

# Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates

---

Heejoon Han  
Dennis Kristensen

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP18/13

# ASYMPTOTIC THEORY FOR THE QMLE IN GARCH-X MODELS WITH STATIONARY AND NON-STATIONARY COVARIATES\*

HEEJOON HAN<sup>†</sup>

DENNIS KRISTENSEN<sup>‡</sup>

MAY 2013

## Abstract

This paper investigates the asymptotic properties of the Gaussian quasi-maximum-likelihood estimators (QMLE's) of the GARCH model augmented by including an additional explanatory variable - the so-called GARCH-X model. The additional covariate is allowed to exhibit any degree of persistence as captured by its long-memory parameter  $d_x$ ; in particular, we allow for both stationary and non-stationary covariates. We show that the QMLE's of the parameters entering the volatility equation are consistent and mixed-normally distributed in large samples. The convergence rates and limiting distributions of the QMLE's depend on whether the regressor is stationary or not. However, standard inferential tools for the parameters are robust to the level of persistence of the regressor with  $t$ -statistics following standard Normal distributions in large sample irrespective of whether the regressor is stationary or not.

## 1 Introduction

To better model and forecast the volatility of economic and financial time series, empirical researchers and practitioners often include exogenous regressors in the specification of volatility dynamics. One particularly popular model within this setting is the so-called GARCH-X model where the basic GARCH specification of Bollerslev (1986) is augmented by adding exogenous regressors to the volatility equation:

$$y_t = \sigma_t(\vartheta) \varepsilon_t, \quad (1)$$

---

\*This paper benefited from discussions with Joon Y. Park, Peter Phillips, Anders Rahbek and Qiyang Wang. We also wish to thank participants at seminars at University of Nottingham, the NBER-NSF time series conference 2012 at Texas A&M, and a time series workshop at University of Copenhagen for comments and suggestions. Kristensen gratefully acknowledges financial support from the European Research Council through a starting grant (No 312474), Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, and the Danish Research Council through its funding of CREATES. This work was supported by a grant from the Kyung Hee University (KHU 2012-1676).

<sup>†</sup>Department of Economics, Kyung Hee University. E-mail: [heejoon@khu.ac.kr](mailto:heejoon@khu.ac.kr).

<sup>‡</sup>Corresponding author. Dep. of Economics, UCL, CREATES and IFS. E-mail: [d.kristensen@ucl.ac.uk](mailto:d.kristensen@ucl.ac.uk).

where  $\varepsilon_t$  is the error process while  $\sigma_t^2(\vartheta)$  is the volatility process given by

$$\sigma_t^2(\vartheta) = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + \pi x_{t-1}^2, \quad (2)$$

for some observed covariate  $x_t$  which is squared to ensure that  $\sigma_t^2(\vartheta) > 0$ , and where  $\vartheta = (\omega, \theta)'$ ,  $\theta = (\alpha, \beta, \pi)'$ , is the vector of parameters. The inclusion of the additional regressor  $x_t$  often helps explaining the volatilities of stock return series, exchange rate returns series or interest rate series and tend to lead to better in-sample fit and out-of sample forecasting performance. Choices of covariates found in empirical studies using the GARCH-X model span a wide range of various economic or financial indicators. Examples include interest rate levels (Brenner et al., 1996; Glosten et al, 1993; Gray, 1996), bid-ask spreads (Bollerslev and Melvin, 1994), interest rate spreads (Dominguez, 1998; Hagiwara & Herce, 1999), forward-spot spreads (Hodrick, 1989), futures open interest (Girma and Mougoue, 2002), information flow (Gallo and Pacini, 2000), and trading volumes (Fleming et al, 2008; Lamoureux and Lastrapes, 1990; Marsh and Wagner, 2005). More recently, various realized volatility measures constructed from high frequency data have been adopted covariates in the GARCH-type models with the rapid development seen in the field of realized volatility; see Barndorff-Nielsen and Shephard (2007), Engle (2002), Engle and Gallo (2006), Hansen et al. (2010), Hwang and Satchell (2005), and Shephard and Sheppard (2010).

While the GARCH-X model and its associated quasi-maximum likelihood estimator (QMLE) have found widespread empirical use, the theoretical properties of the estimator are not fully understood. In particular, given the wide range of different choices of covariates, it is of interest to analyze how the persistence of the chosen covariate influences the QMLE. As shown in Table 1 in Appendix C, the degree of persistence varies a lot across some popular covariates used in GARCH-X specifications. The table reports log-periodogram estimates of memory parameter,  $d_x$ , and estimates of the first-order autocorrelation,  $\rho_1$ , for some time series used as covariates in the literature. For example, interest rate levels and bond yield spreads are highly persistent with estimates of  $d_x$  being mostly larger than 0.8 and  $\rho_1$  estimates close to unity, thereby suggesting unit root type behaviour. Meanwhile, realized volatility measures (realized variance) of various stock index and exchange rate return series are less persistent with estimates of  $d_x$  ranging between 0.3 and 0.6 while the estimates of  $\rho_1$  are relatively small and taking values between 0.64 to 0.88; formal unit root tests clearly reject unit root hypotheses for these time series. A natural concern would be that different degrees of persistence of the chosen covariates would lead to different behaviour of the QMLE and associated inferential tools.

We provide a unified asymptotic theory for the QMLE of the parameters allowing for both stationary and non-stationary regressors. In the case of non-stationary regressors, we model  $x_t$  as an  $I(d_x)$  process with  $1/2 < d_x < 3/2$ . This allows for a wide-range of persistence as captured by the long-memory parameter  $d_x$ , including unit root processes ( $d_x = 1$ ) but also processes with either weaker ( $d_x < 1$ ) or stronger dependence ( $d_x > 1$ ).

Our main results show that to a large extent applied researchers can employ the same techniques when drawing inference regarding model parameters regardless of the degree of persistence of the

regressors. We first show that QMLE consistently estimates  $\vartheta_0$  whether  $x_t$  is stationary or not, but that its limiting distribution is non-standard in the non-stationary case. At the same time, we also demonstrate that the large sample distributions of  $t$ -statistics are invariant to the degree of persistence and always follow  $N(0, 1)$  distributions. As consequence, standard inference tools are applicable whether the regressors are stationary or not, and so researchers do not have to conduct any preliminary analysis of a given covariate before carrying out inference. A simulation study confirms our theoretical findings, with the distribution of standard  $t$ -statistics showing little sensitivity to the degree of persistence of the included covariate.

Our theoretical results have important antecedents in the literature. Our theoretical results for the non-stationary case rely on results developed in Han (2011) who analyzes the time series properties of GARCH-X models with long-memory regressors. He shows how the GARCH-X process explains stylized facts of financial time series such as the long memory property in volatility, leptokurtosis and IGARCH. Kristensen and Rahbek (2005) provided theoretical results for the QMLE in the linear ARCH-X models in the case of stationary regressors. We extend their theoretical results to allow for lagged values of the volatility in the specification and non-stationary regressors. Jensen and Rahbek (2004) and Francq and Zakořan (2012) analyzed the QMLE in the pure GARCH model (i.e., no covariates included,  $\pi = 0$ ) and showed that the QMLE of  $(\alpha, \beta)$  remained consistent and  $\sqrt{n}$ -asymptotically normally distributed even when  $\sigma_t^2(\vartheta)$  was explosive. On the other hand, they found that  $\omega$  is not identified when the volatility process is non-stationary. Our results for the QMLE of  $\theta$  are similar: It remains consistent and  $\sqrt{n}$ -asymptotically normally distributed independently of whether  $x_t^2$ , and thereby  $\sigma_t^2(\vartheta)$ , is explosive or not. However, in contrast to the pure GARCH model, it is possible to identify and consistently estimate  $\omega$  in the GARCH-X model even when  $x_t$  is non-stationary. The contrasting results are due to the fact that the dynamics of a non-stationary pure GARCH process are very different from those of a GARCH-X process with non-stationarity being induced through  $x_t$ .

Finally, Han and Park (2012), henceforth HP2012, established the asymptotic theory of the QMLE for a GARCH-X model where a nonlinear transformation of a unit root process was included as exogenous regressor. Our work complements HP2012 in that we allow for a wider range of dependence in the regressor, but on the other hand do not consider general nonlinear transformations of the variable. In the special case with  $d_x = 1$ , our results for the estimation of  $\theta$  coincide with those of HP2012 with their transformation chosen as the quadratic function. At a technical level, we provide a more detailed analysis of the QMLE compared to HP2012. While HP2012 conjectured that  $\omega$  was not identified and so kept the parameter fixed at its true value in their analysis, we here show that in fact  $\omega$  can be consistently estimated from data and derive the large-sample distribution of its QMLE. This important result is derived by extending some novel limit results for non-stationary regression models developed in Wang and Phillips (2009a,b).

The rest of the paper is organized as follows. Section 2 introduces the model and the QMLE. Section 3 derives the asymptotic theory of the QMLE for the stationary and non-stationary case, while Section 4 analyzes the large sample distributions of the corresponding  $t$ -statistics. The results of a simulation study is presented in Section 5. Section 6 concludes. All proofs have been relegated

to Appendix A. Tables and figures have been collected in Appendix B. Before we proceed, a word on notation: Standard terminologies and notations employed in probability and measure theory are used throughout the paper. Notations for various convergences such as  $\rightarrow_{\text{a.s.}}$ ,  $\rightarrow_p$  and  $\rightarrow_d$  frequently appear, where all limits are taken as  $n \rightarrow \infty$  except where otherwise indicated.

## 2 Model and Estimator

The GARCH-X model is given by eqs. (1)-(2) where the parameters are collected in  $\vartheta = (\omega, \theta)$  where  $\theta = (\alpha, \beta, \pi) \in \Theta \subseteq \mathbb{R}^3$  and  $\omega \in \mathcal{W} \subseteq [0, \infty)$ . The chosen decomposition of the full parameter vector into  $\theta$  and the intercept  $\omega$  is due to the special role played by the latter in the non-stationary case. The true, data-generating parameter is denoted  $\vartheta_0 = (\omega_0, \theta_0)'$ , where  $\theta_0 = (\alpha_0, \beta_0, \pi_0)'$  and the associated volatility process  $\sigma_t^2 = \sigma_t^2(\vartheta_0)$ . We will throughout assume that  $\mathbb{E}[\log(\alpha_0 \varepsilon_t^2 + \beta_0)] < 0$  so that non-stationarity can only be induced by  $x_t$ . In particular, if  $x_t$  is stationary then  $\sigma_t^2$  and  $y_t$  are stationary; see Section 3.1 for details. In the stationary case, we impose no restrictions on its time series dynamics. On the other hand, in the non-stationary case, we focus on the case where  $x_t^2$  is explosive and model it as a long-memory processes of the form

$$x_t = x_{t-1} + \xi_t, \quad (3)$$

where, for a sequence  $\{v_t\}$  which is i.i.d.  $(0, \sigma_v^2)$ ,

$$(1 - L)^d \xi_t = v_t, \quad -1/2 < d < 1/2. \quad (4)$$

Hence,  $x_t$  is an  $I(d_x)$  process with  $d_x = d + 1 \in (1/2, 3/2)$ . Note that  $\{\varepsilon_t\}$  and  $\{v_t\}$  are allowed to be dependent. Hence, the model can accommodate leverage effects catered for by the GJR-GARCH model if  $\{\varepsilon_t\}$  and  $\{v_t\}$  are negatively correlated. See Han (2011) for more details on the model and its time series properties.

Whether  $x_t$  is stationary or not, we will require it to be exogeneous in the sense that  $\mathbb{E}[\varepsilon_t | x_{t-1}] = 0$  and  $\mathbb{E}[\varepsilon_t^2 | x_{t-1}] = 1$ . This restricts the choices of  $x_t$ ; for example, in most situations, the exogeneity assumption will be violated if  $y_t$  is a stock return, say,  $r_{1,t}$  and  $x_{t-1} = r_{2,t}$  is another return series since these will in general be contemporaneously correlated. This in turn will generate simultaneity biases in the estimation of the GARCH-X model similar to OLS in simultaneous equations models. If instead  $x_{t-1} = r_{2,t-1}$ , the GARCH-X model can be thought of as a restricted version of a bivariate GARCH model where lags of  $r_{1,t}$  do not affect the volatility of  $r_{2,t}$  and only the first lag of  $r_{2,t}$  affects the volatility of  $r_{1,t}$ . This restriction may in some cases be implausible. On the other hand, GARCH-X models is a lot simpler to estimate compared to a bivariate GARCH model: The former only contains four parameters while a bivariate BEKK-GARCH(1,1) contains twelve parameters.

Dittmann and Granger (2002) analyzed the properties of  $x_t^2$  given  $x_t$  is fractionally integrated and showed that, when  $x_t$  is a Gaussian fractionally integrated process of order  $d_x$ , then  $x_t^2$  is asymptotically also a long memory process of order  $d_{x^2} = d_x$ . Hence, for  $1/2 < d_x < 3/2$ , the

covariate  $x_t^2$  is non-stationary long memory, including the case of unit root-type behaviour. Considering that the range of memory parameter for real data used as covariates in the literature seldom exceeds unity, the range of  $d_x$  we consider is wide enough to cover all covariates used in the empirical literature.

Our model is related to the one considered in HP2012 given by  $\sigma_t^2(\vartheta) = \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\vartheta) + f(x_{t-1}, \gamma)$ , where  $x_t$  is integrated or near-integrated, and  $f(x_{t-1}, \gamma)$  is a positive, asymptotically homogeneous function as introduced by Park and Phillips (1999).<sup>1</sup> If we let  $d_x = 1$  in our model,  $x_t$  is integrated and our model belongs to the model considered by HP2012 with  $f(x_{t-1}, \gamma) = \omega + \pi x_{t-1}^2$ . While their model allows for more general nonlinear function of  $x_t$ , our analysis includes more general dependence structure of  $x_t$ : It is either stationarity or it is fractionally integrated process with  $1/2 < d_x < 3/2$ . As shown in Table 1 (Appendix C), these are empirically relevant types of dynamic behaviour.

Let  $(y_t, x_{t-1})$  for  $t = 0, \dots, n$ , be  $n + 1 \geq 2$  observations from (1)-(2). We collect the unknown data-generating parameter values in  $\vartheta_0 = (\omega_0, \theta_0) \in \mathcal{W} \times \Theta$  which we wish to estimate. We propose to do so through the Gaussian log-likelihood with  $\varepsilon_t \sim \text{i.i.d. } N(0, 1)$ :

$$L_n(\vartheta) = \sum_{t=1}^n \ell_t(\vartheta), \quad \ell_t(\vartheta) = -\log \sigma_t^2(\vartheta) - \frac{y_t^2}{\sigma_t^2(\vartheta)},$$

where  $\sigma_t^2(\vartheta) = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + \pi x_{t-1}^2$  is the volatility process induced by a given parameter value  $\vartheta$ . It is assumed to be initialized at some fixed parameter independent value  $\bar{\sigma}_0^2 > 0$ ,  $\sigma_0^2(\vartheta) = \bar{\sigma}_0^2$ . We will not restrict  $\varepsilon_t$  to be normally distributed and hence  $L_n(\vartheta)$  is a quasi-log likelihood. The QMLE of  $\vartheta_0$  is then defined as:

$$\hat{\vartheta} = (\hat{\omega}, \hat{\theta}) = \arg \max_{(\omega, \theta) \in \mathcal{W} \times \Theta} L_n(\omega, \theta). \quad (5)$$

The intercept  $\omega_0$  plays a special role since  $\hat{\omega}$  will have radically different behaviour depending on whether  $x_t$  is stationary or not. In fact, HP2012 conjectured that, in the case where  $x_t$  is non-stationary,  $\omega$  could not be identified. This conjecture is supported by the analysis of the QMLE in pure GARCH models ( $\pi = 0$ ) by Jensen and Rahbek (2004) and Francq and Zakořan (2012). They found that when the volatility process is nonstationary,  $\omega_0$  is not identified. Jensen and Rahbek (2004) and HP2012 resolved this issue by fixing  $\omega$  at its true value and only estimating the remaining parameters,  $\theta$ . However, as we shall see, when non-stationarity is generated by an exogenous regressor,  $\omega_0$  in the GARCH-X model can be consistently estimated by the QMLE. However, the rate of convergence of the QMLE of  $\omega$  is slower in the non-stationary case, while  $\hat{\theta}$  converges with  $\sqrt{n}$ -rate independently of the level of persistence of  $x_t$ .

