TESTING GARCH-X TYPE MODELS

RASMUS SØNDERGAARD PEDERSEN AND ANDERS RAHBEK*

Date: August 16, 2017

ABSTRACT

We present novel theory for testing for reduction of GARCH-X type models with an exogenous (X) covariate to standard GARCH type models. To deal with the problems of potential nuisance parameters on the boundary of the parameter space as well as lack of identification under the null, we exploit a noticeable property of specific zero-entries in the inverse information of the GARCH-X type models. Specifically, we consider sequential testing based on two likelihood ratio tests and as demonstrated the structure of the inverse information implies that the proposed test neither depends on whether the nuisance parameters lie on the boundary of the parameter space, nor on lack of identification. Our general results on GARCH-X type models are applied to Gaussian based GARCH-X models, GARCH-X models with Student's t-distributed innovations as well as the integer-valued GARCH-X (PAR-X) models.

KEYWORDS: Testing on the boundary; Likelihood-ratio test; Non-identification; GARCH-X; PAR-X; GARCH models; Integer-valued GARCH; Poisson autoregression.
JEL CLASSIFICATION: C32.

1 Introduction

Conditional volatility models with exogenous explanatory variable(s), or GARCH-X type models, have recently received much attention, see Han and Kristensen (2014) for real-valued variables and Agosto et al. (2016) for integer-valued variables (and references in these). Of particular interest in these models is to formally test if the exogenous variable can be omitted whereby the models can be reduced to pure conditional volatility models. However, the testing problem is highly non-standard, as under the null of no covariate, nuisance parameters appear in the limiting distribution of standard test statistics. In particular, one, or more, nuisance parameters may be on the boundary of the parameter space, and may also be non-identified under the

*Department of Economics, University of Copenhagen, Denmark.

null, which leads to a testing problem in GARCH-X type models not covered by existing literature.

We propose to solve this and thereby deal with the problems of potential nuisance parameters on the boundary of the parameter space as well as lack of identification under the null, by a sequential testing strategy based on two likelihood ratio (LR) tests. We demonstrate that the proposed sequential test neither depends on whether the nuisance parameters lie on the boundary of the parameter space, nor on lack of identification. In order to show this, we derive and exploit a particular property of zero-entries of the *inverse* information in GARCH-X type models.

The first LR test, or rather sup-LR test, addresses the issue of non-identification by testing for no conditional heteroskedasticity. Provided that the null of no conditional heteroskedasticity is rejected, in the second step, the significance of the exogenous covariate is tested. The second test is a LR test where parameters are allowed on the boundary. The null of no exogenous covariate is tested under the assumption of conditional heteroskedasticity (as no conditional heteroskedasticity was rejected in the first step). All parameters are identified, and by exploiting a specific structure of the information matrix for GARCH-X models, we show that the LR statistic is asymptotically pivotal and thus does not depend on whether nuisance parameters are on the boundary or in the interior of the parameter space. Note that if one is willing to assume a priori that the series investigated are conditionally heteroskedastic one can omit the first stage of the sequential test and focus on our new results for the second stage test.

In terms of presenting the results, we first discuss the widely applied Gaussian-based GARCH-X model. Next, we extend the theory to the GARCH-X model with Student's t-distributed innovations, and finally consider the integer-valued (Poisson) GARCH-X model – the PAR-X model – in Agosto et al. (2016).

In terms of existing literature, Han and Kristensen (2014) (see also Han and Park, 2012) consider the asymptotic properties of the (quasi-)maximum likelihood estimator for the GARCH-X model under the assumption that the true parameter value lies in the interior of the parameter spaces, which in particular excludes testing for the presence of exogenous covariates. More recently, Francq and Thieu (2015) consider the asymptotic properties of the (quasi-)maximum likelihood estimator in GARCH-X type models where the true parameter value is a boundary point. However, the assumptions in Francq and Thieu (2015) rule out the possibility of nuisance parameters on the boundary as allowed here. In terms of pure (G)ARCH models (i.e. GARCH models with no exogenous covariates) the general issue of testing with parameters on the boundary of the parameter space has been consid-

Electronic copy available at: https://ssrn.com/abstract=3024648

ered for ARCH(q) models by Silvapulle and Silvapulle (1995) and Demos and Sentana (1998), by Andrews (2001) for the GARCH(1,1) model, and by Francq and Zakoïan (2007,2009) for general GARCH(p, q) models.

The body of literature on constrained M-estimation and testing is vast and dates back to Chernoff (1954). A general theory on estimation and testing on the boundary of the parameter space can be found in Andrews (1999,2001). We refer to Pedersen (2017) for additional references.

The remainder of the paper is organized as follows. In Section 2 we present the GARCH-X model, present the sequential testing scheme, and derive the asymptotic distributions of the LR statistics used in the testing scheme. In Section 3 we discuss the applicability of the testing scheme in the context of the GARCH-X model with Student's t-distributed noise and the integer-valued GARCH-X model. All proofs can be found in the Appendix.

The following notation is applied throughout: For a matrix $x \in \mathbb{R}^{m\times n}$, $\|x\| = \sqrt{\mathrm{tr}(x'x)}$, where $\mathrm{tr}(\cdot)$ denotes the trace, and x' denotes the transpose of x. Unless stated otherwise, all limits are taken as the sample size tends to infinity, that is $T \to \infty$. Lastly, " $\xrightarrow{w}$ " and " $\xrightarrow{p}$ " denote convergence in distribution and probability, respectively.

2 The real-valued GARCH-X Model

As in Han and Kristensen (2014), consider the real-valued GARCH-X model,

$$\begin{aligned} y_t &= \sigma_t z_t, \quad z_t \sim IID(0,1), \\ \sigma_t^2 &= (1-\beta)\omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma x_{t-1}^2, \end{aligned} \tag{2.1}$$

where x_t is an exogenous ergodic covariate. The parameters of the model are given by $\theta = (\gamma, \alpha, \omega)'$ and β, where $\theta \in \Theta$ and $\beta \in \Theta_\beta$ defined by

$$\Theta = \{(\gamma, \alpha, \omega)' \in \mathbb{R}^3 : 0 \le \gamma \le \overline{\gamma}, 0 \le \alpha \le \overline{\alpha}, \underline{\omega} \le \omega \le \overline{\omega}\}, \tag{2.2}$$

for some $0 < \overline{\gamma} < \infty$, $0 < \overline{\alpha} < \infty$, $0 < \underline{\omega} < \overline{\omega} < \infty$, and

$$\Theta_\beta = \{\beta \in \mathbb{R} : 0 \le \beta \le \overline{\beta}\}, \tag{2.3}$$

for some $0 < \overline{\beta} < 1$. We let $\theta_0 \in \Theta$ and $\beta_0 \in \Theta_\beta$ denote the true parameters, and assume throughout that $\underline{\omega} < \omega_0 < \overline{\omega}$, $\alpha_0 < \overline{\alpha}$, and $\beta_0 < \overline{\beta}$ such that $\alpha_0 = \beta_0 = 0$ is allowed.

As mentioned we wish to test whether the covariate x_t is significant for the conditional variance σ_t^2 of y_t. That is, to test the simple hypothesis,

$$H_0 : \gamma = 0,$$

against the alternative where $\gamma > 0$. While empirically of key interest in most applications of models with exogenous covariates such as for the GARCH-X model, testing H_0 is non-standard. Under H_0 we allow for the possibility that the nuisance parameters α (the "ARCH parameter") and β (the "GARCH parameter") lie on the boundary of their respective parameter spaces, that is $\alpha_0 = 0$ and/or $\beta_0 = 0$ is allowed. Additionally, under H_0 and if $\alpha_0 = 0$ then, as well-known, β is non-identified which leads to sup-type tests, see Andrews (2001). Stated differently, the (quasi-)likelihood ratio statistic of H_0 in the GARCH-X model will have different limiting distributions depending on whether the parameters α and β lie on the boundary or not. In particular, the usual likelihood ratio test is asymptotically non-pivotal.

We propose to circumvent the issues by applying a sequential test, while at the same time exploiting a noticeable structure of the inverse information in this testing problem. More precisely, the idea is to replace the likelihood ratio test by a sequential test based on two likelihood ratio (LR) tests: one first tests, by a sup-LR test, the joint hypothesis

$$H_0^* : \gamma = \alpha = 0,$$

and, provided rejection, one next tests by a LR-test the hypothesis $H_0 : \gamma = 0$. Thus $\gamma = 0$ may be rejected provided one rejects initially the joint hypothesis of (conditional) homoskedasticity. This way, we obtain a test which asymptotically does not depend on the α and β parameters. What is crucial here is that the second test is asymptotically pivotal. This is non-trivial as we allow $\beta_0 \geq 0$ and hence different limiting distributions would be expected depending on whether $\beta_0 = 0$ or not. However, as detailed below, a particular zero-entry of the inverse information matrix ensures that indeed the limiting distribution of the second LR statistic is the same whether $\beta_0 = 0$ or not.

We present the two tests in the next two subsections. The first test is the sup-(quasi-)LR test for the hypothesis H_0^* and the test statistic is denoted by LR_T^*. The second test is the (quasi-)LR test of the hypothesis H_0, with the test statistic denoted by LR_T. We emphasize that if one is willing to assume a priori that y_t is not conditionally homoskedastic, i.e. that H_0^* is false, one can skip the first step sup-LR test and move directly to testing H_0.

2.1 Testing H_0^*

With observations $\{(y_t, x_t) : t = 0, ..., T\}$, consider the Gaussian conditional quasi-log-likelihood function given by,

$$\mathcal{L}_T(\theta, \beta) = \sum_{t=1}^{T} l_t(\theta, \beta), \quad l_t(\theta, \beta) = \log\left(\frac{1}{\sqrt{2\pi h_t(\theta, \beta)}} \exp\left\{-\frac{y_t^2}{2h_t(\theta, \beta)}\right\}\right), \tag{2.4}$$

$$h_t(\theta, \beta) = (1-\beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta, \beta) + \gamma x_{t-1}^2, \quad t = 1, ..., T$$

and initial value $h_0(\theta, \beta) = \omega$.

As β is not identified under H_0^*, we consider the sup-LR statistic LR_T^* when testing for H_0^*. See Andrews (2001) for a general theory when testing in the presence of non-identified parameters. Define therefore first the quasi-maximum likelihood estimator (QMLE) for θ for fixed values of β, i.e.

$$\hat{\theta}_\beta := \arg\max_{\theta \in \Theta} \mathcal{L}_T(\theta, \beta), \quad \beta \in \Theta_\beta. \tag{2.5}$$

Likewise, we also define the constrained estimator,

$$\hat{\theta}_\beta^* := \arg\max_{\theta \in \Theta_0^*} \mathcal{L}_T(\theta, \beta), \quad \beta \in \Theta_\beta, \tag{2.6}$$

where $\Theta_0^* = \{\theta \in \Theta : \alpha = \gamma = 0\}$. The standard sup-LR test is given by, $2[\sup_{\beta\in\Theta_\beta} \mathcal{L}_T(\hat{\theta}_\beta, \beta) - \sup_{\beta\in\Theta_\beta} \mathcal{L}_T(\hat{\theta}_\beta^*, \beta)]$. To allow for non-Gaussian innovations z_t, following Andrews (2001, Section 5), we consider the rescaled sup-LR statistic, defined by

$$LR_T^* = \frac{2}{\hat{c}^*}[\sup_{\beta\in\Theta_\beta} \mathcal{L}_T(\hat{\theta}_\beta, \beta) - \sup_{\beta\in\Theta_\beta} \mathcal{L}_T(\hat{\theta}_\beta^*, \beta)], \tag{2.7}$$

where the scaling factor $\hat{c}^*$ is defined by

$$\hat{c}^* = \hat{\kappa}_4^*/2, \quad \hat{\kappa}_4^* = T^{-1}\sum_{t=1}^{T}(y_t^2/\hat{\omega}^* - 1)^2, \tag{2.8}$$

with $\hat{\omega}^* := \hat{\omega}_\beta^*$, the restricted estimator for ω. Under assumptions stated below $\hat{c}^*$ and $\hat{\kappa}_4^*$ have probability limits $c = \kappa_4/2$ and κ_4 respectively, where the kurtosis κ_4 of z_t is given by $\kappa_4 = E[(z_t^2 - 1)^2]$.

To state the limiting distribution of the LR_T^* statistic, we make the following assumptions:

Assumption 2.1 *The process $\{(y_t, x_t) : t \in \mathbb{Z}\}$ is stationary and ergodic.*

Assumption 2.2 *With $\mathcal{F}_t$ the natural filtration generated by $\{(z_s, x_s) : s \leq t\}$, z_t and $\mathcal{F}_{t-1}$ are independent. Moreover, $\kappa_4 := E[(z_t^2 - 1)^2] < \infty$.*

Assumption 2.3 $E[x_t^4] < \infty$.

Assumption 2.4 *For any vector $(a, b)' \in \mathbb{R}^2 \setminus \{0\}$, $az_t^2 + bx_t^2 | \mathcal{F}_{t-1}$ is non-degenerate.*

Remark 2.1 *Assumptions 2.1 and 2.4 are standard regularity conditions. In relation to Assumption 2.1, observe that Han and Kristensen (2014, Lemma 1) state a sufficient condition for the existence of a stationary and ergodic solution to the GARCH-X model which includes the case of $\alpha_0, \beta_0 \geq 0$. In line with Han and Kristensen (2014) and Francq and Thieu (2015), one can relax Assumption 2.2 and the underlying assumption of z_t being IID$(0,1)$. Indeed, one could instead assume that z_t is a martingale difference sequence with respect to $\mathcal{F}_t$ with constant conditional higher-order moments, see Han and Kristensen (2014, Assumptions 1(i) and 2(i)). Relaxing Assumption 2.2 this way implies that one needs to impose finite higher-order moments of z_t and x_t, as discussed in Francq and Thieu (2015). Assumption 2.3 imposes a finite fourth-order moment of x_t, which can be motivated by considering the ratio appearing in the score (and Hessian),*

$$\frac{\partial h_t(\theta, \beta)/\partial \gamma}{h_t(\theta, \beta)} = \frac{x_{t-1}^2 + \beta \partial h_{t-1}(\theta, \beta)/\partial \gamma}{(1-\beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta, \beta) + \gamma x_{t-1}^2}.$$

For $\alpha, \gamma > 0$, that is, with α and γ interior points, the fraction is bounded by a constant, and hence integrable with no further requirements on finite moments (see e.g. the arguments given in Jensen and Rahbek (2004a,2004b) for the non-stationary (G)ARCH model). If, as under H_0^*, $\alpha = \gamma = 0$, the denominator reduces to ω, such that finite second (fourth) order-moments of x_t are needed in order to show that the fraction is (square) integrable. Note also in this respect that Francq and Zakoïan (2009, Assumption A5) assume finite sixth-order moments of y_t when deriving asymptotic properties of the QMLE and related test statistics for the GARCH(p,q) model.

