Asymptotic normality of the QMLE of possibly nonstationary GARCH with serially dependent innovations[*]

Christian M. Dahl[†]
Department of Economics
Purdue University

Emma M. Iglesias
Department of Economics
Michigan State University

May 19, 2006

Abstract

This paper proposes a new parametric volatility model that introduces serially dependent innovations in GARCH specifications. We first prove the asymptotic normality of the QML estimator in this setting, allowing for possible explosive and nonstationary behavior of the GARCH process. We show that this model can generate an alternative measure of risk premium relative to the GARCH-M. Finally, we provide evidence of the usefulness and advantages of our approach relative to competing volatility models through a Monte Carlo experiment and by an application to US treasury bill spot rates. In particular, we illustrate the consequences of dynamic misspecification and demonstrate that the new volatility model can improve upon the fit in-sample as well as out-of-sample relative to traditional GARCH-type specifications.

1 Introduction

Since Engle (1982) proposed the ARCH (autoregressive conditional heteroskedastic) model, and Bollerslev (1986) generalized it to GARCH, there has been a substantial interest in conditional volatility modelling. If y_t denotes the time series of interest, and its conditional variance is given by σ_t^2, the GARCH(1,1) model is specified as

$$y_t\left(\theta^*\right) = \sigma_t \epsilon_t; \qquad \sigma_t^2\left(\theta^*\right) = w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2\left(\theta^*\right),$$

---

[*] Comments from Jeffrey Wooldridge, Nour Meddahi and participants at the conference: *Symposium on Econometric Theory and Applications*, Taipei, Taiwan 2005, are gratefully acknowledged.

[†] Corresponding author. Address: 403 West State Street, Purdue University, West Lafayette, IN 47907-2056. E-mail: dahlc@mgmt.purdue.edu. Phone: 765-494-4503. Fax: 765-496-1778.

with $\theta^* = (w, \alpha, \beta, \sigma_0^2(\theta^*))'$. Following Drost and Nijman (1993), GARCH models can be classified into three categories: strong GARCH (where the innovation process ϵ_t is i.i.d. with zero mean and unit variance); the semi-strong form (where the innovation process is a martingale difference sequence), and finally the weak form, where also the martingale difference sequence assumption is relaxed.

Asymptotic normality of the Quasi Maximum Likelihood (QML) estimator has been established under a wide range of alternative assumptions for strong-GARCH models (see e.g. Lumsdaine (1996)). One of the most recent contribution is by Jensen and Rahbek (2004a, 2004b), who show that the QMLE is asymptotically normal both in the stationary and the nonstationary region. Lee and Hansen (1994) shows the asymptotic normality of the QMLE for the semi-strong form. In relation to weak-form GARCH models, the asymptotic normality of the QML estimator has not yet been established (see, for example, Linton and Mammen (2005, page 787), where they clearly state that).

Weak-form GARCH processes have mainly been treated semiparametrically in the literature, see, e.g., Meddahi and Renault (2004), Linton and Mammen (2005) and Dahl and Levine (2005). However, common for the semiparametric approaches is that they do not permit nonstationary and explosive GARCH processes. The *first* main contribution of our paper is to analyze a fully parametrically specified GARCH process, where the innovations are neither i.i.d. nor a martingale difference sequence and by allowing for explosive and nonstationary behavior of the GARCH process. We are hereby extending the work of Jensen and Rahbek (2004a, 2004b) from strong-GARCH processes to GARCH with serially dependent innovations. Importantly, we do not need to impose the existence of the fourth moment of the dependent variable y_t (which according to Linton and Mammen (2005) is one of the main challenges in relation to modelling weak-form GARCH parametrically) to establish asymptotic normality of the QMLE along the lines of what has been done so far for strong-GARCH specifications (as in Jensen and Rahbek (2004a, 2004b)).

