Asymptotic Theory for the QMLE in GARCH-X Models with Stationary and Non-Stationary Covariates

Heejoon Han and Dennis Kristensen

CREATES Research Paper 2012-25

Asymptotic Theory for the QMLE in GARCH-X Models with Stationary and Non-Stationary Covariates[*]

Heejoon Han[†]   Dennis Kristensen[‡]

May 2012

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 regression coefficients entering the volatility equation are consistent and normally distributed in large samples independently of the degree of persistence. This implies that standard inferential tools, such as t-statistics, do not have to be adjusted to the level of persistence. On the other hand, the intercept in the volatility equation is not identified when the covariate is non-stationary which is akin to the results of Jensen and Rahbek (2004, *Econometric Theory* 20) who develop similar results for the pure GARCH model with explosive volatility.

JEL classification: C22, C50, G12.

Keywords: GARCH; Persistent covariate; Fractional integration; Quasi-maximum likelihood estimator; Asymptotic distribution theory.

[†]Department of Economics and Risk Management Institute, National University of Singapore. E-mail: ecshhj@nus.edu.sg.

[‡]Corresponding author.UCL, CREATES and IFS. E-mail: d.kristensen@ucl.ac.uk.

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 the 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; see, amongst others, Brenner et al. (1996) Fleming et al. (2008) and Han (2011). While the GARCH-X model and its associated quasi-maximum likelihood estimator (QMLE) has found widespread empirical use, the theoretical properties the estimator have not been fully explored. In particular, there is no existing literature that establishes the asymptotic properties of the QMLE for the GARCH-X model. This paper fills this gap.

We provide a unified asymptotic theory for the QMLE for the GARCH-X model allowing for both stationary and nonstationary regressors. 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. In particular, we show that in general all parameters except for the intercept in the GARCH specification can be consistently estimated with the QMLE's following a normal distribution in large samples. In addition, if the regressors are stationary, the intercept estimator also follows a normal distribution. As such, our results imply that standard errors and confidence intervals for most parameters can be computed in a standard fashion whether the regressors are stationary or not. Moreover, our asymptotic results are established under weak moment restrictions on the covariates.

The robustness of the QMLE towards the persistence and moments properties of the chosen covariate is a very attractive feature since it implies that the researcher does not have to conduct any preliminary analysis of a given covariate before estimating the corresponding GARCH model. A simulation study confirms our theoretical findings, with the distribution of the QMLE showing very little sensitivity to the degree of persistence of the included covariate.

In the analysis of the QMLE, we allow for both stationary and non-stationary covariates. In the case of non-stationary regressors, we model the regressor as a (squared) $I(d_x)$ process with $1/2 < d_x < 9/8$. This allows for a wide-range of persistence as captured by the long-memory parameter d_x. In particular, our analysis includes regressors that contains unit roots ($d_x = 1$) but also includes processes with either weaker ($d_x < 1$) or stronger dependence ($d_x > 1$).

This high level of generality in terms of persistence is an important feature of our analysis since economic or financial time series used as covariates in the GARCH-X models display varying degrees of persistent. 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 (Lamoureux and Lastrapes, 1990; Marsh and Wagner, 2005). More recently, various realized volatility measures constructed from high frequency data have been

adopted as covariates in the GARCH-X models with the rapid development seen in the field of realized volatility; see Barndorff-Nielsen and Shephard (2007), Cipollini et al. (2007), Engle (2002), Engle and Gallo (2006), Hansen et al. (2010), Hwang and Satchell (2005), and Shephard and Sheppard (2010).

Table 1 in Appendix C provides log-periodogram estimates of memory parameter d_x and estimates of the first-order autocorrelation for some time series used as covariates in the literature. For example, interest rate levels and bond yield spreads are more persistent: log-periodogram estimates of memory parameter d_x are mostly larger than 0.8 and the autocorrelation estimates are close to unity, which suggest unit root processes. Meanwhile, realized volatility measures (realized variance) of various stock index and exchange rate return series are less persistent: log-periodogram estimates of memory parameter d_x are between 0.3 and 0.6 and the first-order autocorrelations are much smaller than unity ranging from 0.64 to 0.88, which clearly reject unit root hypotheses. The asymptotic theory established in this paper covers all of these choices of regressors.

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 when the regressor is a squared $I(d_x)$ process for $-1/2 < d_x < 1/2$ or $1/2 < d_x < 3/2$. 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) analyzed the QMLE in the pure GARCH model (without any covariates) and showed that the estimated parameters (except for the intercept) remained consistent and asymptotically normally distributed even when in the non-stationary regime. Similarly, we show that the QMLE in the GARCH-X model exhibit the same behaviour whether stationary and non-stationary regressors are included.

Finally, Han and Park (2011) 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 Han and Park (2011) 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 coincide with those of Han and Park (2011) with their transformation chosen as the quadratic function.

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. The results of a simulation study is presented in Section 4. Section 5 concludes the paper. All proofs have been relegated to Appendix A while Appendix B contains tables and figures. Before we proceed, a word on notation: Standard terminologies and notations employed in probability and measure theory are used throughout the paper. In particular, notations for various convergences such as $\rightarrow_{a.s.}$, $\rightarrow_p$ 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

$$y_t = \sigma_t \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ is the error process while $\{\sigma_t^2\}$ is the volatility process. The volatility dynamics are given as

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

where $\{x_t\}$ is an observed covariate which is squared to ensure that $\sigma_t^2 > 0$. The model parameters are $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\pi \geq 0$. We collect the parameters in $\vartheta = (\theta, \omega)$ where $\theta = (\alpha, \beta, \pi)$. The chosen decomposition of the full parameter vector into θ and the intercept ω is due to the special role played by the latter in the non-stationary case.

The regressor x_t will be allowed to be either strictly stationary or a non-stationary long-memory process. In the non-stationary case, we focus on the case where x_t is a long-memory process defined as

$$x_t = x_{t-1} + \xi_t, \tag{3}$$

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

$$(1 - L)^d \xi_t = v_t, \quad -1/2 < d < 1/8. \tag{4}$$

Hence, x_t is an $I(d_x)$ process with $d_x = d + 1 \in (1/2, 9/8)$. Note that $\{\varepsilon_t\}$ and $\{v_t\}$ are allowed to be dependent. Hence, the model can accommodate the 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.

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 $1/2 < d_x < 9/8$, the covariate x_t^2 is nonstationary 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 Han and Park (2011, henceforth HP) given by

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

where x_t is integrated or near-integrated, and $f(x_{t-1}, \pi)^1$ is a positive nonlinear function; more specifically it is an asymptotically homogeneous function as introduced by Park and Phillips (1999). If we let $d_x = 1$ in our model, x_t is integrated and our model belongs to the model considered by HP with $f(x_{t-1}, \pi) = \pi_1 + \pi_2 x_{t-1}^2$. While their model allows more general nonlinear function of x_t, our model allows for more general dependence structure of x_t - either it is stationary or it

---

[1]Note a notational difference in HP. Instead of $f(x_{t-1}, \pi)$, HP use $f(x_t, \pi)$ where (x_t) is adapted to $(\mathcal{F}_{t-1})$.

is fractionally integrated process with $1/2 < d_x < 9/8$. In particular, our model and results are applicable to regressors that exhibit long-memory such as realized volatility measures; this is not covered by HP's model. As shown in Table 1 (Appendix C), unit root hypothesis is clearly rejected for realized volatility measures.

Let (y_t, x_{t-1}) for $t = 1, ..., n$, be $n \geq 1$ observations from the model and let $\vartheta_0 = (\theta_0, \omega_0)$ denote the true data-generating parameter value. Given data, we wish to estimate ϑ_0. We propose to do so through the Gaussian log-likelihood with $\varepsilon_t \sim$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

$$\sigma_t^2(\vartheta) = \omega + \alpha y_{t-1}^2 + \pi x_{t-1}^2 + \beta \sigma_{t-1}^2(\vartheta). \tag{5}$$

The volatility process $\sigma_t^2(\vartheta)$ 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 ε_t to be normally distributed and hence $L_n(\vartheta)$ is a quasi-log likelihood.

Depending on whether we are in the stationary or non-stationary case, the intercept ω is estimable. In the non-stationary case, ω becomes unidentified since, as explained in Han and Park (2011), the constant term is then dominated by the nonstationary covariate or nonstationary volatility and become unimportant asymptotically. A similar finding is reported in Jensen and Rahbek (2004) who analyze the QMLE in the pure GARCH model (without covariates included) when the volatility process is nonstationary. So in the non-stationary case, we can fix ω at an arbitrary value, say, $\bar{\omega} > 0$, and only estimate the remaining parameters, θ. Thus, the two estimators - depending on whether x_t is stationary or non-stationary - become:

$$x_t \text{ stationary:} \qquad \hat{\vartheta} = (\hat{\theta}, \hat{\omega}) = \arg\min_{\vartheta \in \Theta} L_n(\vartheta), \tag{6}$$

$$x_t \text{ non-stationary:} \qquad \hat{\theta}(\bar{\omega}) = \arg\min_{\theta \in \Theta} L_n(\theta, \bar{\omega}). \tag{7}$$

In practice, it should not be necessary to differentiate between the stationary and non-stationary case when computing the estimator, and so we recommend using $\hat{\vartheta}$ in both scenarios. The above definition of $\hat{\theta}(\bar{\omega})$ is only used in the formal analysis of the estimators in the non-stationary case. To formally prove that $\hat{\vartheta}$ is a valid estimator in both scenarios, we would need to show that the score and hessian of the quasi-log likelihood converge uniformly in ω in the non-stationary case in which case the θ-component of $\hat{\vartheta}$, $\hat{\theta}$, is asymptotically first-order equivalent to $\hat{\theta}(\bar{\omega})$. This is technically very demanding though, and so we leave the proof of this for future research.

The main goal of the paper is then to derive the asymptotic distribution of the QMLE in both scenarios.