Theorem 2.1 *Consider the GARCH-X model given by (2.1) with log-likelihood function in (2.4). Under Assumptions 2.1-2.4 and H_0^*, with LR_T^* the rescaled* sup-*LR statistic defined in (2.7), it holds that*

$$LR_T^* \xrightarrow{w} \sup_{\beta \in \Theta_\beta} \left\{ \lambda_\beta' (cKJ_\beta^{-1}K')^{-1} \lambda_\beta \right\}. \tag{2.9}$$

Here $c = \kappa_4/2$, $K = [I_2 : 0] \in \mathbb{R}^{2\times 3}$, J_β *is a constant positive definite matrix defined in (A.3), and*

$$\lambda_\beta = \arg\inf_{\eta\in\mathbb{R}^2_+} \left\{(\eta - Z_\beta)'(KJ_\beta^{-1}K')^{-1}(\eta - Z_\beta)\right\}, \quad Z_\beta \sim N(0, cKJ_\beta^{-1}K').$$

The proof of Theorem 2.1 is given in the Appendix. The limiting distribution in (2.9) is non-standard, in particular so as J_β depends on β, and requires simulations as discussed in Andrews (2001). Also note that e.g. Andrews (2001) and Francq and Zakoïan (2009) provide a geometric interpretation of λ_β as the projection of Z_β onto $\mathbb{R}^2_+$.

2.2 Testing H_0

In the following we consider testing H_0 when H_0^* is rejected, that is, we test H_0 under the assumption that $\alpha_0 > 0$ and β_0 may be on the boundary of the parameter space.

Our results rely on a general result for testing on the boundary of the parameter space. The result formulated in Lemma 2.1 below states that the LR test is asymptotically nuisance parameter free even when a nuisance parameter is allowed to be on the boundary of the parameter space. The lemma relies on the specific condition (A.iv) below on the inverse expected information, which can be verified for the GARCH-X model, in addition to standard high-level conditions (A.i)–(A.iii) for testing on the boundary.

We formulate the lemma in terms of a general likelihood function $\mathcal{L}_T(\tau)$ in terms of the parameter τ.

Lemma 2.1 *Consider a likelihood function* $\mathcal{L}_T(\tau)$ *in terms of the parameter* $\tau = (\gamma, \beta, \eta')'$, *where* $\gamma \in \Theta_\gamma = [0, \bar{\gamma}], \beta \in \Theta_\beta = [0, \bar{\beta}]$, $0 < \bar{\gamma}, \bar{\beta} < \infty$, *and* $\eta \in \Theta_\eta$ *with* Θ_η *a compact subset of* $\mathbb{R}^n$ *for some* $n \in \mathbb{N}$. *With* $\hat{\tau} = \arg\max_{\theta\in\Theta_\gamma\times\Theta_\beta\times\Theta_\eta} \mathcal{L}_T(\tau)$ *and* $\tilde{\tau} = \arg\max_{\theta\in\{0\}\times\Theta_\beta\times\Theta_\eta} \mathcal{L}_T(\tau)$, *define the likelihood ratio statistic* LR_T *for the hypothesis that* $\gamma = 0$, *by* $LR_T = 2(\mathcal{L}_T(\hat{\tau}) - \mathcal{L}_T(\tilde{\tau}))$. *With true value* $\tau_0 = (0, \beta_0, \eta_0)$ *where* $\beta_0 \in \Theta_\beta$ *and* $\eta_0 \in int\Theta_\eta$ *make the following assumptions:*

(A.i) $\hat{\tau}, \tilde{\tau} \xrightarrow{p} \theta_0$.

(A.ii) $\frac{1}{\sqrt{T}}\frac{\partial \mathcal{L}_T(\tau_0)}{\partial\tau} \xrightarrow{w} G$, *where* $G \sim N(0, J)$, *and* $-\frac{1}{T}\frac{\partial^2\mathcal{L}_T(\tau_0)}{\partial\tau\partial\tau'} \xrightarrow{p} J$, *with* J *positive definite.*[1]

[1] Throughout, when the partial derivative of a (likelihood) function in the direction γ (or β) is evaluated at a point where $\gamma = 0$ ($\beta = 0$), the derivative is given in terms of the right-derivative.

(A.iii) *$\mathcal{L}_T(\tau)$ is twice continuously differentiable on $\Theta_\gamma \times \Theta_\beta \times \Theta_\eta$, and for all sequences (γ_T), $\gamma_T \to 0$,*

$$\sup_{\tau \in \Theta_\gamma \times \Theta_\beta \times \Theta_\eta : \|\tau - \tau_0\| \leq \gamma_T} \left\| \frac{1}{T} \frac{\partial^2 \mathcal{L}_T(\tau)}{\partial \tau \partial \tau'} - \frac{1}{T} \frac{\partial^2 \mathcal{L}_T(\tau_0)}{\partial \tau \partial \tau'} \right\| \xrightarrow{p} 0.$$

(A.iv) *Either $\beta_0 \in int\Theta_\beta$, or if $\beta_0 = 0$, $(J^{-1})_{\gamma,\beta} = 0$ where $(\cdot)_{\gamma,\beta}$ denotes the corresponding entry of J^{-1}.*

Then $LR_T \xrightarrow{w} [\max(U, 0)]^2$ with U a standard Gaussian distributed random variable.

Lemma 2.1 follows by careful application of Andrews (2001, Theorem 4) under the additional Assumption (A.iv).

Next, we apply Lemma 2.1 to test H_0 in the GARCH-X model. Consider the standard (quasi-)LR statistic for testing H_0 based on estimation of the parameter $\tau := (\gamma, \alpha, \omega, \beta)' \in \Theta_\tau := \Theta \times \Theta_\beta$, with Θ and Θ_β defined in (2.2) and (2.3), respectively. The Gaussian-based conditional quasi-log-likelihood function is

$$\mathcal{L}_T(\tau) := \sum_{t=1}^{T} l_t(\tau), \quad l_t(\tau) = \log\left(\frac{1}{\sqrt{2\pi h_t(\tau)}} \exp\left\{-\frac{y_t^2}{2h_t(\tau)}\right\}\right), \tag{2.10}$$

$$h_t(\tau) = (1-\beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\tau) + \gamma x_{t-1}^2, \quad t = 1, ..., T,$$

with initial value $h_0(\tau) = \omega$. The QMLE for τ is defined as

$$\hat{\tau} := \arg\max_{\tau \in \Theta_\tau} \mathcal{L}_T(\tau),$$

and the constrained estimator under H_0 is

$$\tilde{\tau} := \arg\max_{\tau \in \Theta_{\tau,0}} \mathcal{L}_T(\tau), \quad \Theta_{\tau,0} := \{(\gamma, \alpha, \omega, \beta)' \in \Theta_\tau : \gamma = 0\}.$$

The rescaled (quasi-)LR statistic LR_T is,

$$LR_T := 2[\mathcal{L}_T(\hat{\tau}) - \mathcal{L}_T(\tilde{\tau})]/\tilde{c}, \tag{2.11}$$

with

$$\tilde{c} := \tilde{\kappa}_4/2, \qquad \tilde{\kappa}_4 := T^{-1} \sum_{t=1}^{T} (y_t^2/h_t(\tilde{\tau}) - 1)^2. \tag{2.12}$$

In order to state the limiting distribution of the LR_T statistic in (2.11), we impose an additional Assumption 2.5 about the dependence between the processes (x_t) and (y_t) under the hypothesis H_0. It is needed in order to ensure condition (A.iv) holds and is in particular implied by assuming that x_t and $\mathcal{F}_t^z$ are independent, with $\mathcal{F}_t^z$ the natural filtration generated by $(z_s : s \leq t)$.

Assumption 2.5 *With* $\beta_0 \in \Theta_\beta$*, then for* $\beta_0 = 0$*,* $E\left(x_t^k | y_t, y_{t-1}\right) = E\left(x_t^k\right)$ *for* $k = 2, 4$*.*

Moreover, we impose moment conditions on x_t and y_t:

Assumption 2.6 *There exists* $\delta > 0$ *such that* $E[\| (x_t^2, y_t^2)' \|^{2(1+\delta)}] < \infty$ *and* $E[z_t^{2(1+\delta)/\delta}] < \infty$*.*

Theorem 2.2 *Consider the GARCH-X model given by (2.1) with log-likelihood function in (2.10). Suppose that Assumptions 2.1, 2.2 and Assumptions 2.4–2.6 hold. Then under* H_0*, with* $\alpha_0 > 0$ *and* $\beta_0 \geq 0$*, it holds that with* LR_T *defined in (2.11),*

$$LR_T \xrightarrow{w} [\max(U, 0)]^2, \quad \text{with } U \sim N(0, 1). \tag{2.13}$$

Remark 2.2 *Note that it is crucial for the condition in (A.iv) to hold that the parameter* ω *is estimated. That is, fixing* ω *at* $\omega = \omega_0$*, then (A.iv) does not hold, that is,* $\left(J^{-1}\right)_{\gamma,\beta} \neq 0$*.*

Remark 2.3 *The results on testing* H_0 *are derived under the assumption that the covariate,* x_t*, is strictly stationary and ergodic (Assumption 2.1). Han and Kristensen (2014), see also Han and Park (2012) and Han (2015), consider the properties of the QMLE of the GARCH-X model in the case where* x_t *is non-stationary under the crucial assumption that the parameters* γ*,* α *and* β *are bounded away from zero and thus the theory cannot be applied to test* H_0*. Also, while much emphasis has been given to the condition (A.iv) of Lemma 2.1, we emphasize that our theory requires the score to be asymptotically Gaussian and the information constant. For the GARCH-X – under non-stationarity of* x_t *– as established by Han and Kristensen (2014), the limit of the score is non-Gaussian, and the limiting information random. Consequently, it is a non-trivial task to derive the limiting distribution of the LR-statistic when* x_t *is non-stationary, and we leave this task for future investigation.*

3 Other GARCH-X Type Models

We show by two examples that we can extend the theory for the GARCH-X model to hold for the case where we either replace the likelihood defining distributional assumption, or replace y_t to be integer-valued as opposed to real-valued. The two examples are the GARCH-X model with Student's t-distributed innovations (t-GARCH-X) and the integer valued Poisson autoregressive model with exogenous variables (PAR-X).

When presenting the two models, we focus on the second test of H_0, while the first step in the sequential testing is omitted for brevity. Thus our focus is to show that for the PAR-X and t-GARCH-X models that the condition (A.iv) on the inverse information holds.

3.1 Student's t-GARCH-X

We show that Lemma 2.1 applies to the t-GARCH-X model. In particular, we give details on establishing condition (A.iv) for the inverse information, and also establish in Theorem 3.1 a novel result on consistency of the t-GARCH-X MLE for condition (A.i). Theory for the t-GARCH model with no exogenous covariate is considered in Berkes and Horváth (2004), Straumann (2005, Ch.6), and Pedersen and Rahbek (2016).

The t-GARCH-X model is,

$$y_t = \sigma_t z_t, \quad \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma x_{t-1}^2, \quad \omega > 0, \alpha, \beta, \gamma \geq 0, \tag{3.1}$$

where (z_t) IID scaled t_ν-distributed with degrees of freedom $\nu > 2$. Specifically, with $\tilde{z}_t$ Student's t-distributed with $\nu > 2$ degrees of freedom, $z_t = \sqrt{(\nu-2)/\nu}\tilde{z}_t$, such that $E[z_t] = 0$ and $E[z_t^2] = 1$. With ν an additional parameter when compared to the GARCH-X model, define the model parameter $\theta := (\gamma, \alpha, \omega, \beta, \nu)' \in \Theta$, where

$$\Theta = \left\{(\gamma, \alpha, \omega, \beta, \nu)' \in \mathbb{R}^5 : 0 \leq \gamma \leq \overline{\gamma}, 0 \leq \alpha \leq \overline{\alpha}, \underline{\omega} \leq \omega \leq \overline{\omega}, 0 \leq \beta \leq \overline{\beta}, \right. \\ \left. \text{and } \underline{\nu} \leq \nu \leq \overline{\nu}\right\},$$

for some $0 < \overline{\gamma} < \infty$, $0 < \overline{\alpha} < \infty$, $0 < \underline{\omega} < \overline{\omega} < \infty$, $0 < \overline{\beta} < 1$, and $2 < \underline{\nu} < \overline{\nu} < \infty$. As before $\theta_0 = (\gamma_0, \alpha_0, \omega_0, \beta_0, \nu_0)' \in \Theta$ denotes the true parameter, and we assume throughout that $\underline{\omega} < \omega_0 < \overline{\omega}$, $\alpha_0 < \overline{\alpha}$, $\beta_0 < \overline{\beta}$, and $\underline{\nu} < \nu_0 < \overline{\nu}$. The Student's t-log-likelihood function is,

$$\mathcal{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = -\frac{1}{2}\log[h_t(\theta)] + \log\{g_\nu[y_t/\sqrt{h_t(\theta)}]\}, \tag{3.2}$$
$$h_t(\theta) = (1-\beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta) + \gamma x_{t-1}^2, \quad t = 1, ..., T,$$

with initial value $h_0(\theta) = \omega$ and

$$g_\nu(x) = \frac{\eta(\nu)}{[(\nu-2)\pi]^{1/2}}\left(1 + \frac{x^2}{\nu-2}\right)^{-\left(\frac{\nu+1}{2}\right)}, \text{ with } \eta(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)},$$

with $\Gamma(\cdot)$ denoting the gamma function.

With $\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}_T(\theta)$, and the constrained estimator $\tilde{\theta} = \arg\max_{\theta \in \Theta_0} \mathcal{L}_T(\theta)$, $\Theta_0 = \{(\gamma, \alpha, \omega, \beta, \nu)' \in \Theta : \gamma = 0\}$ we first state result for consistency of the estimators which is needed to verify condition (A.i) in Lemma 2.1.

Theorem 3.1 *Consider the t-GARCH-X model given by (3.1) with log-likelihood function in (3.2). Suppose that Assumptions 2.1, 2.2, and 2.4 hold, with $E[\|(y_t, x_t)'\|^s] < \infty$ for some $s > 0$. If $(\alpha_0, \gamma_0) \neq 0$, then $\hat{\theta} = \theta_0 + o_p(1)$. If $\gamma_0 = 0$ and $\alpha_0 > 0$, then $\tilde{\theta} = \theta_0 + o_p(1)$.*

Theorem 3.1 implies that under $H_0 : \gamma = 0$ and $\alpha_0 > 0$, the MLE's $\hat{\theta}$ and $\tilde{\theta}$ are consistent. Next, we state the equivalent of Theorem 2.2, where we in the proof in the Appendix establish that conditions (A.i)–(A.iv) of Lemma 2.1 hold for the t-GARCH-X model.

Theorem 3.2 *Consider the t-GARCH-X model given by (3.1) with log-likelihood function in (3.2). Assume that Assumptions 2.1, 2.2, 2.4–2.5 hold and $E[\|(y_t, x_t)'\|^4] < \infty$. Then under H_0, with $\alpha_0 > 0$ and $\beta_0 \geq 0$,*

$$LR_T := 2[\mathcal{L}_T(\hat{\theta}) - \mathcal{L}_T(\tilde{\theta})] \xrightarrow{w} [\max(U, 0)]^2, \text{ where } U \sim N(0, 1).$$

Theorem 3.2 illustrates in particular, that even for the t-GARCH-X model with the t-likelihood and an extra parameter (degrees of freedom ν) to be estimated, the condition (A.iv) still applies.

3.2 The integer-valued GARCH-X model: Poisson Autoregression with Exogenous covariate (PAR-X)

Consider next the Poisson integer-valued GARCH-X model, the PAR-X model as considered in Agosto et al. (2016). We show here that a result similar to Theorem 2.2 (and Theorem 3.2), applies to the PAR-X model. Theory for the pure PAR model is given in Fokianos et al. (2009), and Ahmad and Francq (2016).

