

A new hyperbolic GARCH model

Muyi Li^a, Wai Keung Li^b and Guodong Li^{b,*}

^a*School of Economics and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China*

^b*Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong*

Abstract

There are two commonly used hyperbolic GARCH processes, the FIGARCH and HYGARCH processes, in modeling the long-range dependence in volatility. However, the FIGARCH process always has infinite variance, and the HYGARCH model has a more complicated form. This paper builds a simple bridge between a common GARCH model and an integrated GARCH model, and hence a new hyperbolic GARCH model along the lines of FIGARCH models. The new model remedies the drawback of FIGARCH processes by allowing the existence of finite variance as in HYGARCH models, while it has a form nearly as simple as the FIGARCH model. Two inference tools, including the Gaussian QMLE and a portmanteau test for the adequacy of the fitted model, are derived, and an easily implemented test for hyperbolic memory is also constructed. Their finite sample performances are evaluated by simulation experiments, and an empirical example gives further support to our new model.

JEL Classification: C15; C22

Keywords: ARCH(∞), Hyperbolic GARCH, Long-range dependence, QMLE.

*Correspondence to: Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong. Tel.:+852 3917 1986; fax:+852 2858 9041. Email addresses: limuyi1981@gmail.com (M. Li), hrntlwk@hku.hk(W.K. Li), gdli@hku.hk (G. Li).

1 Introduction

Ding et al. (1993) and Dacorogna et al. (1993) first observed long-range dependence in squared returns, as well as other power transformed absolute returns, of asset prices; see also Ding and Granger (1996). Due to the popularity of ARCH-type models (Engle, 1982; Bollerslev, 1986), it is convenient to discuss such type of long memory feature under the framework of ARCH(∞) models (Robinson, 1991). We may follow the classical approach and define that a sequence of returns $\{y_t\}$ exhibits long memory in volatility if the autocovariance function $\text{cov}(y_t^2, y_{t-k}^2)$ is not absolutely summable. However, under the condition of $E(y_t^4) < \infty$, ARCH(∞) processes always have absolutely summable autocovariances; see Giraitis et al. (2000), Giraitis and Surgailis (2002) and Zaffaroni (2004). In order to measure the memory of an ARCH(∞) process more precisely, Davidson (2004) suggested the concept of hyperbolic memory instead of long memory, and the process is said to have hyperbolic (geometric) memory if its coefficients decay hyperbolically (geometrically). Note that the autocovariance function $\text{cov}(y_t^2, y_{t-k}^2)$ also decays hyperbolically (geometrically) for a hyperbolic (geometric) memory ARCH(∞) process. Consider the GARCH model,

$$y_t = \varepsilon_t \sqrt{h_t}, \quad h_t = \gamma + \sum_{i=1}^Q \alpha_i y_{t-i}^2 + \sum_{j=1}^P \beta_j h_{t-j} \quad \text{or} \quad h_t = \frac{\gamma}{\beta(1)} + \left\{ 1 - \frac{\delta^*(B)}{\beta(B)} \right\} y_t^2, \quad (1)$$

where $\{\varepsilon_t\}$ are identically and independently distributed (*i.i.d.*) with mean zero and variance one, $\gamma > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, B is the back shift operator, $\beta(x) = 1 - \sum_{j=1}^P \beta_j x^j$, $\alpha(x) = \sum_{i=1}^Q \alpha_i x^i$ and $\delta^*(x) = \beta(x) - \alpha(x)$; see Bollerslev (1986). It is obvious that model (1) has geometric memory. This paper will focus on the hyperbolic memory ARCH(∞) models originating from the GARCH model (1), and calls them hyperbolic GARCH models for simplicity.

For the integrated GARCH model, its conditional variance (1) has the form

$$h_t = \frac{\gamma}{\beta(1)} + \left\{ 1 - \frac{\delta_F(B)}{\beta(B)} (1 - B) \right\} y_t^2,$$

where $\delta^*(x) = \delta_F(x)(1 - x)$ since $\sum_{i=1}^Q \alpha_i + \sum_{j=1}^P \beta_j = 1$. By replacing $(1 - B)$ with $(1 - B)^d$ on the right hand side of the above equation, Baillie et al. (1996) proposed the fractionally integrated GARCH (FIGARCH) model, where $0 < d < 1$ and $(1 - B)^d =$

$1 - \sum_{j=1}^{\infty} \pi_j B^j$ with

$$\pi_j = \frac{d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)} = O(j^{-1-d}). \quad (2)$$

This is the first hyperbolic GARCH model in the literature. However, as for the integrated GARCH process, the FIGARCH process always has infinite variance, and this limits its applications. To overcome this drawback, Davidson (2004) proposed a hyperbolic GARCH (HYGARCH) model,

$$y_t = \varepsilon_t \sqrt{h_t}, \quad h_t = \frac{\gamma}{\beta(1)} + \left\{ 1 - \frac{\delta_H(B)}{\beta(B)} [1 - \phi + \phi(1-B)^d] \right\} y_t^2, \quad (3)$$

where $\phi > 0$. This model will reduce to the FIGARCH model if $\phi = 1$, and the variance of y_t is finite when $1 - (1 - \phi)\delta_H(1)/\beta(1) < 1$. Note that the polynomial $\delta^*(x)$ in (1) has a unique root on R^+ , say $1/\phi$, and in terms of ϕ the conditional variance of the GARCH model has the form

$$h_t = \frac{\gamma}{\beta(1)} + \left\{ 1 - \frac{\delta_H(B)}{\beta(B)} (1 - \phi B) \right\} y_t^2 = \frac{\gamma}{\beta(1)} + \left\{ 1 - \frac{\delta_H(B)}{\beta(B)} [1 - \phi + \phi(1-B)] \right\} y_t^2,$$

i.e. we can arrive at the HYGARCH model after replacing $(1-B)$ with $(1-B)^d$ in the above equation; see Li et al. (2011). Note that coefficients π_j s in (2) have more persistence as d decreases, and the memory of $(1-B)^d$ is continuous at $d = 1$ (Davidson, 2004). By letting $0 < d \leq 1$, the HYGARCH model can then be extended to encompass the common GARCH model with geometric memory (Li et al., 2011). Robinson and Zaffaroni (2006) also considered a hyperbolic GARCH model, however, it has not attracted much attention so far.

For an ARCH(∞) model, it is necessary to restrict all its coefficients to be non-negative such that the conditional variance $h_t \geq 0$ with probability one. Non-negativity conditions for FIGARCH and HYGARCH models were discussed respectively by Conrad and Haag (2006) and Conrad (2010). It is obvious that the HYGARCH model has more complicated restrictions than those of the FIGARCH model. Consider the conditional variance of the HYGARCH process in (3). It can be rewritten as $h_t = (1 - \phi)h_{1t} + \phi h_{2t}$, where

$$h_{1t} = \frac{\gamma}{\beta(1)} + \left\{ 1 - \frac{\delta_H(B)}{\beta(B)} \right\} y_t^2 \quad \text{and} \quad h_{2t} = \frac{\gamma}{\beta(1)} + \left\{ 1 - \frac{\delta_H(B)}{\beta(B)} (1-B)^d \right\} y_t^2; \quad (4)$$

see Li et al. (2011) and Li et al. (2013). It can be seen that the FIGARCH component h_{2t} already has the multiplicative form of a geometric decaying pattern $\delta_H(B)/\beta(B)$ and a hyperbolic decaying pattern $(1 - B)^d$, i.e. the conditional variance of the HYGARCH process may have an unnecessarily complicated form. Moreover, the parameters d and ϕ are both related to the memory of a HYGARCH process. This motivates us to consider a new decomposition of the GARCH model with a simpler structure, and it then leads to a new hyperbolic GARCH model in section 2. The proposed model allows the existence of finite variance as in HYGARCH models, while it has a form nearly as simple as FIGARCH models.