One popular extension of GARCH models is the GARCH-in-Mean (GARCH-M) model proposed by Engle, Lilien and Robins (1987), where the conditional volatility function (or functionals of the volatility such as the logarithm) is introduced into the conditional mean equation. This appears to be a very easy way to model the risk premium. However, GARCH-M models have been shown to be very difficult to handle from a theoretical point of view and very little is known about the sample properties of the GARCH-M estimation procedures (see e.g. Engle and Kroner (1995)). The *second* main contribution of our approach is that by allowing for dynamics in the innovation term we can create a pattern in the mean equation that introduces iteration effects between the standardized conditional volatility and lags of the dependent variable (see Section 2 for analytical details). Consequently, the model can be viewed as an alternative specification to the GARCH-M modelling of the risk premium. Contrary to the GARCH-M specification we demonstrate that our model permits a very straightforward proof of the asymptotic properties of the QML estimator. Furthermore, our approach and asymptotic results permit having a function of the conditional volatility in the mean equation as well as having nonstationary behavior in the GARCH process. The properties of the QML estimates associated with GARCH-M models with this type of nonstationarity in the GARCH process has not been established, again possibly due to the theoretical difficulties the GARCH-M specification generates. In addition, we show that the GARCH model with serially dependent innovations actually is a generalization of the double autoregressive model proposed recently by Ling (2004), where the conditional mean function is augmented with a function of the volatility function hereby allowing for a risk premium effect.

The *third* contribution of our model, is that it suggests a particular functional form relationship between the conditional volatility and the conditional mean. In particular, the model implies that it is the ratio of the conditional volatility at time t over the conditional volatility at $(t-1)$ that should enter the mean function. This type of functional relationship between the mean and the volatility is not possible to capture by the applying the traditional GARCH-M. In fact, our model can be combined with the GARCH-M, and in that respect our approach provides empirical researchers with an even richer menu of choices to model volatility and the risk premium. Finally, we offer an example showing that the new volatility model can provide improvements in fit in-sample as well as out-of-sample in relation to other competing GARCH type specifications.

The plan of the paper is as follows. Section 2 presents the general model and the results on consistency and asymptotic normality of the QML estimator. Section 3 presents some illustrations of our approach: First we provide a Monte Carlo simulation study on the consequences of dynamic misspecification of the GARCH model and secondly we present an empirical application illustrating how our new model can be potentially useful for applied researchers. Finally, Section 4 concludes.

2 The model and the asymptotics of the QMLE

2.1 The GARCH(1,1)-AR(1)

Following Jensen and Rahbek (2004b), we specify a GARCH(1,1) in the conditional variance equation

$$y_t = \sigma_t\left(\theta^*\right)\epsilon_t; \qquad \sigma_t^2\left(\theta^*\right) = w + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2\left(\theta^*\right), \tag{1}$$

and to get more generality in some of our results, we begin by simply assuming that ϵ_t is ergodic and strictly stationary (so this may include AR(p) or other much more general processes). If all expressions involved are well defined, we assume that the initial value of ϵ_t is drawn from the strictly stationary distribution and the unobserved variance is parameterized by $\gamma = \sigma_0^2$. The parameters

w, α, β, γ are all positive and $t = 1, ..., T$. Also, the true parameter vector is defined as $\theta_0^* = (w_0, \alpha_0, \beta_0, \gamma_0)'$ and $\sigma_t^2(\theta_0^*) = \sigma_t^2$. First, one important aspect is to find the conditions under which the process given by (1) is strictly stationary. These results are new in the literature (extending the results of Bougerol and Picard (1992b)), and are formulated in the following Lemma:

Lemma 1 *A necessary and sufficient condition for strict stationarity of the GARCH(1,1) where* ϵ_t *is ergodic and strictly stationary, is given by*

$$E \log\left(\alpha_0 \epsilon_t^2 + \beta_0\right) < 0.$$

Proof: Given in Appendix 1.

An additional and important new result (extending the results in Nelson (1990)), under the broad definition of ϵ_t, is given as

Lemma 2 *If* $E \log(\alpha_0 \epsilon_t^2 + \beta_0) \geq 0$ *and* ϵ_t *is ergodic and strictly stationary, as* $t \to \infty$, *then* $\sigma_t^2 \xrightarrow{a.s.} \infty$.

Proof: Given in Appendix 1.