Let $y_t \in \mathbb{N} \cup \{0\}$, $t = 0, 1, ...$ be a time series of counts, and x_t, $t = 0, 1, ...$, as before an ergodic covariate. With $\mathcal{F}_t$ the natural filtration of $\{(y_i, x_i)' : i \leq t\}$, the PAR-X model in Agosto et al. (2016) is given by

$$y_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t), \quad t = 1, ..., T, \tag{3.3}$$

with time-varying (conditional) intensity, $\lambda_t > 0$,

$$\lambda_t = (1 - \beta)\,\omega + \alpha y_{t-1} + \beta \lambda_{t-1} + \gamma f(x_{t-1}).$$

Here $\omega > 0$, $\alpha, \gamma, \beta \geq 0$, and $f(\cdot)$ is a non-negative link function, $f : \mathbb{R} \to [0, \infty)$, see Agosto el al. (2016) for details.

As for the GARCH-X model, we consider testing the hypothesis $H_0 : \gamma = 0$, and as for the GARCH-X model, under H_0 the test is non-pivotal and furthermore β is not identified if $\alpha = 0$. Thus the problem of testing H_0 is identical to the (t-)GARCH-X models and we show that the same approach is indeed applicable.

With parameter $\theta = (\gamma, \alpha, \omega, \beta)'$, the conditional Poisson-log-likelihood function is given by,

$$\mathcal{L}_T(\theta) = \sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = y_t \log[\lambda_t(\theta)] - \lambda_t(\theta), \tag{3.4}$$

$$\lambda_t(\theta) = (1-\beta)\omega + \alpha y_{t-1} + \beta\lambda_{t-1}(\theta) + \gamma f(x_{t-1}), \quad t = 1, ..., T,$$

with initial condition $\lambda_0(\theta) = \omega$. The MLE $\hat{\theta}$, $\hat{\theta} = \arg\max_{\theta\in\Theta} \mathcal{L}_T(\theta)$, where Θ is given by

$$\Theta = \{(\gamma, \alpha, \omega, \beta)' \in \mathbb{R}^4 : 0 \leq \gamma \leq \overline{\gamma}, 0 \leq \alpha \leq \overline{\alpha}, \underline{\omega} \leq \omega \leq \overline{\omega}, 0 \leq \beta \leq \overline{\beta}\},$$

for some $0 < \overline{\gamma} < \infty$, $0 < \overline{\alpha} < \infty$, $0 < \underline{\omega} < \overline{\omega} < \infty$, and $0 < \overline{\beta} < 1$. The constrained MLE is given by,

$$\tilde{\theta} = \arg\max_{\theta\in\Theta_0} \mathcal{L}_T(\theta), \quad \text{with } \Theta_0 = \{(\gamma, \alpha, \omega, \beta)' \in \Theta : \gamma = 0\}.$$

As before let $\theta_0 \in \Theta$ denote the true parameter and assume throughout that $\underline{\omega} < \omega_0 < \overline{\omega}$, $\alpha_0 < \overline{\alpha}$, and $\beta_0 < \overline{\beta}$. Similar to Theorem 2.2, it follows from the proof in the appendix that the limiting information satisfies $(J^{-1})_{\alpha,\gamma} = 0$, that is condition (A.iv) holds.

Assumption 3.1 *(Agosto et al., 2016) The joint process $\{(y_t, \lambda_t, x_t)' : t \in \mathbb{Z}\}$ is stationary and ergodic with $E[\|[y_t, \lambda_t, f(x_t)]'\|^3] < \infty$.*

Assumption 3.2 *With $z_t := y_t/\lambda_t$, it holds that z_t and $\mathcal{F}_{t-1}$ are independent.*

Assumption 3.3 *For $\beta_0 = 0$, $E[f^k(x_t)|y_t, y_{t-1}] = E[f^k(x_t)]$ for $k = 1, 2$.*

We note that Assumption 3.3 holds if the covariate x_t is independent of $\mathcal{F}_t^z$, where $\mathcal{F}_t^z$ denotes the natural filtration generated by $(z_s : s \leq t)$, with $z_t = y_t/\lambda_t$.

Assumption 3.4 *For any $(a, b) \neq (0, 0)$, $ay_t + bf(x_t)|\mathcal{F}_{t-1}$ has a non-degenerate distribution.*

Theorem 3.3 *Consider the PAR-X model given by (3.3) with log-likelihood (3.4). Suppose that Assumptions 3.1-3.4 are satisfied. Then under H_0, with $\alpha_0 > 0$ and $\beta_0 \geq 0$,*

$$LR_T := 2[\mathcal{L}_T(\hat{\theta}) - \mathcal{L}_T(\tilde{\theta})] \xrightarrow{w} [\max(U,0)]^2, \text{ where } U \sim N(0,1).$$

References

Agosto, A., G. Cavaliere, D. Kristensen, and A. Rahbek (2016). Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX), *Journal of Empirical Finance* 38, 640–663.

Ahmad, A. and C. Francq (2016). Poisson QMLE of count time series models. *Journal of Time Series Analysis* 37, 291-314.

Andrews, D.W.K. (1999). Estimation when a parameter is on a boundary, *Econometrica* 67, 1341–1383.

Andrews, D.W.K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis, *Econometrica* 69, 683–734.

Berkes, I. and L. Horváth (2004). The efficiency of the estimators of the parameters in GARCH processes, *Annals of Statistics* 32, 633–655.

Bierens, H.J. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests, *Econometrica* 65, 1129–1151.

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

Chernoff, A. (1954). On the distribution of the likelihood ratio, *The Annals of Mathematical Statistics* 25, 573–578.

Demos, A. and E. Sentana (1998). Testing for GARCH effects: A one-sided approach, *Journal of Econometrics* 86, 97-127.

Fokianos, K., A. Rahbek and D. Tjostheim (2009). Poisson autoregression, *Journal of the American Statistical Association* 104, 1430-1439.

Francq, C. and L.Q. Thieu (2015). Qml inference for volatility models with covariates, MPRA working paper, No. 63198.

Francq, C. and J.M. Zakoïan (2007). Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero, *Stochastic Processes and Their Applications* 117, 1265–1284.

Francq, C. and J.M. Zakoïan (2009). Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons, *Journal of the American Statistical Association* 104, 313–324.

Francq, C. and J.M. Zakoïan (2010). GARCH Models: Structure, Statistical Inference and Financial Applications, Wiley, Chichester.

Han, H. (2015). Asymptotic properties of GARCH-X processes, *Journal of Financial Econometrics* 13, 188-221.

Han, H. and D. Kristensen (2014). Asymptotic for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates, *Journal of Business & Economic Statistics* 32, 416–429.

Han, H. and J.Y. Park (2012). ARCH/GARCH with persistent covariate: Asymptotic theory of MLE, *Journal of Econometrics* 167, 95-112.

Jensen, S.T. and A. Rahbek (2004a). Asymptotic inference for nonstationary GARCH, *Econometric Theory* 20, 1203-1226.

Jensen, S.T. and A. Rahbek (2004b). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case, *Econometrica* 72, 641-646.

Pedersen, R.S. (2017). Inference and testing on the boundary in extended constant conditional correlation GARCH models, *Journal of Econometrics* 196, 23-36.

Pedersen, R.S. and A. Rahbek (2016). Nonstationary GARCH with t-distributed innovations, *Economics Letters* 138, 19-21.

Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications, *The Annals of Mathematical Statistics* 33, 659–680.

Silvapulle M.J. and P. Silvapulle (1995). A score test against one-sided alternatives, *Journal of the American Statistical Association* 90, 342-349.

Straumann, D. (2005). Estimation in Conditionally Heteroscedastic Time Series Models, In: *Lecture Notes in Statistics*, Springer, Heidelberg.

White, H. (2001). Asymptotic Theory for Econometricians: Revised Edition, Emerald Group Publishing Ltd., Bingley.

Appendix

Throughout, we let $0 < C < \infty$ and $0 < \rho < 1$ denote generic constants.

A Proofs of Theorems 2.1 and 2.2

Proof of Theorem 2.1. Consider initially the ergodic version $\mathcal{L}_T^* (\theta, \beta)$ of the log-likelihood function $\mathcal{L}_T (\theta, \beta)$ in (2.4). Specifically, in light of Assumption 2.1, for any $\theta \in \Theta$ and $\beta \in \Theta_\beta$, let

$$\mathcal{L}_T^* (\theta, \beta) = \sum_{t=1}^{T} l_t^*(\theta, \beta), \tag{A.1}$$

$$l_t^*(\theta, \beta) = \log \left(\frac{1}{\sqrt{2\pi h_t^*(\theta, \beta)}} \exp \left\{ -\frac{y_t^2}{2h_t^*(\theta, \beta)} \right\} \right), \quad t \in \mathbb{Z}, \tag{A.2}$$

$$h_t^*(\theta, \beta) = (1 - \beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}^*(\theta, \beta) + \gamma x_{t-1}^2, \quad t \in \mathbb{Z}.$$

We verify the regularity condition of Andrews (2001, Theorem 5):

(i) $\hat{\theta}_\beta = \theta_0 + o_p(1)$ and $\hat{\theta}_\beta^* = \theta_0 + o_p(1)$.

(ii) For fixed $\beta \in \Theta_\beta$, $\mathcal{L}_T^*(\theta, \beta)$ is twice continuously differentiable on Θ. For all sequences (γ_T), $\gamma_T \to 0$,

$$\sup_{\theta \in \Theta : \|\theta - \theta_0\| \leq \gamma_T} \left\| \frac{1}{T} \frac{\partial^2 \mathcal{L}_T^*(\theta, \beta)}{\partial \theta \partial \theta'} - \frac{1}{T} \frac{\partial^2 \mathcal{L}_T^*(\theta_0, \beta)}{\partial \theta \partial \theta'} \right\| \xrightarrow{p} 0.$$

(iii) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial l_t^*(\theta_0, \cdot)}{\partial \theta} \Rightarrow G.$ for some Gaussian process $\{G_\beta : \beta \in \Theta_\beta\}$ that has bounded continuous sample paths almost surely.

(iv) $-\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t^*(\theta_0, \beta)}{\partial \theta \partial \theta'} = J_\beta + o_p(1)$ for all $\beta \in \Theta_\beta$, with J_β positive definite uniformly on Θ_β.

(v) For any $\beta \in \Theta_\beta$, $G_\beta \sim N(0, cJ_\beta)$ with $c = \kappa_4/2$.

(vi) With $\hat{c}^*$ defined in (2.8), $\hat{c}^* = c + o_p(1)$.

(vii) $\sup_{\beta \in \Theta_\beta} \|\partial \mathcal{L}_T^*(\theta_0, \beta)/\partial\theta - \partial \mathcal{L}_T(\theta_0, \beta)/\partial\theta\| = o_p(T^{1/2})$ and $\sup_{(\theta',\beta)' \in \mathcal{V}(\theta_0)\times\Theta_\beta} \|\partial^2 \mathcal{L}_T^*(\theta, \beta)/\partial\theta\partial\theta' - \partial^2 \mathcal{L}_T(\theta, \beta)/\partial\theta\partial\theta'\| = o_p(T)$, with $\mathcal{V}(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| < \varepsilon\}$ for some $\varepsilon > 0$.

Here *(i)* follows by Lemma A.1, while *(ii)* holds by Lemma A.2 and the uniform law of large numbers (ULLN) for ergodic processes (see e.g. Ranga Rao, 1962). *(iii)* holds by Lemma A.3. For *(iv)*, note that by the ergodic theorem and Lemma A.2

$$-\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l_t^*(\theta_0, \beta)}{\partial\theta\partial\theta'} = J_\beta + o_p(1), \quad \text{with } J_\beta = -E\left[\frac{\partial^2 l_t^*(\theta_0, \beta)}{\partial\theta\partial\theta'}\right] \text{ and}$$

$$\frac{\partial^2 l_t^*(\theta, \beta)}{\partial\theta\partial\theta'} = \frac{1}{2h_t^{*2}(\theta, \beta)}\left(2\frac{y_t^2}{h_t^*(\theta, \beta)} - 1\right)\frac{\partial h_t^*(\theta, \beta)}{\partial\theta}\frac{\partial h_t^*(\theta, \beta)}{\partial\theta'}.$$

Hence, with $V_{t,\beta}$ defined in (A.19),

$$J_\beta = E\left[\frac{1}{2\omega_0^2}(2z_t^2 - 1)V_{t,\beta}V_{t,\beta}'\right] = \frac{1}{2\omega_0^2}E\left[V_{t,\beta}V_{t,\beta}'\right]. \tag{A.3}$$

In order to show that the matrix J_β is positive definite, we note that $V_{t,\beta}V_{t,\beta}'$ is positive semidefinite. For $k = (k_1, k_2, k_3) \in \mathbb{R}^3$, $k'J_\beta k = 0$ if and only if

$$k'V_{t,\beta} = k_1\sum_{i=0}^{\infty}\beta^i x_{t-1-i}^2 + k_2\sum_{i=0}^{\infty}\beta^i y_{t-1-i}^2 + k_3 = 0 \quad \text{a.s.} \tag{A.4}$$

Due to Assumption 2.4, we have that (A.4) is true if and only if $k = 0$. We conclude that J_β is positive definite. This establishes *(iv)*. Next, *(v)* follows directly from Lemma A.3 and (A.3), and *(vi)* by Lemma A.4. Finally, *(vii)* is implied by Lemma A.5. ■

Proof of Theorem 2.2. Similar to the proof of Theorem 2.1, we consider the ergodic quasi-log-likelihood function for $\tau \in \Theta_\tau$,

$$\mathcal{L}_T^*(\tau) = \sum_{t=1}^{T} l_t^*(\tau), \tag{A.5}$$

$$l_t^*(\tau) = \log\left(\frac{1}{\sqrt{2\pi h_t^*(\tau)}}\exp\left\{-\frac{y_t^2}{2h_t^*(\tau)}\right\}\right), \quad t \in \mathbb{Z}, \tag{A.6}$$

$$h_t^*(\tau) = (1-\beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}^*(\tau) + \gamma x_{t-1}^2, \quad t \in \mathbb{Z}.$$

We start out by verifying the following conditions which allows to use a modified version of Lemma 2.1:

1. The estimators $\hat{\tau}$ and $\tilde{\tau}$ satisfy condition (A.i) of Lemma 2.1.

2. Condition (A.ii) of Lemma 2.1 is satisfied for the ergodic likelihood in (A.5) where the covariance matrix of G is replaced by cJ with $c = \kappa_4/2$ and $\kappa_4 = E[(z_t^2 - 1)^2]$.

3. Condition (A.iii) of Lemma 2.1 is satisfied for the ergodic likelihood in (A.5).

4. With $\mathcal{V}(\tau_0) = \{\tau \in \Theta_\tau : \|\tau - \tau_0\| < \varepsilon\}$ for some small $\varepsilon > 0$,

$$\|\partial \mathcal{L}_T^*(\tau_0)/\partial\tau - \partial \mathcal{L}_T(\tau_0)/\partial\tau\| = o_p(T^{1/2})$$

   and

$$\sup_{\tau \in \mathcal{V}(\tau_0)} \|\partial^2 \mathcal{L}_T^*(\tau)/\partial\tau\partial\tau' - \partial^2 \mathcal{L}_T(\tau)/\partial\tau\partial\tau'\| = o_p(T).$$

5. With $\tilde{\kappa}_4$ defined in (2.12), $\tilde{\kappa}_4 = \kappa_4 + o_p(1)$.

The consistency of $\hat{\tau}$ and $\tilde{\tau}$ follows by arguments given in Han and Kristensen (2014, proof of Theorem 3), using Assumptions 2.1 and 2.4.