Section 3 derives the Gaussian quasi-maximum likelihood estimation (QMLE) for the new model and a portmanteau test for the adequacy of the fitted model. An easily implemented test for the hyperbolic memory is given in section 4. Section 5 conducts several Monte Carlo simulation experiments to study the finite sample performance of the Gaussian QMLE and two tests. An empirical example is reported in section 6. Proofs of theorems are relegated to the Appendix.

2 A new hyperbolic GARCH model

For the polynomial $\delta^*(x)$ in (1), we consider a decomposition, $\delta^*(x) = (1 - \omega)\beta(x) + \omega\delta_N(x)$, such that

$$\frac{\delta^*(B)}{\beta(B)} = (1 - \omega) + \omega \frac{\delta(B)}{\beta(B)}(1 - B),$$

where $\delta_N(1) = 0$ and $\delta_N(x) = \delta(x)(1 - x)$. After some calculation, we have that $\omega = \alpha(1)/\beta(1)$, and the above decomposition is also unique. As a result, the conditional variance of the GARCH model has the form

$$h_t = \frac{\gamma}{\beta(1)} + \left\{ 1 - \left[(1 - \omega) + \omega \frac{\delta(B)}{\beta(B)}(1 - B) \right] \right\} y_t^2 = \frac{\gamma}{\beta(1)} + \omega \left\{ 1 - \frac{\delta(B)}{\beta(B)}(1 - B) \right\} y_t^2. \quad (5)$$

To make it more clear, let $h_t^* = \omega^{-1}h_t$, and then $h_t^* = \gamma/\omega + \sum_{i=1}^Q (\alpha_i/\omega)y_{t-i}^2 + \sum_{j=1}^P \beta_j h_{t-j}^*$, where h_t is the conditional variance of the GARCH model in (1), and $\sum_{i=1}^Q (\alpha_i/\omega) + \sum_{j=1}^P \beta_j = 1$. As a result, h_t^* is the conditional variance of an integrated GARCH pro-

cess, and $h_t = \omega h_t^*$, i.e. we have built a simple bridge in form between conditional variances of a common GARCH process and an integrated GARCH process.

As for FIGARCH and HYGARCH models, by replacing $(1 - B)$ with $(1 - B)^d$ on the right hand side of (5), we have a new hyperbolic GARCH model,

$$y_t = \varepsilon_t \sqrt{h_t}, \quad h_t = \frac{\gamma}{\beta(1)} + \omega \left\{ 1 - \frac{\delta(B)}{\beta(B)} (1 - B)^d \right\} y_t^2, \quad (6)$$

where $0 < d \leq 1$, $\omega > 0$, $\gamma > 0$ and $\beta(x) = 1 - \sum_{j=1}^p \beta_j x^j$, and $\delta(x) = 1 - \sum_{i=1}^q \delta_i x^i$. Note that $p = P$ and $q = \max\{P, Q\} - 1$. We denote this model by HGARCH(q, d, p) for simplicity.

When $d = 1$, from its evolution, model (6) will become a general GARCH model, and an integrated GARCH model if we further restrict that $\omega = 1$. It is obvious that the new model will reduce to the FIGARCH model as $\omega = 1$. For the nonnegativity condition of the HGARCH model, it is independent of ω , and the restrictions on the other parameters, d , δ_i s and β_j s, are exactly the same as those of the FIGARCH model in Conrad and Haag (2006). Hence, they are less complicated than those of the HYGARCH model.

The conditional variance in (6) can be rewritten into the form of ARCH(∞) models,

$$h_t = \frac{\gamma}{\beta(1)} + \sum_{j=1}^{\infty} b_j y_{t-j}^2, \quad (7)$$

where $(1 - B)^d$ is defined as in (2). Davidson (2004) mentioned that hyperbolic memory ARCH(∞) models have two salient features, the amplitude and the memory. As expected, the parameter d is the memory parameter since it controls the decaying pattern of b_j s in (7). The parameter ω is just the amplitude parameter since it can be verified that $\omega = \sum_{j=1}^{\infty} b_j$.

Zaffaroni (2004) considered a type of ARCH(∞) models with the variance of ε_t being a parameter, and model (6) can be rewritten into the following FIGARCH form,

$$y_t = \varepsilon_t^* \sqrt{h_t^*}, \quad h_t^* = \frac{\gamma}{\beta(1)\omega} + \left\{ 1 - \frac{\delta(B)}{\beta(B)} (1 - B)^d \right\} y_t^2,$$

where $\varepsilon_t^* = \sqrt{\omega} \varepsilon_t$. However, this setting is seldom considered in estimating the parameters of an ARCH-type model since it may suffer from the problem of identifiability. Moreover, from the evolution of the HGARCH model at the beginning of this section,

the GARCH model always has the form of an integrated GARCH model, and then it is misleading to treat model (6) as a special FIGARCH model.

For the existence of a causal stationary solution of ARCH(∞) models, Kazakevicius and Leipus (2003) provided a rigorous result for models with geometric decaying coefficients, and Douc et al. (2008) gave a sufficient condition, which can be applied to FIGARCH models as well as model (6). By Theorem 1 of Douc et al. (2008), we can state the following results for HGARCH models without proof.

Theorem 1. *Suppose that there exists $m \in (0, 1]$ such that $E(|\varepsilon_t|^{2m}) \sum_{j=1}^{\infty} b_j^m < 1$. Then there exists a unique strictly stationary solution of the ARCH(∞) model in (7) with $E(|y_t|^{2m}) < \infty$, and the conditional variance has the form of*

$$h_t = \frac{\gamma}{\beta(1)} + \frac{\gamma}{\beta(1)} \sum_{k=1}^{\infty} \sum_{j_1, \dots, j_k \geq 1} b_{j_1} \cdots b_{j_k} \varepsilon_{t-j_1}^2 \cdots \varepsilon_{t-j_1-j_2-\dots-j_k}^2.$$

When $\omega < 1$, it holds that $E(|\varepsilon_t|^2) \sum_{j=1}^{\infty} b_j = \omega < 1$, and then there always exists a strictly stationary solution to model (6) with $E(y_t^2) < \infty$. However, when $\omega \geq 1$, it is impossible for a stationary solution to exist with finite variance, and the existence of a stationary solution will depend on the distribution of ε_t as for the common GARCH model (Bougerol and Picard, 1992).

Theorem 2. *Suppose that there exists a stationary HGARCH process $\{y_t\}$. If $\omega < [E(|\varepsilon_t|^{2m})]^{-1/m}$, then $E(|y_t|^{2m}) < \infty$, where $m = 1, 2, 3$ or 4 .*

It can be seen that the amplitude parameter controls the higher order moments of the HGARCH process. Note that, by Hölder's inequality, $[E(|\varepsilon_t|^8)]^{-1/4} \leq [E(|\varepsilon_t|^6)]^{-1/3} \leq [E(|\varepsilon_t|^4)]^{-1/2} \leq [E(|\varepsilon_t|^2)]^{-1} = 1$.

3 Statistical inference

3.1 Quasi-maximum likelihood estimation

The Gaussian quasi-maximum likelihood estimation (QMLE) has become a popular approach, and its basic idea is to maximize the likelihood function written under the assumption that innovations ε_t are Gaussian. When the underlying innovation distribution is inappropriately assumed, the Gaussian QMLE can be nearly efficient, while

others such as the Student t QMLE may be even inconsistent (Francq and Zakoian, 2009; Newey and Steigerwald, 1997). This motivates us to consider the Gaussian QMLE of the HGARCH model in (6).