Lemmas 1 and 2 generalize the results in the literature regarding strict stationarity of strong-GARCH processes, to weak-GARCH processes. To obtain the main asymptotic results of the paper we will, in what follows, parameterize the process generating ϵ_t. We will begin by presenting an easy-to-read proof of the asymptotic normality (as done elegantly by Jensen and Rahbek (2004a)), under the simplifying assumption that the innovation term follows an AR(1) process. Later we will generalize all the result to the AR(p) case. Consequently, we will start by considering the following representation

$$y_t = \sigma_t(\theta) \epsilon_t(\theta); \qquad \epsilon_t(\theta) = \rho \epsilon_{t-1}(\theta) + v_t; \qquad \sigma_t^2(\theta) = w + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta), \tag{2}$$

where, if all expressions above are well defined, we condition on the observed value y_0 (as in Jensen and Rahbek (2004a, 2004b)), the initial value of ϵ_t is drawn from the strictly stationary distribution, the unobserved variance is parameterized by $\gamma = \sigma_0^2$, and it holds that

Assumption 1: $v_t \sim iid(0, 1)$ with $E(v_t^4 - 1) = \zeta < \infty$. Also $|\rho_0| < 1$.

The true parameter vector is now defined as $\theta_0 = (w_0, \alpha_0, \beta_0, \gamma_0, \rho_0)'$ and $\theta = (w, \alpha, \beta, \gamma, \rho)'$. The parameters w, α, β, γ are all positive, $\epsilon_t(\theta_0) = \epsilon_t$ and $\sigma_t^2(\theta_0) = \sigma_t^2$. We also impose a second assumption to explicitly define nonstationarity in the GARCH process. By using Lemma 1 we have

Assumption 2: *The true parameters satisfy* $E\log\left(\alpha_0\epsilon_t^2+\beta_0\right)\geq 0$.

A second important issue is to analyze the size of the unconditional volatility that the process given by (2) is capable of generating. We have immediately that

$$E\left(\sigma_t^2\right)=w_0+\alpha_0 E\left(\sigma_{t-1}^2\epsilon_{t-1}^2\right)+\beta_0 E\left(\sigma_{t-1}^2\right),$$

where

$$\begin{aligned}E\left(\sigma_{t-1}^2\epsilon_{t-1}^2\right) &= E\left(\sigma_{t-1}^2\left(\rho_0\epsilon_{t-2}+v_{t-1}\right)^2\right)\\ &= \rho_0^2 E\left(\sigma_{t-1}^2\epsilon_{t-2}^2\right)+E\left(\sigma_{t-1}^2\right).\end{aligned}$$

In order to compare with the traditional expression of the unconditional volatility in the GARCH(1,1) model, we assume that there is a strictly stationary solution, such that

$$E\left(\sigma_t^2\right)=\frac{w_0+\alpha_0\rho_0^2 E\left(\sigma_{t-1}^2\epsilon_{t-2}^2\right)}{\left(1-\alpha_0-\beta_0\right)}.$$

Consequently, it appears to be the term $\alpha_0\rho_0^2 E\left(\sigma_{t-1}^2\epsilon_{t-2}^2\right)/\left(1-\alpha_0-\beta_0\right)$ that invokes a difference relative to the outcome based on the traditional GARCH(1,1) process. The GARCH(1,1)-AR(1) model will be able to generate larger or smaller unconditional variance compared to the GARCH(1,1) depending on the sign of $\alpha_0\rho_0^2 E\left(\sigma_{t-1}^2\epsilon_{t-2}^2\right)/\left(1-\alpha_0-\beta_0\right)$. Note, however, that we do not require stationarity in our framework, so the unconditional variance may not exist in the GARCH(1,1)-AR(1) setting. This feature introduces additional flexibility in the volatility process modelling relative to the traditional GARCH(1,1) model.