Turning to point 2, we note that

$$\frac{\partial l_t^*(\tau_0)}{\partial\tau} = \frac{1}{2}(z_t^2 - 1)V_t,$$

with

$$V_t := \frac{\partial h_t^*(\tau_0)/\partial\tau}{h_t^*(\tau_0)}, \tag{A.7}$$

$$\partial h_t^*(\tau)/\partial\tau = \left(\sum_{i=0}^{\infty}\beta^i x_{t-1-i}^2, \sum_{i=0}^{\infty}\beta^i y_{t-1-i}^2, 1, \sum_{i=1}^{\infty} i\beta^{i-1}(\alpha y_{t-1-i}^2 + \gamma x_{t-1-i}^2)\right)'. \tag{A.8}$$

By Assumption 2.6, $E[\|V_t\|^2] < \infty$, so using that z_t and V_t are independent, and that $E[z_t^4] < \infty$ by Assumption 2.2 , $E[\|\partial l_t^*(\tau_0)/\partial\tau\|^2] < \infty$. Noting that V_t is $\mathcal{F}_{t-1}$-measurable, we have that $\partial l_t^*(\tau_0)/\partial\tau$ is a martingale difference with respect to $\mathcal{F}_t$ and with finite variance. Hence, using that $\partial l_t^*(\tau_0)/\partial\tau$ is ergodic, by the CLT by Brown (1971), we have $T^{-1/2}\partial \mathcal{L}_T^*(\tau_0)/\partial\tau \xrightarrow{w}$

$N(0,\Omega)$, with $\Omega = (\kappa_4/4)E[V_tV_t']$. By the ergodic theorem and Lemma A.6, $-T^{-1}\partial^2\mathcal{L}_T^*(\tau_0)/\partial\tau\partial\tau' = J + o_p(1)$, where, by (A.28),

$$J = \frac{1}{2}E[V_tV_t']. \tag{A.9}$$

Clearly, $\Omega = cJ$ with $c = \kappa_4/2$. By arguments similar to the ones given in the proof of Theorem 2.1 to show that J_β is positive definite, we conclude that $E[V_tV_t']$ is positive definite.

Point 3 follows by Lemma A.6 and the ULLN for ergodic processes. Point 4 follows by arguments similar to the ones given in the proof of Lemma A.5. Point 5 follows by Lemma A.7. If $\beta_0 > 0$ condition (A.iv) of Lemma 2.1 is satisfied. The limiting distribution of LR_T is then immediate from Lemma 2.1, using point 4, that $\tilde{c} = c + o_p(1)$ (by point 5), and Slutzky's Lemma. In the case $\beta_0 = 0$, following condition (A.iv) of Lemma 2.1 we verify $(J^{-1})_{\gamma,\beta} = 0$. From (A.7),(A.8), and (A.9),

$$J = \frac{1}{2}E\left[\frac{\partial h_t^*(\tau_0)/\partial\tau}{h_t^*(\tau_0)}\frac{\partial h_t^*(\tau_0)/\partial\tau'}{h_t^*(\tau_0)}\right],$$

with $\partial h_t^*(\tau_0)/\partial\tau = [x_{t-1}^2, y_{t-1}^2, 1, \alpha_0 y_{t-2}^2]'$, and $h_t^*(\tau_0) = \omega_0 + \alpha_0 y_{t-1}^2$. Hence, using Assumption 2.5,

$$\begin{aligned} J &= \frac{1}{2}E\left[\frac{1}{(\omega_0+\alpha_0 y_{t-1}^2)^2}\begin{pmatrix} x_{t-1}^4 & y_{t-1}^2x_{t-1}^2 & x_{t-1}^2 & \alpha_0 x_{t-1}^2 y_{t-2}^2 \\ & y_{t-1}^4 & y_{t-1}^2 & \alpha_0 y_{t-1}^2 y_{t-2}^2 \\ & & 1 & \alpha_0 y_{t-2}^2 \\ & & & \alpha_0^2 y_{t-2}^4 \end{pmatrix}\right] \\ &= \frac{1}{2}\begin{pmatrix} a\kappa_{4,x} & b\kappa_{2,x} & a\kappa_{2,x} & d\kappa_{2,x} \\ b\kappa_{2,x} & c & b & f \\ a\kappa_{2,x} & b & a & d \\ d\kappa_{2,x} & f & d & g \end{pmatrix}, \end{aligned}$$

where

$$a = E\left[\frac{1}{(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad b = E\left[\frac{y_{t-1}^2}{(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad c = E\left[\frac{y_{t-1}^4}{(\omega_0+\alpha_0 y_{t-1}^2)^2}\right] \tag{A.10}$$

$$d = E\left[\frac{\alpha_0 y_{t-2}^2}{(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad f = E\left[\frac{\alpha_0 y_{t-1}^2 y_{t-2}^2}{(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad g = E\left[\frac{\alpha_0^2 y_{t-2}^4}{(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \tag{A.11}$$

$$\kappa_{2,x} = E[x_t^2], \quad \text{and} \quad \kappa_{4,x} = E[x_t^4]. \tag{A.12}$$

It holds that

$$J^{-1}=\begin{pmatrix} \frac{2}{a\kappa_{4,x}-a\kappa_{2,x}^2} & 0 & \frac{-2}{a\kappa_{4,x}-a\kappa_{2,x}^2}\kappa_{2,x} & 0 \\ 0 & \frac{2d^2-2ag}{gb^2-2bdf+cd^2+af^2-acg} & \frac{2bg-2df}{gb^2-2bdf+cd^2+af^2-acg} & \frac{2af-2bd}{gb^2-2bdf+cd^2+af^2-acg} \\ \frac{-2}{a\kappa_{4,x}-a\kappa_{2,x}^2}\kappa_{2,x} & \frac{2bg-2df}{gb^2-2bdf+cd^2+af^2-acg} & \xi & \frac{2cd-2bf}{gb^2-2bdf+cd^2+af^2-acg} \\ 0 & \frac{2af-2bd}{gb^2-2bdf+cd^2+af^2-acg} & \frac{2cd-2bf}{gb^2-2bdf+cd^2+af^2-acg} & \frac{2b^2-2ac}{gb^2-2bdf+cd^2+af^2-acg} \end{pmatrix} \tag{A.13}$$

with

$$\xi = \frac{-1}{a\left(\kappa_{2,x}^2-\kappa_{4,x}\right)(gb^2-2bdf+cd^2+af^2-acg)} \times \left(2gb^2\kappa_{2,x}^2-4bdf\kappa_{2,x}^2+2cd^2\kappa_{2,x}^2+2a\kappa_{4,x}f^2-2acg\kappa_{4,x}\right).$$

We note that J^{-1} has entries zero with respect to β and γ, and the limiting distribution of LR_T follows by Lemma 2.1, using that $\tilde{c}=c+o_p(1)$ (by point 5), and Slutzky's Lemma. ■

A.1 Lemmas related to the proof of Theorem 2.1

Lemma A.1 *Under Assumptions 2.1-2.4, and H_0^*, $\hat{\theta}_\beta=\theta_0+o_p(1)$ and $\hat{\theta}_\beta^*=\theta_0+o_p(1)$.*

Proof. We start out by showing that $\hat{\theta}_\beta$ is consistent. The proof follows the steps given in Han and Kristensen (2014, Proof of Theorem 3). Since Θ is compact and $\theta\mapsto l_t^*(\theta,\beta)$ is continuous almost surely on Θ for fixed $\beta\in\Theta_\beta$, it suffices to show that (i) $\frac{1}{T}\sum_{t=1}^T l_t^*(\theta,\beta)=E[l_t^*(\theta,\beta)]+o_p(1)$, where $E[l_t^*(\theta,\beta)]$ exists for all $(\theta',\beta)'\in\Theta\times\Theta_\beta$, (ii) $E[l_t^*(\theta_0,\beta)]>E[l_t^*(\theta,\beta)]$ for all $\theta\in\Theta\setminus\{\theta_0\}$ and fixed $\beta\in\Theta_\beta$, (iii) $E[\sup_{\theta\in\Theta}l_t^*(\theta,\beta)]<\infty$ for fixed $\beta\in\Theta_\beta$ and (iv) $\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}|\mathcal{L}_T^*(\theta,\beta)-\mathcal{L}_T(\theta,\beta)|=o_p(T)$.

We note that

$$E[\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta} l_t^*(\theta,\beta)]\leq -\frac{1}{2}\log(\underline{\omega})<\infty. \tag{A.14}$$

Hence (i) follows by the ergodic theorem.

Turning to (ii), since $\gamma_0=\alpha_0=0$, H_0^*, $h_t^*(\theta_0,\beta)=\omega_0$ a.s for all t. Hence,

$$E[|l_t^*(\theta_0,\beta)|]\leq\frac{1}{2}E\{|\log[h_t^*(\theta_0,\beta)]|\}+C<\infty.$$

Using that $\log(x)\leq x-1$ for all $x>0$ and with equality if and only if $x=1$, we have that

$$E[l_t^*(\theta_0,\beta)]-E[l_t^*(\theta,\beta)]\geq\frac{1}{2}E\left\{-\log\left[\frac{h_t^*(\theta_0,\beta)}{h_t^*(\theta,\beta)}\right]-1+\frac{h_t^*(\theta_0,\beta)}{h_t^*(\theta,\beta)}\right\}\geq 0, \tag{A.15}$$

with equality if and only if $h_t^*(\theta_0, \beta) = h_t^*(\theta, \beta)$ a.s. Using again that H_0^*, $h_t^*(\theta_0, \beta) = \omega_0$ a.s, (A.15) holds if and only if $\omega_0 = (1-\beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}^*(\theta, \beta) + \gamma x_{t-1}^2$ a.s.,or equivalently,

$$1 - \frac{\omega}{\omega_0} = \sum_{i=0}^{\infty} \beta^i \left(\alpha z_{t-1-i}^2 + \frac{\gamma}{\omega_0} x_{t-1-i}^2 \right) \quad \text{a.s.} \tag{A.16}$$

Suppose $(\alpha, \gamma) = (0, 0)$, then clearly $\omega = \omega_0$. On the other hand, suppose that $(\alpha, \gamma) \neq (0, 0)$. Then due to (A.16), $\sum_{i=0}^{\infty} \beta^i \left(\alpha z_{t-1-i}^2 + \frac{\gamma}{\omega_0} x_{t-1-i}^2 \right)$ is degenerate, which is ruled out by Assumption 2.4. We conclude that (A.16) holds if and only if $\theta = \theta_0 = (0, 0, \omega_0)'$, and hence that (ii) holds.

We note that (A.14) implies (iii).

It remains to verify (iv). From Francq and Zakoïan (2010, p.157),

$$\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} |l_t^*(\theta, \beta) - l_t(\theta, \beta)| \leq \overline{\beta}^t \frac{1}{2} \left(\frac{1}{\underline{\omega}} + \frac{y_t^2}{\underline{\omega}^2} \right) [\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} h_0^*(\theta, \beta) + \overline{\omega}].$$

Using that $y_t^2 = \omega_0 z_t^2$, we have that for some $r \in (0, 1)$,

$$\begin{aligned} & E \left[\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} |l_t^*(\theta, \beta) - l_t(\theta, \beta)|^r \right] \\ & \leq E \left[\left| \overline{\beta}^t \frac{1}{2} \left(\frac{1}{\underline{\omega}} + \frac{\omega_0 z_t^2}{\underline{\omega}^2} \right) [\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} h_0^*(\theta, \beta) + \overline{\omega}] \right|^r \right] = O(\rho^t), \end{aligned} \tag{A.17}$$

where we have used Assumption 2.2 and that $E[\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} |h_0^*(\theta, \beta)|^r] < \infty$ as $E[\|(x_t, y_t)\|^4] < \infty$ in light of Assumptions 2.2 and 2.3 and the fact that $\overline{\beta} < 1$. Using (A.17) and Markov's inequality, for any $\varepsilon > 0$, $\sum_{t=1}^{\infty} P[\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} |l_t^*(\theta, \beta) - l_t(\theta, \beta)| > \varepsilon] < \infty$. By the Borel-Cantelli Lemma, we conclude that $\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} |l_t^*(\theta, \beta) - l_t(\theta, \beta)| \to 0$ a.s. as $t \to \infty$. As $\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} |\mathcal{L}_T^*(\theta, \beta) - \mathcal{L}_T(\theta, \beta)| \leq \sum_{t=1}^{T} \sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} |l_t^*(\theta, \beta) - l_t(\theta, \beta)|$, we conclude that (iv), using Cesaro's Lemma.