---

<sup>1</sup>Note a notational difference in HP2012: Instead of  $f(x_{t-1}, \gamma)$ , HP2012 use  $f(x_t, \gamma)$  where  $x_t$  is adapted to  $\mathcal{F}_{t-1}$ .

### 3 Asymptotic Theory

The main arguments used to establish the asymptotic distribution of the QMLE are identical for the two cases - stationary or non-stationary regressors. The technical details used to establish the main arguments differ in the two cases though, and so we provide separate proofs for them. But first, we outline the proof strategy for consistency and asymptotic normality of the QMLE to emphasise similarities and differences in the analysis of the two different cases.

To present the arguments in a streamlined fashion, it proves useful to redefine  $\ell_t(\vartheta)$  as a normalized version of the log-likelihood function by subtracting the log-likelihood evaluated at  $\vartheta_0$ ,

$$\ell_t(\vartheta) := \left\{ -\log \sigma_t^2(\vartheta) - \frac{y_t^2}{\sigma_t^2(\vartheta)} \right\} - \left\{ -\log \sigma_t^2 - \frac{y_t^2}{\sigma_t^2} \right\} = -\log(r_t(\vartheta)) - \left\{ \frac{1}{r_t(\vartheta)} - 1 \right\} \varepsilon_t^2$$

where  $\sigma_t^2$  denotes the true data-generating volatility process,

$$\sigma_t^2 = \omega_0 + \alpha_0 y_{t-1}^2 + \beta_0 \sigma_{t-1}^2 + \pi_0 x_{t-1}^2, \quad (6)$$

and  $r_t(\vartheta)$  is a variance-ratio process defined as

$$r_t(\vartheta) := \frac{\sigma_t^2(\vartheta)}{\sigma_t^2}. \quad (7)$$

This normalization does not affect the QMLE since  $-\log \sigma_t^2 - y_t^2/\sigma_t^2$  is parameter independent. Note that the process  $r_t(\vartheta)$  is in general not stationary since  $\sigma_t^2(\vartheta)$  has been initialized at some fixed value and  $x_t$  may be non-stationary. For consistency, the main argument involves showing that the normalized version of the log-likelihood satisfies

$$\sup_{\vartheta \in \mathcal{W} \times \Theta} \frac{1}{n} \|L_n(\vartheta) - L_n^*(\vartheta)\| \xrightarrow{P} 0, \quad (8)$$

where  $L_n^*(\vartheta)$  is given by

$$L_n^*(\vartheta) = \sum_{t=1}^n \ell_t^*(\vartheta), \quad \ell_t^*(\vartheta) = -\log(r_t^*(\vartheta)) - \left\{ \frac{1}{r_t^*(\vartheta)} - 1 \right\} \varepsilon_t^2, \quad (9)$$

and  $r_t^*(\vartheta)$  is a stationary sequence which is asymptotically equivalent to  $r_t(\vartheta)$ . We can now appeal to a uniform LLN for stationary and ergodic sequences to obtain that  $L_n^*(\vartheta)/n \rightarrow_p L^*(\vartheta) := \mathbb{E}[\ell_t^*(\vartheta)]$  uniformly in  $\vartheta$ . The precise definition of  $r_t^*(\vartheta)$ , and thereby  $L^*(\vartheta)$ , depends on whether  $x_t$  is stationary or not. In particular, in the stationary case  $\vartheta_0 = \arg \max_{\vartheta} L^*(\vartheta)$  is uniquely identified and so  $\hat{\vartheta} \rightarrow_p \vartheta_0$  globally, while in the nonstationary case  $L^*(\vartheta) = L^*(\theta)$  is constant w.r.t.  $\omega$  and so we can only conclude that  $\hat{\theta} \rightarrow_p \theta_0$ . This would seem to indicate that in the non-stationary case  $\hat{\omega}$  is inconsistent which would be similar to the explosive pure GARCH model as analyzed by Jensen and Rahbek (2004) and Francq and Zakoian (2012). However, in our case, this conclusion is not correct and is an artifact of normalizing  $L_n(\vartheta)$  by  $1/n$ . By analyzing the

local behaviour of  $L_n(\vartheta)$  in a shrinking neighbourhood of  $\vartheta_0$ , we find that in the non-stationary case  $\hat{\omega}$  remains consistent but converges at a slower rate compared to  $\hat{\theta}$ .

To derive the asymptotic distribution of  $\hat{\vartheta}$ , we proceed to analyze the score and hessian of the quasi-log likelihood. We denote the score vector by  $S_n(\vartheta) = (S_{n,\omega}(\vartheta), S_{n,\theta}(\vartheta))' \in \mathbb{R}^4$ , where  $S_{n,\omega}(\vartheta) = \partial L_n(\vartheta) / (\partial \omega) \in \mathbb{R}$  and  $S_{n,\theta}(\vartheta) = \partial L_n(\vartheta) / (\partial \theta) \in \mathbb{R}^3$  and the Hessian matrix by

$$H_n(\vartheta) = \begin{bmatrix} H_{n,\omega\omega}(\vartheta) & H_{n,\omega\theta}(\vartheta) \\ H_{n,\theta\omega}(\vartheta) & H_{n,\theta\theta}(\vartheta) \end{bmatrix} \in \mathbb{R}^{4 \times 4}, \quad (10)$$

where  $H_{n,\theta\omega}(\vartheta) = \partial^2 L_n(\vartheta) / (\partial \theta \partial \omega) \in \mathbb{R}^3$  and the other components are defined similarly. A standard first order Taylor expansion of the score vector yields  $0 = S_n(\hat{\vartheta}) = S_n(\vartheta_0) + H_n(\bar{\vartheta})(\hat{\vartheta} - \vartheta_0)$ , where  $\bar{\vartheta}$  lies on the line segment connecting  $\hat{\vartheta}$  and  $\vartheta_0$ . Assuming that  $\vartheta_0$  lies in the interior of the parameter space,  $\hat{\vartheta}$  must be an interior solution with probability approaching one (w.p.a.1). That is,  $S_n(\hat{\vartheta}) = 0$  w.p.a.1. What remains is to derive the limiting distribution of  $S_n(\vartheta_0)$  and  $H_n(\bar{\vartheta})$ .

In the stationary case, we can appeal to LLN and CLT for stationary and ergodic sequences and show that

$$S_n(\vartheta_0) / \sqrt{n} \rightarrow_d N(0, \Sigma^{\text{st}}), \quad -H_n(\bar{\vartheta}) / n \rightarrow_p H^{\text{st}} > 0, \quad (11)$$

where  $\Sigma^{\text{st}} \in \mathbb{R}^{4 \times 4}$  and  $H^{\text{st}} \in \mathbb{R}^{4 \times 4}$  are constant. This implies that

$$\sqrt{n}(\hat{\vartheta} - \vartheta_0) \rightarrow_d N(0, \Omega^{\text{st}}), \quad \Omega^{\text{st}} = (H^{\text{st}})^{-1} \Sigma^{\text{st}} (H^{\text{st}})^{-1}. \quad (12)$$

In the non-stationary case, the score and hessian, and thereby the QMLE's, have different asymptotic behaviour. First and foremost,  $\hat{\omega}$  and  $\hat{\theta}$  converge at different rates which we collect in the matrix  $V_n$ ,

$$V_n := \begin{bmatrix} n^{1/4-d/2} & O_{1 \times 3} \\ O_{3 \times 1} & n^{1/2} I_3 \end{bmatrix} \in \mathbb{R}^{4 \times 4}, \quad (13)$$

where  $O_{k \times m} \in \mathbb{R}^{k \times m}$  denotes the matrix of zeros. We show that

$$V_n^{-1} S_n(\vartheta_0) \rightarrow_d MN(0, \Sigma^{\text{nst}}), \quad -V_n^{-1} H_n(\bar{\vartheta}) V_n^{-1} \rightarrow_d H^{\text{nst}} > 0, \quad (14)$$

where  $MN(0, \Sigma^{\text{nst}})$  denotes a mixed-normal distribution with (random) covariance matrix  $\Sigma^{\text{nst}} \in \mathbb{R}^{4 \times 4}$ , and  $H^{\text{nst}} \in \mathbb{R}^{4 \times 4}$  is also random. The proof of eq. (14) employs generalized versions of limit results for fractionally integrated processes developed in Wang and Phillips (2009a) that we have collected in Lemma 6 below. Having established (14), it follows by standard arguments that

$$V_n(\hat{\vartheta} - \vartheta_0) \rightarrow_d MN(0, \Omega^{\text{nst}}), \quad \Omega^{\text{nst}} = (H^{\text{nst}})^{-1} \Sigma^{\text{nst}} (H^{\text{nst}})^{-1}. \quad (15)$$

In particular,  $\hat{\theta}$  is  $\sqrt{n}$ -asymptotically normally distributed while  $\hat{\omega}$  converges with a slower rate of  $n^{1/4-d/2}$  and follows a mixed-normal distribution. So, in comparison to pure explosive GARCH models where  $\omega_0$  is not identified, we can still conduct inference about  $\omega_0$  when the explosiveness is induced by a non-stationary regressor.



In conclusion, the asymptotic distribution of  $\hat{\vartheta}$  depends on whether  $x_t$  is stationary or not. Fortunately, the distribution is in both cases mixed-normal and so standard test statistics prove to be robust to the degree of persistence of  $x_t$ . In particular, we show that standard  $t$ -statistics follow  $N(0, 1)$  distributions irrespective of the regressor's level of persistence.

Since the assumptions and techniques used to establish the above results differ depending on whether  $x_t$  is stationary or not, we consider the two cases in turn: The following subsection covers the stationary case, while the subsequent one focuses on the non-stationary case.

### 3.1 Stationary Case

We first show that the QMLE is globally consistent under the following conditions with  $\mathcal{F}_t$  denoting the natural filtration:

#### Assumption 1

- (i)  $\{(\varepsilon_t, x_t)\}$  is stationary and ergodic with  $\mathbb{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0$  and  $\mathbb{E}[\varepsilon_t^2|\mathcal{F}_{t-1}] = 1$ .
- (ii)  $\mathbb{E}[\log(\alpha_0\varepsilon_t^2 + \beta_0)] < 0$  and  $\mathbb{E}[x_t^{2q}] < \infty$  for some  $0 < q < \infty$ .
- (iii)  $\Theta = \{\vartheta : \underline{\omega} \leq \omega \leq \bar{\omega}, 0 \leq \alpha \leq \bar{\alpha}, 0 \leq \beta \leq \bar{\beta}, 0 \leq \pi \leq \bar{\pi}\}$ , where  $0 < \underline{\omega} \leq \bar{\omega} < \infty$ ,  $\bar{\alpha} < \infty$ ,  $\bar{\beta} < 1$  and  $\bar{\pi} < \infty$ . The true value  $\vartheta_0 \in \Theta$  with  $(\alpha_0, \pi_0) \neq (0, 0)$ .
- (iv) For any  $(a, b) \neq (0, 0)$ :  $a\varepsilon_t^2 + bx_t^2|\mathcal{F}_{t-1}$  has a nondegenerate distribution.

Assumption 1(i) is a generalization of the conditions found in Escanciano (2009) who derives the asymptotic properties of QMLE for pure GARCH processes (that is, no exogenous covariates are included) with martingale difference errors. The assumption is weaker than the i.i.d. assumption imposed in Kristensen and Rahbek (2005). The moment conditions in Assumption 1(ii) implies that a stationary solution to eqs. (1)-(2) at the true parameter value  $\vartheta_0$  exists and has a finite polynomial moment, c.f. Lemma 1 below. We here allow for integrated GARCH processes ( $\alpha + \beta = 1$ ), and impose very weak moment restrictions on the regressor. We do however rule out explosive volatility when  $x_t$  is stationary; we expect that the arguments of Jensen and Rahbek (2004) can be extended to GARCH-X models with  $\mathbb{E}[\log(\alpha_0\varepsilon_{t-1}^2 + \beta_0)] > 0$ , thereby showing that  $\hat{\theta}$  is consistent while  $\hat{\omega}$  is inconsistent. The compactness condition in Assumption 1(iii) should be possible to weaken by following the arguments of Kristensen and Rahbek (2005); this will lead to more complicated proofs though and so we maintain the compactness assumption here for simplicity. The requirement that  $(\alpha_0, \pi_0) \neq (0, 0)$  is needed to ensure identification of  $\beta_0$  since in the case where  $(\alpha_0, \pi_0) = (0, 0)$ ,  $\sigma_t^2 = \sigma_t^2(\vartheta_0) \rightarrow_{a.s.} \omega_0/(1 - \beta_0)$  and so we would not be able to jointly identify  $\omega_0$  and  $\beta_0$ . The non-degeneracy condition in Assumption 1(iv) is also needed for identification. It rules out (dynamic) collinearity between  $y_{t-1}^2$  and  $x_t^2$ . It is similar to the no-collinearity restriction imposed in Kristensen and Rahbek (2005).

To derive the asymptotic properties of  $\hat{\vartheta}$ , we establish some preliminary results. The first lemma states that a stationary solution to the model at the true parameter values exists:

**Lemma 1** *Under Assumption 1: There exists a stationary and ergodic solution to eqs. (1)-(2) at  $\vartheta_0$  satisfying  $\mathbb{E}[\sigma_t^{2s}] < \infty$  and  $\mathbb{E}[y_t^{2s}] < \infty$  for some  $0 < s < 1$ .*

We will in the following work under the implicit assumption that we have observed the stationary solution. Next, we show that for any value of  $\vartheta$  in the parameter space, the volatility-ratio process  $r_t(\vartheta)$  is well-approximated by a stationary version:

**Lemma 2** *Under Assumption 1: With  $s > 0$  given in Lemma 1, there exists some  $K_s < \infty$  such that*

$$\mathbb{E} \left[ \sup_{\vartheta \in \mathcal{W} \times \Theta} |r_t(\vartheta) - r_t^*(\vartheta)|^s \right] \leq K_s \beta^{st},$$

where

$$r_t^*(\vartheta) := \frac{\sigma_{0,t}^2(\vartheta)}{\sigma_t^2}, \quad \sigma_{0,t}^2(\vartheta) := \sum_{i=1}^{\infty} \beta^{i-1} (\omega + \alpha y_{t-i}^2 + \pi x_{t-i}^2). \quad (16)$$

The process  $\sigma_{0,t}^2(\vartheta)$  is stationary and ergodic with  $\mathbb{E}[\sup_{\vartheta \in \mathcal{W} \times \Theta} \sigma_{0,t}^{2s}(\vartheta)] < \infty$ .

Note that, in particular,  $\sigma_t^2 = \sigma_{0,t}^2(\vartheta_0)$ . This in turn implies that eq. (8) holds with  $r_t^*(\vartheta)$  defined in the previous lemma. With these results in hand, we are now ready to show the first main result of this section:

**Theorem 3** *Under Assumption 1, the QMLE  $\hat{\vartheta}$  is consistent.*

Having shown that the QMLE is consistent, we proceed to verify eq. (11) under the following additional assumption:

**Assumption 2**

- (i)  $\kappa_4 = \mathbb{E}[(\varepsilon_t^2 - 1)^2 | \mathcal{F}_{t-1}] < \infty$  is constant.
- (ii)  $\vartheta_0$  is in the interior of  $\Theta$ .