A third important issue that deserves attention, is to note that the conditional volatility $\left(\sigma_t^2\left(\theta\right)\right)$ we are modelling is not the volatility of $\sigma_t\left(\theta\right)v_t$, but rather the volatility of y_t. To understand this better, we compare our framework with a traditional AR(1)-ARCH(1) model (see Ling and McAleer (2003), page 283) where

$$y_t+\phi y_{t-1}=\epsilon_t;\qquad \sigma_t^2\left(\theta\right)=w+\alpha\epsilon_{t-1}^2. \tag{3}$$

Note, that $\sigma_t^2\left(\theta\right)$ is not the volatility of y_t in the AR(1)-ARCH(1) given by (3). A similar feature is also true for the double autoregressive model of Ling (2004), where

$$y_t+\phi y_{t-1}=\epsilon_t;\qquad \sigma_t^2\left(\theta\right)=w+\alpha y_{t-1}^2.$$

Therefore, in volatility models with rather complex structures, the volatility of y_t may not always enter the conditional variance equation. This is also true in our modelling framework as the GARCH(1,1) with serially dependent innovations can be shown to be a generalization of a double autoregressive model with a function of the volatility in the mean equation (i.e. allowing for a risk premium effect).

We concentrate now on the establishing the limiting properties of the QML estimators. The previous model can be re-written as

$$y_t - \rho\sigma_t(\theta)\,\sigma_{t-1}^{-1}(\theta)\,y_{t-1} = \sigma_t(\theta)\,v_t; \qquad \sigma_t^2(\theta) = w + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\theta), \tag{4}$$

where, because of the properties of v_t, it holds that

$$\begin{aligned} E\left(\left(y_t - \rho\sigma_t(\theta)\,\sigma_{t-1}^{-1}(\theta)\,y_{t-1}\right)^2/I_{t-1}\right) &= \sigma_t^2(\theta), \\ E\left(\left(y_t - \rho\sigma_t(\theta)\,\sigma_{t-1}^{-1}(\theta)\,y_{t-1}\right)/I_{t-1}\right) &= 0. \end{aligned}$$

From (4) is becomes clear how by specifying an AR(1) process for the innovation term, a measure of the volatility is introduced into the mean equation of the GARCH(1,1) process. It is also obvious that it provides an alternative approach to modelling the risk premium relative to the traditional GARCH-M proposed by Engle, Lilien and Robins (1987). Also, if $\rho = 0$ (which we can test for in practice), the specification equals the traditional GARCH(1,1) process. We now present the result of the asymptotic behavior of the QMLE in this setting. The QML estimators will maximize the quasi log-likelihood (with $\theta = (w_0, \alpha, \beta, \gamma_0, \rho)'$) given as

$$l_T(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\left(w_0 + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2(\theta)\right) + \frac{\left(y_t - \rho\sigma_t(\theta)\,\sigma_{t-1}^{-1}(\theta)\,y_{t-1}\right)^2}{\sigma_t^2(\theta)}\right). \tag{5}$$

We then have,

Theorem 1 *With (w, γ) fixed at their true values, $(w_0, \gamma_0)'$, consider the model given by the quasi log-likelihood function given as (5). Assume that the true parameter, y_t satisfies Assumption 2. Assume also that Assumption 1 holds. Under these assumptions, there exists a fixed open neighborhood $U = U(\alpha_0, \beta_0, \rho_0)$ of $(\alpha_0, \beta_0, \rho_0)'$ such that with probability tending to one as $T \longrightarrow \infty$, $l_T(\alpha, \beta, \rho)$ has a unique maximum point $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}\right)'$ in U. In addition, the QML estimator $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}\right)'$ is consistent and asymptotically normal,*

$$\sqrt{T}\left[\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}\right) - (\alpha_0, \beta_0, \rho_0)\right]' \xrightarrow{d} N\left(0, \Omega^{-1}\right),$$

with $\mu_i = E\left(\beta_0/\left(\alpha_0\epsilon_t^2 + \beta_0\right)\right)^i, i = 1, 2$ and where

$$\Omega = \begin{pmatrix} \frac{1}{\zeta\alpha_0^2} & \frac{\mu_1}{\zeta\alpha_0\beta_0(1-\mu_1)} & 0 \\ \frac{\mu_1}{\zeta\alpha_0\beta_0(1-\mu_1)} & \frac{(1+\mu_1)\mu_2}{\zeta\beta_0^2(1-\mu_1)(1-\mu_2)} & 0 \\ 0 & 0 & \frac{1}{\left(1-\rho_0^2\right)} \end{pmatrix}.$$

Proof: Given in Appendix 1.