Turning to the consistency of $\hat{\theta}_\beta^*$, we note that $\hat{\theta}_\beta^* = (0, 0, \hat{\omega}_\beta^*)'$ with $\hat{\omega}_\beta^* = \arg\max_{\omega \in [\underline{\omega}, \overline{\omega}]} -\frac{1}{2} \sum_{t=1}^{T} [\log(\omega) + y_t^2/\omega] = \underline{\omega} \mathbf{1} \left\{ T^{-1} \sum_{t=1}^{T} y_t^2 < \underline{\omega} \right\} + \overline{\omega} \mathbf{1} \left\{ T^{-1} \sum_{t=1}^{T} y_t^2 > \overline{\omega} \right\}$ $+ T^{-1} \sum_{t=1}^{T} y_t^2 \mathbf{1} \left\{ \underline{\omega} \leq T^{-1} \sum_{t=1}^{T} y_t^2 \leq \overline{\omega} \right\}$. By assumption, $\omega_0 \in (\underline{\omega}, \overline{\omega})$, such that, using the ergodic theorem, $\hat{\omega}_\beta^* = T^{-1} \sum_{t=1}^{T} y_t^2 + o_p(1) = \omega_0 + o_p(1)$. Hence, $\hat{\theta}_\beta^* = \theta_0 + o_p(1)$. ■

Lemma A.2 *Under Assumptions 2.1-2.3 and H_0^*, with $l_t^*(\theta,\beta)$ defined in (A.2),*

$$E\left[\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\left\|\frac{\partial^2 l_t^*(\theta,\beta)}{\partial\theta\partial\theta'}\right\|\right]<\infty. \tag{A.18}$$

Proof. With θ_i the ith entry of $\theta=(\gamma,\alpha,\omega)'$,

$$\frac{\partial^2 l_t^*(\theta,\beta)}{\partial\theta_i\partial\theta_j'}=-\frac{1}{2}(2\omega_0 z_t^2-1)\frac{1}{h_t^*(\theta,\beta)}\frac{\partial h_t^*(\theta,\beta)/\partial\theta_i}{h_t^*(\theta,\beta)}\frac{\partial h_t^*(\theta,\beta)/\partial\theta_j}{h_t^*(\theta,\beta)}.$$

As $\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\partial h_t^*(\theta,\beta)/\partial\gamma\le\sum_{i=0}^\infty\overline{\beta}^i x_{t-1-i}^2$, $\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\partial h_t^*(\theta,\beta)/\partial\alpha\le\sum_{i=0}^\infty\overline{\beta}^i\omega_0 z_{t-1-i}^2$, $\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\partial h_t^*(\theta,\beta)/\partial\omega=1$, and $\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}h_t^{*-1}(\theta,\beta)\le\underline{\omega}^{-1}$, (A.18) follows by Hölder's inequality and Assumptions 2.2-2.3. ■

Lemma A.3 *Under Assumptions 2.1-2.4 and H_0^*, with $l_t^*(\theta,\beta)$ defined in (A.2),*

$$G_{T,\cdot}:=\frac{1}{\sqrt{T}}\sum_{t=1}^T\frac{\partial l_t^*(\theta_0,\cdot)}{\partial\theta}\Rightarrow G.$$

for some Gaussian process $\{G_\beta:\beta\in\Theta_\beta\}$ that has bounded continuous sample paths almost surely. Moreover, with $\beta_1,\beta_2\in\Theta_\beta$, the process $\{G_\beta:\beta\in\Theta_\beta\}$ has kernel

$$\Sigma_{\beta_1\beta_2}=\frac{\kappa_4}{4\omega_0^2}E[V_{t,\beta_1}V_{t,\beta_2}'],\quad\kappa_4=E[(z_t^2-1)^2],$$

where

$$V_{t,\beta}=\left(\sum_{i=0}^\infty\beta^i x_{t-1-i}^2,\sum_{i=0}^\infty\beta^i y_{t-1-i}^2,1\right)'. \tag{A.19}$$

Proof. Following Andrews (2001, p.730), and noting that $\Theta_\beta=[0,\overline{\beta}]$ is totally bounded, it suffices to show that (i) any finite dimensional distributions of $G_{T,\cdot}$ converge to those of $G_\cdot$ and (ii) $\{G_{T,\cdot}:T\ge1\}$ is tight.
We start out by proving (i). For $\beta\in\Theta_\beta$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T\frac{\partial l_t^*(\theta_0,\beta)}{\partial\theta}=\frac{1}{\sqrt{T}}\sum_{t=1}^T\frac{1}{2\omega_0}(z_t^2-1)V_{t,\beta}, \tag{A.20}$$

with $V_{t,\beta}$ defined in (A.19). By Assumptions 2.2-2.3 and since $\overline{\beta}<1$, $E[\|V_{t,\beta}\|^2]<\infty$, so using that $\frac{1}{2\omega_0}(z_t^2-1)V_{t,\beta}$ is a martingale difference sequence with respect to $\mathcal{F}_{t-1}$, it follows by Brown (1971) that $T^{-1/2}\sum_{t=1}^T\partial l_t^*(\theta_0,\beta)/\partial\theta\xrightarrow{w}N(0,\Sigma_{\beta\beta})$.

Next, let $\beta_1, \beta_2 \in \Theta_\beta$, and $k_1, k_2 \in \mathbb{R}^3$. Using the same arguments as above,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (k_1, k_2) \left(\frac{\partial l_t^*(\theta_0, \beta_1)}{\partial \theta}, \frac{\partial l_t^*(\theta_0, \beta_1)}{\partial \theta} \right)' \xrightarrow{w} N(0, k_1' \Sigma_{\beta_1 \beta_1} k_1 + k_2' \Sigma_{\beta_2 \beta_2} k_2 + k_1' \Sigma_{\beta_1 \beta_2} k_2),$$

where $\Sigma_{\beta_i \beta_j} = \frac{\kappa_4}{4\omega_0^2} E[V_{t,\beta_i} V_{t,\beta_j}']$, $i, j = 1, 2$, with $k_1' \Sigma_{\beta_1 \beta_1} k_1 + k_2' \Sigma_{\beta_2 \beta_2} k_2 + k_1' \Sigma_{\beta_1 \beta_2} k_2 \geq 0$. An application of the Cramer-Wold theorem yields that (i) holds.

Next, we verify (ii) by relying on Bierens and Ploberger (1997, Lemma A.1). We consider (A.20), and note that $E[(2\omega_0)^{-1}(z_t^2 - 1)|\mathcal{F}_{t-1}] = 0$ and $E[(2\omega_0)^{-2}(z_t^2 - 1)^2] < \infty$. Moreover, for any $\beta_1, \beta_2 \in \Theta_\beta$ $\|V_{t,\beta_1} - V_{t,\beta_2}\| \leq |\beta_1 - \beta_2| K_t$, with

$$K_t = \left\| \left(\sum_{i=0}^{\infty} \overline{\beta}^i x_{t-2-i}^2, \sum_{i=0}^{\infty} \overline{\beta}^i y_{t-2-i}^2 \right)' \right\| = \left\| \left(\sum_{i=0}^{\infty} \overline{\beta}^i x_{t-2-i}^2, \sum_{i=0}^{\infty} \overline{\beta}^i \omega_0 z_{t-2-i}^2 \right)' \right\|$$

which is $\mathcal{F}_{t-1}$-measurable. Following Bierens and Ploberger (1997, Lemma A.1), it suffices to show that $\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[(2\omega_0)^{-2}(z_t^2 - 1)^2 K_t^2] < \infty$, which is immediate from Assumptions 2.2-2.3. ■

Lemma A.4 *With $\hat{\kappa}_4^*$ defined in (2.8), suppose that Assumptions 2.1-2.4 and H_0^* hold. Then $\hat{\kappa}_4^* = \kappa_4 + o_p(1)$.*

Proof. We have that $\kappa_4 = E[z_t^4] - 1$, and $\hat{\kappa}_4^* = T^{-1} \sum_{t=1}^{T} y_t^4 / \hat{\omega}^{*2} + 1 - 2T^{-1} \sum_{t=1}^{T} y_t^2 / \hat{\omega}^*$. Note that $y_t^2/\hat{\omega}^* - z_t^2 = (\omega_0/\hat{\omega}^* - 1) z_t^2$ and $y_t^4/\hat{\omega}^{*2} - z_t^4 = [(\omega_0/\hat{\omega}^*)^2 - 1] z_t^4$. Hence, by Lemma A.1 and the ergodic theorem, $\hat{\kappa}_4^* = T^{-1} \sum_{t=1}^{T} z_t^4 + 1 - 2T^{-1} \sum_{t=1}^{T} z_t^2 + o_p(1) = E[z_t^4] - 1 + o_p(1) = \kappa_4 + o_p(1)$. ■

Lemma A.5 *Under Assumptions 2.1, 2.3 and H_0^*, with $\mathcal{L}_T^*(\theta, \beta)$ defined in (A.1) and $\mathcal{L}_T(\theta, \beta)$ defined in (2.4),*

$$\sup_{\beta \in \Theta_\beta} \|\partial \mathcal{L}_T^*(\theta_0, \beta)/\partial\theta - \partial \mathcal{L}_T(\theta_0, \beta)/\partial\theta\| = o_p(T^{1/2}) \tag{A.21}$$

and

$$\sup_{(\theta', \beta)' \in \Theta \times \Theta_\beta} \|\partial^2 \mathcal{L}_T^*(\theta, \beta)/\partial\theta\partial\theta' - \partial^2 \mathcal{L}_T(\theta, \beta)/\partial\theta\partial\theta'\| = o_p(T). \tag{A.22}$$

Proof. We start out by showing that

$$E\left[\sup_{\beta \in \Theta_\beta} \|\partial l_t^*(\theta_0, \beta)/\partial\theta - \partial l_t(\theta_0, \beta)/\partial\theta\|^r \right] = O(\rho^t), \tag{A.23}$$

for some sufficiently small $r > 0$. We have that

$$\partial l_t^*(\theta_0, \beta)/\partial\theta - \partial l_t(\theta_0, \beta)/\partial\theta = -\frac{1}{2\omega_0}(z_t^2 - 1)\left[\partial h_t^*(\theta_0, \beta)/\partial\theta - \partial h_t(\theta_0, \beta)/\partial\theta\right].$$

Note that

$$\partial h_t^*(\theta, \beta)/\partial\theta = V_{t,\beta} \quad \text{and} \quad \partial h_t(\theta, \beta)/\partial\theta = \left(\sum_{i=0}^{t-1} \beta^i x_{t-1-i}^2, \sum_{i=0}^{t-1} \beta^i y_{t-1-i}^2, 1\right)'. \tag{A.24}$$

Hence,

$$\sup_{\beta\in\Theta_\beta} \|\partial h_t^*(\theta, \beta)/\partial\theta - \partial h_t(\theta, \beta)/\partial\theta\| \le \overline{\beta}^t \left\| \left(\sum_{i=0}^{\infty} \overline{\beta}^i x_{-1-i}^2, \sum_{i=0}^{\infty} \overline{\beta}^i y_{-1-i}^2\right)' \right\|, \tag{A.25}$$

and we have that for some small $r \in (0, 1)$,

$$\begin{aligned}
&E\left[\sup_{\beta\in\Theta_\beta} \|\partial l_t^*(\theta_0, \beta)/\partial\theta - \partial l_t(\theta_0, \beta)/\partial\theta\|^r\right] \\
&\le E\left[\left(\frac{1}{2\omega_0}\right)^r |z_t^2 - 1|^r \overline{\beta}^{rt} \left\| \left(\sum_{i=0}^{\infty} \overline{\beta}^i x_{-1-i}^2, \sum_{i=0}^{\infty} \overline{\beta}^i y_{-1-i}^2\right)' \right\|^r\right] \\
&= E\left[\left(\frac{1}{2\omega_0}\right)^r |z_t^2 - 1|^r \overline{\beta}^{rt} \left\| \left(\sum_{i=0}^{\infty} \overline{\beta}^i x_{-1-i}^2, \sum_{i=0}^{\infty} \overline{\beta}^i \omega_0 z_{-1-i}^2\right)' \right\|^r\right] = O(\rho^t)
\end{aligned}$$

where we have used Assumption 2.3 and that $\overline{\beta} < 1$. Hence, (A.23) holds. By the Markov inequality and the c_r inequality (see e.g. White, 2001, Proposition 3.8), we have for any $\varepsilon > 0$ and some $r \in (0, 1)$,

$$\begin{aligned}
&P\left[T^{-1/2} \sup_{\beta\in\Theta_\beta} \|\partial \mathcal{L}_T^*(\theta_0, \beta)/\partial\theta - \partial \mathcal{L}_T(\theta_0, \beta)/\partial\theta\| > \varepsilon\right] \\
&\le T^{-r/2}\varepsilon^{-r} \sum_{t=1}^{T} E\left[\sup_{\beta\in\Theta_\beta} \|\partial l_t^*(\theta_0, \beta)/\partial\theta - \partial l_t(\theta_0, \beta)/\partial\theta\|^r\right] = o(1),
\end{aligned}$$

as $T \to \infty$, where we have used (A.23). We conclude that (A.21) holds.

In order to show (A.22), we start out by showing that for any $\theta_i, \theta_j \in \{\gamma, \alpha, \omega\}$, for some sufficiently small $r \in (0, 1)$,

$$E\left[\sup_{(\theta', \beta)'\in\Theta\times\Theta_\beta} |\partial^2 l_t^*(\theta, \beta)/\partial\theta_i\partial\theta_j - \partial^2 l_t(\theta, \beta)/\partial\theta_i\partial\theta_j|^r\right] = O(\rho^t), \tag{A.26}$$

From Francq and Zakoïan (2010, p.167), suppressing the dependence on θ, β,

$$
\begin{aligned}
&\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta} |\partial^2 l_t^*(\theta,\beta)/\partial\theta_i\partial\theta_j - \partial^2 l_t(\theta,\beta)/\partial\theta_i\partial\theta_j|^r \\
&\leq \frac{1}{2}\left| \left(2\frac{y_t^2}{h_t^*} - 2\frac{y_t^2}{h_t}\right)\left(\frac{1}{h_t^*}\frac{\partial h_t^*}{\partial\theta_i}\right)\left(\frac{1}{h_t^*}\frac{\partial h_t^*}{\partial\theta_j}\right) \right. \\
&+ \left(2\frac{y_t^2}{h_t} - 1\right)\left[\left(\frac{1}{h_t^*} - \frac{1}{h_t}\right)\frac{\partial h_t^*}{\partial\theta_i} + \frac{1}{h_t}\left(\frac{\partial h_t^*}{\partial\theta_i} - \frac{\partial h_t}{\partial\theta_i}\right)\right]\left(\frac{1}{h_t^*}\frac{\partial h_t^*}{\partial\theta_j}\right) \\
&\left. + \left(2\frac{y_t^2}{h_t} - 1\right)\left(\frac{1}{h_t}\frac{\partial h_t}{\partial\theta_i}\right)\left[\left(\frac{1}{h_t^*} - \frac{1}{h_t}\right)\frac{\partial h_t^*}{\partial\theta_j} + \frac{1}{h_t}\left(\frac{\partial h_t^*}{\partial\theta_j} - \frac{\partial h_t}{\partial\theta_j}\right)\right]\right|
\end{aligned}
$$

Note that $y_t^2 = \omega_0 z_t^2$, $\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta} h_t^{-1} \leq \underline{\omega}^{-1}$, and $\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta} h_t^{*-1} \leq \underline{\omega}^{-1}$. Moreover,

$$
\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\left|\frac{1}{h_t^*} - \frac{1}{h_t}\right| \leq \frac{1}{\underline{\omega}^2}\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}|h_t^* - h_t| \leq \frac{1}{\underline{\omega}^2}\overline{\beta}^t\left(\sum_{i=0}^{\infty}\overline{\beta}^i x_{-1-i}^2 + \sum_{i=0}^{\infty}\overline{\beta}^i y_{-1-i}^2\right).
$$

Using Assumption 2.3, we conclude that for some small $0 < r^* < 1$,

$$
E\left[\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\left|\frac{1}{h_t^*} - \frac{1}{h_t}\right|^{r^*}\right] = O(\rho^t).
$$

Likewise, in light of (A.24)

$$
E\left[\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\left|\frac{\partial h_t}{\partial\theta_i}\right|^{r^*}\right] = O(\rho^t) \quad \text{and} \quad E\left[\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\left|\frac{\partial h_t^\star}{\partial\theta_i}\right|^{r^*}\right] = O(\rho^t),
$$

and using (A.25),

$$
E\left[\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\left|\frac{\partial h_t^*}{\partial\theta_i} - \frac{\partial h_t}{\partial\theta_i}\right|^{r^*}\right] = O(\rho^t).
$$

Combining these properties, and applying Hölder's inequality repeatedly, we conclude that (A.26) holds for some sufficiently small $0 < r < 1$. By arguments identical to the ones given above, we conclude that for any $\varepsilon > 0$,

$$
P\left[T^{-1}\sup_{(\theta',\beta)'\in\Theta\times\Theta_\beta}\|\partial^2\mathcal{L}_T^*(\theta,\beta)/\partial\theta_i\partial\theta_j - \partial^2\mathcal{L}_T(\theta,\beta)/\partial\theta_i\partial\theta_j\| > \varepsilon\right] = o(1).
$$