Assumption 2(i) is used to show that the variance of the score exists. It could be weakened to allow for  $\mathbb{E}[(\varepsilon_t^2 - 1)^2 | \mathcal{F}_{t-1}]$  to be time-varying as in Escanciano (2009), but for simplicity and to allow for easier comparison with the results in the non-stationary case, we maintain Assumption 2(i). Assumption 2(ii) is needed in order to ensure that  $S_n(\hat{\vartheta}) = 0$  w.p.a.1.

As a first step towards eq. (11), the following lemma proves useful. It basically shows that the derivatives of the volatility-ratio process  $r_t^*(\vartheta)$  are stationary with suitable moments:

**Lemma 4** *Under Assumptions 1-2:  $\partial r_t^*(\vartheta) / (\partial \vartheta)$  and  $\partial^2 r_t^*(\vartheta) / (\partial \vartheta \partial \vartheta')$  are stationary and ergodic for all  $\vartheta \in \mathcal{W} \times \Theta$ . Moreover, there exists stationary and ergodic sequences  $B_{k,t} \in \mathcal{F}_{t-1}$ ,  $k = 0, 1, 2$ , which are independent of  $\vartheta$  such that*

$$\frac{1}{r_t^*(\vartheta)} \leq B_{0,t}, \quad \frac{\|\partial r_t^*(\vartheta) / (\partial \vartheta)\|}{r_t^*(\vartheta)} \leq B_{1,t}, \quad \frac{\|\partial^2 r_t^*(\vartheta) / (\partial \vartheta \partial \vartheta')\|}{r_t^*(\vartheta)} \leq B_{2,t},$$

for all  $\vartheta$  in a neighbourhood of  $\vartheta_0$ , where  $\mathbb{E}[B_{1,t} + B_{2,t}^2] < \infty$  and  $\mathbb{E}[B_{0,t} \{B_{1,t} + B_{2,t}^2\}] < \infty$ .

This lemma is used to construct suitable bounds for the score and hessian that allow us to appeal to CLT and LLN for stationary and ergodic sequences, and thereby establishing eq. (11):

**Theorem 5** *Under Assumptions 1-2, the QMLE  $\hat{\vartheta}$  satisfies eq. (12) where, with  $\kappa_4$  given in Assumption 2 and  $r_t^*(\vartheta)$  in eq. (16),  $\Sigma^{st} = \kappa_4 H^{st}$  and  $H^{st} = \mathbb{E} \left[ \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta} \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta'} \right]$ .*

### 3.2 Non-stationary Case

For consistency, we follow a similar strategy to develop the asymptotic properties of the QMLE when  $x_t^2$  is explosive, except a different variance-ratio approximation has to be used. To develop this variance-ratio approximation, we utilize some results derived in Han (2011). We impose the following conditions on the model which are slightly stronger than the ones imposed in the stationary case, but on the other hand allow for non-stationary regressors:

#### Assumption 3

- (i)  $\{\varepsilon_t\}$  and  $\{v_t\}$  are i.i.d., mutually independent, and satisfies  $\mathbb{E}[\varepsilon_t] = \mathbb{E}[v_t] = 0$ ,  $\mathbb{E}[\varepsilon_t^2] = 1$ , and  $\mathbb{E}[|v_t|^p] < \infty$  for some  $p \geq 2$ .
- (ii)  $\Theta = \{\theta \in \mathbb{R}^3 : \underline{\alpha} \leq \alpha \leq \bar{\alpha}, \underline{\beta} \leq \beta \leq \bar{\beta}, \underline{\pi} \leq \pi \leq \bar{\pi}\}$  and  $\mathcal{W} = [\underline{\omega}, \bar{\omega}]$  where  $0 < \underline{\alpha} < \bar{\alpha} < \infty$ ,  $0 < \underline{\beta} < \bar{\beta} < 1$ ,  $0 < \underline{\pi} < \bar{\pi} < \infty$  and  $0 < \underline{\omega} < \bar{\omega} < \infty$ .
- (iii)  $\{x_t\}$  solves eqs. (3)-(4) with  $d \in (-1/2, 1/2)$ .
- (iv)  $\mathbb{E}[|\varepsilon_t|^q] < \infty$  and  $\mathbb{E}[(\beta_0 + \alpha_0 \varepsilon_t^2)^{q/2}] < 1$  for some  $q > 4$ .
- (v)  $1/p + 2/q < 1/2 + d$ .

Assumption 3(i) requires the errors driving the model to be i.i.d. which is stronger than Assumption 1(i). We expect that it could be weakened to allow for some dependence, but this would greatly complicate the analysis. Similarly, the mutual independence of  $\{\varepsilon_t\}$  and  $\{v_t\}$  is a technical assumption and only used to establish the LLN and CLT in Lemma 6. Since Lemma 6 is only used in the analysis of  $\hat{\omega}$ , the proof of consistency and asymptotic normality of  $\hat{\theta}$  is valid without the independence assumption. We conjecture that Lemma 6, and thereby the asymptotic properties of  $\hat{\omega}$  as stated below, holds under weaker assumptions than independence, but this requires a different proof technique; see Wang (2013). Assumption 3(ii) restricts the parameters to be strictly positive; this is used when showing that  $r_t(\vartheta)$  is well-approximated by a stationary version uniformly over  $\vartheta$ . A similar restriction is found in Francq and Zakoian (2012). Assumption 3(iii) precisely defines the covariate  $\{x_t\}$  as an  $I(d_x)$  process with  $1/2 < d_x < 3/2$ . This restriction on  $d_x$  is imposed in order to employ the results of Han (2011) and the limit results in Lemma 6 below.

Assumptions 3(iv)-(v) correspond to Assumptions 2(b)-(c) in HP2012. Assumption 3(iv) introduces some moments conditions for the innovation sequences  $\{v_t\}$  and  $\{\varepsilon_t\}$ . It is stronger than  $\mathbb{E}[\log(\beta + \alpha \varepsilon_t^2)] < 0$  as imposed in Assumption 1(ii). In particular, while  $\alpha + \beta = 1$  is allowed for the stationary case in the previous section, (iv) rules this out in the nonstationary case. We

do not find this restrictive though since, when  $x_t$  is an I(1) process and  $\alpha + \beta = 1$ ,  $y_t^2$  has I(2) type behaviour which is not very likely for most economic and financial time series. Moreover, in most applications, when additional regressors are included, it is usually found that  $\alpha + \beta < 1$  so this restriction does not appear restrictive from an empirical point of view. Together Assumptions 3(iv)-(v) can lead to quite strong moment restrictions. For example, if  $d$  is close to  $-1/2$ , then  $p$  and  $q$  have to be chosen very large for the inequality in Assumption 3(v) to hold. These are used when developing the stationary approximation of the volatility ratio process  $r_t(\vartheta)$  which relies on the existence of certain moments. We conjecture that our theory would go through under weaker moment restrictions, but unfortunately we have not been able to demonstrate this here.

For the proof of the non-stationary case, we first present some additional notation and useful results. Let  $D[0, 1]$  be the space of *cadlag* functions on  $[0, 1]$  equipped with the uniform metric, and  $\Rightarrow$  denote weak convergence on  $D[0, 1]$ . Also, let  $L_{W_d}(t, x)$  denote the local time of a fractionally integrated Brownian motion and  $K > 0$  a normalizing constant (see Wang and Phillips, 2009a for precise definitions). Then the following lemma proves fundamental in establishing the necessary limit results for the score and hessian:

**Lemma 6** *Let  $\{x_t\}$  satisfy Assumption 3(iii) and  $f(x)$  be an integrable function.*

- (i) *Suppose  $\{w_t\}$  is stationary, independent of  $\{x_t\}$ , and satisfies  $\sum_{t=1}^{\infty} |\text{Cov}(w_0, w_t)| < \infty$ . Then,*

$$\frac{1}{n^{1/2-d}} \sum_{t=1}^{\lfloor ns \rfloor} f(x_{t-1})w_t \Rightarrow L_{W_d}(s, 0) \times K \mathbb{E}[w_t] \int_{-\infty}^{\infty} f(x)dx \text{ on } D[0, 1].$$

- (ii) *Suppose in addition that  $u_t$  is a martingale difference sequence w.r.t. a filtration  $\mathcal{F}_t$  that  $(x_{t-1}, w_t)$  is adapted to;  $\{x_t\}$  and  $\{u_t\}$  are independent,  $\mathbb{E}[u_t^2 | \mathcal{F}_{t-1}] = \sigma_u^2 > 0$  and  $\sup_{t \geq 1} \mathbb{E}[|u_t w_t|^{q_u}] < \infty$  a.s. for some  $q_u > 2$ ;  $\sum_{t=1}^{\infty} |\text{Cov}(w_0^2, w_t^2)| < \infty$ . Then,*

$$\frac{1}{n^{1/4-d/2}} \sum_{t=1}^{\lfloor ns \rfloor} f(x_{t-1})w_t u_t \Rightarrow \sqrt{L_{W_d}(s, 0)} G(s),$$

where  $G(s)$  is a Gaussian process which is independent of  $L_{W_d}(s, 0)$  and with covariance kernel  $(s_1 \wedge s_2) K \mathbb{E}[w_t^2] \sigma_u^2 \int_{-\infty}^{\infty} f^2(x)dx$ .

**Remark 7** *A sufficient condition for the assumptions on  $\{w_t\}$  in (i) and (ii) to hold is that it is stationary and  $\beta$ -mixing such that, for some  $\delta > 0$ ,  $\mathbb{E}[|w_t|^{2(1+\delta)}] < \infty$  and its mixing coefficients satisfy  $\sum_{t=1}^{\infty} \beta_t^{\delta/(1+\delta)} < \infty$ ; see, for example, Yoshihara (1976, Lemma 1).*

The above lemma is a generalization of the LLN and CLT established in Wang and Phillips (2009a) to allow for inclusion of a stationary component,  $w_t$ . It is the fundamental tool in our analysis of the score and hessian w.r.t.  $\omega$  since the first and second derivative of  $r_t(\vartheta)$  w.r.t.  $\omega$  can be written on the form  $f(x_{t-1})w_t$  for a suitable choice of  $f$  and  $w_t$ . Employing results in Han (2011), we also develop a stationary approximation of the variance ratio  $r_t(\vartheta) = \sigma_t^2(\vartheta) / \sigma_t^2$  that is used in the asymptotic analysis of the score and hessian w.r.t.  $\theta$ .

**Lemma 8** *Under Assumption 3,*

$$\sup_{\vartheta \in \mathcal{W} \times \Theta} \max_{1 \leq t \leq n} |r_t(\vartheta) - r_t^*(\theta)| = o_p(1), \quad (17)$$

where, with  $z_t = z_t(\theta_0)$ ,

$$r_t^*(\theta) := \frac{z_t(\theta)}{z_t}, \quad z_t(\theta) = \alpha \sum_{i=1}^{\infty} \beta^{i-1} z_{t-i} \varepsilon_{t-i}^2 + \frac{\pi}{\pi_0} \frac{1}{1-\beta}. \quad (18)$$

The sequence  $r_t^*(\theta)$  is stationary and ergodic with  $\mathbb{E} \left[ \sup_{\theta} r_t^*(\theta)^{-k} \right] < \infty$  for any  $k \in \mathbb{R}$ . Moreover,  $\sup_{\vartheta} \{ \sigma_t^2(\vartheta_0) \sigma_t^{-2}(\vartheta) \} \leq W_t$ , where  $W_t$  is stationary and ergodic with  $\mathbb{E} [W_t^k] < \infty$  for any  $k > 0$ .

Lemma 8 is used to establish eq. (8). It is important to note that  $r_t^*(\theta)$  does not depend on the regressor  $x_t$  (and so is stationary), but still contains information about its regression coefficient,  $\pi$ . On the other hand,  $r_t^*(\theta)$ , and thereby  $L_n^*(\vartheta) = L_n^*(\theta)$ , is independent of  $\omega$  and so asymptotically the log-likelihood contains no information about this parameter in large samples. We are therefore only able to show global consistency of  $\hat{\theta}$ . However, a local analysis of  $L_n(\vartheta)$ , where Lemma 6 is used to verify the high-level conditions in Kristensen and Rahbek (2010, Lemma 11), shows that  $\hat{\omega}$  is locally consistent but converges at a slower than standard rate:

**Theorem 9** *Under Assumption 3,  $\hat{\theta} \rightarrow_p \theta_0$ . Moreover, with probability tending to one as  $n \rightarrow \infty$ , there exists a unique maximum point  $\hat{\vartheta} = (\hat{\omega}, \hat{\theta})$  of  $L_n(\vartheta)$  in  $\{ \vartheta : |\omega - \omega_0| \leq \epsilon, \quad n^{1/4+d/2} \|\theta - \theta_0\| \leq \epsilon \}$ , for some  $\epsilon > 0$ , that satisfies  $\hat{\omega} = \omega_0 + o_p(1)$  and  $\hat{\theta} = \theta_0 + o_p(1/n^{1/4+d/2})$ .*

To avoid additional notation, we here use  $\hat{\theta}$  to denote both the global and local estimator. In finite samples, these two could differ if there exists a local maximum in a neighbourhood of  $\theta_0$ . Moreover, the stated rate result only holds for the local estimator. Ideally, we would have carried out a global analysis of  $\hat{\omega}$  as well. However, to our knowledge, there exists no results for global consistency for estimators in non-standard settings where components of the estimator converge at different rates; see e.g. Kristensen and Rahbek (2010).

Next, we analyze the asymptotic distribution of  $\vartheta$  where we apply the general result of Kristensen and Rahbek (2010, Lemma 12) to our specific estimator:

**Theorem 10** *Let Assumption 3 hold. Then eq. (15) holds with  $\Sigma^{nst} = \kappa_4 H^{nst}$  and*

$$H^{nst} = \begin{bmatrix} H_{\omega\omega}^{nst} & O_{1 \times 3} \\ O_{3 \times 1} & H_{\theta\theta}^{nst} \end{bmatrix} \in \mathbb{R}^{4 \times 4}.$$

where

$$\begin{aligned} H_{\omega\omega}^{nst} &= K \frac{\mathbb{E} [1/z_t^2]}{(1-\beta_0)^2} \int_{-\infty}^{\infty} \left( \frac{1}{\omega_0 + \pi_0 s^2} \right)^2 ds \times L_{W_d}(1, 0), \\ H_{\theta\theta}^{nst} &= \mathbb{E} \left[ \frac{\partial r_t^*(\theta_0)}{\partial \theta} \frac{\partial r_t^*(\theta_0)}{\partial \theta'} \right] \in \mathbb{R}^{3 \times 3}. \end{aligned}$$

## 4 Robust Inference

Comparing Theorems 5 and 10, we see that the large-sample distribution of the QMLE changes quite substantially when we move from the stationary case to the non-stationary one. One could therefore fear that, for a chosen regressor, inference would be dependent on whether  $x_t$  is stationary or not. However, in both cases, the limiting distribution of the QMLE is mixed normal with the (possibly random) covariance matrix being the product of limits of the (appropriately scaled) score and hessian. Whether  $x_t$  is stationary or not, a natural estimator of the covariance matrix is

$$\hat{\Omega} = H_n^{-1}(\hat{\vartheta})\Sigma_n(\hat{\vartheta})H_n^{-1}(\hat{\vartheta}), \text{ where } \Sigma_n(\vartheta) = \sum_{t=1}^n \frac{\partial \ell_t(\vartheta)}{\partial \vartheta} \frac{\partial \ell_t(\vartheta)}{\partial \vartheta'}, \quad (19)$$

and  $H_n(\vartheta)$  is defined in eq. (10). As we shall see,  $\hat{\Omega}$  automatically adjust to the level of persistence and converges to the correct asymptotic limit in both cases. As a consequence, for example, standard  $t$ -statistic will be normally distributed in large samples whether  $x_t$  is stationary or non-stationary. That is, standard inferential procedures regarding  $\vartheta_0$  are robust to the persistence of  $x_t$ . We conjecture that similar results hold for other statistics such as the likelihood-ratio statistic.