Furthermore,

Theorem 2 *If* $E\log\left(\alpha_0\epsilon_t^2+\beta_0\right)>0$, *then the results in Theorem 1 hold for any arbitrary value of* $w>0$ *and* $\gamma>0$.

Proof: Theorem 2 is proved by a direct application of Theorem 2 of Jensen and Rahbek (2004b).

In order to make the proofs easier to follow, we have structured the derivatives in the Appendix such that comparisons to the results of Jensen and Rahbek (2004a, 2004b) are straightforward to make. It is also important to underline that the results for the score and the information in each of the proofs carry over to the stationary case by using the ergodic theorem for the observed information as in Jensen and Rahbek (2004b, page 1212). The results for the third order derivatives, apply directly both in the stationary and the nonstationary case.

2.2 The GARCH(1,1)-AR(p)

We now generalize the results in Theorems 1 and 2 to the AR(p) case. Consider now the GARCH(1,1)-AR(p) model given as

$$y_t=\sigma_t\left(\theta\right)\epsilon_t\left(\theta\right);\qquad \epsilon_t\left(\theta\right)=\sum_{i=1}^{p}\rho_i\epsilon_{t-i}\left(\theta\right)+v_t;\qquad \sigma_t^2\left(\theta\right)=w+\alpha y_{t-1}^2+\beta\sigma_{t-1}^2\left(\theta\right),\tag{6}$$

where, if all expressions involved are well defined, we condition on the observed value y_0 (as Jensen and Rahbek (2004a, 2004b)), and the initial value of ϵ_t is drawn from the strictly stationary distribution and $\gamma=\sigma_0^2$. The true parameter vector is defined as $\theta_0=\left(w_0,\alpha_0,\beta_0,\gamma_0,\rho_{10},...,\rho_{p0}\right)'$ and $\theta=\left(w,\alpha,\beta,\gamma,\rho_1,...,\rho_p\right)'$. The parameters w,α,β,γ are all positive and $\sigma_t^2\left(\theta_0\right)=\sigma_t^2$. Moreover, we impose

Assumption 3: $v_t\sim iid\left(0,1\right)$ with $E\left(v_t^4-1\right)=\zeta<\infty$. In addition, the roots of the autoregressive polynomial $\left(1-\rho_{10}L-....-\rho_{p0}L\right)$ lie outside the unit circle (with L being the lag operator).

Lemmas 1 and 2 in the previous section also hold in this framework. To determine the unconditional volatility of this process, we first take expectations

$$E\left(\sigma_t^2\right)=w_0+\alpha_0E\left(\sigma_{t-1}^2\epsilon_{t-1}^2\right)+\beta_0E\left(\sigma_{t-1}^2\right),$$

where

$$\begin{aligned} E\left(\sigma_{t-1}^{2}\epsilon_{t-1}^{2}\right) &= E\left(\sigma_{t-1}^{2}\left(\rho_{10}\epsilon_{t-2}+\rho_{20}\epsilon_{t-3}+...+\rho_{p0}\epsilon_{t-p-1}+v_{t-1}\right)^{2}\right) \\ &= \rho_{10}^{2}E\left(\sigma_{t-1}^{2}\epsilon_{t-2}^{2}\right)+...+\rho_{p0}^{2}E\left(\sigma_{t-1}^{2}\epsilon_{t-p-1}^{2}\right) \\ &\quad +E\left(\sigma_{t-1}^{2}\right)+2\rho_{10}\rho_{20}E\left(\sigma_{t-1}^{2}\epsilon_{t-2}\epsilon_{t-3}\right)+...+2\rho_{(p-1)0}\rho_{p0}E\left(\sigma_{t-1}^{2}\epsilon_{t-p}\epsilon_{t-p-1}\right). \end{aligned}$$