■

A.2 Lemmas related to the proof of Theorem 2.2

Lemma A.6 *Under Assumptions 2.1, 2.2, 2.4, 2.5, 2.6, H_0 and $\alpha_0 \in (0, \overline{\alpha})$, with $l_t^*(\tau)$ defined in (A.6) and $\mathcal{V}(\tau_0) = \{\tau \in \Theta_\tau : \|\tau - \tau_0\| < \varepsilon\}$ for some small $\varepsilon > 0$,*

$$E\left[\sup_{\tau \in \mathcal{V}(\tau_0)} \|\partial^2 l_t^*(\tau)/\partial\tau\partial\tau'\|\right] < \infty. \tag{A.27}$$

Proof. For $\tau_i, \tau_j \in \{\gamma, \alpha, \omega, \beta\}$,

$$\begin{aligned}\frac{\partial^2 l_t^*(\tau)}{\partial\tau_i\partial\tau_j} &= \frac{-1}{2}\left(1 - \frac{y_t^2}{h_t^*(\tau)}\right)\left(\frac{\partial^2 h_t^*(\tau)/\partial\tau_i\partial\tau_j}{h_t^*(\tau)}\right) \\ &\quad - \frac{1}{2}\left(2\frac{y_t^2}{h_t^*(\tau)} - 1\right)\left(\frac{\partial h_t^*(\tau)/\partial\tau_i}{h_t^*(\tau)}\right)\left(\frac{\partial h_t^*(\tau)/\partial\tau_j}{h_t^*(\tau)}\right),\end{aligned} \tag{A.28}$$

so we seek to show that

$$E\left[\sup_{\tau \in \mathcal{V}(\tau_0)} \left|\left(\frac{y_t^2}{h_t^*(\tau)}\right)\left(\frac{\partial^2 h_t^*(\tau)/\partial\tau_i\partial\tau_j}{h_t^*(\tau)}\right)\right|\right] < \infty$$

and

$$E\left[\sup_{\tau \in \mathcal{V}(\tau_0)} \left|\left(\frac{y_t^2}{h_t^*(\tau)}\right)\left(\frac{\partial h_t^*(\tau)/\partial\tau_i}{h_t^*(\tau)}\right)\left(\frac{\partial h_t^*(\tau)/\partial\tau_j}{h_t^*(\tau)}\right)\right|\right] < \infty.$$

By Hölder's inequality, it suffices to show that for some $p, q > 1$ satisfying $q^{-1} + p^{-1} = 1$,

$$E\left[\sup_{\tau \in \mathcal{V}(\tau_0)} \left|\frac{y_t^2}{h_t^*(\tau)}\right|^q\right] < \infty, \tag{A.29}$$

and

$$E\left[\sup_{\tau \in \mathcal{V}(\tau_0)} \left|\left(\frac{\partial h_t^*(\tau)/\partial\tau_i}{h_t^*(\tau)}\right)\left(\frac{\partial h_t^*(\tau)/\partial\tau_j}{h_t^*(\tau)}\right)\right|^p\right] < \infty, \quad E\left[\sup_{\tau \in \mathcal{V}(\tau_0)} \left|\frac{\partial^2 h_t^*(\tau)/\partial\tau_i\partial\tau_j}{h_t^*(\tau)}\right|^p\right] < \infty. \tag{A.30}$$

Note that

$$\frac{y_t^2}{h_t^*(\tau)} = z_t^2 \frac{h_t^*(\tau_0)}{h_t^*(\tau)}. \tag{A.31}$$

Choosing $\mathcal{V}(\tau_0)$ such that α is bounded away from zero on $\mathcal{V}(\tau_0)$, by Francq and Zakoïan (2010, p.164),

$$\left|\frac{h_t^*(\tau_0)}{h_t^*(\tau)}\right| \leq C + \frac{\alpha_0}{\alpha}\sum_{i=0}^{\infty}\frac{\beta_0^i \alpha y_{t-1-i}^2}{\omega + \beta^i \alpha y_{t-1-i}^2}.$$

Clearly, if $\beta_0 = 0$, $|h_t^*(\tau_0)/h_t^*(\tau)| \leq C$, uniformly on $\mathcal{V}(\tau_0)$. Hence, with $\delta > 0$ defined in Assumption 2.6, using (A.31) and Assumption 2.2

$$E\left[\sup_{\tau\in\mathcal{V}(\tau_0)}\left|\frac{y_t^2}{h_t^*(\tau)}\right|^{(1+\delta)/\delta}\right] < \infty, \quad \text{if } \beta_0 = 0.$$

If $\beta_0 > 0$ we may choose $\mathcal{V}(\tau_0)$ such that β is bounded away from zero on $\mathcal{V}(\tau_0)$. In that case, using Francq and Zakoïan (2010, p.164), for $s \in (0,1)$

$$\left|\frac{h_t^*(\tau_0)}{h_t^*(\tau)}\right| \leq C + \frac{\alpha_0}{\alpha}\sum_{i=0}^{\infty}\left(\frac{\beta_0}{\beta^{1-s}}\right)^i\left(\frac{\alpha y_{t-1-i}^2}{\omega}\right)^s. \qquad \text{(A.32)}$$

In light of (A.31) and (A.32), choosing s sufficiently small and $\mathcal{V}(\tau_0)$ such that $\beta_0/\beta^{1-s} < 1$ uniformly on $\mathcal{V}(\tau_0)$ and $E[y_t^{2s(1+\delta)/\delta}] < \infty$, we have by Assumption 2.2 and repeated use of Minkowski's inequality,

$$E\left[\sup_{\tau\in\mathcal{V}(\tau_0)}\left|\frac{y_t^2}{h_t^*(\tau)}\right|^{(1+\delta)/\delta}\right] < \infty, \quad \text{if } \beta_0 > 0.$$

Hence, (A.29) holds for $q = (1+\delta)/\delta > 1$. Turning to (A.30), note that in particular

$$\frac{\partial h_t^*(\tau)/\partial\beta}{h_t^*(\tau)} \leq \omega^{-1}\sum_{i=1}^{\infty} i\beta^{i-1}(\alpha y_{t-1-i}^2 + \gamma x_{t-1-i}^2).$$

By Assumption 2.6 and Minkowski's inequality,

$$E\left[\sup_{\tau\in\mathcal{V}(\tau_0)}\left|\left(\frac{\partial h_t^*(\tau)/\partial\beta}{h_t^*(\tau)}\right)^2\right|^{1+\delta}\right] < \infty.$$

By similar arguments, we conclude that (A.30) holds for $p = 1+\delta$. ■

Lemma A.7 *Under Assumptions 2.1, 2.2, 2.4, 2.5, 2.6, H_0 and $\alpha_0 \in (0, \overline{\alpha})$, with $\tilde{\kappa}_4$ defined in (2.12), $\tilde{\kappa}_4 = \kappa_4 + o_p(1)$.*

Proof. By definition

$$\tilde{\kappa}_4 = \frac{1}{T}\sum_{t=1}^{T}\tilde{z}_t^4 + 1 - 2\frac{1}{T}\sum_{t=1}^{T}\tilde{z}_t^2, \qquad \text{(A.33)}$$

and we will start out by focusing on the first term. It holds that

$$\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{z}_t^4-\frac{1}{T}\sum_{t=1}^{T}z_t^4\right|\leq\frac{1}{T}\sum_{t=1}^{T}z_t^4\left|\left[\frac{h_t^*(\tau_0)}{h_t(\tilde{\tau})}\right]^2-1\right|$$

$$\leq\frac{1}{T}\sum_{t=1}^{T}z_t^4\left(X_{1t}+X_{2t}+X_{3t}+X_{4t}+X_{5t}\right),$$

where

$$X_{1t}=\left[\frac{h_t^*(\tau_0)-h_t^*(\tilde{\tau})}{h_t(\tilde{\tau})}\right]^2,\quad X_{2t}=\left[\frac{h_t^*(\tilde{\tau})-h_t(\tilde{\tau})}{h_t(\tilde{\tau})}\right]^2,\quad X_{3t}=2\frac{|h_t^*(\tau_0)-h_t^*(\tilde{\tau})|}{h_t(\tilde{\tau})}$$

$$X_{4t}=2\frac{|h_t^*(\tilde{\tau})-h_t(\tilde{\tau})|}{h_t(\tilde{\tau})},\quad\text{and}\quad X_{5t}=2\left[\frac{|h_t^*(\tau_0)-h_t^*(\tilde{\tau})|}{h_t(\tilde{\tau})}\right]\left[\frac{|h_t^*(\tilde{\tau})-h_t(\tilde{\tau})|}{h_t(\tilde{\tau})}\right].$$

Using that $h_t(\tilde{\tau})\geq\underline{\omega}$ on $\Theta_{\tau,0}$ and a Taylor expansion $T^{-1}\sum_{t=1}^{T}z_t^4X_{1t}\leq\underline{\omega}^{-2}\|\tilde{\tau}-\tau_0\|^2T^{-1}\sum_{t=1}^{T}z_t^4\sup_{\tau\in\Theta_{\tau,0}}\|\partial h_t^*(\tau)/\partial\tau\|^2$, where $\partial h_t^*(\tau)/\partial\tau$ is given in (A.8). By Assumption 2.6, and ULLN for ergodic processes, $T^{-1}\sum_{t=1}^{T}z_t^4$ $\times\sup_{\tau\in\Theta_{\tau,0}}\|\partial h_t^*(\tau)/\partial\tau\|^2\xrightarrow{p}E[z_t^4\sup_{\tau\in\Theta_{\tau,0}}\|\partial h_t^*(\tau)/\partial\tau\|^2]<\infty$. Using that $\tilde{\tau}=\tau_0+o_p(1)$, we have $T^{-1}\sum_{t=1}^{T}z_t^4X_{1t}=o_p(1)$. By a similar argument, we conclude that $T^{-1}\sum_{t=1}^{T}z_t^4X_{3t}=o_p(1)$. Noting that $h_t^*(\tau)-h_t(\tau)=\beta^t(\sum_{i=0}^{\infty}\beta^i\alpha y_{-1-i}^2)$, choosing $s\in(0,1)$ sufficiently small, using the c_r inequality and Assumption 2.6, $E[z_t^{4s}\sup_{\tau\in\Theta_{\tau,0}}|h_t^*(\tau)-h_t(\tau)|^{2s}]=O(\rho^t)$. Hence for any $\varepsilon>0$

$$\sum_{t=1}^{\infty}P\left(|z_t^4X_{2t}|>\varepsilon\right)\leq\sum_{t=1}^{\infty}P\left\{|\underline{\omega}^{-2}z_t^4\sup_{\tau\in\Theta_{\tau,0}}|h_t^*(\tau)-h_t(\tau)|^2|>\varepsilon\right\}$$

$$\leq\varepsilon^{-s}\sum_{t=1}^{\infty}E\left\{\underline{\omega}^{-2s}z_t^{4s}\sup_{\tau\in\Theta_{\tau,0}}|h_t^*(\tau)-h_t(\tau)|^{2s}\right\}<\infty.$$

By the Borel-Cantelli theorem and Cesaro's lemma, we have that $T^{-1}\sum_{t=1}^{T}z_t^4X_{2t}=o_p(1)$. By similar arguments, $T^{-1}\sum_{t=1}^{T}z_t^4X_{4t}=o_p(1)$ and $T^{-1}\sum_{t=1}^{T}z_t^4X_{5t}=o_p(1)$, and we have that $T^{-1}\sum_{t=1}^{T}\tilde{z}_t^4=T^{-1}\sum_{t=1}^{T}z_t^4+o_p(1)$. By the LLN, $T^{-1}\sum_{t=1}^{T}\tilde{z}_t^4=E[z_t^4]+o_p(1)$. By similar arguments $T^{-1}\sum_{t=1}^{T}\tilde{z}_t^2=E[z_t^2]+o_p(1)$, and in light of (A.33) we conclude that $\tilde{\kappa}_4=\kappa_4+o_p(1)$. ■

B Proofs of Theorems 3.1, 3.2, and 3.3

Proof of Theorem 3.1. The proof mimics Han and Kristensen (2014, Proof of Theorem 3) and arguments given in Straumann (2005, Proof of Theorem

6.1.1). We will focus on the consistency of $\hat{\theta}$ and note that the consistency of $\tilde{\theta}$ is proved by identical arguments.

First, we introduce the ergodic log-likelihood function

$$\mathcal{L}_T^*(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t^*(\theta), \quad l_t(\theta) = -\frac{1}{2}\log[h_t^*(\theta)] + \log\{g_\nu[y_t/\sqrt{h_t^*(\theta)}]\}, \tag{B.1}$$

$$h_t^*(\theta) = (1-\beta)\omega + \alpha y_{t-1}^2 + \beta h_{t-1}^*(\theta) + \gamma x_{t-1}^2, \quad t \in \mathbb{Z}. \tag{B.2}$$

Since Θ is compact and $\theta \mapsto l_t(\theta)$ is continuous almost surely on Θ, it suffices to show that (i) $\frac{1}{T}\sum_{t=1}^{T} l_t^*(\theta) = E[l_t^*(\theta)] + o_p(1)$, where $E[l_t^*(\theta)]$ exists for all $\theta \in \Theta$, (ii) $E[l_t^*(\theta_0)] > E[l_t^*(\theta)]$ for all $\theta \in \Theta \setminus \{\theta_0\}$, (iii) $E[\sup_{\theta\in\Theta} l_t^*(\theta)] < \infty$, and (iv) $\sup_{\theta\in\Theta} |\mathcal{L}_T^*(\theta) - \mathcal{L}_T(\theta)| = o_p(T)$.