**Theorem 11** *Under either Assumptions 1-2 or Assumption 3, with  $\hat{\Omega}$  defined in eq. (19),*

$$t = \hat{\Omega}^{-1/2}\{\hat{\vartheta} - \vartheta_0\} \rightarrow_d N(0, I_4),$$

## 5 Simulation Study

To investigate the relevance and usefulness of our asymptotic results, we conduct a simulation study to see whether standard  $t$ -statistics are sensitive towards the level of persistence,  $d_x$ , in finite samples. Our simulation design is based on the GARCH-X model with the exogenous regressor  $x_t$  being generated by  $x_t = (1 - L)^{-d_x} v_t$ . The data-generating GARCH parameter values are set to be  $\omega_0 = 0.01$ ,  $\alpha_0 = 0.05$ ,  $\beta_0 = 0.6$  and  $\pi_0 = 0.1$ . These parameter values are similar to the estimates reported in Shephard and Sheppard (2010) where  $x_t^2$  is a realized volatility measure. The innovation processes  $\{\varepsilon_t\}$  and  $\{v_t\}$  are chosen to be i.i.d. standard normal and mutually independent.<sup>2</sup> The initial values are set  $x_0 = 0$  and  $\sigma_0^2 = 0.01$ . We consider the following four data generating processes depending on  $d_x$  in  $x_t$ .

stationary cases		nonstationary cases	
DGP 1	$d_x = 0.0$	DGP 3	$d_x = 0.7$
DGP 2	$d_x = 0.3$	DGP 4	$d_x = 1.0$

The null distributions of each of the  $t$ -statistics associated with  $\omega$ ,  $\alpha$ ,  $\beta$  and  $\pi$  are simulated for  $n = 500$  and  $5,000$  with  $10,000$  iterations. The simulation results are reported in Figures 1 and 2 in Appendix C. Figure 1 reports the results for the stationary cases and show that the large

<sup>2</sup>We also tried the case for  $v_t = -\varepsilon_t$  and the results are still similar.

sample  $N(0, 1)$  distribution of the  $t$ -statistics is a very good finite-sample approximation. For the non-stationary cases as reported in Figure 2, the asymptotic  $N(0, 1)$  approximation is also precise, albeit less so compared to the stationary case.

Our simulation results show that the empirical distributions of the  $t$ -statistics are close to normal for moderate sample sizes and become more so as the sample size increases. This is true regardless of the value of the memory parameter  $d_x$  in  $x_t$ . In conclusion, the individual  $t$ -statistics of  $(\omega, \alpha, \beta, \pi)$  are robust towards the dependence structure of  $x_t$  in the GARCH-X model. Researchers do not need to determine whether  $x_t$  is stationary or not before they implement the QMLE and associated inferential tools for the GARCH-X model.

## 6 Conclusion

We have here developed asymptotic theory of QMLE's in GARCH models with additional persistent covariates in the variance specification. It is shown that the asymptotic behaviour of the QMLE's depend on whether the regressor is stationary or not. At the same time, standard inferential tools, such as  $t$ -statistics, for the parameters are robust towards the level of persistence. In particular, in contrast to the explosive case in pure GARCH models, one can draw inference about the intercept parameter  $\omega$ .

A number of extensions of the theory would be of interest: For example, to show global consistency of  $\hat{\omega}$  and to analyze the properties of the QMLE in alternative GARCH specifications with persistent regressors.

## A Proofs of Section 3.1

**Proof of Lemma 1.** With  $\rho_t := \alpha_0 \varepsilon_{t-1}^2 + \beta_0 \geq 0$ , and  $b_t := \omega_0 + \pi_0 x_t^2 \geq 0$ , rewrite eq. (6) as  $\sigma_t^2 = \rho_t \sigma_{t-1}^2 + b_t$ . This is a stochastic recursion where  $\{(a_t, b_t)\}$  is a stationary and ergodic sequence. The first part of the result now follows from Brandt (1986) since Assumption 1(ii) implies that the Lyapunov coefficient associated with the above stochastic recursion is negative and that  $\mathbb{E}[\log^+(b_t)] < \infty$ . The stationary solution can be written as  $\sigma_t^2 = b_t + \sum_{i=0}^{\infty} \rho_t \cdots \rho_{t-i} b_{t-i-1}$ . Following Berkes et al (2003, p. 207-208), the negative Lyapunov coefficient implies that  $\mathbb{E}[(\rho_0 \cdots \rho_m)^{2s}] < 1$  for some  $s > 0$  and  $m \geq 1$ ; thus,  $\mathbb{E}[(\rho_t \cdots \rho_{t-i})^{2s}] \leq c \bar{\rho}^i$  for some  $c < \infty$  and  $\bar{\rho} < 1$ . Without loss of generality, we choose  $s < q/2$  with  $q$  given in Assumption 1(ii) such that  $\mathbb{E}[b_t^{2s}] < \infty$ . Thus,

$$\begin{aligned} \mathbb{E}[\sigma_t^{2s}] &\leq \mathbb{E}[b_t^s] + \sum_{i=0}^{\infty} \mathbb{E}[(\rho_t \cdots \rho_{t-i})^s b_t^s] \leq \mathbb{E}[b_t^s] + \sqrt{\mathbb{E}[b_t^{2s}]} \sum_{i=0}^{\infty} \sqrt{\mathbb{E}[(\rho_t \cdots \rho_{t-i})^{2s}]} \\ &= \mathbb{E}[b_t^s] + c \sqrt{\mathbb{E}[b_t^{2s}]} (1 - \bar{\rho})^{-1} < \infty. \end{aligned}$$

That  $\mathbb{E}[y_t^{2s}] < \infty$  follows from eq. (1) together with Assumption 1(ii). ■

**Proof of Lemma 2.** Eq. (2) can be rewritten as  $\sigma_t^2(\vartheta) := \beta \sigma_{t-1}^2(\vartheta) + w_t(\vartheta)$  which is an AR(1) model with stationary errors  $w_t(\vartheta)$ . The first part of the result now follows by Berkes et al (2003, Lemma 2.2). From Lemma 1 together with Assumption 1(ii),  $\mathbb{E}[\sup_{\vartheta \in \Theta} w_t^s(\vartheta)] \leq \bar{\omega}^s + \bar{\alpha}^s \mathbb{E}[y_{t-1}^{2s}] + \bar{\pi}^s \mathbb{E}[x_{t-1}^{2s}] < \infty$ . Thus,

$$\mathbb{E} \left[ \sup_{\vartheta \in \Theta} \sigma_{0,t}^{2s}(\vartheta) \right] \leq \sum_{i=0}^{\infty} \bar{\beta}^{is} \mathbb{E} \left[ \sup_{\vartheta \in \Theta} w_t^s(\vartheta) \right] < \infty.$$

Next, observe that  $\sigma_t^2(\vartheta) - \sigma_{0,t}^2(\vartheta) = \beta^t \{\bar{\sigma}_0^2 - \sigma_{0,0}^2(\vartheta)\}$ , where  $\bar{\sigma}_0^2 > 0$  is the fixed, initial value. The result now follows with  $K_s = \mathbb{E}[\sup_{\vartheta \in \mathcal{D}} |\bar{\sigma}_0^2 - \sigma_{0,0}^2(\vartheta)|^s]$  which is finite since  $\mathbb{E}[\sup_{\vartheta \in \Theta} \sigma_{0,0}^{2s}(\vartheta)] < \infty$ . ■

**Proof of Theorem 3.** Define  $\hat{\vartheta}^* = \arg \max_{\vartheta \in \Theta} L_n^*(\vartheta)$  where  $L_n^*(\vartheta)$  is defined in eq. (9) with  $r_t(\vartheta)$  given in eq. (16). We first show consistency of  $\hat{\vartheta}^*$  by verifying the conditions in Kristensen and Rahbek (2005, Proposition 2):

- (i) The parameter space  $\Theta$  is a compact Euclidean space with  $\vartheta_0 \in \Theta$ .
- (ii)  $\vartheta \mapsto \ell_t^*(\vartheta)$  is continuous almost surely.
- (iii)  $L_n^*(\vartheta)/n \rightarrow_p L^*(\vartheta) := \mathbb{E}[\ell_t^*(\vartheta)]$  where the limit exists,  $\forall \vartheta \in \Theta$ .
- (iv)  $L^*(\vartheta_0) > L^*(\vartheta)$ ,  $\forall \vartheta \neq \vartheta_0$ .
- (v)  $\mathbb{E}[\sup_{\vartheta \in \mathcal{D}} \ell_t^*(\vartheta)] < +\infty$  for any compact set  $\mathcal{D} \subset \Theta$  with  $\theta_0 \notin \mathcal{D}$ .

Condition (i) holds by assumption, while (ii) follows by the continuity of  $\vartheta \mapsto r_t^2(\vartheta)$  as given in eq. (16). Condition (iii) follows by the LLN for stationary and ergodic sequences if the limit  $L^*(\vartheta)$  exists; the limit is indeed well-defined since  $\ell_t^*(\vartheta) \leq -\log(\omega/\omega_0)$  such that  $\mathbb{E}[\ell_t^*(\vartheta)^+] < \infty$ . To prove condition (iv), first observe that  $r_t^*(\vartheta_0) = 1$  which in turn implies that  $L(\vartheta_0) =$



0. Moreover,  $\omega_0 \leq \log(\sigma_{0,t}^2(\vartheta_0))$  such that  $\mathbb{E}[(\log \sigma_{0,t}^2(\vartheta_0))^-] < \infty$ , while  $\mathbb{E}[(\log \sigma_{0,t}^2(\vartheta_0))^+] \leq (\log \mathbb{E}[\sigma_{0,t}^2(\vartheta_0)])^+ / s < \infty$  by Jensen's inequality and Lemma 2. Thus,  $\mathbb{E}[|\ell_t^*(\vartheta_0)|] < \infty$  is well-defined, while either (a)  $L(\vartheta) = -\infty$  or (b)  $L(\vartheta) \in (-\infty, \infty)$ . Now, let  $\vartheta \neq \vartheta_0$  be given: Then, if (a) holds,  $L^*(\vartheta_0) > -\infty = L^*(\vartheta)$ . If (b) holds, the following calculations are allowed:

$$L^*(\vartheta) = -\mathbb{E} \left[ \log(r_t^*(\vartheta)) + \left\{ \frac{1}{r_t^*(\vartheta)} - 1 \right\} \varepsilon_t^2 \right] = -\mathbb{E} \left[ \log(r_t^*(\vartheta)) + \left\{ \frac{1}{r_t^*(\vartheta)} - 1 \right\} \right],$$

where we have used that  $\mathbb{E}[\varepsilon_t^2 | \mathcal{F}_{t-1}] = 1$ .  $L^*(\vartheta) \leq 0 = L^*(\vartheta_0)$  with equality if and only if  $r_t^2(\vartheta) = 1$  a.s. Suppose that  $r_t^2(\vartheta) = 1$  a.s.  $\Leftrightarrow \sigma_{0,t}^2(\vartheta) = \sigma_{0,t}^2(\vartheta_0)$  a.s. or, equivalently,

$$\omega_0 + \sum_{i=1}^{\infty} c_i(\vartheta_0) Z_{t-i} = \omega + \sum_{i=1}^{\infty} c_i(\vartheta) Z_{t-i}, \quad (20)$$

where  $c_i(\vartheta) = (\alpha\beta^{i-1}, \pi\beta^{i-1})'$  and  $Z_{t-1} = (y_{t-1}^2, x_{t-1}^2)'$ . We then claim that  $\omega_0 = \omega$  and  $c_i(\vartheta_0) = c_i(\vartheta)$  for all  $i \geq 1$ ; this in turn implies  $\vartheta = \vartheta_0$ . We show this by contradiction: Let  $m > 0$  be the smallest integer for which  $c_i(\vartheta_0) \neq c_i(\vartheta)$  (if  $c_i(\vartheta_0) = c_i(\vartheta)$  for all  $i \geq 1$ , then  $\omega_0 = \omega$ ). Thus,

$$a_0 y_{t-m}^2 + b_0 x_{t-m}^2 = \omega - \omega_0 + \sum_{i=1}^{\infty} a_i y_{t-m-i}^2 + \sum_{i=1}^{\infty} b_i x_{t-m-i}^2,$$

where  $a_i := \alpha_0 \beta_0^{i-1} - \alpha \beta^{i-1}$  and  $b_i := \pi_0 \beta_0^{i-1} - \pi \beta^{i-1}$ . The right hand side belongs to  $\mathcal{F}_{t-m-1}$ . Thus,  $a_0 y_{t-m}^2 + b_0 x_{t-m}^2 | \mathcal{F}_{t-m-1}$  is constant. This is ruled out by Assumption 1(iv). Finally, condition (v) follows from  $\sup_{\vartheta \in \mathcal{D}} \ell_t^*(\vartheta) \leq -\sup_{\vartheta \in \mathcal{D}} \log(\omega) \leq -\log(\underline{\omega}) < +\infty$ .

Now, return to the actual, feasible QMLE,  $\hat{\vartheta}$ . Using Lemma 2,

$$\begin{aligned} \sup_{\vartheta \in \mathcal{W} \times \Theta} |L_n^*(\vartheta) - L_n(\vartheta)| &\leq \sum_{t=1}^n \sup_{\vartheta \in \Theta} \left\{ \left| \frac{\sigma_t^2(\vartheta) - \sigma_{0,t}^2(\vartheta)}{\sigma_t^2(\vartheta) \sigma_{0,t}^2(\vartheta)} \right| y_{t-1}^2 + \left| \log \left( 1 + \frac{\sigma_t^2(\vartheta) - \sigma_{0,t}^2(\vartheta)}{\sigma_t^2(\vartheta)} \right) \right| \right\} \\ &\leq \frac{K}{\underline{\omega}^2} \sum_{t=1}^n \bar{\beta}^t y_{t-1}^2 + \frac{K}{\underline{\omega}^2} \sum_{t=1}^n \bar{\beta}^t, \end{aligned}$$

where  $\lim_{n \rightarrow \infty} \sum_{t=1}^n \bar{\beta}^t = (1 - \bar{\beta})^{-1} < \infty$  while  $\lim_{n \rightarrow \infty} \sum_{t=1}^n \bar{\beta}^t y_{t-1}^2 < \infty$  by Berkes et al (2003, Lemma 2.2) in conjunction with Lemma 1. Thus,  $\sup_{\vartheta \in \Theta} |L_n^*(\vartheta) - L_n(\vartheta)| / n = o_p(1/n)$ . Combining this with the above analysis of  $L_n^*(\vartheta)$ , it then follows from Kristensen and Shin (2012, Proposition 1) that  $\|\hat{\vartheta}^* - \hat{\vartheta}\| = o_p(1/n)$ . In particular,  $\hat{\vartheta}$  is consistent. ■

**Proof of Lemma 4.** Observe that

$$\begin{aligned} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \omega} &= \sum_{i=0}^{\infty} \beta^i = \frac{1}{1 - \beta}, & \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \alpha} &= \sum_{i=0}^{\infty} \beta^i y_{t-i-1}^2, \\ \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \pi} &= \sum_{i=0}^{\infty} \beta^i x_{t-i-1}^2, & \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \beta} &= \sum_{i=0}^{\infty} \beta^i \sigma_{0,t-i-1}^2(\vartheta). \end{aligned}$$

By the same arguments as in the proof of Lemma 2, these processes are stationary.