3 Asymptotic Theory

The basic arguments used to establish the asymptotic distribution of the QMLE are identical for the two cases - stationary or non-stationary regressors. The only difference is that in the non-stationary case, we keep ω fixed at $\bar{\omega}$ since this is not identified while in this situation, we are able to estimate ω together with the other parameters. We briefly outline the proof strategy: We denote the score vector by $S_{n,\gamma}(\gamma) = \partial L_n(\gamma)/(\partial\gamma) = \sum_{t=1}^{n} s_{\gamma,t}(\gamma)$ and the Hessian matrix by $H_{n,\gamma\gamma}(\gamma) = \partial^2 L_n(\gamma)/(\partial\gamma\partial\gamma') = \sum_{t=1}^{n} h_{\gamma\gamma,t}(\gamma)$, respectively, where $\gamma = \vartheta$ or $\gamma = \theta$ depending on whether we are in the stationary or non-stationary case. The asymptotic distribution of the QMLE, $\hat{\gamma}$, is then obtained from the first order Taylor expansion of the score vector, i.e.,

$$S_{n,\gamma}(\gamma) = S_{n,\gamma}(\gamma_0) + H_{n,\gamma\gamma}(\bar{\gamma})(\hat{\gamma} - \gamma_0), \tag{8}$$

where $\bar{\gamma}$ lies on the line segment connecting $\hat{\gamma}$ and γ_0. If $\hat{\gamma}$ is an interior solution, we have $S_n(\hat{\gamma}) = 0$. Therefore, we may write from (8)

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = -[H_{n,\gamma\gamma}(\bar{\gamma})/n]^{-1} \left[S_{n,\gamma}(\gamma_0)/\sqrt{n} \right]. \tag{9}$$

We then proceed to verify the following conditions ML1-ML2; these will together yield the desired result:

ML1 $S_{n,\gamma}(\gamma_0)/\sqrt{n} \to_d N(0, V_{\gamma\gamma})$.

ML2 $-H_{n,\gamma\gamma}(\bar{\gamma})/n \to_p H_{\gamma\gamma}$ for some $H_{\gamma\gamma} > 0$.

Combining eq. (9) with ML1-ML2, it straightforwardly follows that

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = -[H_{n,\gamma\gamma}(\gamma_0)/n]^{-1} \left[S_{n,\gamma}(\gamma_0)/\sqrt{n} \right] + o_p(1) \to_d N\left(0, H_{\gamma\gamma}^{-1} V_{\gamma\gamma} H_{\gamma\gamma}^{-1}\right).$$

Let

$$V_{\vartheta\vartheta}^{\text{stat}} = \begin{bmatrix} V_{\theta\theta}^{\text{stat}} & V_{\theta\omega}^{\text{stat}} \\ V_{\omega\theta}^{\text{stat}} & V_{\omega\omega}^{\text{stat}} \end{bmatrix} \in \mathbb{R}^{4\times4}, \quad H_{\vartheta\vartheta}^{\text{stat}} = \begin{bmatrix} H_{\theta\theta}^{\text{stat}} & H_{\theta\omega}^{\text{stat}} \\ H_{\omega\theta}^{\text{stat}} & H_{\omega\omega}^{\text{stat}} \end{bmatrix} \in \mathbb{R}^{4\times4},$$

denote the matrices derived in the stationary case, while $V_{\theta\theta}^{\text{nonstat}} \in \mathbb{R}^{3\times3}$ and $H_{\theta\theta}^{\text{nonstat}} \in \mathbb{R}^{3\times3}$ denote the matrices obtained in non-stationary one. As we shall see, depending on whether x_t is stationary or not, the limiting covariance terms $(V_{\theta\theta}^{\text{stat}}, H_{\theta\theta}^{\text{stat}})$ and $(V_{\theta\theta}^{\text{nonstat}}, H_{\theta\theta}^{\text{nonstat}})$ will be different. However, at the same time, as part of our proofs, we show that the following two estimators converge towards the relevant versions of $V_{\theta\theta}$ and $H_{\theta\theta}$ irrespectively of whether we are in the stationary or non-stationary regime:

$$\hat{V}_{\theta\theta} = \frac{1}{n} \sum_{t=1}^{n} s_{\theta,t}(\hat{\theta}) s_{\theta,t}(\hat{\theta})', \quad \hat{H}_{\theta\theta} = \frac{1}{n} \sum_{t=1}^{n} h_{\theta\theta,t}(\hat{\theta}). \tag{10}$$

This implies that standard sample versions of t-statistics and (quasi-)likelihood-ratio statistics will follow the same distributions asymptotically whether x_t is stationary or not.

Since the techniques used to establish ML1-ML2 and the resulting limiting covariance matrices differ between the stationary and non-stationary case, so we split the theoretical results into two parts: The following subsection covers the stationary case, while the second one focuses on the non-stationary case.

3.1 Stationary Case

For the stationary case, we provide a full, global analysis of the QMLE, $\hat{\vartheta}$. As a first step, we show that the estimators 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) $\mathbb{E}[\log(\alpha_0\varepsilon_{t-1}^2 + \beta_0)] < 0$ 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 $a, b \in \mathbb{R}$: $a\varepsilon_t^2 + bx_t^2|\mathcal{F}_{t-1}$ 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. It is identical to the 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 ϑ_0 exists and has a finite polynomial moment, c.f. Lemma 1 below. We here allow for integrated GARCH processes ($\alpha + \beta = 1$). We will in the following work under the implicit assumption that we have observed the stationary solution.

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. Note that we cannot remove the restriction on the parameter space of $\beta < 1$ though since this will lead to $\sigma_t^2(\vartheta)$ being non-stationary. The requirement that $(\alpha_0, \pi_0) \neq (0, 0)$ is needed to ensure identification of β_0 since in the case where $(\alpha_0, \pi_0) = (0, 0)$, $\sigma_t^2 = \sigma_t^2(\vartheta_0) \to_{a.s.} \omega_0/(1 - \beta_0)$ and so we would not be able to jointly identify ω_0 and β_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:

7

Lemma 1 *Under Assumption 1: There exists a stationary and ergodic solution to eqs. (1)-(2) at ϑ_0 satisfying $\mathbb{E}\left[\sigma_t^{2s}\right] < \infty$ and $\mathbb{E}\left[y_t^{2s}\right] < \infty$ for some $0 < s < 1$.*

Next, we show that for any value of θ in the parameter space, the recursion defining $\sigma_t^2(\vartheta)$ has a stationary solution:

Lemma 2 *Under Assumption 1: For any $\vartheta \in \Theta$, eq. (5) has a stationary ergodic solution given by*

$$\sigma_{0,t}^2(\vartheta) := \sum_{i=0}^{\infty} \beta^i w_{t-i}(\vartheta),, \quad w_t(\vartheta) := \omega + \alpha y_{t-1}^2 + \pi x_{t-1}^2, \tag{11}$$

satisfying $\mathbb{E}\left[\sup_{\vartheta \in \Theta} \sigma_{0,t}^{2s}(\vartheta)\right] < \infty$ with $s > 0$ given in Lemma 1.

Finally, we show that the initial value chosen for $\sigma_t^2(\vartheta)$ is asymptotically irrelevant such that $\sigma_t^2(\vartheta)$ is asymptotically first-order equivalent to the corresponding stationary solution:

Lemma 3 *Under Assumption 1: With $s > 0$ given in Lemma 1, there exists some $K_s < \infty$ such that*

$$\mathbb{E}\left[\sup_{\vartheta \in \Theta}\left|\sigma_t^2(\vartheta) - \sigma_{0,t}^2(\vartheta)\right|^s\right] \leq K_s \beta^{st}.$$

With these results in hand, we are now ready to show the first main result of this section:

Theorem 4 *Under Assumption 1, the QMLE $\hat{\vartheta}$ in eq. (6) is consistent.*

Having shown that the QMLE is consistent, we verify ML1-ML2 under the following additional assumption:

Assumption 2

(i) $\kappa_4 = \mathbb{E}[(\varepsilon_t^2 - 1)^2] < \infty$.

(ii) ϑ_0 is in the interior of Θ.

Assumption 2(i) is used to show that the variance of the score exists. Assumption 2(ii) is needed in order to employ the Taylor expansion arguments outlined above when deriving the asymptotic distribution.

To verify ML1-ML2 stated above, the following lemma proves useful. It basically shows that the derivatives of the volatility process $\sigma_t^2(\vartheta)$ has properties that are similar to those of $\sigma_{0,t}^2(\vartheta)$: The recursions defining the derivatives have stationary solutions with suitable moments:

Lemma 5 *Under Assumptions 1-2: The models defining $\partial \sigma_t^2(\vartheta)/(\partial \vartheta)$ and $\partial^2 \sigma_t^2(\vartheta)/(\partial \vartheta \partial \vartheta')$ have stationary and ergodic solutions which we denote $\partial \sigma_{0,t}^2(\vartheta)/(\partial \vartheta)$ and $\partial^2 \sigma_{0,t}^2(\vartheta)/(\partial \vartheta \partial \vartheta')$.*

Moreover, there exists stationary and ergodic sequences $B_{k,t} \in \mathcal{F}_{t-1}$, $k = 0,1,2$, which are independent of ϑ such that

$$\frac{\sigma_{0,t}^2(\vartheta_0)}{\sigma_{0,t}^2(\vartheta)} \leq B_{0,t}, \quad \frac{\left\| \partial \sigma_{0,t}^2(\vartheta)/(\partial \vartheta) \right\|}{\sigma_{0,t}^2(\vartheta)} \leq B_{1,t}, \quad \frac{\left\| \partial^2 \sigma_{0,t}^2(\vartheta)/(\partial \vartheta \partial \vartheta') \right\|}{\sigma_{0,t}^2(\vartheta)} \leq B_{2,t}$$

for all ϑ in a neighbourhood of ϑ_0. The sequences satisfy $\mathbb{E}\left[B_{1,t} + B_{2,t}^2 \right] < \infty$ and $\mathbb{E}\left[B_{0,t} \left\{ B_{1,t} + B_{2,t}^2 \right\} \right] < \infty$.