Denote the parameter vector by $\boldsymbol{\theta} = (\gamma, \boldsymbol{\delta}', \boldsymbol{\beta}', \omega, d)' \in R^{p+q+3}$, where $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)'$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$. The true value of the parameter vector $\boldsymbol{\theta}_0$ is assumed to be an interior point of a compact set $\Theta \subset R^{p+q+3}$. Consider the conditional Gaussian log likelihood function of model (6), $-0.5L_n(\boldsymbol{\theta}) - \log \sqrt{2\pi}$, where

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^n l_t(\boldsymbol{\theta}), \quad l_t(\boldsymbol{\theta}) = \frac{y_t^2}{h_t(\boldsymbol{\theta})} + \log[h_t(\boldsymbol{\theta})],$$

and $h_t(\boldsymbol{\theta}) = \gamma/\beta(1) + \sum_{j=1}^{\infty} b_j(\boldsymbol{\theta})y_{t-j}^2$ with $b_j(\boldsymbol{\theta})$ s being functions of $\boldsymbol{\theta}$. Note that the function $h_t(\boldsymbol{\theta})$ depends on past observations infinitely far away, and hence initial values are needed for y_s^2 with $s \leq 0$. We set them to zero as in Robinson and Zaffaroni (2006), and denote by $\tilde{h}_t(\boldsymbol{\theta})$ the function $h_t(\boldsymbol{\theta})$ with these initial values. Accordingly, we can denote $\tilde{l}_t(\boldsymbol{\theta})$ and $\tilde{L}_n(\boldsymbol{\theta})$. As a result, the Gaussian QMLE can be defined as

$$\hat{\boldsymbol{\theta}}_n = \operatorname{argmin} \tilde{L}_n(\boldsymbol{\theta}).$$

Assumption 1. *There is no common root between polynomials $\delta(x)$ and $\beta(x)$, coefficients $b_j(\boldsymbol{\theta})$ s in the function $h_t(\boldsymbol{\theta}) = \gamma/\beta(1) + \sum_{j=1}^{\infty} b_j(\boldsymbol{\theta})y_{t-j}^2$ are all positive for each $\boldsymbol{\theta} \in \Theta$, and the density function of ε_t satisfies that*

$$f(x) = O(L(|x|^{-1})|x|^{\kappa_1}) \quad \text{as } x \rightarrow 0$$

with $\kappa_1 > -1$ and $L(\cdot)$ being a slowly varying function.

Assumption 2. *There exists a strictly stationary and ergodic solution $\{y_t\}$ to the HGARCH model in (6), $E(|\varepsilon_t|^{2+\kappa_2}) < \infty$ for a $\kappa_2 > 0$, and $E(|y_t|^{2\kappa_3}) < \infty$ for a $\kappa_3 \in ((1 + d_{\min})^{-1}, 1)$ with $d_{\min} = \min_{\boldsymbol{\theta} \in \Theta} d$.*

Assumption 3. *$E(|\varepsilon_t|^4) < \infty$, $0.5 < d_0 < 1$, and $E(|y_t|^{2\kappa_4}) < \infty$ for a $\kappa_4 \in (4/(2d_0 + 3), 1)$, where d_0 is the true value of the memory parameter d .*

Theorem 3. *Suppose that Assumptions 1 and 2 hold. Then $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_0$ in the almost surely sense.*

If Assumption 3 further holds, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightarrow_d N\{0, [E(\varepsilon_t^4) - 1]\Omega^{-1}\},$$

where $\Omega = E[h_t^{-2}(\boldsymbol{\theta}_0)(\partial h_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta})(\partial h_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}')]]$.

Assumptions 2 and 3 do not rule out the case with $E(y_t^2) = \infty$ or $\omega \geq 1$, see Theorem 2 in the previous section. The condition of $d_0 > 0.5$ is necessary to make sure that the initial values for y_s^2 with $s \leq 0$ can be asymptotically ignored in deriving the asymptotic normality (Robinson and Zaffaroni, 2006). For the case with $d_0 = 1$, it is impossible for the QMLE to be asymptotically normal since it is on the boundary of the parameter space, and we will explore it to some extent by considering a score test in section 4.

In practice, we may estimate quantities $E(\varepsilon_t^4)$ and Ω in the above asymptotic variance respectively by

$$\frac{1}{n} \sum_{t=1}^n \frac{y_t^4}{\tilde{h}_t^2(\hat{\boldsymbol{\theta}}_n)} \quad \text{and} \quad \hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{h}_t^2(\hat{\boldsymbol{\theta}}_n)} \frac{\partial \tilde{h}_t(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} \frac{\partial \tilde{h}_t(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}'},$$

and hence the asymptotic variance in Theorem 3. It can be verified that these estimators are consistent.

3.2 Portmanteau test

Following Box-Jenkins' three-stage modeling strategy, it is natural to construct a portmanteau test to check whether the fitted HGARCH model in the previous subsection is adequate, and the squared residual autocorrelations play a key role in diagnostic checking for models with time varying conditional variance; see Li and Mak (1994), Kwan et al. (2011) and Kwan et al. (2012). Following their ideas, we first derive the asymptotic normality of squared residual autocorrelations, and then a portmanteau test with the asymptotic distribution being a chi-squared distribution.

In this subsection, we will assume that y_s^2 with $s \leq 0$ are observable, i.e. $\tilde{h}_t(\boldsymbol{\theta}) = h_t(\boldsymbol{\theta})$, since the initial values for them can be asymptotically ignored under the conditions for the asymptotic normality in Theorem 3. Without confusion, we denote $h_t(\hat{\boldsymbol{\theta}}_n)$ by \hat{h}_t for simplicity, where $\hat{\boldsymbol{\theta}}_n$ is the Gaussian QMLE in the previous subsection. Note that $\{y_t/\hat{h}_t^{1/2}\}$ is the residual sequence from model (6), and it holds that $n^{-1} \sum_{t=1}^n (y_t^2/\hat{h}_t) =$

$1+o_p(1)$. For a positive integer k , we then can define the squared residual autocorrelation at lag k as follows,

$$\widehat{r}_k = \frac{\sum_{t=k+1}^n (y_t^2/\widehat{h}_t - 1)(y_{t-k}^2/\widehat{h}_{t-k} - 1)}{\sum_{t=1}^n (y_t^2/\widehat{h}_t - 1)^2}.$$

For a predetermined K , let $\widehat{R} = (\widehat{r}_1, \dots, \widehat{r}_K)'$. By a method similar to Li and Li (2005) and Kwan et al. (2012), together with the Taylor expansion, the central limit theorem and the Cramér-Wold device, we can derive that

$$\sqrt{n}\widehat{R} \rightarrow_d N(0, \Sigma), \quad (8)$$

where I_K is the K -dimensional identity matrix, $\Omega = E[h_t^{-2}(\boldsymbol{\theta}_0)(\partial h_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta})(\partial h_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}')]'$ is defined as in Theorem 3, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_K)$ with $\mathbf{X}_k = -E\{h_t^{-1}(\boldsymbol{\theta}_0)[y_{t-k}^2/h_{t-k}(\boldsymbol{\theta}_0) - 1][\partial h_t(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}]\}$, and

$$\Sigma = I_K - \frac{1}{E(\varepsilon_t^4) - 1} \mathbf{X}'\Omega^{-1}\mathbf{X}.$$

Let

$$\widehat{\mathbf{X}}_k = -\frac{1}{n} \sum_{t=k+1}^n \frac{1}{\widehat{h}_t} \left(\frac{y_{t-k}^2}{\widehat{h}_{t-k}} - 1 \right) \frac{\partial h_t(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}.$$

and $\widehat{\mathbf{X}} = (\widehat{\mathbf{X}}_1, \dots, \widehat{\mathbf{X}}_K)$. We can show that $\widehat{\mathbf{X}} = \mathbf{X} + o_p(1)$. Together with the estimators for the quantities $E\varepsilon_t^4$ and Ω in the previous subsection, we can obtain a consistent estimator of Σ , denoted by $\widehat{\Sigma}$. Based on the asymptotic normality of \widehat{R} in (8), we can construct the portmanteau test as follows,

$$Q_R(K) = n\widehat{R}'\widehat{\Sigma}^{-1}\widehat{R},$$

which is asymptotically distributed as χ_K^2 , the chi-square distribution with K degrees of freedom, if the model is adequate.