The proof for the second-order partial derivatives w.r.t.  $\omega$ ,  $\alpha$  and  $\beta$  proceeds along the lines of Francq and Zakoïan (2004, p. 619) since these do not involve  $x_t$ . Regarding the second-order derivatives involving  $\pi$ , using the above expressions of the first-order derivatives:

$$\frac{\partial^2 \sigma_{0,t}^2(\vartheta)}{\partial \omega \partial \pi} = \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \alpha \partial \pi} = \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \pi^2} = 0, \quad \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \beta \partial \pi} = \sum_{i=0}^t \beta^i \frac{\partial \sigma_{0,t-i-1}^2(\vartheta)}{\partial \pi}.$$

Again, these are clearly stationary.

Moreover, by the same arguments as in Francq and Zakoïan, 2004, p. 622), there exists constants  $c < \infty$  and  $0 < \rho < 1$  such that for all  $\vartheta$  in a neighbourhood of  $\vartheta_0$  and all  $0 < r \leq s$ ,

$$\frac{1}{r_t^*(\vartheta)} = \frac{\sigma_t^2}{\sigma_{0,t}^2(\vartheta)} \leq c \sum_{i=0}^{\infty} \rho^{ri} \bar{w}_{t-i}^r =: B_{0,t},$$

where  $\bar{w}_t := \bar{\omega} + \bar{\alpha} y_{t-1}^2 + \bar{\pi} x_{t-1}^2$  is stationary and ergodic with  $\mathbb{E}[\bar{w}_t^r] < \infty$ . This in turn implies that  $B_{0,t}$  is stationary and ergodic with first moment. Given the representations of  $\sigma_{0,t}^2(\vartheta_0)$  and  $\partial \sigma_{0,t}^2(\vartheta_0) / (\partial \vartheta)$ , it is easily shown that for some constant  $c < \infty$  the following inequalities hold for all  $\vartheta$  in a neighbourhood of  $\vartheta_0$  (see Francq and Zakoïan, 2004, p. 619):

$$\begin{aligned} \frac{1}{\sigma_{0,t}^2(\vartheta_0)} \frac{\partial \sigma_{0,t}^2(\vartheta_0)}{\partial \omega} &\leq \frac{1}{\omega_0}, & \frac{1}{\sigma_{0,t}^2(\vartheta_0)} \frac{\partial \sigma_{0,t}^2(\vartheta_0)}{\partial \alpha} &\leq \frac{1}{\alpha_0}, \\ \frac{1}{\sigma_{0,t}^2(\vartheta_0)} \frac{\partial \sigma_{0,t}^2(\vartheta_0)}{\partial \pi} &\leq \frac{1}{\pi_0}, & \frac{1}{\sigma_{0,t}^2(\vartheta_0)} \frac{\partial \sigma_{0,t}^2(\vartheta_0)}{\partial \beta} &\leq c \sum_{i=0}^{\infty} \bar{\beta}^{ri} \bar{w}_t^r. \end{aligned}$$

Choosing  $B_{1,t}$  as the Euclidean norm of the above individual bounds,  $\|\partial r_t^*(\vartheta) / (\partial \vartheta)\| / r_t^*(\vartheta) = \|\partial \sigma_{0,t}^2(\vartheta) / (\partial \vartheta)\| / \sigma_{0,t}^2(\vartheta_0) \leq B_{1,t}$ . By the same arguments as in Francq and Zakoïan (2004, p. 620), similar bounds can be established for the second order derivatives. For example, we have  $\|\partial^2 \sigma_{0,t}^2(\vartheta) / (\partial \beta \partial \pi)\| / \sigma_{0,t}^2(\vartheta_0) \leq c \sum_{i=0}^{\infty} \bar{\beta}^{ri} \bar{w}_t^r$ , and we define  $B_{2,t}$  accordingly. By inspection of the definitions of  $B_{0,t}$ ,  $B_{1,t}$  and  $B_{2,t}$ , one finds that stated moment exists by choosing  $r > 0$  sufficiently small. ■

**Proof of Theorem 5.** As shown in the proof of Theorem 3,  $\|\hat{\vartheta}^* - \hat{\vartheta}\| = o_p(1/\sqrt{n})$ ; thus, it suffices to analyze  $\hat{\vartheta}^*$ . The score and hessian are given by

$$\begin{aligned} S_n^*(\vartheta) &= \frac{\partial L_n^*(\vartheta)}{\partial \vartheta} = \sum_{t=1}^n \frac{1}{\sigma_{0,t}^2(\vartheta)} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta} \left\{ \frac{y_t^2}{\sigma_{0,t}^2(\vartheta)} - 1 \right\}, \\ H_n^*(\vartheta) &= \frac{\partial^2 L_n^*(\vartheta)}{\partial \vartheta \partial \vartheta'} = \sum_{t=1}^n h_t^*(\vartheta), \end{aligned}$$

where derivatives w.r.t.  $\sigma_{0,t}^2(\vartheta)$  can be found in the proof of Lemma 4, and

$$h_t^*(\vartheta) = \left\{ \frac{1}{\sigma_{0,t}^2(\vartheta)} \frac{\partial^2 \sigma_{0,t}^2(\vartheta)}{\partial \vartheta \partial \vartheta'} - \frac{1}{\sigma_{0,t}^4(\vartheta)} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta'} \right\} \left\{ \frac{y_t^2}{\sigma_{0,t}^2(\vartheta)} - 1 \right\} - \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta'} \frac{y_t^2}{\sigma_{0,t}^6(\vartheta)}.$$

We now verify the the two convergence results stated in eq. (11): First, we employ the CLT for Martingale differences in Brown (1971, Theorem 2) to show that

$$\frac{1}{\sqrt{n}} S_n^*(\vartheta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta} \{\varepsilon_t^2 - 1\} \rightarrow_d N(0, V_{\vartheta\vartheta}^{\text{stat}}(\vartheta_0)). \quad (21)$$

By Assumption 1(i),  $X_t := \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta} \{\varepsilon_t^2 - 1\}$  is a Martingale difference and  $S_n^*(\vartheta_0)/\sqrt{n}$  has quadratic variation

$$\langle S_n^*(\vartheta_0)/\sqrt{n} \rangle = \kappa_4 \frac{1}{n} \sum_{t=1}^n \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta} \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta'} \rightarrow_p \kappa_4 \mathbb{E} \left[ \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta} \frac{\partial r_t^*(\vartheta_0)}{\partial \vartheta'} \right] < \infty,$$

where we have used Assumption 2(i) and Lemma 4. This shows that eq. (1) in Brown (1971) holds. Eq. (2) of Brown (1971) holds since, by stationarity and  $\mathbb{E}[\|X_t\|^2] < \infty$ ,

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[ \|X_t\|^2 \mathbb{I} \{ \|X_t\| > c\sqrt{n} \} \right] = \mathbb{E} \left[ \|X_t\|^2 \mathbb{I} \{ \|X_t\| > c\sqrt{n} \} \right] \rightarrow 0.$$

For the hessian,  $\|h_t^*(\vartheta)\| \leq \{B_{2,t} + B_{1,t}^2\} \{1 + B_{0,t}\varepsilon_t^2\} + B_{1,t}^2 B_{0,t}\varepsilon_t^2$  for all  $\vartheta$  in some neighbourhood of  $\vartheta_0$ , where the right-hand side has finite first moment, c.f. Lemma 4. It now follows by standard uniform convergence results for averages of stationary sequences (see e.g. Kristensen and Rahbek (2005, Proposition 1) that  $\sup_{\|\vartheta - \vartheta_0\| < \delta} \|H_n^*(\vartheta) - H^{\text{st}}(\vartheta)\| \rightarrow_p 0$ , for some  $\delta > 0$ , where  $H^{\text{stat}}(\vartheta) = \mathbb{E} \left[ h_{\vartheta\vartheta,t}^*(\vartheta) \right]$ . Moreover,  $\vartheta \mapsto H^{\text{st}}(\vartheta)$  is continuous. Since  $\hat{\vartheta}^* \rightarrow_p \vartheta_0$ ,  $\bar{\vartheta} \rightarrow_p \vartheta_0$  and so lies in any arbitrarily small neighbourhood w.p.a.1. To complete the proof, we verify that  $H_{\vartheta\vartheta}^{\text{st}}(\vartheta_0)$  is non-singular: The process  $\Psi_t := \partial \sigma_{0,t}^2(\vartheta_0) / (\partial \vartheta) \in \mathbb{R}^4$  can be written as  $\Psi_t = \beta \Psi_{t-1} + W_t$ , where  $W_t := [1, y_{t-1}, x_{t-1}, \sigma_{0,t-1}^2(\vartheta_0)]'$ . Suppose that there exists  $\lambda \in \mathbb{R}^4 \setminus \{0\}$  and  $t \geq 1$  such that  $\lambda' \Psi_t = 0$  a.s. Since  $\Psi_t$  is stationary, this must hold for all  $t$ . This implies that  $\lambda' W_t = 0$  a.s. for all  $t \geq 1$ . However, this is ruled out by Assumption 1(iv). It must therefore hold that  $\lambda' \Psi_t / \sigma_{0,t}^2(\vartheta_0) = 0$  if and only if  $\lambda = 0$ ; thus,  $H^{\text{st}}(\vartheta_0) = \mathbb{E} \left[ \Psi_t \Psi_t' / \sigma_{0,t}^4(\vartheta_0) \right]$  is non-singular. ■

## B Proofs of Section 3.2

**Proof of Lemma 6.** To prove (i), define  $\psi_n'(s) = n^{-(1/2-d)} \sum_{t=1}^{\lfloor ns \rfloor} f(x_{t-1}) w_t$  and  $\psi_n''(s) = n^{-(1/2-d)} \sum_{t=1}^{\lfloor ns \rfloor} f(x_{t-1}) \mathbb{E}[w_t]$  which both belong to  $D[0,1]$ . First, by Theorem 2.1 in Wang and Phillips (2009a), henceforth WP2009a, and Lemma 1 in Kasparis et. al. (2012),  $\psi_n''(s) \Rightarrow$

$L_{W_d}(s, 0) \times K \mathbb{E}[w_t] \int_{-\infty}^{\infty} f(x) dx$  on  $D[0, 1]$ . We show the following two claims: (i.a)  $|\psi'_n(s) - \psi''_n(s)| = o_p(1)$  and (i.b)  $\psi'_n(s)$  is tight; (i.a) implies that  $\psi'_n(s)$  and  $\psi''_n(s)$  have the same finite dimensional limit distributions which together with (i.b) imply weak convergence of  $\psi'_n(s)$  towards the limit of  $\psi''_n(s)$ . To show (i.a), use independence between  $w_t$  and  $x_t$  to write with  $\mathcal{X}_n = (x_1, \dots, x_n)$ ,

$$\mathbb{E} \left[ |\psi'_n(s) - \psi''_n(s)|^2 \mid \mathcal{X}_n \right] = \frac{\text{Var}(w_t)}{n^{2(1/2-d)}} \sum_{t=1}^n f^2(x_{t-1}) + \frac{1}{n^{2(1/2-d)}} \sum_{t \neq u} f(x_{t-1}) f(x_{u-1}) \text{Cov}(w_t, w_u).$$

Using the covariance condition together with  $f(x) \leq C$  for some  $C < \infty$ , we obtain

$$\begin{aligned} \left| \sum_{t \neq u} f(x_{t-1}) f(x_{u-1}) \text{Cov}(w_t, w_u) \right| &\leq \sum_{t \neq u} |f(x_{t-1})| |f(x_{u-1})| |\text{Cov}(w_t, w_u)| \\ &\leq C \sum_{t=1}^n |f(x_{t-1})| \sum_{u=1}^{\infty} |\text{Cov}(w_t, w_u)| \\ &= C \sum_{t=1}^n |f(x_{t-1})| \times \sum_{u=1}^{\infty} |\text{Cov}(w_0, w_u)| \end{aligned}$$

By WP2009a,  $n^{-1/2+d} \sum_{t=1}^n |f(x_{t-1})|^q = O_p(1)$ ,  $q = 1, 2$ , and so  $\mathbb{E} \left[ |\psi'_n(s) - \psi''_n(s)|^2 \mid \mathcal{X}_n \right] = o_p(1)$ .

By Markov's inequality, this implies that  $P \left( |\psi'_n(s) - \psi''_n(s)|^2 > \delta \mid \mathcal{X}_n \right) = o_p(1)$  for any  $\delta > 0$ .

Thus,  $P \left( |\psi'_n(s) - \psi''_n(s)|^2 > \delta \right) = \mathbb{E} \left[ P \left( |\psi'_n(s) - \psi''_n(s)|^2 > \delta \mid \mathcal{X}_n \right) \right] \rightarrow 0$ . To show (i.b), we apply Theorem 5 in Billingsley (1974) and wish to show that there exists a sequence of  $\alpha_n(\epsilon, \delta)$  satisfying  $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \alpha_n(\epsilon, \delta) = 0$  for each  $\epsilon > 0$  such that, for  $0 \leq s_1 \leq s_2 \leq \dots \leq s_m \leq s \leq 1$ ,  $s - s_m \leq \delta$ , we have

$$P \left( |\psi'_n(s) - \psi'_n(s_m)| \geq \epsilon \mid \psi'_n(s_1), \psi'_n(s_2), \dots, \psi'_n(s_m) \right) \leq \alpha_n(\epsilon, \delta), \quad \text{a.s.} \quad (22)$$

A sufficient conditions for eq. (22) is

$$\sup_{|s_1 - s_2| \leq \delta} P \left( \left| \sum_{t=[ns_1]+1}^{[ns_2]} f(x_{t-1}) w_t \right| \geq \epsilon n^{1/2-d} \mid \psi'_n(s_1), \psi'_n(s_2), \dots, \psi'_n(s_m) \right) \leq \alpha_n(\epsilon, \delta).$$

As before, we first establish a conditional version: Define  $\alpha_n(\mathcal{X}_n, \epsilon, \delta)$  as

$$\alpha_n(\mathcal{X}_n, \epsilon, \delta) := \epsilon^{-2} n^{-2(1/2-d)} \sup_{0 \leq s \leq \delta} \mathbb{E} \left[ \left| \left\{ \sum_{t=1}^{[ns]} f(x_{t-1}) w_t \right\}^2 \right| \mathcal{X}_n \right].$$

Similar to the proof of (i.a), we have that, for large enough  $n$ ,

$$\begin{aligned}\alpha_n(\mathcal{X}_n, \epsilon, \delta) &\leq \epsilon^{-2} n^{-2(1/2-d)} \sum_{t=1}^n f^2(x_{t-1}) \mathbb{E}[w_t^2] + \epsilon^{-2} n^{-2(1/2-d)} \sum_{t_1 \neq t_2} f(x_{t_1-1}) f(x_{t_2-1}) |\mathbb{E}[w_{t_1} w_{t_2}]| \\ &\leq \epsilon^{-2} n^{-(1/2-d)} O_p(1).\end{aligned}$$

This shows that eq. (22) holds in probability conditional on  $\mathcal{X}_n$  which in turn implies that it also holds unconditionally of  $\mathcal{X}_n$ .