This lemma allows us to bound the score and hessian in a suitable way, thereby establishing ML1-ML2:

Theorem 6 *Under Assumptions 1-2, the QMLE in eq. (6) satisfies*

$$\sqrt{n}(\hat{\vartheta} - \vartheta_0) \to_d N\left(0, H_{\vartheta\vartheta}^{-1} V_{\vartheta\vartheta} H_{\vartheta\vartheta}^{-1} \right),$$

where, with κ_4 given in Assumption 2, $V_{\vartheta\vartheta} = \kappa_4 H_{\vartheta\vartheta}$ and

$$H_{\vartheta\vartheta} = \mathbb{E}\left[\frac{1}{\sigma_{0,t}^4(\vartheta)} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta} \frac{\partial \sigma_{0,t}^2(\vartheta)}{\partial \vartheta'} \right].$$

3.2 Non-stationary Case

For the non-stationary case, we cannot apply the same arguments to establish ML1-ML2 as used in the stationary case. Instead, we adopt the techniques of HP, which is based on asymptotic results for nonlinear, nonstationary time series by Park and Phillips (2001). The basic steps to prove the asymptotic theory in this section are similar to those in HP. However, while the covariate x_t in HP is an $I(1)$ process, the covariate in our model is an $I(d_x)$ process with the wide range of memory parameter d_x $(1/2 < d_x < 9/8)$. Hence, it requires some new technical tools as developed 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) $\{(\varepsilon_t, v_t)\}$ is i.i.d, adapted to $\{\mathcal{F}_t\}$, and satisfies $\mathbb{E}\left[\varepsilon_t \right] = \mathbb{E}\left[v_t \right] = 0$, $\mathbb{E}\left[\varepsilon_t^2 \right] = 1$, and $\mathbb{E}\left[|v_t|^p \right] < \infty$ for some $p \geq 2$.

(ii) $\Theta = \left\{ \vartheta : \underline{\omega} \leq \omega \leq \bar{\omega},\ 0 \leq \alpha \leq \bar{\alpha}, 0 \leq \beta \leq \bar{\beta}, 0 \leq \pi \leq \bar{\pi} \right\}$, where $0 < \underline{\omega} \leq \bar{\omega} < \infty$, $\alpha + \beta < 1$ and $\bar{\pi} < \infty$. $\vartheta_0 \in \Theta$ with $(\alpha_0, \pi_0) \neq (0,0)$.

(iii) $\{x_t\}$ solves eqs. (3)-(4).

(iv) $\mathbb{E}\left[|\varepsilon_t|^q \right] < \infty$ and $\mathbb{E}\left[\left(\beta + \alpha \varepsilon_t^2 \right)^{q/2} \right] < 1$ for some $q > 4$.

(v) $1/p + 2/q < 1/2 + d$.

Assumption 3 precisely defines the covariate $\{x_t\}$ as an $I(d_x)$ process for $1/2 < d_x < 9/8$, and introduces moments conditions for the innovation sequences $\{v_t\}$ and $\{\varepsilon_t\}$. While $\alpha + \beta = 1$ is allowed for the stationary case in the previous section, we do not consider this possibility in the nonstationary case. To see why this restriction is not empirically restrictive, note that our model can be rewritten as

$$y_t^2 = \omega + (\alpha + \beta) y_{t-1}^2 + \pi x_{t-1}^2 + (u_t - \beta u_{t-1}),$$

where $u_t = y_t^2 - \sigma_t^2$ is a martingale difference sequence. When x_t is an I(1) process and $\alpha + \beta = 1$, y_t^2 becomes as persistent as an I(2) process, which is not very likely for most economic and financial time series. Hence, the assumption $\alpha + \beta < 1$ appears very plausible for economic data when x_t is non-stationary.

Assumptions 3(iv) is the same as Assumption 2(b) in HP and Assumption 3(v) corresponds to Assumption 2(c) in HP. See HP for detailed explanations. Note that

$$\mathbb{E}\left(\log\left(\beta + \alpha\varepsilon_t^2\right)\right) \le \log\left[\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)\right] \le \frac{2}{q}\log\left[\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)^{q/2}\right]$$

for any $q > 2$, by the successive applications of Jensen's inequality. As a result, it follows from Assumptions 3(iv) that $\mathbb{E}\left[\log\left(\beta + \alpha\varepsilon_t^2\right)\right] < 0$, which is necessary for the stationary case as in Assumptions 1(ii).

For the proof of the nonstationary case, we first present some additional notation and some useful results. Let $\sigma_{n\xi}^2 = \mathbb{E}[(\sum_{t=1}^n \xi_t)^2]$ where ξ_t is defined in eq. (4). It is known that $\sigma_{n\xi}^2 = O_p\left(n^{1+2d}\right)$ for $-1/2 < d < 1/2$. Let $[z]$ denote the integer part of z. Under suitable conditions, it is known that

$$\sigma_{n\xi}^{-1} \sum_{t=1}^{[nr]} \xi_t \to_d W_d(r), \quad r \in [0,1],$$

where W_d is a fractional Brownian motion, defined for $d \in (-1/2, 1/2)$ by

$$W_d(r) = \frac{1}{\Gamma(d+1)K_d^{1/2}}\left(\int_0^r (r-s)^d\, dV_0(s) + \int_{-\infty}^0 \left[(r-s)^d - (-s)^d\right] dB(s)\right). \quad (12)$$

Here, B is the standard Brownian motion and

$$K_d = \frac{1}{\Gamma(d+1)^2}\left(\frac{1}{2d+1} + \int_0^\infty \left((1+\tau)^d - \tau^d\right)^2 d\tau\right).$$

The scale constant K_d is chosen to make $\mathbb{E}[W_d^2(1)] = 1$. When $x_t = x_{t-1} + \xi_t$, it is also known that

$$\frac{1}{n\sigma_{n\xi}}\sum_{t=1}^n x_t \to_d \int_0^1 W_d(r)\,dr, \quad \frac{1}{n\sigma_{n\xi}^2}\sum_{t=1}^n x_t^2 \to_d \int_0^1 (W_d(r))^2\, dr.$$

For details, see the proof of Lemma 1(b) in Han (2011).

We now proceed to analyze the QMLE, $\hat{\theta}(\bar{\omega})$, as defined in eq. (7). Due to the combination of non-stationary variables and parameters entering the model non-linearly, we only provide a local analysis of the estimator. That is, we only give results for a shrinking neighbourhood around θ_0. For a global analysis, we would first need to show that the QMLE is consistent as done in the stationary case. This is however a quite difficult task and so we, as most other papers on non-linear estimators in a non-stationary environment, focus on the local properties.

We first consider the case where $\bar{\omega} = \omega_0$, and the intial value of the volatility process used for estimation, $\bar{\sigma}^2$, has been chosen to be equal to the data-generating intial value, $\bar{\sigma}^2 = \sigma_0^2(\theta_0, \omega_0)$. We then extend our results to the general case where potentially $\bar{\omega} \neq \omega_0$ and $\bar{\sigma}^2 \neq \sigma_0^2(\theta_0, \omega_0)$. In the following, we suppress functional dependence on $\bar{\omega} = \omega_0$ since it remains fixed. As a first step, we first show that $\sigma_t^2(\theta_0)$ is well-approximated by

$$\sigma_{0,t}^2(\theta_0) := \left(\omega_0 + \pi_0 x_{t-1}^2\right) z_t, \quad z_t := 1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \left(\beta_0 + \alpha_0 \varepsilon_{t-i}^2\right) \tag{13}$$

for $t \geq 1$. This is stated in the following lemma:

Lemma 7 *Under Assumption 3, for all n large enough, for all $i \geq 1$, and for any arbitrary value $\bar{\omega} \geq 0$:*

$$\sigma_{n\xi}^{-2} \max_{1 \leq t \leq n} \left|\sigma_t^2(\theta_0) - \sigma_{0,t}^2(\theta_0)\right| = o_p(1).$$

$$\beta^{i-1} \sigma_{n\xi}^{-2} \max_{1 \leq t \leq n} \left|x_{t-1}^2 z_{t-i} - x_{t-i-1}^2 z_{t-i}\right| = o_p(1).$$

The above lemma is the non-stationary counterpart to the error bounds derived in Lemma 3. This allows us to replace $\sigma_t^2(\theta_0)$ by $\sigma_{0,t}^2(\theta_0)$ in the analysis of the score and hessian. In particular, we show that the score w.r.t. θ is first-order equivalent to $M(\theta_0) \sum_{t=1}^{T} u_t(\theta)$ where $u_t(\theta) = (u_{1t}(\theta), u_{2t}(\theta), u_{3t}(\theta))'$ is defined as

$$u_{1t}(\theta) = \sum_{i=1}^{\infty} \beta^{i-1} \frac{z_{t-i} \varepsilon_{t-i}^2}{z_t} \times \{\varepsilon_t^2 - 1\}, \; u_{2t}(\theta) = \sum_{i=1}^{\infty} \beta^{i-1} \frac{z_{t-i}}{z_t} \times \{\varepsilon_t^2 - 1\} \text{ and } u_{3t}(\theta) = \frac{1}{(1-\beta) z_t} \times \{\varepsilon_t^2 - 1\},$$

and

$$M(\theta) = \text{diag}\{1, 1, 1/\pi\}. \tag{14}$$

Note that $\{u_t(\theta)\}$ is a martingale difference sequence. In particular, as shown in Lemma 1 in HP,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t(\theta_0) \to_d U(r), \tag{15}$$

where U is a vector Brownian motion. Its covariance matrix given by $\mathbb{E}\left[u_t(\theta) u_t(\theta)'\right] = \kappa_4 V(\theta)$

where $\kappa_4 = \mathbb{E}[\,(\varepsilon_t^2 - 1)^2\,]$ and

$$\Omega\left(\theta\right) = \mathbb{E}\left[\frac{1}{z_t^2}\left(\begin{array}{ccc}\left(\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i}\varepsilon_{t-i}^2\right)^2 & \sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i}\varepsilon_{t-i}^2\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i} & \frac{1}{1-\beta}\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i}\varepsilon_{t-i}^2 \\[2ex] \sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i}\varepsilon_{t-i}^2\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i} & \left(\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i}\right)^2 & \frac{1}{1-\beta}\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i} \\[2ex] \frac{1}{1-\beta}\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i}\varepsilon_{t-i}^2 & \frac{1}{1-\beta}\sum\limits_{i=1}^{\infty}\beta^{i-1}z_{t-i} & \frac{1}{(1-\beta)^2}\end{array}\right)\right].$$
(16)