4 Test for the hyperbolic memory

For a fitted HGARCH model, the memory parameter usually has a value between zero and one even though the time series is generated from a GARCH model with geometric

memory. It is clearly of interest to construct a test to check whether an observed sequence has hyperbolic memory in volatility,

$$H_0 : d = 1 \quad \text{vs} \quad H_1 : 0 < d < 1.$$

Note that the null hypothesis of $d = 1$ is at the boundary of the parameter space, and therefore a score test is more convenient here.

Let $\boldsymbol{\theta}_1 = (\gamma, \boldsymbol{\delta}', \boldsymbol{\beta}', \omega)'$ and then $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', d)'$. The score function can be defined as

$$S_n(\boldsymbol{\theta}_1) = \frac{1}{\sqrt{n}} \frac{\partial L_n(\boldsymbol{\theta})}{\partial d} \Big|_{d=1} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(1 - \frac{y_t^2}{h_t(\boldsymbol{\theta}_1, 1)} \right) \frac{1}{h_t(\boldsymbol{\theta}_1, 1)} \frac{\partial h_t(\boldsymbol{\theta}_1, 1)}{\partial d},$$

where

$$\frac{\partial h_t(\boldsymbol{\theta}_1, 1)}{\partial d} = \omega \delta(B) \left\{ y_{t-1}^2 - \sum_{k=2}^{\infty} \frac{y_{t-k}^2}{k(k-1)} \right\} + \sum_{j=1}^p \beta_j \frac{\partial h_{t-j}(\boldsymbol{\theta}_1, 1)}{\partial d}.$$

We denote by $\tilde{S}_n(\boldsymbol{\theta}_1)$ when y_s^2 with $s \leq 0$ are replaced by initial values in section 3.1.

Consider the restricted Gaussian QMLE

$$\tilde{\boldsymbol{\theta}}_{1n} = \underset{\boldsymbol{\theta}_1}{\operatorname{argmin}} \tilde{L}_n(\boldsymbol{\theta}_1, 1). \quad (9)$$

Under the null hypothesis of $d = 1$, the statistic $\tilde{S}_n(\tilde{\boldsymbol{\theta}}_{1n})$ is asymptotically normal, and the score test can be constructed based on it.

It is noteworthy that, when $d = 1$, model (6) will reduce to the common GARCH model. Denote by $\boldsymbol{\lambda} = (\gamma, \boldsymbol{\alpha}', \boldsymbol{\beta}')$ the parameter vector of model (1), where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q)'$. We can formalize the observation at the beginning of section 2 with the following mapping,

$$\boldsymbol{\lambda} = (\gamma, \boldsymbol{\alpha}', \boldsymbol{\beta}')' \rightarrow \boldsymbol{\theta}_1 = (\gamma, \boldsymbol{\delta}', \boldsymbol{\beta}', \omega)', \quad (10)$$

where $\omega = \alpha(1)/\beta(1) = \sum_{i=1}^Q \alpha_i / (1 - \sum_{j=1}^P \beta_j)$, and $\delta(x) = 1 - \sum_{i=1}^q \delta_i x^i = [\beta(x) - \omega^{-1} \alpha(x)] / (1 - x)$. Let

$$L_n^*(\boldsymbol{\lambda}) = \sum_{t=1}^n \left\{ \frac{y_t^2}{h_t(\boldsymbol{\lambda})} + \log[h_t(\boldsymbol{\lambda})] \right\},$$

where the function $h_t(\boldsymbol{\lambda})$ satisfies the iterative equation (1), and this notation is a slight abuse since $h_t(\boldsymbol{\theta})$ is reserved for HGARCH models in section 3.1. Similar to the QMLE in section 3.1, the initial values are needed for y_s^2 with $s \leq 0$, and we employ the same

values as in section 3.1. Denote by $\tilde{h}_t(\boldsymbol{\lambda})$ and $\tilde{L}_n^*(\boldsymbol{\lambda})$ the functions of $h_t(\boldsymbol{\lambda})$ and $L_n^*(\boldsymbol{\lambda})$ with these initial values, respectively. We then consider the Gaussian QMLE of the GARCH model as follows,

$$\tilde{\boldsymbol{\lambda}}_n = \operatorname{argmin} \tilde{L}_n(\boldsymbol{\lambda}).$$

Obviously, both $\tilde{L}_n(\boldsymbol{\lambda})$ and $\tilde{L}_n(\boldsymbol{\theta}_1, 1)$ are likelihood functions under H_0 . However, it is much more complicated to directly optimize $\tilde{L}_n(\boldsymbol{\lambda})$ in (9), and the above re-parametrization gives us a chance of performing the test via the QMLE of the common GARCH model.

It holds that $\tilde{\boldsymbol{\theta}}_{1n} = \boldsymbol{\theta}_1(\tilde{\boldsymbol{\lambda}}_n)$ and $\tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n) = \tilde{h}_t(\tilde{\boldsymbol{\theta}}_{1n}, 1)$, where $\boldsymbol{\theta}_1(\cdot)$ refers to the mapping given by (10), see Li et al. (2011). We next derive the score test based on the statistic

$$\tilde{S}_n(\tilde{\boldsymbol{\theta}}_{1n}) = \tilde{S}_n^*(\tilde{\boldsymbol{\lambda}}_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(1 - \frac{y_t^2}{\tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n)} \right) \frac{\tilde{h}_t^d(\tilde{\boldsymbol{\lambda}}_n)}{\tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n)},$$

where the function $h_t^d(\boldsymbol{\lambda})$ satisfies the iterative equation

$$h_t^d(\boldsymbol{\lambda}) = \omega \delta(B) \left\{ y_{t-1}^2 - \sum_{k=2}^{\infty} \frac{y_{t-k}^2}{k(k-1)} \right\} + \sum_{j=1}^p \beta_j h_{t-j}^d(\boldsymbol{\lambda}),$$

ω and $\delta(B)$ are both functions of $\boldsymbol{\lambda}$ as in (10), and $\tilde{h}_t^d(\boldsymbol{\lambda})$ is the function of $h_t^d(\boldsymbol{\lambda})$ with initial values.

Denote by $\boldsymbol{\lambda}_0$ or $\boldsymbol{\theta}_{10}$ the true parameter vector, and it holds that $\boldsymbol{\theta}_{10} = \boldsymbol{\theta}_1(\boldsymbol{\lambda}_0)$. Let

$$\begin{aligned} D &= E \left[\frac{h_t^d(\boldsymbol{\lambda}_0)}{h_t(\boldsymbol{\lambda}_0)} \right]^2, \quad D_n = \frac{1}{n} \sum_{t=1}^n \left[\frac{\tilde{h}_t^d(\tilde{\boldsymbol{\lambda}}_n)}{\tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n)} \right]^2, \\ \mathbf{I} &= E \left\{ \frac{h_t^d(\boldsymbol{\lambda}_0)}{h_t^2(\boldsymbol{\lambda}_0)} \frac{\partial h_t(\boldsymbol{\lambda}_0)}{\partial \boldsymbol{\lambda}} \right\}, \quad \mathbf{I}_n = \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\tilde{h}_t^d(\tilde{\boldsymbol{\lambda}}_n)}{\tilde{h}_t^2(\tilde{\boldsymbol{\lambda}}_n)} \frac{\partial \tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n)}{\partial \boldsymbol{\lambda}} \right\}, \\ \mathbf{J} &= E \left\{ \frac{1}{h_t^2(\boldsymbol{\lambda}_0)} \frac{\partial h_t(\boldsymbol{\lambda}_0)}{\partial \boldsymbol{\lambda}} \frac{\partial h_t(\boldsymbol{\lambda}_0)}{\partial \boldsymbol{\lambda}'} \right\}, \quad \mathbf{J}_n = \frac{1}{n} \sum_{t=1}^n \left\{ \frac{1}{\tilde{h}_t^2(\tilde{\boldsymbol{\lambda}}_n)} \frac{\partial \tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n)}{\partial \boldsymbol{\lambda}} \frac{\partial \tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n)}{\partial \boldsymbol{\lambda}'} \right\}, \end{aligned}$$

