (6.20)$$

Finally, we conclude (6.16) from (6.17)-(6.20), which completes the proof. *Q.E.D.*

7. Final remarks and acknowledgements

The paper essentially solves the problem of testing the conditional distribution in the framework of general (symmetric) GARCH(p,q) models via the omnibus data-driven score test of fit methodology. Composite hypothesis case naturally embraces the simple one by including additional nuisance parameter λ. Obviously, some minor technical readjustments and restatements of the lemmas and propositions proved in [7] and [15] are welcome to absolutely fully complete the testing problem, whereas the simulation study and practical test performance for empirical data could prove the usefulness of the proposed tool. The possible future computational paper can also contain the primary model diagnostics, while the implementation of the test statistic itself goes exactly along the same lines as described and practiced in the above-mentioned former papers, but with taking into account the presented extension which numerically can be handled outright. However, these issues fall beyond the scope of this paper, due to volume constraints, too.

The author would like to thank two referees for their constructive comments and remarks, improving the clarity and transparency of this paper.

References

[1] Bai J., *Testing parametric conditional distributions of dynamic models.* Rev. Econ. Stat. 85, 2003, 531—549.

[2] Bickel P.J., Klaassen C.A.J., Ritov Y., Wellner J.A. *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins Press, London, 1993.

[3] Bollerslev T., *Generalized autoregressive conditional heteroskedasticity.* J. Econometrics 31, 1986, 307—327.

[4] Chen Y.-T., *A new class of characteristic-function-based distribution tests and its application to GARCH Model.* Working paper, Institute for Social Sciences and Philosophy, Academia Sinica, 2002.

[5] Francq C., Zakoïan J.M., *GARCH models. Structure, statistical inference and financial applications.* Wiley, UK. 2010.

[6] Inglot, T., Kallenberg, W.C.M., Ledwina, T. (1997). *Data driven smooth tests for composite hypotheses.* Ann. Stat. 25(3), 1222—1250.

[7] Inglot T., Stawiarski B., *Data-driven score test of fit for conditional distribution in the GARCH(1,1) model.* Prob. Math. Stat. 25, 2005, 331—362.

[8] Kallenberg W.C.M., Ledwina T, *Data driven smooth tests for composite hypotheses: comparison of powers.* J. Stat. Comp. Simul. 59, 1997, 101—121.

[9] Kundu S., Majumdar S., Mukherjee K., *Central Limit Theorems revisited.* Stat. & Prob. Letters 47, 2000, 265—275.

[10] Ledwina, T., *Data driven version of Neyman's smooth test of fit.* J. Amer. Stat. Assoc. 89, 1994, 1000–1005.

[11] Lumsdaine R.L., *Consistency and asymptotic normality of the Quasi Maximum Likelihood Estimator in IGARCH(1,1) and covariance stationary models.* Econometrica 64, 1996, 575—596.

[12] Mittnik S., Paolella M.S., Rachev S., *Unconditional and conditional distributional models for the Nikkei index.* Asia-Pacific Fin. Markets 5, 1998, 99—128.

[13] Neyman J., *Smooth test for goodness of fit.* Skand. Aktuarietidskr. 20, 1937, 149—199.

[14] Rao C. R., *Linear statistical inference and its applications.* Wiley, New York (973.

[15] Stawiarski B., *Score test of fit for composite hypothesis in the GARCH(1,1) model.* JSPI 139, 2009, 593—616.

[16] Varma J.R., *Value at Risk models in the Indian stock market*. Working paper, IIMA, 1999.