This matrix will make up the covariance matrix of the score and hessian in the non-stationary case. Note that it does not contain x_t because its asymptotic impact on the score and hessian is negligible. With $S_{n,\theta}(\theta)$ and $H_{n,\theta\theta}(\theta)$ denoting the elements of the full score and hessian corresponding to θ, this is stated in the following theorem which also gives the asymptotic distribution of the QMLE:

Theorem 8 *Let Assumption 3 hold, $\bar{\omega} = \omega_0$ and $\bar{\sigma}^2 = \sigma_0^2\left(\theta_0, \omega_0\right)$. Then,*

$$\begin{aligned}n^{-1/2}S_{n,\theta}(\theta_0) & \rightarrow_d N\left(0, V_{\theta\theta}\right), \\ -n^{-1}H_{n,\theta\theta}(\theta_0) & \rightarrow_d H_{\theta\theta},\end{aligned}$$

where $V_{\theta\theta} = \kappa_4 H_{\theta\theta}$ and $H_{\theta\theta} = M\left(\theta_0\right)\Omega\left(\theta_0\right)M\left(\theta_0\right)'$ with $M\left(\theta\right)$ and $\Omega\left(\theta\right)$ given in eqs. (14) and (16) respectively.

Therefore, with probability tending to one, there exists a unique minimum point $\hat{\theta}$ of $L_n(\theta, \bar{\omega})$ in the neighbourhood $\{\theta : ||\theta - \theta_0|| < \epsilon\}$ for some $\epsilon > 0$; it satisfies:

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N\left(0, H_{\theta\theta}^{-1}V_{\theta\theta}H_{\theta\theta}^{-1}\right).$$

The QMLE $\hat{\theta}$ is consistent and converges with standard $\sqrt{n}$-rate of convergence towards a Normal distribution. Theorem 8 implies that the standard inference procedure is valid even in the nonstationary case. For example, the t-statistics for all individual parameters have standard normal limit distribution. This means that regardless that the covariate in the GARCH-X model is stationary or nonstationary, we can simply adopt the standard inference procedure for the QMLE of the GARCH-X models.

As noted earlier, the covariance matrices making up the asymptotic variance of $\hat{\theta}$ does not contain x_t when this is non-stationary. In contrast, in the stationary case, the corresponding covariance matrices do contain information about x_t, c.f. Theorem 6. However, the estimators given in eq. (10) converge towards the correct limits in both cases. For example, the finite-sample t-statistic satisfies

$$t = \{\hat{\kappa}_4 H_{n,\theta\theta}(\theta_0)\}^{-1/2}\{\hat{\theta} - \theta_0\} \rightarrow_d N\left(0, I_3\right),$$

where $\hat{\kappa}_4$ is an estimator of κ_4, whether x_t is stationary or non-stationary. As such, standard inferential procedures regarding θ are robust to the persistence of x_t.

When x_t is an $I(1)$ process, our model belongs to the model considered by HP with their volatility function $f(x_t)$ being linear in parameter. Not surprisingly, the asymptotic distribution is

identical, which is expressed in HP as $\nu'_n(\hat{\theta}_n - \theta_0) \to_d N\left(0, \kappa_4 \Omega^{-1}\right)$ where $\nu_n = \sqrt{n}\text{diag}\left\{(1, 1, 1/\pi_0)\right\}$.

Next, we analyze the behaviour of the QMLE when $\bar{\omega} \neq \omega_0$ and $\bar{\sigma}^2 \neq \sigma_0^2(\theta_0, \omega_0)$:

Theorem 9 *Let Assumption 3 hold with $d > -1/4$. Then the conclusions of Theorem 8 hold for any $\bar{\omega} \neq \omega_0$ and $\bar{\sigma}^2 \neq \sigma_0^2(\theta_0, \omega_0)$.*

Unfortunately, we have only been able to extend the result in Theorem 9 to the case with $d > -1/4$. We conjecture that this is caused by some of the inequalities employed in the proof of Theorem 9 not being sharp enough, and that the conclusions also hold for $d \leq -1/4$.

4   Simulation Study

To investigate the relevance and usefulness of our asymptotic theories, we conduct a simulation study. In particular, we wish to investigate the sensitivity of the QMLE towards the level of persistence, d_x. Furthermore, we would like to validate our conjecture that in practice the non-identification of ω is irrelevant for the performance of the estimators of the remaining parameters.

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 $\omega_0 = 0.01$, $\alpha_0 = 0.1$, $\beta_0 = 0.4$ and $\pi_0 = 0.1$. The innovation processes $\{\varepsilon_t\}$ and $\{v_t\}$ are chosen to be i.i.d. standard normal and mutually independent.[2] 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	$d_x = 0$	DGP 3	$d_x = 0.7$
DGP 2	$d_x = 0.3$	DGP 4	$d_x = 0.1$

In our simulation, the null distributions of the t-statistics for individual parameters α, β and π are simulated for $n = 500$ and $5,000$ with $10,000$ iterations. We do not report the result for ω because it is unidentified in the nonstationary case as explained in the previous section. However, note that we estimate all parameters including ω for all DGP's in our simulations.

The simulation results are reported in Figures 1 and 2 in Appendix C. Figure 1 reports the results for the stationary cases. The Gaussian limit distribution theory is effectively demonstrated in Figures 1. It is also true for the nonstationary cases reported in Figure 2. The empirical distributions of the t-statistics are close to normal 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.

Our simulation results imply that the QMLE's of (α, β, π) 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 in the GARCH-X model. The standard inference procedure will be valid regardless of the dependence structure of x_t.

---

[2] We also tried the case for $v_t = -\varepsilon_t$ and the results are still similar.

5 Conclusion

We have here developed asymptotic theory of QMLE's in GARCH models with additional persistent covariates. It is shown that the asymptotic distributions of the regression coefficients - α, β and π - are robust towards the level of persistence, while the intercept - ω - becomes unidentified when the regressor is non-stationary.

A number of extensions of the theory would be of interest: For example, to analyze the properties of the QMLE in alternative GARCH specifications with persistent regressors; provide a formal proof of our conjecture that in the non-stationary case the choice of the unidentified parameter ω is irrelevant for the estimation of the remaining parameters (this is supported by our simulation results); and provide a global analysis of the QMLE in the non-stationary case.

A   Proofs of Section 3.1

Proof of Lemma 1.   With $a_t := \alpha_0 \varepsilon_{t-1}^2 + \beta_0 \geq 0$, and $b_t := \omega_0 + \pi_0 x_t^2 \geq 0$, rewrite eq. (**??**) as

$$\sigma_t^2 = a_t \sigma_{t-1}^2 + b_t.$$

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 $\mathbb{E}\left[\log^+ (b_t)\right] < \infty$. The stationary solution can be written as

$$\sigma_t^2 = b_t + \sum_{i=0}^{\infty} a_t \cdots a_{t-i} b_{t-i-1}. \tag{17}$$

Following Berkes et al (2003, p. 207-208), the negative Lyapunov coefficient implies that $\mathbb{E}[(a_0 \cdots a_m)^{2s}] < 1$ for some $s > 0$ and $m \geq 1$; thus, $\mathbb{E}[(a_t \cdots a_{t-i})^{2s}] \leq c\rho^i$ for some $c < \infty$ and $\rho < 1$. Without loss of generality, we choose $s < q/2$ with q given in Assumption 1(ii) such that $\mathbb{E}\left[b_t^{2s}\right] < \infty$. Thus,

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

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

Proof of Lemma 2.   Eq. (5) can be rewritten as $\sigma_t^2(\vartheta) := \beta \sigma_{t-1}^2(\vartheta) + v_t(\vartheta)$ 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), $\mathbb{E}\left[\sup_{\vartheta \in \Theta} w_t^s(\vartheta)\right] \leq \bar{\omega}^s + \bar{\alpha}^s \mathbb{E}\left[y_{t-1}^{2s}\right] + \bar{\pi}^s \mathbb{E}\left[x_{t-1}^{2s}\right] < \infty$. Thus,

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

∎

Proof of Lemma 3.   Observe that

$$\sigma_t^2(\vartheta) - \sigma_{0,t}^2(\vartheta) = \beta \left\{\sigma_{t-1}^2(\vartheta) - \sigma_{0,t-1}^2(\vartheta)\right\} = ... = \beta^t \left\{\bar{\sigma}_0^2 - \sigma_{0,0}^2(\vartheta)\right\},$$

where $\bar{\sigma}_0^2 > 0$ is the fixed, intial value used to compute the likelihood function. The result now follows with $K_s = E[\sup_{\vartheta \in \mathcal{D}} \left|\bar{\sigma}_0^2 - \sigma_{0,0}^2(\vartheta)\right|^s]$ which is finite according to Lemma 1. ∎

Proof of Theorem 4. Define $\hat{\vartheta}^* = \arg\max_{\vartheta \in \Theta} L_n^*(\vartheta)$ where $L_n^*(\vartheta) = \sum_{t=1}^n \ell_t^*(\vartheta)$, $\ell_t^*(\vartheta) = -\log \sigma_{0,t}^2(\vartheta) - y_t^2/\sigma_{0,t}^2(\vartheta)$, denotes the stationary version of the quasi-log likelihood. We first show consistency of $\hat{\vartheta}^*$ by verifying the conditions in Kristensen and Rahbek (2005, Proposition 2):

(i) The parameter space Θ is a compact Euclidean space with $\vartheta_0 \in \Theta$.