To show (ii), write  $n^{d/2-1/4} \sum_{t=1}^{[ns]} f(x_{t-1}) w_t u_t = \sum_{t=1}^{[ns]} Z_{n,t} w_t u_t$  where  $Z_{n,t} := n^{-(1/4-d/2)} f(x_{t-1})$ . The sequence  $\{Z_{n,t} w_t u_t\}$  is a martingale difference w.r.t.  $\mathcal{F}_t$  with quadratic variation,  $\sigma_u^2 \sum_{t=1}^{[ns]} Z_{n,t}^2 w_t^2$ . By the same arguments as in the proof of part (i) of this lemma,  $\sigma_u^2 \sum_{t=1}^{[ns]} Z_{n,t}^2 w_t^2 = \Lambda_n(s) + o_p(1)$  where

$$\Lambda_n(s) = \sigma_u^2 \mathbb{E}[w_t^2] \times n^{-(1/2-d)} \sum_{t=1}^{[ns]} f^2(x_{t-1}) \Rightarrow K \sigma_u^2 \mathbb{E}[w_t^2] \int_{-\infty}^{\infty} f^2(x) dx \times L_{W_d}(s, 0).$$

As in Proof of Theorem 3.1 in WP2009a, under a suitable probability space there exists an equivalent process  $x_t^*$  of  $x_t$  such that the corresponding quadratic variation  $\Lambda_n^*(s) \rightarrow_p K \sigma_u^2 \mathbb{E}[w_t^2] \int_{-\infty}^{\infty} f^2(x) dx \times L_{W_d}(s, 0)$ . Without loss of generality we assume that  $x_t$  satisfies this. We now wish to show that  $V_n(s) := \Lambda_n^{-1}(s) \sum_{t=1}^{[ns]} Z_{n,t} w_t u_t \Rightarrow G(s)$  on  $D[0, 1]$ , where  $G(s)$  is a Gaussian process with covariance kernel  $(s_1 \wedge s_2)$  along the lines of the proof of eq. 5.21 in WP2009a: First, observe that since  $\{x_t\}$ , and therefore  $\Lambda_n^2(s)$ , is independent of  $\{w_t, u_t\}$ ,  $V_n(s)$  is a martingale conditional on  $\mathcal{X}_n$ . It then follows from Hall and Heyde (1981, Theorem 3.9) that  $\sup_v |P(V_n(s) \leq v | \mathcal{X}_n) - \Phi(v)| \leq A(q_u) \mathcal{L}_n^{1/(1+q_u)}$  a.s., for any  $s \in [0, 1]$ , where  $A(q_u)$  is a constant depending only on  $q_u$  and

$$\begin{aligned}\mathcal{L}_n &= \frac{\sup_{t \geq 1} \mathbb{E}[|u_t w_t|^{q_u}]}{\Lambda_n^{q_u}} n^{-(1/4-d/2)q_u} \sum_{t=1}^n |f(x_{t-1})|^{q_u} \\ &\quad + \frac{\sigma_u^{q_u}}{\Lambda_n^{q_u}} n^{-(1/2-d)q_u} \mathbb{E} \left[ \left| \sum_{t=1}^n f^2(x_{t-1}) \{w_t^2 - \mathbb{E}[w_t^2]\} \right|^{q_u/2} \middle| \mathcal{X}_n \right].\end{aligned}$$

By part (i),  $n^{-(1/4-d/2)q_u} \sum_{t=1}^n |f(x_{t-1})|^{q_u} = o_p(1)$  and so the first term is  $o_p(1)$ . As before, assuming without loss of generality  $q_u \leq 4$ ,

$$\mathbb{E} \left[ \left| \sum_{k=1}^n f^2(x_{t-1}) \{w_t^2 - \mathbb{E}[w_t^2]\} \right|^2 \middle| \mathcal{X}_n \right] \leq C \sum_{t=1}^n f^2(x_{t-1}) \times \sum_{u=1}^{\infty} |\text{Cov}(w_t^2, w_u^2)|,$$

and so the second term of  $\mathcal{L}_n$  is also  $o_p(1)$ . We conclude that

$$\sup_v |P(V_n(s) \leq v) - \Phi(v)| \leq \mathbb{E} \left[ \sup_v |P(V_n(s) \leq v | \mathcal{X}_n) - \Phi(v)| \right] \rightarrow 0.$$

Finally, tightness of  $V_n(s)$  follows by the same arguments as in the proof of (i). ■

**Proof of Lemma 8.** Define  $\sigma_{0,t}^2 := \pi x_{t-1}^2 z_t$  with  $z_t$  given in the lemma. We then claim that, for all  $n$  large enough and for all  $i \geq 1$ :

$$w_{n\xi}^{-2} \max_{1 \leq t \leq n} |\sigma_t^2 - \sigma_{0,t}^2| = o_p(1), \quad \beta^{i-1} w_{n\xi}^{-2} \max_{1 \leq t \leq n} |x_{t-1}^2 z_{t-i} - x_{t-i}^2 z_{t-i}| = o_p(1), \quad (23)$$

where  $w_{n\xi}^{-2} := \mathbb{E}[(\sum_{t=1}^n \xi_t)^2] = O(n^{1+2d})$ . If  $w_{n\xi}^{-2} \max_{1 \leq t \leq n} |\omega z_t| = o_p(1)$ , the first of the above two claims follows from Han (2011, Lemma 5). It is shown in the proof of Lemma B in HP2012 that  $\max_{1 \leq t \leq n} |z_t| = O_p(\tau_n n^{2/q}) + o_p(1)$  where  $\tau_n = n^r$  with  $0 < r < 1/4 + d/2 - 1/2p - 1/q$ . Note in particular that  $\tau_n \rightarrow \infty$  and  $\tau_n^2 n^{-1/2-d+1/p+2/q} = n^{2r-1/2-d+1/p+2/q} \rightarrow 0$ . Therefore, due to  $w_{n\xi}^2 = O_p(n^{1+2d})$ ,

$$w_{n\xi}^{-2} \max_{1 \leq t \leq n} |\omega z_t| = |\omega| w_{n\xi}^{-2} \max_{1 \leq t \leq n} |z_t| \leq |\omega| O_p(\tau_n n^{-1-2d+2/q}) + |\omega| o_p(n^{-1-2d}) = o_p(1).$$

The second claim follows from Lemma 6 in Han (2011).

Now, by definition of  $\sigma_t^2(\vartheta)$ ,

$$\sigma_t^2(\vartheta) = \beta^t \bar{\sigma}_0^2 + \omega \sum_{i=1}^t \beta^{i-1} + \alpha \sum_{i=1}^t \beta^{i-1} y_{t-i}^2 + \pi \sum_{i=1}^t \beta^{i-1} x_{t-i}^2.$$

Thus

$$r_t(\vartheta) = \frac{\sigma_t^2(\vartheta)}{\sigma_t^2} = \frac{\beta^t \bar{\sigma}_0^2 + \omega \sum_{i=1}^t \beta^{i-1}}{\sigma_t^2} + \alpha \sum_{i=1}^t \beta^{i-1} \frac{y_{t-i}^2}{\sigma_t^2} + \pi \sum_{i=1}^t \beta^{i-1} \frac{x_{t-i}^2}{\sigma_t^2}.$$

The first term is negligible since  $1/\sigma_{[rn]}^2 \leq 1/(\omega_0 + \pi_0 x_{[rn]}^2) = O_p(w_{n\xi}^{-2})$  uniformly over  $r \in (0, 1)$ .

Next, note that, using eq. (23),  $w_{n\xi}^{-2} \sigma_t^2 = w_{n\xi}^{-2} \sigma_{0,t}^2 + o_p(1)$  uniformly over  $1 \leq t \leq n$ . Thus,

$$\begin{aligned} \frac{y_{t-i}^2}{\sigma_t^2} &= \frac{\sigma_{t-i}^2 \varepsilon_{t-i}^2}{\sigma_t^2} = \frac{w_{n\xi}^{-2} \sigma_{t-i}^2 \varepsilon_{t-i}^2}{w_{n\xi}^{-2} \sigma_t^2} = \frac{w_{n\xi}^{-2} \sigma_{0,t-i}^2 \varepsilon_{t-i}^2}{w_{n\xi}^{-2} \sigma_{0,t}^2} + o_p(1) \\ &= \frac{\pi_0 x_{t-1-i}^2 z_{t-i}}{\pi_0 x_{t-1}^2 z_t} \varepsilon_{t-i}^2 + o_p(1) = \frac{z_{t-i}}{z_t} \varepsilon_{t-i}^2 + o_p(1), \end{aligned}$$

and, similarly,  $x_{t-i}^2/\sigma_t^2 = x_{t-i}^2/\sigma_{0,t}^2 + o_p(1) = 1/(\pi_0 z_t) + o_p(1)$ . In total, with  $r_t^*(\theta)$  defined in the lemma,

$$r_t(\vartheta) = \left\{ \alpha \sum_{i=1}^t \beta^{i-1} z_{t-i} \varepsilon_{t-i}^2 + \frac{\pi}{\pi_0} \sum_{i=1}^t \beta^{i-1} \right\} \times \frac{1}{z_t} + o_p(1) = r_t^*(\theta) + o_p(1).$$

This equation holds uniformly in  $\vartheta$  and  $1 \leq t \leq n$ .

Next, observe that, by definition,  $z_t$  solves  $z_t = \rho_t z_{t-1} + 1$ ,  $\rho_t := \alpha_0 \varepsilon_{t-1}^2 + \beta_0$ . Since  $E[\log \rho_t] < 0$ ,  $z_t$  is stationary and ergodic implying that  $z_t(\theta)$  and thereby  $r_t^*(\theta)$  is also stationary and ergodic.

We recognize  $z_t(\theta)$  as the volatility process of a pure, stationary GARCH model (i.e., no regressors included) with intercept parameterized as  $\pi/\pi_0$ . We can therefore employ results for pure (stationary) GARCH models to bound  $r_t^{-1}(\theta)$ . Using the same arguments as in, for example, Francq and Zakoian (2004, p. 622), we obtain that for any  $k$ , there exists some  $\rho \in (0, 1)$  and any  $s \in (0, 1)$  and  $\delta > 0$ ,  $r_t^*(\theta)^{-k} \leq c \left[ 1 + \sum_{i=0}^{\infty} (1 + \delta)^i \rho^{si} \bar{y}_{t-1-i}^{2ks} \right]$  uniformly over  $\theta \in \Theta$ . The left hand is stationary and ergodic with first moment for  $s$  chosen small enough relative to  $k$ .

Finally,

$$\frac{\sigma_t^2}{\sigma_t^2(\vartheta)} = \frac{\omega_0 + \alpha_0 \varepsilon_{t-1}^2 \sigma_{t-1}^2 + \beta_0 \sigma_{t-1}^2 + \pi_0 x_{t-1}^2}{\omega + \alpha \varepsilon_{t-1}^2 \sigma_{t-1}^2 + \beta \sigma_{t-1}^2(\vartheta) + \pi x_{t-1}^2} \leq \frac{\omega_0}{\underline{\omega}} + \frac{\alpha_0}{\underline{\alpha}} + \frac{\beta_0}{\underline{\alpha \varepsilon^2}} + \frac{\pi_0}{\underline{\pi}},$$

for  $\varepsilon_{t-1}^2 > \underline{\varepsilon}^2$ , where  $\underline{\omega} > 0$ ,  $\underline{\alpha} > 0$ ,  $\underline{\beta} > 0$  and  $\underline{\pi} > 0$  are given in Assumption 3. We then use the same arguments as in Francq and Zakoian (2012, p. 843-844) to show that  $\sigma_t^2/\sigma_t^2(\vartheta) \leq W_t$  with  $\mathbb{E}[W_t^k] < \infty$  for any  $k > 0$ . ■

**Proof of Theorem 9.** We first show that  $\hat{\theta}^* := \arg \max_{\theta \in \Theta} L_n^*(\theta)$  satisfies  $\hat{\theta}^* \xrightarrow{P} \theta_0$ . This is shown by verifying conditions (i)-(v) as stated in the proof of Theorem 3. Condition (i) holds by assumption, while (ii) follows by the continuity of  $\theta \mapsto r_t^*(\theta)$  as given in eq. (18). Condition (iii) follows by the LLN for stationary and ergodic sequences if the limit  $L^*(\vartheta)$  exists; the limit is indeed well-defined since, by Lemma 8,  $\mathbb{E}[r_t^*(\theta)^{-k}] < \infty$  for any  $k > 0$ . To prove condition (iv), we see that, by the same arguments as in the proof of Theorem 5,  $L^*(\theta_0) \geq L^*(\theta)$  with equality if and only if  $r_t^*(\theta) = 1$  a.s. Suppose that indeed  $r_t^*(\theta) = 1$  a.s. for some  $\theta \in \Theta$ . By definition of  $r_t(\theta)$ , this is equivalent to  $z_t(\theta_0) = z_t$  a.s., where  $z_t(\theta)$  is defined in eq. (18). Observe that with  $\tilde{y}_t = z_t \varepsilon_t$ , we have that the two processes satisfy  $z_t = 1 + \alpha_0 \tilde{y}_{t-1}^2 + \beta_0 z_{t-1}$  and  $z_t(\theta) = \pi/\pi_0 + \alpha \tilde{y}_{t-1}^2 + \beta z_{t-1}(\theta)$ . Thus, the processes correspond to the true and model-implied volatility in a pure GARCH model with intercept  $\tilde{\omega} = \pi/\pi_0$ . We can then employ the same arguments as in the proof of Theorem 3 and obtain that  $z_t(\theta_0) = z_t$  a.s.  $\Leftrightarrow \theta = \theta_0$ . Finally, condition (v) follows from

$$|\ell_t^*(\theta)| \leq |\log r_t^*(\theta)| + \varepsilon_t^2 \left\{ \frac{1}{r_t(\theta)} + 1 \right\} \leq \sup_{\theta \in \Theta} r_t^*(\theta)^s + \sup_{\theta \in \Theta} r_t^*(\theta)^{-1} + \varepsilon_t^2 \left\{ \sup_{\theta \in \Theta} r_t^*(\theta)^{-1} + 1 \right\} =: \bar{\ell}_t^*,$$

where  $\mathbb{E}[\bar{\ell}_t^*] < \infty$  by Lemma 8.

Now, return to the original estimator,  $\hat{\vartheta}$ . Write the log-likelihood as  $L_n(\vartheta) = L_n^*(\theta) + R_n(\vartheta)$ , where

$$R_n(\vartheta) = \frac{1}{n} \sum_{t=1}^n \left[ \varepsilon_t^2 \left\{ \frac{1}{r_t^*(\theta)} - \frac{1}{r_t(\vartheta)} \right\} + \log \left( \frac{r_t^*(\theta)}{r_t(\vartheta)} \right) \right].$$

Using the same arguments as in Francq and Zakoian (2012, p. 844) together with Lemma 8, we obtain that  $R_n(\vartheta) = o_p(1)$  uniformly in  $\vartheta$ . Thus, by the same arguments as in the proof of Theorem 3,  $\|\hat{\vartheta} - \hat{\vartheta}^*\| = o_p(1)$  where  $\hat{\vartheta}^* = \arg \max_{\theta \in \Theta} \tilde{L}_n^*(\vartheta)$  and  $\tilde{L}_n^*(\omega, \theta) = L_n^*(\theta)$  for any  $(\omega, \theta) \in \mathcal{W} \times \Theta$ .