$\kappa_\varepsilon = \operatorname{var}(\varepsilon_t)$, and $\hat{\kappa}_\varepsilon = n^{-1} \sum_{t=1}^n [y_t^2/\tilde{h}_t(\tilde{\boldsymbol{\lambda}}_n) - 1]^2$. It is readily verified that $D_n = D + o_p(1)$, $\mathbf{I}_n = \mathbf{I} + o_p(1)$, $\mathbf{J}_n = \mathbf{J} + o_p(1)$, and $\hat{\kappa}_\varepsilon = \kappa_\varepsilon + o_p(1)$. The score test statistic can be constructed as

$$T_s = \frac{[\tilde{S}_n^*(\tilde{\boldsymbol{\lambda}}_n)]^2}{\hat{\kappa}_\varepsilon (D_n - \mathbf{I}_n' \mathbf{J}_n^{-1} \mathbf{I}_n)}.$$

Theorem 4. *Suppose that the conditions for the asymptotic normality in Theorem 3 hold. Under H_0 , if $E(y_t^4) < \infty$, then $\tilde{S}_n^*(\tilde{\boldsymbol{\lambda}}_n) \rightarrow_d N\{0, \kappa_\varepsilon (D - \mathbf{I}' \mathbf{J}^{-1} \mathbf{I})\}$, and $T_s \rightarrow_d \chi_1^2$, where χ_1^2 is the standard chi-square distribution with one degree of freedom.*

The proof of the above theorem is similar to that of Theorem 1 in Li et al. (2011), and hence omitted here. Similar to Theorem 2 in Li et al. (2011), we can further verify that the score test also has asymptotic nontrivial power.

5 Simulation studies

This section conducts three simulation experiments to evaluate the finite sample performance of the Gaussian QMLE in section 3.1, the portmanteau test in section 3.2 and the test for the hyperbolic memory in section 4, respectively. We consider three sample sizes, $n = 1000, 2000$ and 4000 , in all simulation experiments, and there are 1000 replications for each sample size. As in Baillie et al. (1996), for each generated sequence, the first 2000 observations are discarded in order to mitigate the effect of the initial values, i.e. there are $2000 + n$ observations generated each time.

In the first experiment, we generate the data by a HGARCH(1, d , 1) model as follows,

$$y_t = \varepsilon_t \sqrt{h_t}, \quad h_t = \frac{\gamma}{1 - \beta_1} + \omega \left[1 - \frac{1 - \delta_1 B}{1 - \beta_1 B} (1 - B)^d \right] y_t^2,$$

where the conditional variance has the form of

$$h_t = \gamma + \omega(\pi_1 + \delta_1 - \beta_1) y_{t-1}^2 + \omega \sum_{j=2}^{\infty} (\pi_j - \delta_1 \pi_{j-1}) y_{t-j}^2 + \beta_1 h_{t-1},$$

the π_j s are defined as in (2) and the innovations $\{\varepsilon_t\}$ are *i.i.d.* with mean zero and variance one. The parameter vector $\boldsymbol{\theta} = (\gamma, \delta_1, \beta_1, \omega, d)' = (0.1, 0.2, 0.4, 0.5, 0.6)'$, and we consider two distributions for ε_t , the standard normal distribution, $N(0, 1)$, and the standardized Student's t distribution with seven degrees of freedom, t_7 . Note that, under fixed d , π_j is monotonically decreasing with j , and becomes extremely small when j is large. Hence, as in Baillie et al. (1996) and Lombardi and Gallo (2002), we only keep the first 200 coefficients in calculating $(1 - B)^d$, i.e. $(1 - B)^d \approx 1 - \sum_{j=1}^{200} \pi_j B^j$, and there are no significant changes by considering more coefficients. Tables 1 and 2 give the estimation results based on the Gaussian QMLE, and they include the bias, empirical standard errors (EmpStd) and theoretical standard errors (TheoStd), where the empirical standard error refers to the square roots of mean squared errors, and the theoretical standard error is calculated based on the asymptotic variance in Theorem 3.

From Tables 1 and 2, it can be seen that both biases and empirical standard errors are within the acceptable range and decrease as the sample size increases. The empirical standard errors are close to the corresponding theoretical ones especially when the sample size is as large as $n = 4000$. In Table 2 with ε_t following t_7 , the estimator of ω has a larger bias as $n = 1000$, and there is also a larger difference between the empirical standard error and the theoretical one when $n = 2000$. It may be due to the heavy tails of the generated sample; see Baillie et al. (1996). The situation gets substantial improvement as the sample size becomes as large as 4000.

The second experiment is conducted to evaluate the performance of the portmanteau test $Q_R(K)$ in section 3.2. The data generating process is a HGARCH(2, d , 1) model with the parameter vector

$$\boldsymbol{\theta} = (\gamma, \delta_1, \delta_2, \beta_1, \omega, d) = (0.1, \delta_2, 0.2, 0.4, 0.5, 0.8), \quad (11)$$

and then we estimated the sample by a HGARCH(1, d , 1) model. The value of δ_2 is set to zero to evaluate the size, and 0.2 to evaluate the power. Note that the conditional variance of the HGARCH(2, d , 1) process has the form of

$$h_t = \gamma + \omega \{ (\pi_1 + \delta_1 - \beta_1) y_{t-1}^2 + (\pi_2 - \delta_1 \pi_1 + \delta_2) y_{t-2}^2 + \sum_{j=3}^{\infty} (\pi_j - \delta_1 \pi_{j-1} - \delta_2 \pi_{j-2}) y_{t-j}^2 \} + \beta_1 h_{t-1}.$$

The significance level is set to 0.05, and rejection rates of $Q_R(K)$ are listed in Table 3 with $K = 2, 5, 8, 15$ and 20. Like its classical ARMA counterpart, the $Q_R(K)$ statistic should be most powerful in detecting ignored dependence on the lag-structure.

It can be observed from Table 3 that the test is a little bit conservative when the sample size is small, and its empirical sizes are close to the nominal value of 0.05 when the sample size is as large as $n = 4000$. All other empirical sizes are close to the nominal value. Its empirical powers increase as the sample size increases. As expected, all empirical powers decrease as K increases. An important use of the results in section 3.2 is, of course, to provide correct standard errors for the squared residual autocorrelations individually (Li, 2004). As this fact has been well documented and Table 3 has already provided validation of the asymptotic result jointly, we choose to skip reporting simulation results from this perspective.

The third experiment is to study the performance of the score test in section 4 and, for simplicity, the data generating process is a HGARCH(0, d , 1) with the parameter vector

$$\boldsymbol{\theta} = (\gamma, \beta_1, \omega, d) = (0.1, 0.4, 0.5, d),$$

where the case with $d = 1$ corresponds to the size and that with $d < 1$ to the power. We consider two significance levels, 0.05 and 0.10, and the rejection rates are presented in Table 4. It can be seen that the empirical sizes are all close to the nominal levels. It is remarkable that the empirical powers are not monotonically increasing when the value of d reaches 0.6. Similar phenomena have been observed and well discussed in Li et al. (2011). The reason here could be the over compensation of β to the effect of d .