(i) Follows by Assumption 2.1 and the ergodic theorem, provided that $E[l_t^*(\theta)]$ exists for all $\theta \in \Theta$. Note that by definition, $\log[h_t^*(\theta)] \geq \log(\underline{\omega})$, and hence $\sup_{\theta\in\theta} l_t^*(\theta) \leq C$ such that $E[l_t^*(\theta)^+] < \infty$ for all $\theta \in \Theta$. Turning to (ii), from Han and Kristensen (2014, Proof of Theorem 3), $E|\log[h_t^*(\theta_0)]| < \infty$, such that

$$E[|l_t^*(\theta_0)|] \leq E|\log[h_t^*(\theta_0)]| + E|\log g_{\nu_0}[z_t]| < \infty.$$

Next, following Straumann (2005, Proof of Theorem 6.1.1), let

$$f_t(\theta) = \frac{g_\nu[y_t/\sqrt{h_t^*(\theta)}]}{h_t^*(\theta)} \quad \text{and} \quad r_t(\theta) = \frac{\sqrt{h_t^*(\theta_0)}}{\sqrt{h_t^*(\theta)}}.$$

Using that $\log(x) \leq x - 1$ for all $x > 0$ with equality if and only if, it holds that

$$E[l_t^*(\theta)] - E[l_t^*(\theta_0)] = E\left[\log\frac{f_t(\theta)}{f_t(\theta_0)}\right] \leq E\left[\frac{f_t(\theta)}{f_t(\theta_0)}\right] - 1$$

with equality if and only if $f_t(\theta) = f_t(\theta_0)$ a.s. We have that

$$\frac{f_t(\theta)}{f_t(\theta_0)} = \frac{g_\nu[y_t/\sqrt{h_t^*(\theta)}]\sqrt{h_t^*(\theta_0)}}{g_{\nu_0}[y_t/\sqrt{h_t^*(\theta_0)}]\sqrt{h_t^*(\theta)}} = \frac{g_\nu[r_t(\theta)z_t]r_t(\theta)}{g_{\nu_0}[z_t]}.$$

Note that by Assumption 2.2, z_t and $r_t(\theta)$ are independent. Consider the conditional expectation of $f_t(\theta)/f_t(\theta_0)$ given $r_t(\theta)$:

$$\begin{aligned} E\left[\frac{f_t(\theta)}{f_t(\theta_0)}|r_t(\theta)\right] &= \int \frac{g_\nu[r_t(\theta)z]r_t(\theta)}{g_{\nu_0}[z]} g_{\nu_0}[z]dz \\ &= \int \frac{g_\nu[r_t(\theta)z]r_t(\theta)}{g_{\nu_0}[z]} g_{\nu_0}[z]dz \\ &= \int g_\nu[r_t(\theta)z]r_t(\theta)dz. \end{aligned}$$

Using that $g_\nu[r_t(\theta)z]r_t(\theta_0)$ is the (conditional) density of $z_t/r_t(\theta)$, we conclude that $E\left[f_t(\theta)/f_t(\theta_0)|r_t(\theta)\right] = 1$, such that $E\left[f_t(\theta)/f_t(\theta_0)\right] = 1$. Hence,

$$E[l_t^*(\theta)] \leq E[l_t^*(\theta_0)]$$

with equality if and only if $f_t(\theta) = f_t(\theta_0)$ a.s. So it remains to show that $f_t(\theta) = f_t(\theta_0)$ a.s. implies $\theta = \theta_0$. Observe that $f_t(\theta) = f_t(\theta_0)$ a.s. if and only if

$$g_{\nu_0}[z_t] = g_\nu[r_t(\theta)z_t]r_t(\theta) \quad \text{a.s.} \tag{B.3}$$

Suppose that $(\gamma, \alpha, \omega, \beta) \neq (\gamma_0, \alpha_0, \omega_0, \beta_0)$, then by Assumption 2.4 and arguments given in Han and Kristensen (2014, Proof of Theorem 3),

$$P(r_t(\theta) \neq 1) > 0.$$

By Straumann (2005, Lemma 6.1.2), $P\{g_{\nu_0}[z_t] \neq g_\nu[r_t(\theta)z_t]r_t(\theta)\} > 0$, which contradicts (B.3), so necessarily we must have that $(\gamma, \alpha, \omega, \beta) = (\gamma_0, \alpha_0, \omega_0, \beta_0)$. In light of (B.3), using that necessarily $r_t(\theta) = 1$, it remains to show that $g_{\nu_0}[z_t] = g_\nu[z_t]$ a.s. implies that $\nu = \nu_0$, which is trivial. We conclude that (ii) holds.

(iii) holds by the arguments given in order to establish (i).

Lastly, (iv) is shown by arguments similar to the ones given in Han and Kristensen (2014, Proof of Theorem 3). ■

Proof of Theorem 3.2. We show that the conditions of Lemma 2.1 apply. Due to Theorem 3.1, we have that $\hat{\theta}$ and $\tilde{\theta}$ are consistent for θ, and hence condition (A.i) of Lemma 2.1 is satisfied. For brevity, we focus on establishing (A.ii)-(A.iii) of Lemma 2.1 for the ergodic log-likelihood function in (B.1)-(B.2).

In order show (A.ii), the asymptotic normality of the score is established, using the martingale CLT by Brown (1971). We will rely on some results from the supplementary material to Pedersen and Rahbek (2016). The score contributions are

$$\frac{\partial l_t^*(\theta)}{\partial \theta_i} = \frac{1}{2}\left[\frac{(\nu+1)\, y_t^2/h_t^*(\theta)}{(\nu-2) + y_t^2/h_t^*(\theta)} - 1\right]\frac{\partial h_t^*(\theta)/\partial \theta_i}{h_t^*(\theta)} \quad \text{for } \theta_i \in \{\gamma, \alpha, \omega\},$$

and

$$\frac{\partial l_t^*(\theta)}{\partial \nu} = \frac{\partial \log \eta(\nu)}{\partial \nu} - \tfrac{1}{2}\log\left(1 + \frac{y_t^2/h_t^*(\theta)}{\nu-2}\right) + \frac{1}{2(\nu-2)}\left[\frac{(\nu+1)\, y_t^2/h_t^*(\theta)}{(\nu-2) + y_t^2/h_t^*(\theta)} - 1\right].$$

Consider the score contribution at θ_0, $\mathcal{S}_t = (s_{t,\gamma}, s_{t,\alpha}, s_{t,\omega}, s_{t,\beta}, s_{t,\nu})'$, where

$$s_{t,\gamma} = \frac{\partial l_t^*(\theta_0)}{\partial \gamma}, \quad s_{t,\alpha} = \frac{\partial l_t^*(\theta_0)}{\partial \alpha}, \quad s_{t,\omega} = \frac{\partial l_t^*(\theta_0)}{\partial \omega}, \quad s_{t,\beta} = \frac{\partial l_t^*(\theta_0)}{\partial \beta}, \quad s_{t,\nu} = \frac{\partial l_t^*(\theta_0)}{\partial \nu}.$$

For $\theta_i \in \{\gamma, \alpha, \omega, \beta\}$,

$$s_{t,\theta_i} = \frac{1}{2} z_{1t}^* \frac{\partial h_t^*(\theta_0)/\partial\theta_i}{h_t^*(\theta_0)} \quad \text{and} \quad s_{t,\nu} = z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0 - 2)}, \tag{B.4}$$

where

$$z_{1t}^* = \left[\frac{(\nu_0+1)\, z_t^2}{(\nu_0 - 2) + z_t^2} - 1\right] \quad \text{and} \quad z_{3t}^* = \left[\frac{\partial \log \eta(\nu_0)}{\partial \nu} - \tfrac{1}{2} \log\left(1 + \frac{z_t^2}{\nu_0 - 2}\right)\right].$$

From Pedersen and Rahbek (2016, Lemma A.5 in the supplementary material), $E[z_{1t}^*] = E[z_{3t}^*] = 0$ and $E[z_{1t}^{*2}], E[z_{3t}^{*2}] < \infty$. Hence, using that $E[\|(y_t, x_t)'\|^4] < \infty$ and Assumption 2.2, we have that $\mathcal{S}_t$ is a martingale difference sequence with respect to $\mathcal{F}_t$ with $E[\|\mathcal{S}_t\|^2] < \infty$. Then using the ergodic theorem $\frac{1}{T}\sum_{t=1}^T E[(k'\mathcal{S}_t)^2|\mathcal{F}_{t-1}] \xrightarrow{p} k'\Sigma k < \infty$ for any $k \in \mathbb{R}^5 \setminus \{0\}$ and some constant matrix Σ. Moreover, using the ergodic theorem, for any $\delta > 0$ and any $k \in \mathbb{R}^5$, $\frac{1}{T}\sum_{t=1}^T E[(k'\mathcal{S}_t)^2 \mathbf{1}_{(|k'\mathcal{S}_t| > T^{1/2}\delta)}] = o_p(1)$, verifying the Lindeberg condition. It remains to show that $k'\Sigma k > 0$ for any $k \in \mathbb{R}^5 \setminus \{0\}$, i.e. that Σ is positive definite. We note that $\Sigma = E[\mathcal{S}_t \mathcal{S}_t']$, so Σ is positive semi-definite. Following Straumann (2005, proof of Lemma 6.3.2), suppose that there exists $k = (k_1, k_2, k_3, k_4, k_5)' \in \mathbb{R}^5$ such that $k'\Sigma k = 0$, which is equivalent to

$$k'\mathcal{S}_t = k_1 s_{t,\gamma} + k_2 s_{t,\alpha} + k_3 s_{t,\omega} + k_4 s_{t,\beta} + k_5 s_{t,\nu} = 0 \quad a.s. \tag{B.5}$$

We will argue that it cannot be the case that $k \neq 0$. Suppose that $(k_1, k_2, k_3, k_4)' = 0$ and $k_5 \neq 0$. Then $k'\Sigma k = k_5^2 E[s_{t,\nu}^2]$. From Pedersen and Rahbek (2016, proof of Lemma A.1 in the supplementary material), with $\psi'(\cdot)$ the trigamma function,

$$E[s_{t,\nu}^2] = \frac{1}{4}\left[\psi'\left(\frac{\nu_0}{2}\right) - \psi'\left(\frac{\nu_0+1}{2}\right)\right] + \frac{6}{(\nu_0-2)^2(\nu+1)(\nu_0+3)} > 0,$$

which contradicts $k'\Sigma k = 0$. Suppose that $(k_1, k_2, k_3, k_4)' \neq 0$ and $k_5 = 0$. Using (B.4), that $P(z_{1t}^* \neq 0) = 1$, and that $P[h_t^*(\theta_0) > 0] = 1$, we have that (B.5) is equivalent to

$$k_1 \sum_{i=0}^\infty \beta^i x_{t-1-i}^2 + k_2 \sum_{i=0}^\infty \beta^i y_{t-1-i}^2 + k_3 + k_4 \sum_{i=1}^\infty i\beta^{i-1}(\alpha_0 y_{t-1-i}^2 + \gamma_0 x_{t-1-i}^2) = 0 \quad a.s.,$$

which is ruled out by Assumption 2.4, using that $(\alpha_0, \gamma_0) \neq 0$. Lastly, suppose that $(k_1, k_2, k_3, k_4)' \neq 0$ and $k_5 \neq 0$. Again, using that $P(z_{1t}^* \neq 0) = 1$ and

$P[\sigma_t^2(\theta_0) > 0] = 1$, (B.5) is equivalent to

$$k_1 \sum_{i=0}^{\infty} \beta^i x_{t-1-i}^2 + k_2 \sum_{i=0}^{\infty} \beta^i y_{t-1-i}^2 + k_3 + k_4 \sum_{i=1}^{\infty} i\beta^{i-1}(\alpha_0 y_{t-1-i}^2 + \gamma_0 x_{t-1-i}^2)$$
$$= 2\frac{z_{3t}^*}{z_{1t}^*} + \frac{1}{(\nu_0 - 2)} \quad a.s,$$

which contradicts the fact that z_{3t}^*/z_{1t}^* is non-degenerate and that z_{3t}^*/z_{1t}^* and $\mathcal{F}_{t-1}$ are independent. We conclude that $k'\Sigma k > 0$ for any $k \in \mathbb{R}^5 \backslash \{0\}$. Using (B.8) and the ergodic theorem, $-\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 l_t^*(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} J$, and we conclude that condition (A.ii) is satisfied.

In order to establish (A.iii), we consider the second derivative of the log-likelihood contribution. With $\theta_i, \theta_j \in \{\omega, \alpha, \gamma, \beta\}$,

$$\frac{\partial^2 l_t^*(\theta)}{\partial\theta_i\partial\theta_j} = \frac{1}{2}\left[1 - \frac{(\nu+1)\, y_t^2/h_t^*(\theta)}{(\nu-2) + y_t^2/h_t^*(\theta)} - \frac{(\nu+1)(\nu-2)\, y_t^2/h_t^*(\theta)}{\left[(\nu-2) + y_t^2/h_t^*(\theta)\right]^2}\right]$$
$$\times \left(\frac{\partial h_t^*(\theta)/\partial\theta_i}{h_t^*(\theta)}\right)\left(\frac{\partial h_t^*(\theta)/\partial\theta_j}{h_t^*(\theta)}\right)$$
$$+ \frac{1}{2}\left[\frac{(\nu+1)\, y_t^2/h_t^*(\theta)}{(\nu-2) + y_t^2/h_t^*(\theta)} - 1\right]\left(\frac{\partial^2 h_t^*(\theta)/\partial\theta_i\partial\theta_j}{h_t^*(\theta)}\right)$$

$$\frac{\partial^2 l_t^*(\theta)}{\partial\nu^2} = \frac{\partial^2 \log\eta(\nu)}{\partial\nu\partial\nu} + \frac{1}{(\nu-2)}\frac{y_t^2/h_t^*(\theta)}{(\nu-2) + y_t^2/h_t^*(\theta)} - \frac{1}{2(\nu-2)^2}\left[\frac{(\nu+1)\, y_t^2/h_t^*(\theta)}{(\nu-2) + y_t^2/h_t^*(\theta)} - 1\right]$$
$$- \left(\frac{1}{2(\nu-2)}\right)\frac{(\nu+1)\, y_t^2/h_t^*(\theta)}{\left[(\nu-2) + y_t^2/h_t^*(\theta)\right]^2},$$

and

$$\frac{\partial^2 l_t^*(\theta)}{\partial\theta_i\partial\nu} = \frac{1}{2}\left[\frac{y_t^2/h_t^*(\theta)}{(\nu-2) + y_t^2/h_t^*(\theta)} - \frac{(\nu+1)\, y_t^2/h_t^*(\theta)}{\left[(\nu-2) + y_t^2/h_t^*(\theta)\right]^2}\right]\frac{\partial h_t^*(\theta)/\partial\theta_i}{h_t^*(\theta)}.$$

It holds that

$$V_t(\theta) := \left(\frac{\partial h_t^*(\theta)}{\partial\gamma}, \frac{\partial h_t^*(\theta)}{\partial\alpha}, \frac{\partial h_t^*(\theta)}{\partial\omega}, \frac{\partial h_t^*(\theta)}{\partial\beta}\right)'$$
$$= \left(\sum_{i=0}^{\infty} \beta^i x_{t-1-i}^2, \sum_{i=0}^{\infty} \beta^i y_{t-1-i}^2, 1, \sum_{i=1}^{\infty} i\beta^{i-1}(\alpha y_{t-1-i}^2 + \gamma x_{t-1-i}^2)\right)',$$

$$\frac{\partial^2 l_t^*(\theta)}{\partial\theta_i\partial\theta_j} = 0 \text{ for } \theta_i, \theta_j \in \{\omega, \alpha, \gamma\},$$

and

$$\frac{\partial^2 l_t^*(\theta)}{\partial\gamma\partial\beta} = \sum_{i=1}^{\infty} i\beta^{i-1}x_{t-1-i}^2, \quad \frac{\partial^2 l_t^*(\theta)}{\partial\alpha\partial\beta} = \sum_{i=1}^{\infty} i\beta^{i-1}y_{t-1-i}^2, \quad \frac{\partial^2 l_t^*(\theta)}{\partial\omega\partial\beta} = 0,$$
$$\frac{\partial^2 l_t^*(\theta)}{\partial\beta\partial\beta} = \sum_{i=2}^{\infty} i(i-1)\beta^{i-2}x_{t-1-i}^2.$$

For $\theta_i, \theta_j \in \{\omega, \alpha, \gamma, \beta\}$, on Θ,

$$\left|\frac{\partial^2 l_t^*(\theta)}{\partial\theta_i\partial\theta_j}\right| \leq \frac{1}{2}[1 + 2(\overline{\nu}+1)]\frac{1}{\underline{\omega}^2}|\partial h_t^*(\theta)/\partial\theta_i||\partial h_t^*(\theta)/\partial\theta_j|$$
$$+ \frac{1}{2}[\overline{\nu}+2]|\partial^2 h_t^*(\theta)/\partial\theta_i\partial\theta_2|.$$

and

$$\left|\frac{\partial^2 l_t^*(\theta)}{\partial\theta_i\partial\nu}\right| \leq \frac{1}{2}\left[1 + \frac{(\overline{\nu}+1)}{(\underline{\nu}-2)}\right]\underline{\omega}^{-1}|\partial\sigma_t^2(\theta)/\partial\theta_i|$$

Hence, using that $E[\|(y_t, x_t)'\|^4] < \infty$ and that $|\beta| < 1$ uniformly on Θ,

$$E[\sup_{\theta\in\Theta}|\partial^2 l_t^*(\theta)/\partial\theta_i\partial\theta_j|] < \infty \text{ and } E[\sup_{\theta\in\Theta}|\partial^2 l_t^*(\theta)/\partial\theta_i\partial\nu|] < \infty, \quad \text{for} \quad \theta_i, \theta_j \in \{\omega, \alpha, \gamma, \beta\}. \tag{B.6}$$

Moreover, on Θ,

$$\left|\frac{\partial^2 l_t^*(\theta)}{\partial\nu^2}\right| \leq \sup_{\nu\in[\underline{\nu},\overline{\nu}]}\frac{\partial^2\log\eta(\nu)}{\partial\nu\partial\nu} + \frac{1}{(\underline{\nu}-2)} + \frac{\overline{\nu}+2}{2(\underline{\nu}-2)^2} + \left(\frac{1}{2(\underline{\nu}-2)}\right)\frac{(\overline{\nu}+1)}{(\underline{\nu}-2)} < \infty,$$

and hence

$$E\left[\sup_{\theta\in\Theta}\left|\frac{\partial^2 l_t^*(\theta)}{\partial\nu^2}\right|\right] < \infty. \tag{B.7}$$

By (B.6)-(B.7),

$$E[\sup_{\theta\in\Theta}\|\partial^2 l_t^*(\theta)/\partial\theta\partial\theta'\|] < \infty. \tag{B.8}$$

Using (B.8) and ULLN for ergodic processes, we conclude that condition (A.iii) holds.