(ii) $\vartheta \mapsto \ell_t^*(\vartheta)$ is continuous almost surely.

(iii) $L_n^*(\vartheta)/n \to_p L(\vartheta) := \mathbb{E}[\ell_t^*(\vartheta)]$ 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 $\vartheta \mapsto \sigma_{0,t}^2(\vartheta)$ as given in eq. (11). Condition (iii) follows by the LLN for stationary and ergodic sequences if the limit $L(\vartheta)$ exists; the limit is indeed well-defined since $\ell_t^*(\vartheta) \leq -\log(\omega)$ such that $\mathbb{E}[\ell_t^*(\vartheta)^+] < \infty$. To prove condition (iv), first observe that $L(\vartheta_0) \in (-\infty, \infty)$:

$$\ell_t^*(\vartheta_0) = -\log\left(\sigma_{0,t}^2(\vartheta_0)\right) - \frac{\sigma_{0,t}^2(\vartheta_0)\varepsilon_t^2}{\sigma_{0,t}^2(\vartheta_0)} = -\log\left(\sigma_{0,t}^2(\vartheta_0)\right) - \varepsilon_t^2,$$

where $\mathbb{E}[\varepsilon_t^2] = 1$ by Assumption 1(ii). Moreover, $\omega_0 \leq \log\left(\sigma_{0,t}^2(\vartheta_0)\right)$ such that $\mathbb{E}[\left(\log \sigma_{0,t}^2(\vartheta_0)\right)^-] < \infty$, while $\mathbb{E}[\left(\log \sigma_{0,t}^2(\vartheta_0)\right)^+] \leq \left(\log \mathbb{E}[\sigma_{0,t}^{2s}(\vartheta_0)]\right)^+/s < \infty$ 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:

$$L(\vartheta) = -\mathbb{E}\left[\log\left(\sigma_{0,t}^2(\vartheta)\right) + \frac{y_t^2}{\sigma_{0,t}^2(\vartheta)}\right] = -\mathbb{E}\left[\log\left(\sigma_{0,t}^2(\vartheta)\right) + \frac{\sigma_{0,t}^2(\vartheta_0)}{\sigma_{0,t}^2(\vartheta)}\right],$$

where we have used that $\mathbb{E}[\varepsilon_t^2|y_{t-1}, y_{t-2}, ...] = 1$. From the last equality,

$$L(\vartheta_0) - L(\vartheta) = 1 - \mathbb{E}\left[\log\left(\frac{\sigma_{0,t}^2(\vartheta)}{\sigma_{0,t}^2(\vartheta_0)}\right) + \frac{\sigma_{0,t}^2(\vartheta_0)}{\sigma_{0,t}^2(\vartheta)}\right] \geq 0$$

with equality if and only if $\sigma_{0,t}^2(\vartheta) = \sigma_{0,t}^2(\vartheta_0)$ a.s. Suppose that $\sigma_{0,t}^2(\vartheta) = \sigma_{0,t}^2(\vartheta_0)$ a.s, or equivalently,

$$\omega_0 + \sum_{i=1}^\infty c_i(\vartheta_0) X_{t-i} = \omega + \sum_{i=1}^\infty c_i(\vartheta) X_{t-i}, \tag{18}$$

where $c_i(\vartheta) = \left(\alpha\beta^{i-1}, \pi\beta^{i-1}\right)'$ and $X_{t-1} = \left(y_{t-1}^2, x_{t-1}^2\right)'$. 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,

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

16

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 3,

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

where $\lim_{n \to \infty} \sum_{t=1}^{n} \bar{\beta}^t = (1 - \bar{\beta})^{-1} < \infty$ while $\lim_{n \to \infty} \sum_{t=1}^{n} \bar{\beta}^t y_{t-1}^2 < \infty$ 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 5. Observe that

$$
\begin{aligned}
\frac{\partial \sigma_t^2 (\vartheta)}{\partial \omega} &= 1 + \beta \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \omega} = ... = \sum_{i=0}^{t} \beta^i, \\
\frac{\partial \sigma_t^2 (\vartheta)}{\partial \alpha} &= y_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \alpha} = ... = \sum_{i=0}^{t} \beta^i y_{t-i-1}^2, \\
\frac{\partial \sigma_t^2 (\vartheta)}{\partial \pi} &= x_{t-1}^2 + \beta \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \beta} = ... = \sum_{i=0}^{t} \beta^i x_{t-i-1}^2, \\
\frac{\partial \sigma_t^2 (\vartheta)}{\partial \beta} &= \sigma_{t-1}^2 (\vartheta) + \beta \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \beta} = ... = \sum_{i=0}^{t} \beta^i \sigma_{t-i-1}^2 (\vartheta).
\end{aligned}
\tag{19}
$$

By the same arguments as in the proof of Lemma 2, these equations have stationary solutions.

The proof for the second-order partial derivatives w.r.t. ω, α and β proceeds along the lines of Franq and Zakoian (2004, p. 619) since these do not involve x_t. Regarding the second-order derivatives involving π, using the above expressions of the first-order derivatives:

$$
\begin{aligned}
\frac{\partial^2 \sigma_t^2 (\vartheta)}{\partial \omega \partial \pi} &= \beta \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \omega \partial \pi} = ... = 0 \\
\frac{\partial \sigma_t^2 (\vartheta)}{\partial \alpha \partial \pi} &= \beta \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \alpha \partial \pi} = ... = 0, \\
\frac{\partial \sigma_t^2 (\vartheta)}{\partial \pi^2} &= \beta \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \beta \partial \pi} = ... = 0, \\
\frac{\partial \sigma_t^2 (\vartheta)}{\partial \beta \partial \pi} &= \frac{\partial \sigma_{t-1}^2 (\vartheta)}{\partial \pi} + \beta \frac{\partial^2 \sigma_{t-1}^2 (\vartheta)}{\partial \beta \partial \pi} = ... = \sum_{i=0}^{t} \beta^i \frac{\partial \sigma_{t-i-1}^2 (\vartheta)}{\partial \pi}.
\end{aligned}
\tag{20}
$$

17

Again, there clearly exist stationary solutions to these equations.

Moreover, by the same arguments as in Franq and Zakoïan, 2004, p. 622), there exists constants $c < \infty$ and $0 < \rho < 1$ such that for all ϑ in a neighbourhood of ϑ_0 and all $0 < r \leq s$,

$$\frac{\sigma_{0,t}^2 (\vartheta_0)}{\sigma_{0,t}^2 (\vartheta)} \leq c \sum_{i=0}^{\infty} \rho^{ri} \bar{w}_t^r,$$

where $\bar{w}_t := \bar{\omega} + \bar{\alpha} y_{t-1}^2 + \bar{\pi} x_t^2$ is stationary and ergodic with $\mathbb{E}[\bar{w}_t^r] < \infty$. This in turn implies that $\sum_{i=0}^{\infty} \rho^{is} \bar{w}_t^r$ is stationary and ergodic with first moment. Given the representations of the stationary solutions $\sigma_{0,t}^2 (\vartheta_0)$ and $\partial \sigma_t^2 (\vartheta_0) / (\partial \vartheta)$, it is easily shown that for some constant $c < \infty$ the following inequalities hold for all ϑ in a neighbourhood of ϑ_0.(see Franq and Zakoïan, 2004, p. 619)

$$\frac{1}{\sigma_{0,t}^2 (\vartheta_0)} \frac{\partial \sigma_{0,t}^2 (\vartheta_0)}{\partial \omega} \leq \frac{1}{\omega_0}, \quad \frac{1}{\sigma_{0,t}^2 (\vartheta)} \frac{\partial \sigma_{0,t}^2 (\vartheta)}{\partial \alpha} \leq \frac{1}{\alpha_0},$$

$$\frac{1}{\sigma_{0,t}^2 (\vartheta)} \frac{\partial \sigma_{0,t}^2 (\vartheta)}{\partial \pi} \leq \frac{1}{\pi_0}, \quad \frac{1}{\sigma_{0,t}^2 (\vartheta)} \frac{\partial \sigma_{0,t}^2 (\vartheta)}{\partial \beta} \leq c \sum_{i=0}^{\infty} \bar{\beta}^{ri} \bar{w}_t^r,$$

Finally, by the same arguments as in Franq and Zakoian (2004, p. 620), it also holds that

$$\frac{1}{\sigma_{0,t}^2 (\vartheta)} \frac{\partial \sigma_{0,t}^2 (\vartheta)}{\partial \beta \partial \pi} \leq c \sum_{i=0}^{\infty} \bar{\beta}^{ri} \bar{w}_t^r$$

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. $\blacksquare$

Proof of Theorem 6. As shown in the proof of Theorem 4, $||\hat{\vartheta}^* - \hat{\vartheta}|| = O_P(1/n)$; thus, it suffices to analyze $\hat{\vartheta}^*$. By a first-order Taylor expansion of the first-order condition,

$$0 = S_{n,\vartheta}^* (\vartheta_0) + H_{n,\vartheta\vartheta}^* (\bar{\vartheta}) (\hat{\vartheta}^* - \vartheta_0),$$

where $\bar{\vartheta}$ lies on the line segment connecting $\hat{\vartheta}^*$ and ϑ_0, and

$$S_{n,\vartheta}^* (\vartheta) = \frac{\partial L_n^* (\vartheta)}{\partial \vartheta} = \sum_{t=1}^{n} \frac{1}{\sigma_{0,t}^2 (\vartheta)} \frac{\partial \sigma_{0,t}^2 (\vartheta)}{\partial \vartheta} \left\{ \frac{y_t^2}{\sigma_{0,t}^2 (\vartheta)} - 1 \right\},$$

$$H_{n,\vartheta\vartheta}^* (\vartheta) = \frac{\partial^2 L_n^* (\vartheta)}{\partial \vartheta \partial \vartheta'} = \sum_{t=1}^{n} h_{\vartheta\vartheta,t}^* (\vartheta),$$

where derivatives w.r.t. $\sigma_{0,t}^2(\vartheta)$ can be found in the proof of Lemma 5, and

$$
\begin{aligned}
h_{\vartheta\vartheta,t}^*(\vartheta) \; = \; & \left\{ \frac{1}{\sigma_{0,t}^2(\vartheta_0)} \frac{\partial^2 \sigma_{0,t}^2(\vartheta_0)}{\partial\vartheta\partial\vartheta'} - \frac{1}{\sigma_{0,t}^4(\vartheta_0)} \frac{\partial\sigma_{0,t}^2(\vartheta)}{\partial\vartheta} \frac{\partial\sigma_{0,t}^2(\vartheta)}{\partial\vartheta'} \right\} \left\{ \frac{y_t^2}{\sigma_{0,t}^2(\vartheta)} - 1 \right\} \\
& - \frac{\partial\sigma_{0,t}^2(\vartheta)}{\partial\vartheta} \frac{\partial\sigma_{0,t}^2(\vartheta)}{\partial\vartheta'} \frac{y_t^2}{\sigma_{0,t}^6(\vartheta)}.
\end{aligned}
$$

We now verify the three conditions, ML1-ML2, which in turn will yield the desired result.