Local consistency of  $\hat{\omega}$  and the local rate result for  $\hat{\theta}$  follow as part of the results shown in the proof of Theorem 10. ■

**Proof of Theorem 10.** We first establish some approximations: It follows from eq. (23) that

$$\beta^{i-1} w_{n\xi}^{-2} \sigma_{t-i}^2 = \beta^{i-1} \left( w_{n\xi}^{-2} \pi_0 x_{t-i}^2 \right) z_{t-i} + o_p(1) = \beta^{i-1} \left( w_{n\xi}^{-2} \pi_0 x_{t-1}^2 \right) z_{t-i} + o_p(1)$$

for all  $i \geq 1$  and  $t = 1, \dots, n$ , and note that

$$\max_{1 \leq t \leq n} \left| \frac{1}{\sigma_t^2} - \frac{1}{\sigma_{0,t}^2} \right| \leq \max_{1 \leq t \leq n} \frac{\omega_0}{(\pi_0 x_{t-1}^2)^2} = O_p(w_{n\xi}^{-4}). \quad (24)$$

Thus, by the same arguments as in the proof of Lemma 8,

$$\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\vartheta)}{\partial \omega} = \frac{1}{\sigma_t^2} \sum_{i=1}^t \beta^{i-1} = \frac{1}{\sigma_{0,t}^2} \frac{1}{1-\beta} + O_p(w_{n\xi}^{-4}), \quad (25)$$

$$\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\vartheta)}{\partial \alpha} = \sum_{i=1}^t \beta^{i-1} \frac{y_{t-i}^2}{\sigma_{0,t}^2} + o_p(1) = \frac{\partial r_t^*(\theta)}{\partial \alpha} + o_p(1), \quad (26)$$

$$\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\vartheta_0)}{\partial \beta} = \sum_{i=1}^t \beta^{i-1} \frac{\sigma_{t-i}^2(\vartheta)}{\sigma_{0,t}^2} + o_p(1) = \frac{\partial r_t^*(\theta)}{\partial \beta} + o_p(1), \quad (27)$$

$$\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\vartheta)}{\partial \pi} = \sum_{i=1}^t \beta^{i-1} \frac{x_{t-i}^2}{\sigma_{0,t}^2} + o_p(1) = \frac{\partial r_t^*(\theta)}{\partial \pi} + o_p(1), \quad (28)$$

uniformly in  $t = 1, \dots, n$  and  $\vartheta$ , where  $r_t^*(\theta)$  is defined in eq. (18). In total,

$$\frac{\partial r_t(\vartheta)}{\partial \theta} = \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\vartheta)}{\partial \theta} = \frac{\partial r_t^*(\theta)}{\partial \theta} + o_p(1). \quad (29)$$

It is easily seen that  $\mathbb{E}[\sup_{\theta \in \Theta} \|\partial r_t^*(\theta) / (\partial \theta)\|^{2+\delta}] < \infty$  for some  $\delta > 0$  by the same arguments as in Lemma 8. Similarly, it is easily shown that

$$\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\vartheta)}{\partial \omega \partial \beta} = -\frac{1}{\sigma_{0,t}^2} \frac{1}{(1-\beta)^2} + o_p(1),$$

while  $\partial \sigma_t^2(\vartheta) / (\partial \omega \partial \vartheta_k) = 0$ ,  $k = 1, 2, 3$ , and

$$\frac{\partial^2 r_t(\vartheta)}{\partial \theta \partial \theta'} = \frac{1}{\sigma_t^2} \frac{\partial^2 \sigma_t^2(\vartheta)}{\partial \theta \partial \theta'} = \frac{\partial^2 r_t^*(\theta)}{\partial \theta \partial \theta'} + o_p(1),$$

where  $\mathbb{E}[\sup_{\theta \in \Theta} \|\partial^2 r_t^*(\theta) / (\partial \theta \partial \theta')\|] < \infty$ .

We now verify the conditions in Lemmas 11-12 of Kristensen and Rahbek (2010) which in turn imply local consistency and the claimed asymptotic distribution, respectively. To write our estimation problem in their notation, define  $v_{\omega,n} = n^{1/4-d/2}$  and  $v_{\theta,n} = n^{1/2}$ , so that  $V_n$  defined in eq. (13) can be written as  $V_n = \text{diag}\{v_{\omega,n}, v_{\theta,n} I_3\}$ . Next, we let  $Q_n(\vartheta) = L_n(\vartheta) / v_{\omega,n}^2$  denote the normalized log-likelihood and let  $U_n = V_n / v_{\omega,n} = \text{diag}\{1, n^{1/4+d/2} I_3\}$  be the associated rate matrix.



We then claim that

$$(i) v_{\omega,n} U_n^{-1} \frac{\partial Q_n(\vartheta_0)}{\partial \vartheta} \rightarrow_d MN(0, \Sigma^{\text{nst}}), \quad (ii) -U_n^{-1} \frac{\partial^2 Q_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} U_n^{-1} \rightarrow_p H^{\text{nst}} > 0, \quad (30)$$

and, with  $\mathcal{B}_n(\vartheta_0, \epsilon) = \{\vartheta : \|U_n(\vartheta - \vartheta_0)\| < \epsilon\}$  for some small  $\epsilon > 0$ ,

$$\sup_{\vartheta \in \mathcal{B}_n(\vartheta_0, \epsilon)} \left\| U_n^{-1} \left\{ \frac{\partial^2 Q_n(\vartheta)}{\partial \vartheta \partial \vartheta'} - \frac{\partial^2 Q_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \right\} U_n^{-1} \right\| = O_p(\epsilon). \quad (31)$$

Note that (i) of eq. (30) implies that  $U_n^{-1} \partial Q_n(\vartheta_0) / (\partial \vartheta) = o_p(1)$ .

We first show (ii) of eq. (30): Note that

$$U_n^{-1} \frac{\partial^2 Q_n(\vartheta_0)}{\partial \vartheta \partial \vartheta'} U_n^{-1} = \{v_{\omega,n} U_n\}^{-1} H_n(\vartheta_0) \{v_{\omega,n} U_n\}^{-1} = \begin{bmatrix} n^{d-1/2} H_{n,\omega\omega} & n^{d/2-3/4} H_{n,\omega\theta} \\ n^{d/2-3/4} H_{n,\theta\omega} & n^{-1} H_{n,\theta\theta} \end{bmatrix}.$$

We analyze the four elements of  $H_n(\vartheta_0)$  separately. First, using the above approximations,  $h_{\theta\theta,t}(\vartheta) = h_{\theta\theta,t}^*(\theta) + o_p(1)$  where

$$h_{\theta\theta,t}^*(\theta) := \left\{ \frac{\partial^2 r_t^*(\theta)}{\partial \theta \partial \theta'} - \frac{\partial r_t^*(\theta)}{\partial \theta} \frac{\partial r_t^*(\theta)}{\partial \theta'} \right\} \left\{ \frac{\varepsilon_t^2}{r_t^*(\theta)} - 1 \right\} - \frac{\partial r_t^*(\theta)}{\partial \theta} \frac{\partial r_t^*(\theta)}{\partial \theta'} \frac{\varepsilon_t^2}{r_t^*(\theta)}.$$

The process  $h_{\theta\theta,t}^*(\theta)$  is stationary and ergodic with  $E[\sup_{\theta \in \Theta} \|h_{\theta\theta,t}^*(\theta)\|] < \infty$ . It therefore follows from the uniform Law of Large Numbers that  $\sup_{\vartheta} \|H_{n,\theta\theta}(\vartheta)/n - H_{\theta\theta}^{\text{nst}}(\theta)\| \rightarrow_p 0$  where  $H_{\theta\theta}^{\text{nst}}(\theta) = \mathbb{E}[h_{\theta\theta,t}^*(\theta)]$ . Next, using eq. (25),

$$\begin{aligned} -n^{d-1/2} H_{n,\omega\omega}(\vartheta_0) &= \frac{1}{(1-\beta_0)^2} \times \frac{1}{n^{1/2-d}} \sum_{t=1}^n \frac{2\varepsilon_t^2 - 1}{\sigma_t^4} + o_p(1) \\ &= \frac{1}{(1-\beta_0)^2} \times \frac{1}{n^{1/2-d}} \sum_{t=1}^n \frac{2\varepsilon_t^2 - 1}{(\omega_0 + \pi_0 x_{t-1}^2)^2 z_t^2} + o_p(1) \\ &= \frac{1}{(1-\beta_0)^2} \times \frac{1}{n^{1/2-d}} \sum_{t=1}^n \frac{w_t}{(\omega_0 + \pi_0 x_{t-1}^2)^2} + o_p(1) \end{aligned}$$

where  $w_t := (2\varepsilon_t^2 - 1) / z_t^2$  is stationary and geometrically  $\beta$ -mixing, c.f. Carrasco and Chen (2002). Since  $w_t$  and  $x_t$  are independent and  $f(x) = 1 / (\omega_0 + \pi_0 x^2)^2$  is integrable, we can employ Lemma 6(i) to obtain  $-n^{d-1/2} H_{n,\omega\omega}(\vartheta_0) \rightarrow_d H_{\omega\omega}^{\text{nst}}$ . Similarly,

$$\begin{aligned} -n^{d-1/2} H_{n,\omega\alpha}(\vartheta_0) &= \frac{1}{n^{1/2-d}} \sum_{t=1}^n \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \omega} \frac{\partial \sigma_t^2}{\partial \alpha} \{2\varepsilon_t^2 - 1\} \\ &= \frac{1}{1-\beta} \times \frac{1}{n^{1/2-d}} \sum_{t=1}^n \frac{2\varepsilon_t^2 - 1}{(\omega_0 + \pi_0 x_{t-1}^2)^2} \frac{\partial r_t^*(\theta)}{z_t} \frac{\partial r_t^*(\theta)}{\partial \alpha} + o_p(1) \\ &\rightarrow {}_d K \times L_{W_d}(1, 0) \int_{-\infty}^{\infty} \frac{1}{\omega_0 + \pi_0 s^2} ds \frac{1}{1-\beta_0} \mathbb{E} \left[ \frac{\partial r_t^*(\theta)}{\partial \alpha} z_t^{-1} \right]. \end{aligned}$$

In particular,  $n^{d/2-3/4}H_{n,\omega\alpha}(\vartheta_0) = n^{-1/4-d/2} \times \{n^{d-1/2}H_{n,\omega\alpha}(\vartheta_0)\} = o_p(1)$  since  $-1/2 < d < 1/2$ . The other cross-terms involving  $\omega$  are shown to be  $o_p(1)$  in the same manner.

Next, we show (i) of eq. (30): Observe that  $V_n^{-1}S_n(\vartheta_0) = [n^{d/2-1/4}S_{n,\omega}(\vartheta_0), n^{-1/2}S_{n,\theta}(\vartheta_0)]'$ . It follows from Lemma 6(ii) that  $n^{d/2-1/4}S_{n,\omega}(\vartheta_0) \rightarrow_d MN(0, \Sigma_{\omega\omega}^{\text{nst}})$  while, employing the same arguments as in the proof of Theorem 5 together with the stationary approximation results derived above,  $n^{-1/2}S_{n,\theta}(\vartheta_0) \rightarrow_d N(0, \Sigma_{\theta\theta}^{\text{nst}})$ . The convergence is joint since the martingale difference,  $\varepsilon_t^2 - 1$ , is common to the two components of the score, and it is easily checked, by the same arguments as for the hessian, that  $\Sigma_{\omega\theta}^{\text{nst}} = O_{1 \times 3}$ .

Finally, we verify eq. (31): We have already proved that this holds for  $H_{n,\theta\theta}(\vartheta)$ . What remains is to show that it also holds for the components involving  $\omega$ . We only show the result for  $\partial^2 Q_n(\vartheta) / (\partial\omega^2)$  since the proof for the other partial derivatives follows along the same lines. For  $\vartheta \in \mathcal{B}_n(\vartheta_0, \epsilon)$ ,  $\|\theta - \theta_0\| \leq n^{-1/4-d/2}\epsilon$  and  $\|\omega - \omega_0\| \leq \epsilon$ . Thus, by the mean-value theorem, for some  $\bar{\vartheta}$  on the line segment connecting  $\vartheta$  and  $\vartheta_0$ ,

$$\begin{aligned} \left\| \frac{\partial^2 Q_n(\vartheta)}{\partial\omega^2} - \frac{\partial^2 Q_n(\vartheta_0)}{\partial\omega^2} \right\| &= n^{-1/2+d} \|H_{n,\omega\omega}(\vartheta) - H_{n,\omega\omega}(\vartheta_0)\| \\ &\leq \left\| n^{-1/2+d} \frac{\partial H_{n,\omega\omega}(\bar{\vartheta})}{\partial\theta} \right\| \|\theta - \theta_0\| + \left\| n^{-1/2+d} \frac{\partial H_{n,\omega\omega}(\bar{\vartheta})}{\partial\omega} \right\| \|\omega - \omega_0\| \\ &\leq \left\| n^{-3/4+d/2} \frac{\partial H_{n,\omega\omega}(\bar{\vartheta})}{\partial\theta} \right\| \epsilon + \left\| n^{-1/2+d} \frac{\partial H_{n,\omega\omega}(\bar{\vartheta})}{\partial\omega} \right\| \epsilon. \end{aligned}$$

We then wish to show that  $n^{-3/4+d/2} \partial H_{n,\omega\omega}(\bar{\vartheta}) / (\partial\theta) = O_p(1)$  and  $n^{-1/2+d} \partial H_{n,\omega\omega}(\bar{\vartheta}) / (\partial\omega) = O_p(1)$ . The third-order derivative is  $\partial H_{n,\omega\omega}(\vartheta) / (\partial\vartheta) = \sum_{t=1}^n \partial h_{\omega\omega,t}(\vartheta) / (\partial\vartheta)$  where, using that  $\partial^2 \sigma_t^2(\vartheta) / (\partial\omega^2) = \partial^3 \sigma_t^2(\vartheta) / (\partial\omega^2 \partial\vartheta) = 0$ ,

$$\begin{aligned} \frac{h_{\omega\omega,t}(\vartheta)}{\partial\vartheta_k} &= \frac{2}{\sigma_t^6(\vartheta)} \left( \frac{\partial\sigma_t^2(\vartheta)}{\partial\omega} \right)^2 \frac{\partial\sigma_t^2(\vartheta)}{\partial\vartheta_k} \left\{ \frac{\sigma_t^2 \varepsilon_t^2}{\sigma_t^2(\vartheta)} - 1 \right\} \\ &\quad + 2 \left( \frac{\partial\sigma_t^2(\vartheta)}{\partial\omega} \right)^2 \frac{\sigma_t^2 \varepsilon_t^2}{\sigma_t^8(\vartheta)} \frac{\partial\sigma_t^2(\vartheta)}{\partial\vartheta_k}. \end{aligned}$$

As shown in the proof of Theorem 9,  $\sigma_t^2 / \sigma_t^2(\vartheta) \leq W_t$  with  $\mathbb{E}[W_t^k] < \infty$  for any  $k > 0$ , and so

$$\begin{aligned} \left| \frac{h_{\omega\omega,t}(\vartheta)}{\partial\vartheta_k} \right| &\leq \frac{2}{\sigma_t^6(\vartheta)} \frac{1}{(1-\beta)^2} \left\| \frac{\partial\sigma_t^2(\vartheta)}{\partial\vartheta_k} \right\| \{W_t \varepsilon_t^2 + 1\} \\ &\quad + 2 \frac{1}{(1-\beta)^2} \frac{1}{\sigma_t^6(\vartheta)} \left\| \frac{\partial\sigma_t^2(\vartheta)}{\partial\vartheta_k} \right\| W_t \varepsilon_t^2 \\ &\leq C \frac{1}{\sigma_t^6(\vartheta)} \left\| \frac{\partial\sigma_t^2(\vartheta)}{\partial\vartheta_k} \right\| \{W_t \varepsilon_t^2 + 1\}. \end{aligned}$$

Employing the same arguments as in the analysis of the hessian, we obtain the desired result.  $\blacksquare$

**Proof of Theorem 11.** For both the stationary and non-stationary case, we have already shown as part of the proofs of Theorems 5 and 10 that  $\sup_{\|\vartheta - \vartheta_0\| < \delta} \|V_n^{-1}H_n(\vartheta)V_n^{-1} - H(\vartheta)\| \rightarrow_p 0$ .