Next, for the case $\beta_0 > 0$, there are no nuisance parameters on the boundary, and the limiting distribution of the LR statistic is immediate from Lemma 2.1. We then turn to the case $\beta_0 = 0$. Using that $J = \Sigma = E[\mathcal{S}_t\mathcal{S}_t']$ and that $h_t^*(\theta_0) = \omega_0 + \alpha_0 y_{t-1}^2$ under H_0, $J = E[\mathcal{S}_t\mathcal{S}_t']$, where

$$\mathcal{S}_t = \left(\frac{z_{1t}^* x_{t-1}^2}{2\sigma_t^2(\theta_0)}, \frac{z_{1t}^* y_{t-1}^2}{2\sigma_t^2(\theta_0)}, \frac{z_{1t}^*}{2\sigma_t^2(\theta_0)}, \frac{\alpha_0 z_{1t}^* y_{t-2}^2}{2\sigma_t^2(\theta_0)}, z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0-2)}\right)'.$$

Hence, using that $E[\|(y_t, x_t)'\|^4] < \infty$ and $E[z_{1t}^{*2}] = 2\nu_0/(\nu_0+3)$, $E[z_{1t}^* z_{3t}^*] = -(\nu_0+1)^{-1}$ (Pedersen and Rahbek, 2016, Lemma A.5 in the supplementary material),

$$
\begin{aligned}
J &= E\begin{pmatrix}
\frac{z_{1t}^{*2}x_{t-1}^4}{4h_t^{*2}(\theta_0)} & \frac{z_{1t}^{*2}x_{t-1}^2 y_{t-1}^2}{4h_t^{*2}(\theta_0)} & \frac{z_{1t}^{*2}x_{t-1}^2}{4h_t^{*2}(\theta_0)} & \frac{\alpha_0 z_{1t}^{*2}x_{t-1}^2 y_{t-2}^2}{4h_t^{*2}(\theta_0)} & \frac{z_{1t}^{*}x_{t-1}^2}{2h_t^{*}(\theta_0)}\left(z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0-2)}\right) \\
 & \frac{z_{1t}^{*2}y_{t-1}^4}{4h_t^{*2}(\theta_0)} & \frac{z_{1t}^{*2}y_{t-1}^2}{4h_t^{*2}(\theta_0)} & \frac{\alpha_0 z_{1t}^{*2}y_{t-1}^2 y_{t-2}^2}{4h_t^{*2}(\theta_0)} & \frac{z_{1t}^{*}y_{t-1}^2}{2h_t^{*}(\theta_0)}\left(z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0-2)}\right) \\
 & & \frac{z_{1t}^{*2}}{4h_t^{*2}(\theta_0)} & \frac{\alpha_0 z_{1t}^{*2}y_{t-2}^2}{4h_t^{*2}(\theta_0)} & \frac{z_{1t}^{*}}{2h_t^{*}(\theta_0)}\left(z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0-2)}\right) \\
 & & & \frac{\alpha_0^2 z_{1t}^{*2}y_{t-2}^4}{4h_t^{*2}(\theta_0)} & \frac{\alpha_0 z_{1t}^{*}y_{t-2}^2}{2h_t^{*}(\theta_0)}\left(z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0-2)}\right) \\
 & & & & \left(z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0-2)}\right)^2
\end{pmatrix} \\
&= \begin{pmatrix}
\eta a\kappa_{4,x} & \eta b\kappa_{2,x} & \eta a\kappa_{2,x} & \eta d\kappa_{2,x} & j\phi\kappa_{2,x} \\
\eta b\kappa_{2,x} & \eta c & \eta b & \eta f & k\phi \\
\eta a\kappa_{2,x} & \eta b & \eta a & \eta d & j\phi \\
\eta d\kappa_{2,x} & \eta f & \eta d & \eta g & m\phi \\
j\phi\kappa_{2,x} & k\phi & j\phi & m\phi & \xi
\end{pmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
&a = E\left[\frac{1}{4(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad b = E\left[\frac{y_{t-1}^2}{4(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad c = E\left[\frac{y_{t-1}^4}{4(\omega_0+\alpha_0 y_{t-1}^2)^2}\right] \\
&d = E\left[\frac{\alpha_0 y_{t-2}^2}{4(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad f = E\left[\frac{\alpha_0 y_{t-1}^2 y_{t-2}^2}{4(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \quad g = E\left[\frac{\alpha_0^2 y_{t-2}^4}{4(\omega_0+\alpha_0 y_{t-1}^2)^2}\right], \\
&j = E\left[\frac{1}{2(\omega_0+\alpha_0 y_{t-1}^2)}\right], \quad k = E\left[\frac{y_{t-1}}{2(\omega_0+\alpha_0 y_{t-1}^2)}\right], \quad m = E\left[\frac{\alpha_0 y_{t-2}}{2(\omega_0+\alpha_0 y_{t-1}^2)}\right] \\
&\phi = \{\nu_0/[(\nu_0+3)(\nu_0-2)] - 1/(\nu_0+1)\}, \quad \xi = E\left[\left(z_{3t}^* + \frac{z_{1t}^\star}{2(\nu_0-2)}\right)^2\right], \quad \eta = E[z_{1t}^{*2}]
\end{aligned}
$$

$\kappa_{2,x} = E[x_t^2]$, and $\kappa_{4,x} = E[x_t^4]$.

It holds that

$$
J^{-1} = \begin{pmatrix}
\frac{-1}{a\eta\kappa_{2,x}^2 - a\eta\kappa_{4,x}} & 0 & \frac{1}{a\eta\kappa_{2,x}^2 - a\eta\kappa_{4,x}}\kappa_{2,x} & 0 & 0 \\
0 & & & & \\
\frac{1}{a\eta\kappa_{2,x}^2 - a\eta\kappa_{4,x}}\kappa_{2,x} & & \mathcal{J}_{22} & & \\
0 & & & & \\
0 & & & &
\end{pmatrix},
$$

with some positive definite 4×4 matrix $\mathcal{J}_{22}$. Hence, condition (A.iv) of Lemma 2.1 holds, and the limiting distribution of the LR statistic then follows by Lemma 2.1. ■

Proof for Theorem 3.3. We show that the conditions of Lemma 2.1 are satisfied. Under Assumptions 3.1, 3.2, and 3.4, the consistency of $\hat{\theta}$ and $\tilde{\theta}$ holds by Agosto et al. (2016, Theorem 2), and hence condition (A.i) of Lemma 2.1 is satisfied. For brevity, we focus on establishing conditions (A.ii)-(A.iii) of Lemma 2.1 for the ergodic log-likelihood function,

$$\begin{aligned}\mathcal{L}_T^*(\theta) &= \sum_{t=1}^{T} l_t^*(\theta), \quad l_t^*(\theta) = y_t \log[\lambda_t^*(\theta)] - \lambda_t^*(\theta),\\ \lambda_t^*(\theta) &= (1-\beta)\omega + \alpha y_{t-1} + \beta\lambda_{t-1}^*(\theta) + \gamma f(x_{t-1}), \quad t \in \mathbb{Z}.\end{aligned}$$

Consider the score contribution, $\partial l_t^*(\theta_0)/\partial\theta = (y_t/\lambda_t^*(\theta_0) - 1)(\partial\lambda_t^*(\theta_0)/\partial\theta)$. By Agosto et al. (2016, Section A.4.1), $T^{-1/2}\sum_{t=1}^{T}\partial l_t^*(\theta_0)/\partial\theta \xrightarrow{w} N(0,J)$, with $J = -E[\partial^2 l_t^*(\theta_0)/\partial\theta\partial\theta']$, which is positive definite due to Agosto et al. (2016, Section A.4.2), using Assumption 3.4. Moreover, due to Agosto et al. (2016, Section A.4.2), $T^{-1}\sum_{t=1}^{T}\partial^2 l_t^*(\theta_0)/\partial\theta\partial\theta' = -J + o_p(1)$. We conclude that condition (A.ii) of Lemma 2.1 holds.

Turning to condition (A.iii), consider the second derivative of the log-likelihood contribution,

$$-\frac{\partial^2 l_t^*(\theta)}{\partial\theta\partial\theta'} = \frac{y_t}{\lambda_t^{*2}(\theta)}\frac{\partial\lambda_t^*(\theta)}{\partial\theta}\frac{\partial\lambda_t^*(\theta)}{\partial\theta'} - \left(\frac{y_t}{\lambda_t^*(\theta)} - 1\right)\frac{\partial\lambda_t^*(\theta)}{\partial\theta\partial\theta'},$$

where $\partial\lambda_t^*(\theta)/\partial\theta = \{\sum_{i=0}^{\infty}\beta^i f(x_{t-1-i}), \sum_{i=0}^{\infty}\beta^i y_{t-1-i}, 1, \sum_{i=1}^{\infty} i\beta^{i-1}[\alpha y_{t-1-i} + \gamma f(x_{t-1-i})]\}'$, and

$$\frac{\partial\lambda_t^*(\theta)}{\partial\theta\partial\theta'}\begin{pmatrix} 0 & 0 & 0 & \sum_{i=1}^{\infty} i\beta^{i-1} f(x_{t-1-i}) \\ & 0 & 0 & \sum_{i=1}^{\infty} i\beta^{i-1} y_{t-1-i} \\ & & 0 & 0 \\ & & & \sum_{i=2}^{\infty} i(i-1)\beta^{i-2}[\alpha y_{t-1-i} + \beta f(x_{t-1-i})] \end{pmatrix}.$$

Since, $\lambda_t^*(\theta) \geq \underline{\omega} > 0$ on Θ and $\sup_{\theta\in\Theta}\beta \leq \overline{\beta} < 1$, it holds that $E[\sup_{\theta\in\Theta}\|\partial^2 l_t^*(\theta)/\partial\theta\partial\theta'\|]$ $< \infty$ due to Assumption 3.1 and Hölder's inequality. By the ULLN for ergodic processes, we then have that condition (A.iii) of Lemma 2.1 is satisfied.

For the case $\beta_0 > 0$, the limiting distribution of LR_T follows directly from

Lemma 2.1. In the case $\beta_0 = 0$, $\lambda_t^*(\theta_0) = \omega_0 + \alpha_0 y_{t-1}$. Here

$$
\begin{aligned}
J &= -E\left[\frac{\partial^2 l_t^*(\theta_0)}{\partial\theta\partial\theta'}\right] \\
&= E\left[\frac{1}{\omega_0 + \alpha_0 y_{t-1}}\begin{pmatrix} f^2(x_{t-1}) & y_{t-1}f(x_{t-1}) & f(x_{t-1}) & \alpha_0 f(x_{t-1})y_{t-2} \\ & y_{t-1}^2 & y_{t-1} & \alpha_0 y_{t-1}y_{t-2} \\ & & 1 & \alpha_0 y_{t-2} \\ & & & \alpha_0^2 y_{t-2}^2 \end{pmatrix}\right] \\
&= \begin{pmatrix} a\kappa_{2,x} & b\kappa_{1,x} & a\kappa_{1,x} & d\kappa_{1,x} \\ b\kappa_{1,x} & c & b & f \\ a\kappa_{1,x} & b & a & d \\ d\kappa_{1,x} & f & d & g \end{pmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
a &= E\left[\frac{1}{\omega_0 + \alpha_0 y_{t-1}}\right], \quad b = E\left[\frac{y_{t-1}}{\omega_0 + \alpha_0 y_{t-1}}\right], \quad c = E\left[\frac{y_{t-1}^2}{\omega_0 + \alpha_0 y_{t-1}}\right] \\
d &= E\left[\frac{\alpha_0 y_{t-2}}{\omega_0 + \alpha_0 y_{t-1}}\right], \quad f = E\left[\frac{\alpha_0 y_{t-1} y_{t-2}}{\omega_0 + \alpha_0 y_{t-1}}\right], \quad g = E\left[\frac{\alpha_0^2 y_{t-2}^2}{\omega_0 + \alpha_0 y_{t-1}}\right], \\
\kappa_{1,x} &= E[f(x_t)], \quad \text{and} \quad \kappa_{2,x} = E[f^2(x_t)],
\end{aligned}
$$

and where we have used Assumption 3.3. It holds that

$$
J^{-1} = \begin{pmatrix} \frac{1}{a\kappa_{2,x} - a\kappa_{1,x}^2} & 0 & \frac{-\kappa_{1,x}}{a\kappa_{2,x} - a\kappa_{1,x}^2} & 0 \\ 0 & \frac{d^2 - ag}{\zeta} & \frac{bg - df}{\zeta} & \frac{af - bd}{\zeta} \\ \frac{-\kappa_{1,x}}{a\kappa_{2,x} - a\kappa_{1,x}^2} & \frac{bg - df}{\zeta} & \xi & \frac{cd - bf}{\zeta} \\ 0 & \frac{af - bd}{\zeta} & \frac{cd - bf}{\zeta} & \frac{b^2 - ac}{\zeta} \end{pmatrix},
$$

with $\zeta := gb^2 - 2bdf + cd^2 + af^2 - acg$ and

$$
\xi := \frac{gb^2\kappa_{1,x}^2 - 2bdf\kappa_{1,x}^2 + cd^2\kappa_{1,x}^2 + a\kappa_{2,x}f^2 - acg\kappa_{2,x}}{-a\left(\kappa_{1,x}^2 - \kappa_{2,x}\right)\left(gb^2 - 2bdf + cd^2 + af^2 - acg\right)}.
$$

We note that $(J^{-1})_{\gamma,\beta} = 0$, and hence the limiting distribution of LR_T follows by Lemma 2.1. ■