Regarding ML1: By the CLT for stationary Martingale differences (Brown, 1971),

$$
\frac{1}{\sqrt{n}} S_{n,\vartheta}^*(\vartheta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sigma_{0,t}^{-2}(\vartheta_0) \frac{\partial\sigma_{0,t}^2(\vartheta_0)}{\partial\vartheta} \left\{ \varepsilon_t^2 - 1 \right\} \to_d N(0, V_{\vartheta\vartheta}(\vartheta_0)), \tag{21}
$$

since $\mathbb{E}[\sigma_{0,t}^{-4}(\vartheta_0) \left\| \partial\sigma_{0,t}^2(\vartheta_0)/\partial\vartheta \right\|^2] < \infty$ by Lemma 5.

Regarding ML2: Observe that by Lemma 5,

$$
\|h_{\vartheta\vartheta,t}^*(\vartheta)\| \leq \left\{ B_{2,t} + B_{1,t}^2 \right\} \left\{ 1 + B_{0,t}\varepsilon_t^2 \right\} + B_{1,t}^2 B_{0,t}\varepsilon_t^2.
$$

for all ϑ in some neighbourhood of ϑ_0, where the right-hand side has finite first moment. It now follows by standard uniform convergence results for averages of stationary sequences (see e.g. Kristensen and Rahbek (2005, Proposition 1) that $\sup_{\|\vartheta-\vartheta_0\|<\delta} \left\| H_{n,\vartheta\vartheta}^*(\vartheta)/n - H_{\vartheta\vartheta}(\vartheta) \right\| \to_p 0$, for some $\delta > 0$, where $H_{\vartheta\vartheta}(\vartheta)$ is defined in the theorem. Moreover, $\vartheta \mapsto H_{\vartheta\vartheta}(\vartheta)$ is continuous. Since $\hat{\vartheta}^* \to_p \vartheta_0$, $\bar{\vartheta} \to_p \vartheta_0$ and so lies in any arbitrarily small neighbourhood with probability approaching one. To complete the proof, we verify that $H_{\vartheta\vartheta}(\vartheta_0)$ is non-singular: The process $\Psi_t := \partial\sigma_{0,t}^2(\vartheta_0)/(\partial\vartheta) \in \mathbb{R}^4$ can be written as

$$
\Psi_t = \beta\Psi_{t-1} + W_t, \quad W_t := \left[1, y_{t-1}, x_{t-1}, \sigma_{0,t-1}^2(\vartheta_0) \right]'.
$$

Suppose that there exists $\lambda \in \mathbb{R}^4 \setminus \{0\}$ and $t \geq 1$ such that $\lambda'\Psi_t = 0$ a.s. Since Ψ_t is stationary, this must hold for all t. Given the above equation, this implies that $\lambda'W_t = 0$ 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, $H_{\vartheta\vartheta}(\vartheta_0) = \mathbb{E}\left[\Psi_t\Psi_t'/\sigma_{0,t}^4(\vartheta_0) \right]$ is non-singular. ∎

B   Proofs of Section 3.2

Proof of Lemma 7.   If $\sigma_{n\xi}^{-2} \max_{1\leq t\leq n} |\omega_0 z_t| = o_p(1)$, the stated result follows from Han (2011, Lemma 5). It is shown in the proof of Lemma B in HP that

$$
\max_{1\leq t\leq n} |z_t| = O_p\left(\tau_n n^{2/q} \right) + o_p(1)
$$

where $\tau_n = n^r$ with $0 < r < 1/4 + d/2 - 1/2p - 1/q$. Note in particular that

$$\tau_n \to \infty \quad \text{and} \quad \tau_n^2 n^{-1/2-d+1/p+2/q} = n^{2r-1/2-d+1/p+2/q} \to 0.$$

Therefore, due to $\sigma_{n\xi}^2 = O_p\left(n^{1+2d}\right)$,

$$\sigma_{n\xi}^{-2} \max_{1 \leq t \leq n} |\omega_0 z_t| = |\omega_0| \sigma_{n\xi}^{-2} \max_{1 \leq t \leq n} |z_t| \leq |\omega_0| O_p\left(\tau_n n^{-1-2d+2/q}\right) + |\omega_0| o_p\left(n^{-1-2d}\right) = o_p(1).$$

This completes the proof.

The second part of the stated result follows from Lemma 6 in Han (2011). ■

Proof of Theorem 8. We verify ML1-ML2. The strategy of proof and the arguments employed are similar to those in the proof of Lemma 2 in HP, and so we only provide details where the arguments differ.[3]

To prove ML1, let

$$S_{n,\theta}(\theta) = \frac{\partial L_n(\theta)}{\partial \theta} = \frac{1}{n} \sum_{t=1}^n \sigma_t^{-2}(\theta) \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \left\{ \frac{y_t^2}{\sigma_t^2(\theta)} - 1 \right\}$$

be the score function w.r.t. θ at $\omega = \omega_0$. That is, it contains the first three components of $S_n^*(\vartheta)$ as given in the proof of Theorem 6. We first show that

$$n^{-1/2} S_{n,\theta}(\theta_0) \to_d \left(U_1(1), U_2(1), \frac{1}{\pi_0} U_3(1) \right)', \tag{22}$$

where $U(r) = (U_1(r), U_2(r), U_3(r))$ was defined in eq. (15).

It follows from Lemma 7 that

$$\beta^{i-1} \sigma_{n\xi}^{-2} \sigma_{t-i}^2(\theta_0) = \beta^{i-1} \left(\sigma_{n\xi}^{-2} \pi_0 x_{t-i-1}^2 \right) z_{t-i} + o_p(1) = \beta^{i-1} \left(\sigma_{n\xi}^{-2} \pi_0 x_{t-1}^2 \right) z_{t-i} + o_p(1)$$

for all $i \geq 1$ uniformly in $t = 1, \ldots, n$, and note that

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

These in turn yield

$$\frac{y_{t-i}^2}{\sigma_{0,t}^2} = \frac{\beta^{i-1} \sigma_{n\xi}^{-2} \sigma_{t-i}^2(\theta_0, \omega_0)}{\beta^{i-1} \sigma_{n\xi}^{-2} \sigma_{0,t}^2} \varepsilon_{t-i}^2 = \frac{z_{t-i} \varepsilon_{t-i}^2}{z_t} + o_p(1) \tag{23}$$

$$\frac{x_{t-i-1}^2}{\sigma_{0,t}^2} = \frac{\beta^{i-1} \sigma_{n\xi}^{-2} x_{t-1}^2}{\beta^{i-1} \sigma_{n\xi}^{-2} \left(\pi_0 x_{t-1}^2\right)} \frac{1}{z_t} + o_p(1) = \frac{1}{\pi_0 z_t} + o_p(1) \tag{24}$$

---

[3] Note a notatonal difference in HP. Instead of x_{t-1}, HP use x_t that is adapted to $(\mathcal{F}_{t-1})$.

uniformly in $t = 1, \ldots, n$. Eq. (22) now follows from (15), (23) and (24). For example,

$$
\begin{aligned}
n^{-1/2} S_{n,\pi}(\theta_0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=0}^{\infty} \beta_0^i \frac{\beta^{i-1} \sigma_{n\xi}^{-2} x_{t-1}^2}{\beta^{i-1} \sigma_{n\xi}^{-2} \left(\pi_0 x_{t-1}^2\right)} \frac{1}{z_t} (\varepsilon_t^2 - 1) + o_p(1) \\
&= \sum_{i=0}^{\infty} \beta_0^i \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{\pi_0 z_t} (\varepsilon_t^2 - 1) + o_p(1) \\
&\to_d \sum_{i=0}^{\infty} \beta_0^i \int_0^1 \frac{1}{\pi_0} d\left(1 - \beta_0\right) U_3(r) = \frac{1}{\pi_0} U_3(1).
\end{aligned}
\tag{25}
$$

Consequently,

$$
n^{-1/2} S_n(\theta_0) \to_d MU(1) = N\left(0, \kappa_4 M(\theta_0) \Omega(\theta_0) M(\theta_0)'\right).
$$

Next, to verify ML2, we show the following two claims:

Claim 1: $-H_n(\theta_0)/n \to_p M(\theta_0) \Omega(\theta_0) M(\theta_0)'$.

Claim 2: There exists a sequence $\mu_n > 0$ such that $\mu_n/\sqrt{n} \to 0$, and such that

$$
\sup_{\vartheta \in \mathcal{N}_n} \left\| \mu_n^{-2}[H_n(\theta) - H_n(\theta_0)] \right\| \to_p 0,
$$

where $\mathcal{N}_n = \{\theta \in \Theta : \|\mu_n(\theta - \theta_0)\| \leq 1\}$ is a sequence of shrinking neighborhoods of θ_0.