6 Empirical examples

In this section, we apply the proposed HGARCH model as well as FIGARCH and HYGARCH models to the volatilities of two financial time series: daily closing prices of Heng Seng index and daily exchange rates of Korean Won again US dollar.

We focus on the centered log returns in percentage, denoted by $\{r_t\}$, and employ the ARFIMA(1, d , 0) model to remove the possible mean structure. The details are given below.

- The daily closing prices of Heng Seng index are from December 31, 1986 to January 7, 2010, and there are 5713 observations in total. The fitted ARFIMA(1, d , 0) model has the form of $(1 - 0.02B)(1 - B)^{-4.1403 \times 10^{-4}} r_t = y_t$.
- The daily exchange rates of Korean Won again US dollar are from April 17, 1990 to December 31, 2010, and there are 5400 observations in total. The fitted ARFIMA(1, d , 0) model has the form of $(1 - 0.0596B)(1 - B)^{-0.0684} r_t = y_t$.

It is noteworthy to point out that both sequences span covers the Asian financial crisis from 1997 to 1998 and the global financial crisis from 2007 to 2009.

Three hyperbolic GARCH models, including HGARCH(1, d , 1), FIGARCH(1, d , 1) and HYGARCH(1, d , 1) models, are then applied to the residuals, denoted by $\{y_t\}$, from

the ARFIMA(1, d , 0) model. To compare the empirical performance of these three models from both fitting and forecasting aspects, each sequence is divided into two parts: the first part is used to for in-sample fitting, while the second part is used to for out-of-sample forecasts.

Table 5 gives in-sample fitting results of three hyperbolic GARCH models. The FIGARCH model has the worse performance in terms of the value of the maximized likelihood function (LLF), especially for the sequence of exchange rates with the parameter of ω (or ϕ) being significantly different from one. The performances of HGARCH and HYGARCH models are similar, and their corresponding estimated parameters are close to each other. The estimated memory parameter d in HGARCH and HYGARCH models are larger than that in the FIGARCH model if the estimated amplitude parameter ω is less than 1, and vice versa. This matches the simulation results in Section 5 and estimation results in Conrad (2010).

We next compare these three hyperbolic GARCH models from the forecasting aspect. The forecast horizon is set to 1, 5, 10, or 22, and the corresponding Value-at-Risks (VaRs) are calculated. Note that 1- and 5-day-ahead forecasts can be treated as the short term forecasts, while 10- and 22-day-ahead forecasts are the long term forecasts. For each sequence, we calculate the out-of-sample coverage rate of the lower and upper 95% predictive intervals under the three models. The unconditional coverage test statistic in Kupiec (1995) and the conditional coverage test statistic in Christoffersen (1998) together with their p -values are also calculated with rolling estimation. The results are reported in Tables 6 and 7, and we have three findings given below.

First, for the sequence of daily closing prices, it can be seen that, under the 5% significance level, these three models pass almost all of the unconditional and conditional tests for both lower and upper tails with different forecasts steps. For the sequence of exchange rates, a few very small p -values occur when the FIGARCH model is used to calculate the coverage tests at the lower tail for different forecasts steps. Additionally, the coverage rates for both lower and upper tails under HGARCH and HYGARCH models are somewhat closer to the nominal value of 95% than those under the FIGARCH model. From these viewpoints, both HGARCH and HYGARCH model outperform the FIGARCH model.

Second, most of the empirical results for HGARCH and HYGARCH models are almost the same. Note that the 1-step-ahead forecast of the HGARCH(1, d , 1) model is

$$\hat{h}_t(1) = \gamma + \omega[\pi_1 + \delta_1 - \beta_1]y_t^2 + \omega \sum_{j=2}^{\infty} (\pi_j - \delta_1 \pi_{j-1})y_{t-j}^2 + \beta_1 h_t,$$

and the l -step-ahead with $l \geq 2$ forecast is

$$\begin{aligned} \hat{h}_t(l) &= \gamma + [\omega(\pi_1 + \delta_1) + (1 - \omega)\beta_1]\hat{h}_t(l - 1) \\ &+ \omega \sum_{j=2}^{l-1} (\pi_j - \delta_1 \pi_{j-1})\hat{h}_t(j - 1) + \omega \sum_{j=l}^{\infty} (\pi_j - \delta_1 \pi_{j-1})y_{t+l-j}^2. \end{aligned}$$

For the HYGARCH(1, d , 1) model, the 1-step-ahead forecast is

$$\hat{h}_t(1) = \gamma + [\omega\pi_1 + \delta_1 - \beta_1]y_t^2 + \omega \sum_{j=2}^{\infty} (\pi_j - \delta_1 \pi_{j-1})y_{t-j}^2 + \beta_1 h_t,$$

and the l -step-ahead with $l \geq 2$ forecast is

$$\begin{aligned} \hat{h}_t(l) &= \gamma + [\omega(\pi_1 + \delta_1)]\hat{h}_t(l - 1) \\ &+ \omega \sum_{j=2}^{l-1} (\pi_j - \delta_1 \pi_{j-1})\hat{h}_t(j - 1) + \omega \sum_{j=l}^{\infty} (\pi_j - \delta_1 \pi_{j-1})y_{t+l-j}^2. \end{aligned}$$

The only difference lies in the second term on the right hand side of the equation for $\hat{h}_t(l)$. Therefore, if the estimations of the parameters are near to each other, then the forecasts values should be close. It is thus not surprising that the corresponding test statistics calculated based on these forecasts are very close to each other.

Finally, although the proposed HGARCH model can not outperform the HYGARCH model from the fitting and forecasting aspects, both models have better performance than the FIGARCH model. What is more, from the nonnegative constraints of the parameter space and the specification of the conditional variance, the proposed HGARCH provide a more straightforward parameterization and better interpretation than the HYGARCH model.

Acknowledgement

We thank the three guest editors and the two anonymous referees for valuable comments that led to the substantial improvement of this paper. W.K. Li would like to thank the

participants of the conference “Frontiers of Time Series Analysis and Related Fields” for their useful discussions. This research was partially supported by the National Natural Science Foundation of China (NSFC) under Grant (No. 11301433), the Social Science Foundation of Fujian Province of China (No. 2013C056) and the Hong Kong Research Grants Council grants HKU 702908P and HKU 703711P.

Appendix: technical details

This appendix gives the proofs of Theorems 2 and 3.

Proof of Theorem 2. Note that the case with $m = 1$ is implied by Theorem 1, and the condition of $\omega < 1$ actually is also necessary for finite second moment. From Giraitis et al. (2000) and Davidson (2004), the condition of $\omega < [E(|\varepsilon_t|^4)]^{-1/2}$ is sufficient for finite fourth moment. We next show the case of $E(|y_t|^{2m}) < \infty$ with $m = 3$ and 4.

Denote $M_m = E|y_t|^{2m}$ and $\mu_m = E|\varepsilon_t|^{2m}$. Note that $M_3 = \mu_3 E h_t^3$, and

$$\begin{aligned} E h_t^3 &= \left[\frac{\gamma}{\beta(1)} \right]^3 + \frac{3\gamma^2}{[\beta(1)]^2} \sum_{k=1}^{\infty} b_k E y_{t-k}^2 + \frac{3\gamma}{\beta(1)} \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} b_{k_1} b_{k_2} E y_{t-k_1}^2 y_{t-k_2}^2 \\ &\quad + \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \sum_{k_3=1}^{\infty} b_{k_1} b_{k_2} b_{k_3} E y_{t-k_1}^2 y_{t-k_2}^2 y_{t-k_3}^2. \end{aligned}$$

By Hölder’s inequality, it holds that $E(y_{t-k_1}^2 y_{t-k_2}^2) \leq M_2$ and $E(y_{t-k_1}^2 y_{t-k_2}^2 y_{t-k_3}^2) \leq M_3$.