In the nonstationary case,  $V_n$  is defined in eq. (13),  $U_n = V_n/v_{\omega,n}$  and  $H(\vartheta) = H^{\text{nst}}(\vartheta)$ ; in the stationary case,  $V_n = \sqrt{n}I_4$ ,  $U_n = I_4$  and  $H(\vartheta) = H^{\text{st}}(\vartheta)$ . We now analyze  $\hat{\Sigma} = \Sigma_n(\hat{\vartheta})$  where  $\Sigma_n(\vartheta) = \sum_{t=1}^n s_t(\vartheta)s_t(\vartheta)'$  and  $s_t(\vartheta) = \partial \ell_t(\vartheta) / (\partial \vartheta)$ . First consider the stationary case: As part of the proof of Theorem 5, it was also shown that  $s_t(\vartheta) = \frac{1}{\sigma_{0,t}^2(\vartheta)} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta} \left\{ \frac{y_t^2}{\sigma_{0,t}^2(\vartheta)} - 1 \right\} + o_p(1)$ . The first term on the right hand side is continuous w.r.t.  $\vartheta$  and, by Lemma 4, is uniformly bounded by a stationary sequence with second moment. It therefore follows by the uniform LLN, that  $\sup_{\|\vartheta - \vartheta_0\| < \delta} \|\Sigma_n(\vartheta)/n - \Sigma^{\text{st}}(\vartheta)\| \rightarrow_p 0$  where  $\vartheta \mapsto \Sigma^{\text{st}}(\vartheta)$  is continuous; in particular,  $\Sigma_n(\hat{\vartheta})/n \rightarrow_p \Sigma^{\text{st}}$ . In conclusion,  $n\hat{\Omega} \rightarrow_p \Omega^{\text{st}}$  and so  $\hat{\Omega}^{-1/2}\{\hat{\vartheta} - \vartheta_0\} = (\hat{\Omega}/n)^{-1/2}\sqrt{n}\{\hat{\vartheta} - \vartheta_0\} \rightarrow_d N(0, I_4)$ .

For the non-stationary case, we proceed as in the analysis of the hessian: First, write

$$\Sigma_n(\vartheta) = \begin{bmatrix} \Sigma_{n,\omega\omega}(\vartheta) & \Sigma_{n,\omega\theta}(\vartheta) \\ \Sigma_{n,\theta\omega}(\vartheta) & \Sigma_{n,\theta\theta}(\vartheta) \end{bmatrix} = \sum_{t=1}^n \begin{bmatrix} s_{t,\omega}^2(\vartheta) & s_{t,\omega}(\vartheta)s_{t,\theta}(\vartheta)' \\ s_{t,\omega}(\vartheta)s_{t,\theta}(\vartheta) & s_{t,\theta}(\vartheta)s_{t,\theta}(\vartheta)' \end{bmatrix},$$

where  $s_{t,\omega}(\vartheta)$  and  $s_{t,\theta}(\vartheta)$  denote the partial derivatives of  $\ell_t(\vartheta)$  w.r.t.  $\omega$  and  $\theta$ , respectively. Observe that  $s_{t,\theta}(\vartheta)$  has a stationary approximation, and so, similar to the stationary case, we can appeal to a uniform LLN for stationary and ergodic sequences to obtain  $\Sigma_{n,\theta\theta}(\hat{\vartheta})/n \rightarrow_p \Sigma^{\text{nst}}$ . Next,

$$n^{d-1/2} \sum_{t=1}^n s_{t,\omega}^2(\vartheta_0) = \frac{1}{(1 - \beta_0)^2} \times n^{d-1/2} \sum_{t=1}^n \frac{1}{(\omega_0 + \pi_0 x_{t-1}^2)^2 z_t^2} \{\varepsilon_t^2 - 1\}^2 + o_p(1) \rightarrow_d \Sigma_{\omega\omega}^{\text{nst}},$$

and, similar to the proof of eq. (31),  $\sup_{\|U_n(\vartheta - \vartheta_0)\| < \delta} \|n^{d-1/2} \sum_{t=1}^n \{s_{t,\omega}^2(\vartheta) - s_{t,\omega}^2(\vartheta_0)\}\| = o_p(1)$ . Similarly, we can show that  $n^{d/2-3/4} \sum_{t=1}^n s_{t,\omega}(\vartheta)s_{t,\theta}(\vartheta)' = o_p(1)$ . In conclusion,  $V_n^{-1}\hat{\Omega}V_n^{-1} \rightarrow_p \Omega^{\text{st}}$  and so  $\hat{\Omega}^{-1/2}\{\hat{\vartheta} - \vartheta_0\} = (V_n^{-1}\hat{\Omega}V_n^{-1})^{-1/2}V_n\{\hat{\vartheta} - \vartheta_0\} \rightarrow_d N(0, I_4)$ . ■

## C Tables and Figures

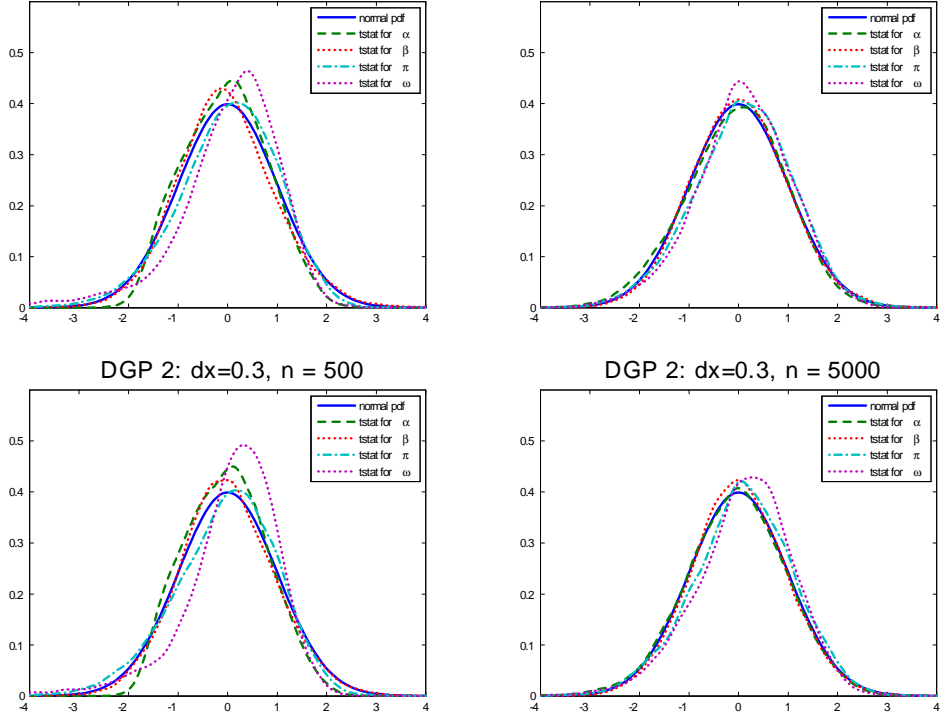
**Table 1.** Estimates of memory parameter  $d_x$  and AR(1) coefficient for various time series

time series	$\hat{d}_x$	AR coefficient	sample period	$T$
3M treasury bill rate level	0.94	1.00	1996/01/02 – 2009/02/27	3434
Bond yield spread (AAA-BAA)	0.88	0.99	1987/11/02 – 2003/06/30	3938
RV of Dow Jones Industrials	0.46	0.66	1996/01/03 – 2009/02/27	3261
RV of CAC 40	0.44	0.66	1996/01/03 – 2009/02/27	3301
RV of FTSE 100	0.42	0.64	1996/01/03 – 2009/02/27	2844
RV of German DAX	0.42	0.66	1996/01/03 – 2009/02/27	3296
RV of British Pound	0.56	0.88	1999/01/04 – 2009/03/01	2576
RV of Euro	0.34	0.67	1999/01/04 – 2009/03/01	2592
RV of Swiss Franc	0.43	0.69	1999/01/04 – 2009/03/01	2571
RV of Japanese Yen	0.47	0.70	1999/01/04 – 2009/03/01	2590

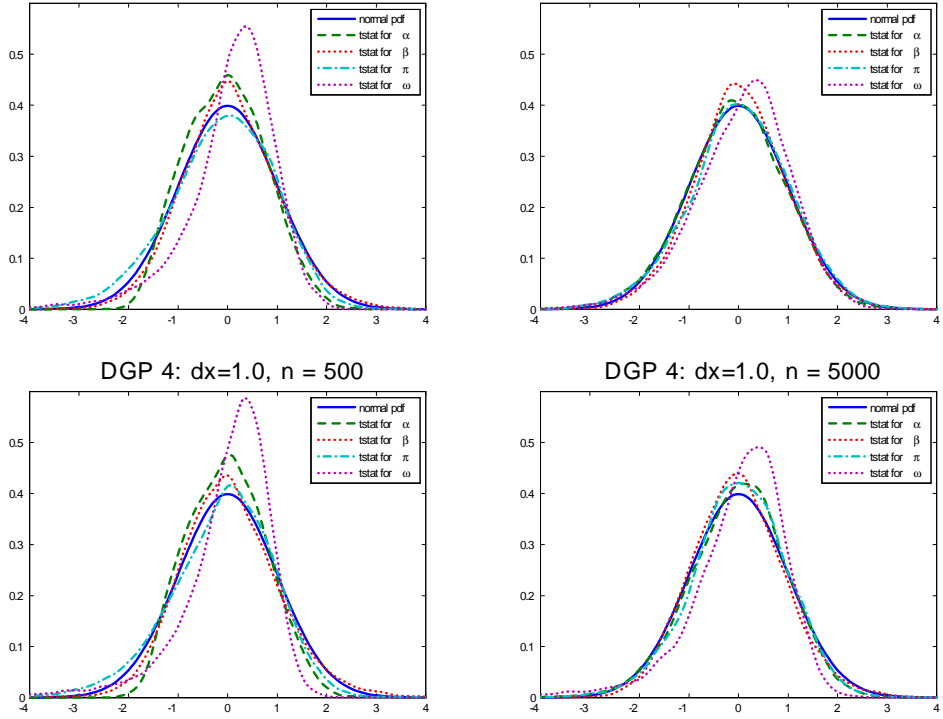
Notes:  $\hat{d}_x$  is the log periodogram estimate of the memory parameter  $d_x$  and  $T$  is the number of observations. RV represents the realized variance of return series. All realized variance series are from ‘Oxford-Man Institute’s realised library’, produced by Heber et al. (2009).<sup>3</sup> All time series are at the daily frequency.

<sup>3</sup>See <http://realized.oxford-man.ox.ac.uk/>.

**Figure 1.** The simulated densities of  $t$ -statistics for the stationary cases  
 DGP 1:  $dx=0.0$ ,  $n = 500$                       DGP 1:  $dx=0.0$ ,  $n = 5000$



**Figure 2.** The simulated densities of  $t$ -statistics for the nonstationary cases  
 DGP 3:  $dx=0.7$ ,  $n = 500$                       DGP 3:  $dx=0.7$ ,  $n = 5000$



## References

- Barndorff-Nielsen, O.E. and N. Shephard (2007) Variation, jumps and high frequency data in financial econometrics. in R. Blundell, T. Persson and W.K. Newey, (Eds.), *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, Econometric Society Monographs, 328-372. Cambridge University Press.
- Berkes, I., L. Horváth and P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9(2), 201–227.
- Billingsley, P. (1974) Conditional distributions and tightness. *Annals of Probability* 2, 480-485.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T. and M. Melvin (1994) Bid-Ask spreads and the volatility in the foreign exchange market: An empirical analysis. *Journal of International Economics* 36, 355-372.
- Brandt, A. (1986) The stochastic equation  $Y_{n+1} = A_n Y_n + B_n$  with stationary coefficients. *Advances in Applied Probability* 18, 211-220.
- Brenner, R.J., R.H. Harjes and K.F. Kroner (1996) Another look at models of the short-term interest rate. *Journal of Financial and Quantitative Analysis* 31, 85-107.
- Brown, B. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59-66.
- Carrasco, M. and X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17-39.
- Dittmann, I. and C.W.J. Granger (2002) Properties of nonlinear transformations of fractionally integrated processes. *Journal of Econometrics* 110, 113-133.
- Dominguez, K. (1998) Central bank intervention and exchange rate volatility. *Journal of International Money and Finance* 17, 161–190.
- Engle, R.F. (2002) New frontiers for ARCH models. *Journal of Applied Econometrics*, 17, 425-446.
- Engle, R.F. and G.M. Gallo (2006) A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131, 3-27.
- Escanciano, J.M. (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561-570.
- Fleming, J., C. Kirby and B. Ostdiek (2008) The specification of GARCH models with stochastic covariates. *Journal of Futures Markets*, 28, 911-934.
- Franq, C. and J.-M. Zakořan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

- Franq, C. and J.-M. Zakoïan (2012) Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroskedasticity models. *Econometrica* 80, 821–861.
- Gallo, G., and B. Pacini (2000). The effects of trading activity on market volatility. *European Journal of Finance* 6, 163–175.
- Girma, P., and M. Mougoue (2002) An empirical examination of the relation between futures spreads volatility, volume, and open interest. *Journal of Futures Markets* 22, 1083–1102.
- Hagiwara, M., and M. Herce (1999) Endogenous exchange rate volatility, trading volume and interest rate differentials in a model of portfolio selection. *Review of International Economics* 7, 202–218.
- Hall, P. and C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.
- Han, H. (2011) Asymptotic properties of GARCH-X processes. Manuscript, National University of Singapore.
- Han, H. and J.Y. Park (2012) ARCH/GARCH with persistent covariates: Asymptotic theory of MLE. *Journal of Econometrics* 167, 95–112.
- Hansen, P.R., Z. Huang and H.H. Shek (2010) Realized GARCH: A complete model of returns and realized measures of volatility. Manuscript, Department of Economics, Stanford University.
- Hodrick, R. (1989). Risk, uncertainty, and exchange rates. *Journal of Monetary Economics*, 23, 433–459.
- Hwang, S. and S.E. Satchell (2005) GARCH Model with cross-sectional volatility: GARCHX models. *Applied Financial Economics* 15, 203–216.
- Heber, G., A. Lunde, N. Shephard and K. Sheppard (2009) *OMI's realised library*. Version 0.1, Oxford-Man Institute, University of Oxford.
- Jensen, S.T. and A. Rahbek (2004) Asymptotic inference for nonstationary GARCH. *Econometric Theory*, 20, 1203–1226.
- Kasparis, I., E. Andreou and P.C.B. Phillips (2012) Nonparametric predictive regression. Manuscript, University of Cyprus.
- Kristensen, D. and A. Rahbek (2005) Asymptotics of the QMLE for a class of ARCH(q) models. *Econometric Theory* 21, 946–961.
- Kristensen, D. and A. Rahbek (2010) Likelihood-Based Inference for Cointegration with Nonlinear Error-Correction. *Journal of Econometrics* 158, 78–94.

- Kristensen, D. and Y. Shin (2012) Estimation of dynamic models with nonparametric simulated maximum likelihood. *Journal of Econometrics* 167, 76–94.
- Lamoureux, C. G. and W. D. Lastrapes (1990) Heteroskedasticity in stock return data: Volume versus GARCH effects. *Journal of Finance* 45, 221–229.
- Marsh, T. A., and N. Wagner (2005) Surprise volume and heteroskedasticity in equity market returns. *Quantitative Finance*, 5, 153–168.
- Park, J.Y. and P.C.B. Phillips (1999) Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, 15, 269-298.
- Park, J.Y. and P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica*, 69, 117-161.
- Shephard, N. and K. Sheppard (2010) Realising the future: forecasting with high frequency based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197-231.
- Wang, Q. (2013) Martingale limit theorem revisited and nonlinear cointegrating regression. Forthcoming in *Econometric Theory*.
- Wang, Q. and P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710-738.
- Wang, Q. and P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901-1948.
- Yoshihara K. (1976) Limiting behaviour of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 35, 237-252.