Proof of Claim 1: The Hessian $H_{n,\theta\theta}(\theta) = \partial^2 L_n(\theta)/\left(\partial\theta\partial\theta'\right)$ is given in the Proof of Theorem 6 with $\partial\sigma_t^2(\theta)/\partial\theta$ and $\partial^2\sigma_t^2(\theta)/\left(\partial\theta\partial\theta'\right)$ given in eqs. (19)-(20) with $\omega = \omega_0$. We write $H_{n,\theta\theta}(\theta) = H_{n,\theta\theta}^a(\theta) + H_{n,\theta\theta}^b(\theta)$, where

$$
H_{n,\theta\theta}^a(\theta) = \sum_{t=1}^{n} \left[\left(1 - \frac{2y_t^2}{\sigma_t^2(\theta)}\right) \frac{1}{\sigma_t^4(\theta)} \frac{\partial\sigma_t^2(\theta)}{\partial\theta} \frac{\partial\sigma_t^2(\theta)}{\partial\theta'}\right] = \left[H_{n,ij}^a(\theta_0)\right]_{i,j=1,2,3}
$$

and

$$
H_{n,\theta\theta}^b(\theta_0) = \sum_{t=1}^{n} \left[\left(\frac{y_t^2}{\sigma_t^2(\theta)} - 1\right) \frac{1}{\sigma_t^2(\theta)} \frac{\partial^2\sigma_t^2(\theta)}{\partial\theta\partial\theta'}\right] = \left[H_{n,ij}^b(\theta_0)\right]_{i,j=1,2,3}.
$$

First, by the same arguments as in the second step of the proof of Lemma 2 in HP, it follows that

$$
-n^{-1} H_{n,\theta\theta}^a(\theta_0) \to_p M(\theta_0) V(\theta_0) M(\theta_0)'.
$$

For example,

$$
\begin{aligned}
-n^{-1} H_{n,\pi\pi}^a(\theta_0) &= \frac{1}{n} \sum_{t=1}^{n} \left(2\varepsilon_t^2 - 1\right) \frac{\sum_{i=0}^{\infty} \beta_0^i \sigma_{n\xi}^{-2} x_{t-i-1}^2}{\sigma_{n\xi}^{-2} \sigma_t^2(\theta_0)} \frac{\sum_{i=0}^{\infty} \beta_0^i \sigma_{n\xi}^{-2} x_{t-i-1}^2}{\sigma_{n\xi}^{-2} \sigma_t^2(\theta_0)} + o_p(1) \\
&= \frac{1}{n} \sum_{t=1}^{n} \frac{1}{z_t^2} \left(\sum_{i=0}^{\infty} \beta_0^i \frac{\beta^{i-1} \sigma_{n\xi}^{-2} x_{t-1}^2}{\beta^{i-1} \sigma_{n\xi}^{-2} \left(\pi_0 x_{t-1}^2\right)}\right)^2 + o_p(1) \\
&\to_d \mathbb{E}\left[\frac{1}{(1 - \beta_0)^2 z_t^2}\right] \frac{1}{\pi_0^2}.
\end{aligned}
$$

21

Next, since $\{\varepsilon_t^2 - 1\}$ is a martingale difference sequence,

$$n^{-1/2} H_{n,\theta\theta}^b(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[(\varepsilon_t^2 - 1) \frac{1}{\sigma_t^2(\theta_0)} \frac{\partial^2 \sigma_t^2(\theta_0)}{\partial\theta_0 \partial\theta_0'} \right] + o_p(1) = O_p(1); \qquad (26)$$

see Han (2011, Remark A1) for details. For example,

$$
\begin{aligned}
n^{-1/2} H_{n,\beta\pi}^b(\theta_0) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\varepsilon_t^2 - 1) \frac{1}{\sigma_t^2(\theta_0)} \sum_{i=1}^{\infty} \beta_0^{i-1} \sum_{k=0}^{\infty} \beta_0^k x_{t-i-k-1}^2 + o_p(1) \\
&= \sum_{i=1}^{\infty} \beta_0^{i-1} \sum_{k=0}^{\infty} \beta_0^k \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\beta_0^{i-1} \sigma_{n\xi}^{-2} x_{t-1}^2}{\beta_0^{i-1} \sigma_{n\xi}^{-2} (\pi_0 x_{t-1}^2)} \frac{1}{z_t} (\varepsilon_t^2 - 1) + o_p(1) \\
&\to_d \sum_{i=1}^{\infty} \beta_0^{i-1} \sum_{k=0}^{\infty} \beta_0^k \int_0^1 \frac{1}{\pi_0} d(1 - \beta_0) U_3(r) = \frac{1}{\pi_0(1 - \beta_0)} U_3(1).
\end{aligned}
$$

Note that we do not need the fourth step in the proof of Lemma 2 in HP due to $\partial^2 \sigma_t^2(\theta)/(\partial\pi)^2 = 0$ and (26). We conclude that

$$-n^{-1} H_{n,\theta\theta}(\theta_0) = -n^{-1} H_{n,\theta\theta}^a(\theta_0) + o_p(1) \to_p M(\theta_0) \Omega(\theta_0) M(\theta_0)',$$

which establishes Claim 1.

Proof of Claim 2: We first note that, for any $\bar{s} > 0$, there exists $\delta > 0$ such that

$$(\sqrt{n})^{-1+\delta} \left\| \sigma_{n\xi}^4 \sup_{|s| \leq \bar{s}} \left(\sup_{\pi \in \mathcal{N}_n} (\pi\sqrt{n}s)^{-4} \right) \right\| \to 0. \qquad (27)$$

Since $\sigma_{n\xi}^2 = O_p(n^{1+2d})$,

$$(\sqrt{n})^{-1+\delta} \sigma_{n\xi}^4 (\sqrt{n})^{-4} = O_p\left(n^{-1/2+\delta/2+4d_x} \right) = o_p(1)$$

if $\delta < 1 - 8d$. Eq. (27) in turn implies that

$$(\sqrt{n})^{-1+\delta} \left\| \sigma_{n\xi}^2 \sup_{|s| \leq \bar{s}} \left(\sup_{\pi \in \mathcal{N}_n} (\pi\sqrt{n}s)^{-2} \right) \right\| \to 0. \qquad (28)$$

Note that (27) and (28) correspond to Assumption 4 in HP.

To establish Claim 2, we now choose $\tilde{\delta}$ such that $0 < \tilde{\delta} < \delta/6$, and $\mu_n = n^{1/2-\tilde{\delta}}$ so that $\mu_n/\sqrt{n} \to 0$ as required. In particular,

$$\left\| n^{1/2-\tilde{\delta}} (\theta - \theta_0) \right\| \leq 1 \qquad (29)$$

for all $\theta \in N_n$. To establish

$$\mu_n^{-2} \{H_{n,\theta\theta}(\theta) - H_{n,\theta\theta}(\theta_0)\} = o_p(1),$$

we only need to show

$$n^{-1+2\tilde{\delta}} \{H_{n,\theta\theta}^a(\theta) - H_{n,\theta\theta}^a(\theta_0)\} \to_p 0 \tag{30}$$

uniformly for all $\theta \in \mathcal{N}_n$, because (26) implies $\left\| \mu_n^{-2} \{H_{n,\theta\theta}^b(\theta) - H_{n,\theta\theta}^b(\theta_0)\} \right\| = o_p(1)$. The verification of (30) is done by combining the arguments of the third step in the proof of Lemma 2 in HP with the limit results in Han (2011). For example, with $g_t(x_t, \sigma_{n\xi}, \theta) = \sum_{i=0}^{\infty} \beta^i \sigma_{n\xi}^{-4} x_{t-i-1}^4$,

$$n^{-1} \sum_{t=1}^{n} g_t(x_t, \sigma_{n\xi}, \theta) \to_d \frac{1}{1-\beta} \int_0^1 (W_d(r))^4 \, dr$$

by Lemmas 1(b) and 4 in Han (2011). This in turn implies that

$$
\begin{aligned}
&\left| n^{-1+2\tilde{\delta}} \{H_{n,\pi\pi}^a(\theta) - H_{n,\pi\pi}^a(\theta_0)\} \right| \\
\leq\ & \left| \frac{1}{n^{1-2\tilde{\delta}}} \sum_{t=1}^{n} \frac{1}{\sigma_t^4(\theta)} (\alpha_0 - \alpha) g(x_t, \sigma_{n\xi}, \theta) \right| \\
<\ & \frac{n^{3\tilde{\delta}}}{\sqrt{n}} \sigma_{n\xi}^4 \sup_{|s| \leq \bar{s}} \sup_{\theta \in N_n} \left| \frac{1}{(\pi(\sqrt{n}s)^2)^2} \right| \frac{1}{n} \sum_{t=1}^{n} |g_t(x_t, \sigma_{n\xi}, \theta)| \\
=\ & o_p(1),
\end{aligned}
\tag{31}
$$

where the third line follows from (29) and $0 < \sigma_t^{-2}(\theta) < 1/(\pi x_{t-1}^2)$ for $\pi > 0$ and $x_{t-1}^2 > 0$, while the fourth one is a consequence of (27). By similar arguments, it is shown that the remaining elements of the matrix in (30) go to zero in probability uniformly in $\theta \in \mathcal{N}_n$. This completes the proof. ∎