Then we have that

$$M_3 \leq \mu_3 \{ \gamma^3 / [\beta(1)]^3 + 3\omega M_1 \gamma^2 / [\beta(1)]^2 + 3\omega^2 M_2 \gamma / \beta(1) + \omega^3 M_3 \},$$

which implies that

$$M_3 \leq \frac{\mu_3}{1 - \mu_3 \omega^3} \{ \gamma^3 / [\beta(1)]^3 + 3\omega M_1 \gamma^2 / [\beta(1)]^2 + 3\omega^2 M_2 \gamma / \beta(1) \}.$$

Thus, if $\mu_3 < \infty$, then the condition $\omega < \mu_3^{-1/3}$ is sufficient for the existence of the sixth order moment of y_t , i.e. $M_3 < \infty$.

For the eighth order moment, we can show that $M_4 = \mu_4 E h_t^4$ and

$$\begin{aligned} E h_t^4 &= \frac{\gamma^4}{[\beta(1)]^4} + \frac{4\gamma^3}{[\beta(1)]^3} \sum_{k=1}^{\infty} b_k E y_{t-k}^2 + \frac{6\gamma^2}{[\beta(1)]^2} \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} b_{k_1} b_{k_2} E y_{t-k_1}^2 y_{t-k_2}^2 \\ &+ \frac{4\gamma}{\beta(1)} \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \sum_{k_3=1}^{\infty} b_{k_1} b_{k_2} b_{k_3} E y_{t-k_1}^2 y_{t-k_2}^2 y_{t-k_3}^2 \\ &+ \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \sum_{k_3=1}^{\infty} \sum_{k_4=1}^{\infty} b_{k_1} b_{k_2} b_{k_3} b_{k_4} E y_{t-k_1}^2 y_{t-k_2}^2 y_{t-k_3}^2 y_{t-k_4}^2. \end{aligned}$$

By Hölder's inequality again, it holds that $E(y_{t-k_1}^2 y_{t-k_2}^2 y_{t-k_3}^2 y_{t-k_4}^2) \leq M_4$. As a result, we have that

$$M_4 \leq \mu_4 \{ \gamma^4 / [\beta(1)]^4 + 4\omega M_1 \gamma^3 / [\beta(1)]^3 + 6\omega^2 M_2 \gamma^2 / [\beta(1)]^2 + 4\omega^3 M_3 \gamma / \beta(1) + \omega^4 M_4 \},$$

which implies that

$$M_4 \leq \frac{\mu_4}{1 - \mu_4 \omega^4} \{ \gamma^4 / [\beta(1)]^4 + 4\omega M_1 \gamma^3 / [\beta(1)]^3 + 6\omega^2 M_2 \gamma^2 / [\beta(1)]^2 + 4\omega^3 M_3 \gamma / \beta(1) \}.$$

Thus, if $\mu_4 < \infty$, then the condition $\omega < \mu_4^{-1/4}$ is sufficient for the existence of the eighth order moment of y_t , i.e. $M_4 < \infty$.

□

Proof of Theorem 3. We apply Theorems 1 and 2 of Robinson and Zaffaroni (2006) to prove this theorem, and it is sufficient to verify their Assumptions A-H. Note that Assumptions A-E and H can be directly implied by assumptions in this theorem. By a method similar to the proof of Corollary 1 in Robinson and Zaffaroni (2006), we can further show Assumptions F and G, and hence complete the proof. □

References

- Baillie, R. T., Bollerslev, T., Mikkeslen, H. O., 1996. Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 74, 3–30.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.
- Bougerol, P., Picard, N., 1992. Stationarity of garch processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.

- Christoffersen, P. F., 1998. Evaluating interval forecasts. *International economic review* 39, 841–862.
- Conrad, C., 2010. Non-negativity conditions for the hyperbolic GARCH model. *Journal of Econometrics* 157, 441–457.
- Conrad, C., Haag, B., 2006. Inequality constraints in the fractionally integrated garch model. *Journal of Financial Econometrics* 4, 413–449.
- Dacorogna, M., Müller, U., Nagler, R., Olsen, R., Pictet, O., 1993. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12, 413–438.
- Davidson, J., 2004. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business & Economic Statistics* 22, 16–29.
- Ding, Z. X., Grange, C. W. J., Engle, R. F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Ding, Z. X., Granger, C. W. J., 1996. Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics* 73, 185–215.
- Douc, R., Roueff, F., Soulier, P., 2008. On the existence of some ARCH(∞) processes. *Stochastic Processes and their Applications* 118, 755–761.
- Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Francq, C., Zakoian, J.-M., 2009. A tour in the asymptotic theory of GARCH estimation. In: Andersen, T. G., Davis, R. A., Krieb, J.-P., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Springer.
- Giraitis, L., Kokoszka, P., Leipus, R., 2000. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* 16, 3–22.
- Giraitis, L., Surgailis, D., 2002. ARCH-type bilinear models with double long memory. *Stochastic Processes and their Applications* 100, 275–300.

- Kazakevicius, V., Leipus, R., 2003. A new theorem on the existence of invariant distributions with applications to arch processes. *Journal of Applied probability* 40, 147–162.
- Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3 (2), 73–84.
- Kwan, W., Li, W. K., Li, G., 2011. On the threshold hyperbolic GARCH models. *Statistics and Its Interface* 4, 159–166.
- Kwan, W., Li, W. K., Li, G., 2012. On the estimation and diagnostic checking of the arfima-hygarch model. *Computational Statistics and Data Analysis* 56, 3632–3644.
- Li, G., Li, W. K., 2005. Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika* 92, 691–701.
- Li, M., Li, G., Li, W. K., 2011. Score tests for hyperbolic GARCH models. *Journal of Business & Economic Statistics* 29, 579–586.
- Li, M., Li, W. K., Li, G., 2013. On mixture memory GARCH models. *Journal of Time Series Analysis* 34, 606–624.
- Li, W. K., 2004. *Diagnostic Checks in Time Series*. Chapman & Hall.
- Li, W. K., Mak, T. K., 1994. On the squared residual autocorrelations in non-linear series with conditional heteroskedasticity. *Journal of Time Series Analysis* 15, 627–636.
- Lombardi, M. J., Gallo, G. M., 2002. Analytic Hessian matrices and the computation of FIGARCH estimates. *Statistical Methods and Applications* 11, 247–264.
- Newey, W. K., Steigerwald, D. G., 1997. Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65, 587–599.
- Robinson, P. M., 1991. Testing for strong serial correlation and dynamic conditional heteroscedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.
- Robinson, P. M., Zaffaroni, P., 2006. Pseudo-maximum likelihood estimation of ARCH(∞) models. *The Annals of Statistics* 34, 1049–1074.

Zaffaroni, P., 2004. Stationarity and memory of ARCH(∞) models. *Econometric Theory* 20, 147–160.