Proof of Theorem 9. To analyze the impact of $\bar{\omega} \neq \omega_0$, we proceed along the same lines as in Jensen and Rahbek (2004, Proof of Lemma 13):

$$\frac{y_t^2}{\sigma_t^2(\theta_0, \bar{\omega})} = \varepsilon_t^2 + \frac{(\omega_0 - \bar{\omega})\varepsilon_t^2}{\bar{\omega} + \pi_0 x_{t-1}^2},$$

and so the score satisfies

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sigma_t^{-2}(\theta_0, \bar{\omega}) \frac{\partial \sigma_t^2(\theta_0, \bar{\omega})}{\partial \theta_0} \left(\frac{y_t^2}{\sigma_t^2(\theta_0, \bar{\omega})} - \varepsilon_t^2 \right) = o_p(1), \tag{32}$$

for $d > -1/4$. For example,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{i=0}^{\infty} \beta_0^i \frac{x_{t-i-1}^2}{\sigma_{0,t}^2} \frac{(\omega_0 - \bar{\omega}) \varepsilon_t^2}{\bar{\omega} + \pi_0 x_{t-1}^2} \leq \frac{1}{1 - \beta_0} \sqrt{n} \max_{1 \leq t \leq n} \left| \frac{1}{x_{t-1}^2} \right| \frac{1}{n} \sum_{t=1}^{n} \frac{|\omega_0 - \bar{\omega}|}{\pi_0^2 z_t} \varepsilon_t^2 + o_p(1)$$

$$= \sqrt{n} \sigma_{n\xi}^{-2} O_p(1) = O_p\left(n^{-1/2-2d}\right).$$

Similar to eq. (32), we have that

$$n^{-1} \sum_{t=1}^{n} \left[\left(\frac{y_t^2}{\sigma_t^2(\theta_0, \bar{\omega})} - \varepsilon_t^2 \right) \frac{1}{\sigma_{0,t}^4} \frac{\partial \sigma_t^2(\theta_0, \bar{\omega})}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0, \bar{\omega})}{\partial \theta'} \right] = o_p(1),$$

and

$$n^{-1/2} \sum_{t=1}^{n} \left[\left(\frac{y_t^2}{\sigma_t^2(\theta_0, \bar{\omega})} - \varepsilon_t^2 \right) \frac{1}{\sigma_{0,t}^2} \frac{\partial^2 \sigma_t^2(\theta_0, \bar{\omega})}{\partial \theta \partial \theta'} \right] = o_p(1),$$

and so the impact of $\bar{\omega} \neq \omega_0$ on the hessian is also negiglible.

The proof of the particular choice of the initial value $\bar{\sigma}^2$ being asymptotically negiglible follows along similar lines and so are left out; see also Jensen and Rahbek (2004, Lemma 14). ∎

C   Tables and Figures

Table 1. Estimates of memory parameter d_x and AR 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).[4] All time series are at the daily frequency.

---

[4] See http://realized.oxford-man.ox.ac.uk/.

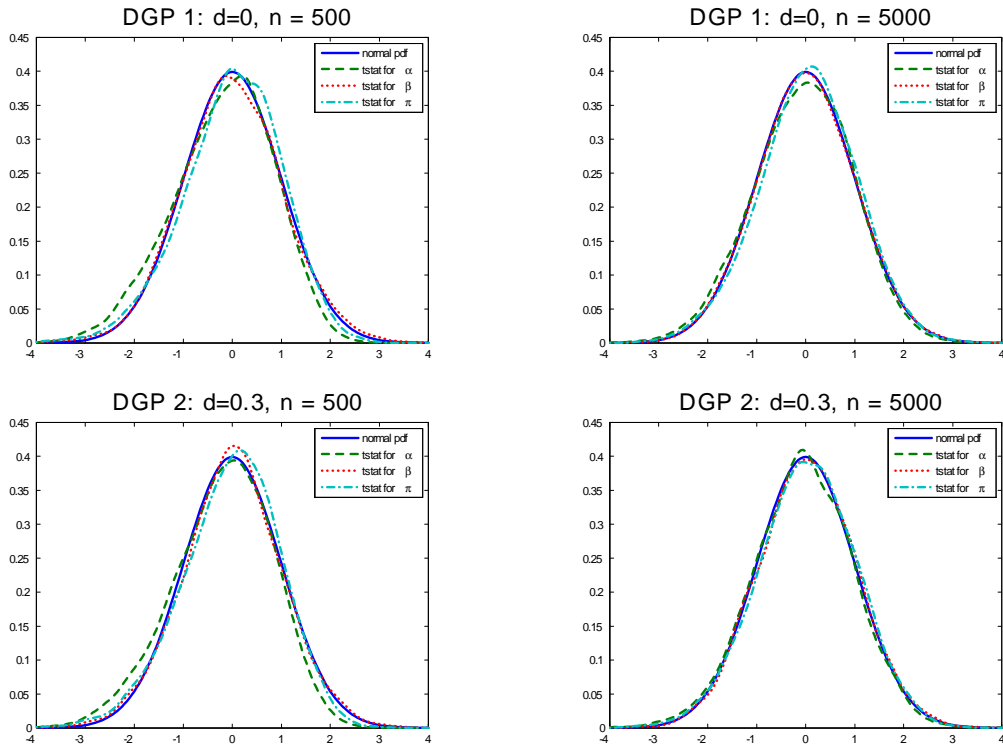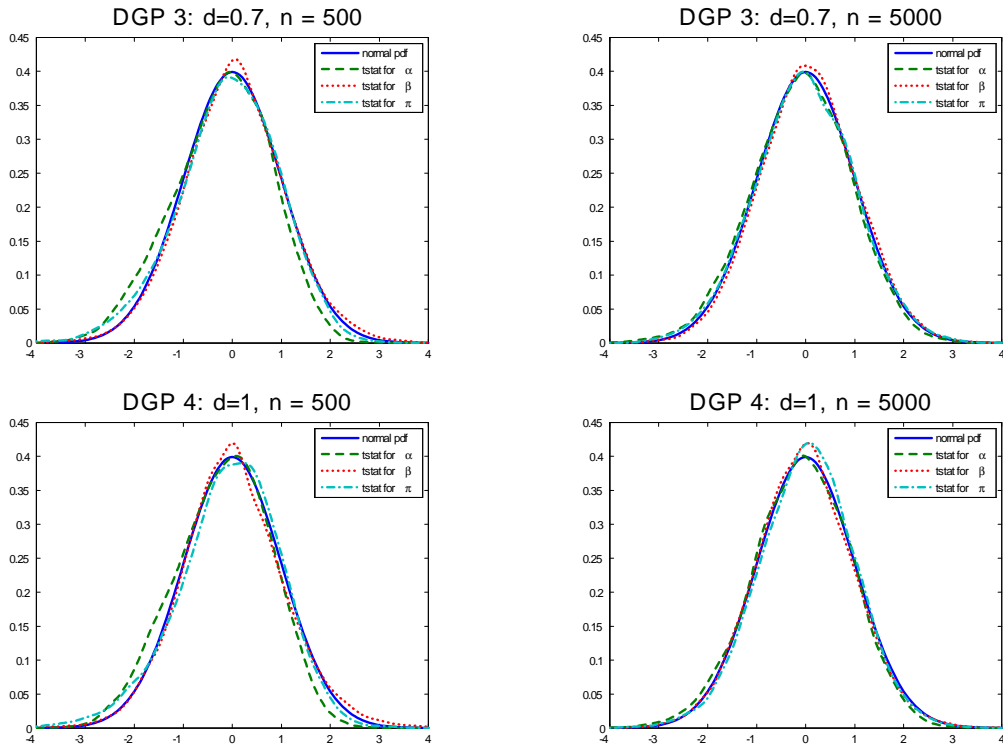Figure 1. The simulated densities of t-statistics for the stationary cases

Figure 2. The simulated densities of t-statistics for the nonstationary cases

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.

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.

Brown, B. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59-66.

Cipollini, F., R.F. Engle and G. Gallo (2007) A model for multivariate non-negative valued processes in financial econometrics. Manuscript, Stern School of Business, New York University.

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. Zakoian (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

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

Han, H. (2011) Asymptotic properties of GARCH-X processes. Manuscript, National University of Singapore.

Han, H. and J.Y. Park (2011) 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.

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 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.

Research Papers
2012

2012-08:   Anne Opschoor, Michel van der Wel, Dick van Dijk and Nick Taylor: On the Effects of Private Information on Volatility

2012-09:   Annastiina Silvennoinen and Timo Teräsvirta: Modelling conditional correlations of asset returns: A smooth transition approach

2012-10:   Peter Exterkate: Model Selection in Kernel Ridge Regression

2012-11:   Torben G. Andersen, Nicola Fusari and Viktor Todorov: Parametric Inference and Dynamic State Recovery from Option Panels

2012-12:   Mark Podolskij and Katrin Wasmuth: Goodness-of-fit testing for fractional diffusions

2012-13:   Almut E. D. Veraart and Luitgard A. M. Veraart: Modelling electricity day-ahead prices by multivariate Lévy

2012-14:   Niels Haldrup, Robinson Kruse, Timo Teräsvirta and Rasmus T. Varneskov: Unit roots, nonlinearities and structural breaks

2012-15:   Matt P. Dziubinski and Stefano Grassi: Heterogeneous Computing in Economics: A Simplified Approach

2012-16:   Anders Bredahl Kock and Laurent A.F. Callot: Oracle Inequalities for High Dimensional Vector Autoregressions

2012-17:   Eric Hillebrand, Huiyu Huang, Tae-Hwy Lee and Canlin Li: Using the Yield Curve in Forecasting Output Growth and Inflation

2012-18:   Eric Hillebrand and Tae-Hwy Lee: Stein-Rule Estimation and Generalized Shrinkage Methods for Forecasting Using Many Predictors

2012-19:   Bent Jesper Christensen, Morten Ørregaard Nielsen and Jie Zhu: The impact of financial crises on the risk-return tradeoff and the leverage effect

2012-20:   Hendrik Kaufmann, Robinson Kruse and Philipp Sibbertsen: On tests for linearity against STAR models with deterministic trends

2012-21:   Andrey Launov, Olaf Posch and Klaus Wälde: On the estimation of the volatility-growth link

2012-22:   Peter O. Christensen and Zhenjiang Qin: Information and Heterogeneous Beliefs: Cost of Capital, Trading Volume, and Investor Welfare

2012-23:   Zhenjiang Qin: Heterogeneous Beliefs, Public Information, and Option Markets

2012-24:   Zhenjiang Qin: Continuous Trading Dynamically Effectively Complete Market with Heterogeneous Beliefs

2012-25:   Heejoon Han and Dennis Kristensen: Asymptotic Theory for the QMLE in GARCH-X Models with Stationary and Non-Stationary Covariates