Table 1: Estimation results of the HGARCH(1, d , 1) model with ε_t following $N(0, 1)$.

n		γ	δ_1	β_1	ω	d
1000	Bias	-0.0034	-0.0227	0.0942	-0.0448	0.1364
	EmpStd	0.0483	0.1959	0.2323	0.1058	0.1985
	TheoStd	0.0568	0.2415	0.2918	0.1156	0.2612
2000	Bias	-0.0018	-0.0258	0.0567	-0.0295	0.0942
	EmpStd	0.0383	0.1493	0.2013	0.0772	0.1569
	TheoStd	0.0404	0.1774	0.2249	0.0739	0.1698
4000	Bias	-0.0011	-0.0134	0.0306	-0.0156	0.0505
	EmpStd	0.0283	0.1210	0.1624	0.0539	0.1113
	TheoStd	0.0285	0.1281	0.1670	0.0522	0.1096

Table 2: Estimation results of the HGARCH(1, d , 1) model with ε_t following t_7 .

n		γ	δ_1	β_1	ω	d
1000	Bias	-0.0049	-0.0394	0.0913	0.1136	0.1450
	EmpStd	0.0510	0.1876	0.2471	1.4168	0.2319
	TheoStd	0.0690	0.3038	0.3538	1.3853	0.3103
2000	Bias	-0.0051	-0.0161	0.0687	0.0093	0.0975
	Empstd	0.0424	0.1841	0.2240	0.5442	0.1914
	TheoStd	0.0522	0.2421	0.2882	0.2385	0.2032
4000	Bias	-0.0028	-0.0116	0.0353	-0.0055	0.0549
	EmpStd	0.0336	0.1523	0.1925	0.1633	0.1410
	TheoStd	0.0373	0.1811	0.2182	0.1629	0.1319

Table 3: Empirical sizes and powers of the portmanteau test $Q_R(K)$.

n	K				
	2	5	8	15	20
Size					
1000	0.0619	0.0639	0.0670	0.0784	0.1072
2000	0.0602	0.0643	0.0723	0.0743	0.0884
4000	0.0450	0.0540	0.0480	0.0690	0.0690
Power					
1000	0.3380	0.2500	0.2020	0.1590	0.1430
2000	0.5900	0.4440	0.3600	0.2860	0.2520
4000	0.8650	0.7870	0.6810	0.5650	0.5180

Table 4: Rejection rates of the score test T_s with the null hypothesis of $d = 1$.

n	$d = 1.0$		$d = 0.9$		$d = 0.75$		$d = 0.6$	
	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10
1000	0.0420	0.0890	0.0980	0.1750	0.1450	0.2240	0.1100	0.1940
2000	0.0460	0.1020	0.1730	0.2650	0.2680	0.3650	0.2060	0.3100
4000	0.0460	0.0930	0.3160	0.4240	0.4880	0.5920	0.3430	0.4290

Table 5: Estimation results of three hyperbolic GARCH models on closing prices of Heng Seng index (HSI) and exchange rates of Korean Won against US dollar (KRW-USD). Theoretical standard errors are given in the parentheses, and LLF refers to the value of the likelihood function.

	FIGARCH	HGARCH	HYGARCH
	HSI		
γ	0.0773 (0.0249)	0.0795 (0.0250)	0.0808 (0.0244)
δ_1	0.2418 (0.0627)	0.2038 (0.0665)	0.1933 (0.0653)
β_1	0.7662 (0.0517)	0.7992 (0.0497)	0.7966 (0.0491)
ω (or ϕ)	– (–)	0.9179 (0.0680)	0.9789 (0.0178)
d	0.7106 (0.0760)	0.7888 (0.0834)	0.7976 (0.0822)
LLF value	–10188	–10186	–10186
	KRW-USD		
γ	0.0065 (0.0024)	0.062 (0.0028)	0.0059 (0.0027)
δ_1	0.3598 (0.0949)	0.4258 (0.1354)	0.4515 (0.1472)
β_1	0.7399 (0.0746)	0.6267 (0.1259)	0.6501 (0.1056)
ω (or ϕ)	– (–)	1.2259 (0.0919)	1.1454 (0.0608)
d	0.6542 (0.0871)	0.4566 (0.0700)	0.4453 (0.0401)
LLF value	–4128.0	–4112.1	–4112.5

Table 6: Out-of-sample coverage rates of the lower and upper 95% predictive intervals, the unconditional coverage test statistic (UCTS) and the conditional coverage test statistic (CCTS) for daily closing prices of Heng Seng index (HSI). The p -values are given in the parentheses.

	UCTS		CCTS		coverage rates(%)	
	Lower	Upper	Lower	Upper	Lower	Upper
1-day-ahead						
FIGARCH	0.6592 (0.4169)	0.1645 (0.6850)	0.7221 (0.6970)	2.3881 (0.3030)	94.2	95.4
HGARCH	1.0134 (0.3141)	0.1645 (0.6850)	1.0369 (0.5954)	2.3881 (0.3030)	94	95.4
HYGARCH	1.4386 (0.2304)	0.1645 (0.6850)	1.4418 (0.4863)	2.3881 (0.3030)	93.8	95.4
5-day-ahead						
FIGARCH	0.0242 (0.8765)	0.3336 (0.5635)	2.3449 (0.3096)	2.3808 (0.3041)	95.16	95.56
HGARCH	0.2094 (0.6472)	0.2094 (0.6472)	1.6156 (0.4458)	0.3985 (0.8193)	94 .56	94.56
HYGARCH	0.2094 (0.6472)	0.2094 (0.6472)	1.6156 (0.4458)	0.3985 (0.8193)	94 .56	94.56
10-day-ahead						
FIGARCH	0.5519 (0.4575)	0.9257 (0.3360)	4.0405 (0.1326)	2.7154 (0.2572)	95.72	95.72
HGARCH	0.0949 (0.7581)	0.2777 (0.5982)	3.9644 (0.1378)	2.4432 (0.2948)	94.70	95.32
HYGARCH	0.0949 (0.7581)	0.2777 (0.5982)	3.9644 (0.1378)	2.4432 (0.2948)	94.70	95.32
22-day-ahead						
FIGARCH	0.3855 (0.5347)	5.0393 (0.0248)	3.7612 (0.1525)	5.9456 (0.0512)	95.62	96.87
HGARCH	0.1891 (0.6637)	0.1632 (0.6863)	6.7769 (0.0338)	2.3858 (0.3033)	94.57	95.20
HYGARCH	0.1891 (0.6637)	0.1632 (0.6863)	6.7769 (0.0338)	2.3858 (0.3033)	94.57	95.20

Table 7: Out-of-sample coverage rates of the lower and upper 95% predictive intervals, the unconditional coverage test statistic (UCTS) and the conditional coverage test statistic (CCTS) for daily exchange rates of Korean Won against US dollar. The p -values are given in the parentheses.

1-day-ahead	UCTS		CCTS		coverage rates(%)	
	Lower	Upper	Lower	Upper	Lower	Upper
FIGARCH	6.2367 (0.0125)	0.0001 (0.9918)	12.5085 (0.0019)	2.7469 (0.2532)	92.40	94.80
HGARCH	1.9325 (0.1645)	0.6941 (0.4048)	6.3214 (0.0424)	2.6298 (0.2685)	93.60	95.60
HYGARCH	1.9325 (0.1645)	0.6941 (0.4048)	6.3214 (0.0424)	2.6298 (0.2685)	93.60	95.60
5-day-ahead						
FIGARCH	2.6326 (0.1047)	1.1011 (0.2940)	7.3509 (0.0253)	1.5904 (0.4515)	93.35	93.95
HGARCH	0.6291 (0.4277)	2.8562 (0.0910)	2.4905 (0.2879)	3.1218 (0.2099)	95.77	96.57
HYGARCH	0.6291 (0.4277)	2.8562 (0.0910)	2.4905 (0.2879)	3.1218 (0.2099)	95.77	96.57
10-day-ahead						
FIGARCH	1.2167 (0.2700)	2.2135 (0.1368)	5.2676 (0.0718)	2.2180 (0.3299)	93.69	93.48
HGARCH	0.0986 (0.7535)	2.6946 (0.1007)	2.3650 (0.3065)	2.9512 (0.2286)	95.32	96.54
HYGARCH	0.0986 (0.7535)	2.6946 (0.1007)	2.3650 (0.3065)	2.9512 (0.2286)	95.32	96.54
22-day-ahead						
FIGARCH	6.5225 (0.0107)	0.0525 (0.8187)	6.5300 (0.0382)	0.4096 (0.8148)	92.28	94.78
HGARCH	0.1632 (0.6863)	3.0954 (0.0785)	1.0076 (0.6042)	3.4440 (0.1787)	95.41	96.66
HYGARCH	0.1632 (0.4169)	3.0954 (0.5367)	1.0076 (0.5842)	3.4440 (0.2994)	95.41	96.66