Consequently, if there exists a strictly stationary solution,

$$\begin{aligned} E\left(\sigma_{t}^{2}\right) &= \frac{w_{0}}{\left(1-\alpha_{0}-\beta_{0}\right)} \\ &\quad +\frac{\alpha_{0}\left[\rho_{10}^{2}E\left(\sigma_{t-1}^{2}\epsilon_{t-2}^{2}\right)+...+\rho_{p0}^{2}E\left(\sigma_{t-1}^{2}\epsilon_{t-p-1}^{2}\right)\right]}{\left(1-\alpha_{0}-\beta_{0}\right)} \\ &\quad +\frac{\alpha_{0}\left[2\rho_{10}\rho_{20}E\left(\sigma_{t-1}^{2}\epsilon_{t-2}\epsilon_{t-3}\right)+...+2\rho_{(p-1)0}\rho_{p0}E\left(\sigma_{t-1}^{2}\epsilon_{t-p}\epsilon_{t-p-1}\right)\right]}{\left(1-\alpha_{0}-\beta_{0}\right)}. \end{aligned}$$

It is again clear that there are some new extra-terms in the unconditional volatility that would not be present given that the data was generated from the traditional GARCH(1,1) process. Hence, more flexibility is introduced. In order to deriving the limiting properties of the QML estimators, consider rewriting the previous representation as

$$y_{t}-\sum_{i=1}^{p}\rho_{i}\sigma_{t}\left(\theta\right)\sigma_{t-i}^{-1}\left(\theta\right)y_{t-i}=\sigma_{t}\left(\theta\right)v_{t};\qquad \sigma_{t}^{2}\left(\theta\right)=w+\alpha y_{t-1}^{2}+\beta\sigma_{t-1}^{2}\left(\theta\right).$$

As in the previous section, we see that the AR(p) structure in the innovation process generates a measure of risk premium in the mean equation. For the estimation of the model, the QML estimators will maximize the quasi log-likelihood (with $\theta=\left(w_{0},\alpha,\beta,\gamma_{0},\rho_{1},...,\rho_{p}\right)'$) given as

$$(7)\qquad l_{T}\left(\theta\right)=-\frac{1}{2}\sum_{t=1}^{T}\left(\log\left(w_{0}+\alpha y_{t-1}^{2}+\beta\sigma_{t-1}^{2}\left(\theta\right)\right)+\frac{\left(y_{t}-\sum_{i=1}^{p}\rho_{i}\sigma_{t}\left(\theta\right)\sigma_{t-i}^{-1}\left(\theta\right)y_{t-i}\right)^{2}}{\sigma_{t}^{2}}\right).$$

We can now state the most important results of the paper.

Theorem 3 *Let data be generated according to the GARCH(1,1)-AR(p) process given by (6) and let Assumptions 2 and 3 hold. Then the results of Theorem 1 will still hold. The Ω matrix in this case will be block diagonal and the lower block will consist of the information matrix of an AR(p) process as given in Hamilton (1994, page 125) using the Galbraith and Galbraith (1974) equation.*

Proof: Given in Appendix 2.

As in the GARCH(1,1)-AR(1) case we have the following companion result to Theorem 3.

Theorem 4 *If $E\log\left(\alpha_{0}\epsilon_{t}^{2}+\beta_{0}\right)>0$, then the results in Theorem 3 hold for any arbitrary value of $w>0$ and $\gamma>0$.*

Proof: Theorem 4 is proved by a direct application of Theorem 2 of Jensen and Rahbek (2004b).

Again, the results for the score and the information in each of the proofs carry over to the stationary case by using the ergodic theorem for the observed information as in Jensen and Rahbek (2004b, page 1212). The results for the third order derivatives, apply directly both in the stationary and the nonstationary case.

By establishing Theorems 3 and 4 we have generalized the main results of Jensen and Rahbek (2004a,2004b) to a much richer class of volatility models. In the next section we will turn to the subject of the empirical relevance of our new approach.

3 Illustrations

In this section we illustrate the importance and need of carefully considering whether to specify the dynamics of the GARCH process in the conditional mean function (i.e., by choosing the AR-GARCH specification), a GARCH-M model or to specify the dynamics in the innovation process by working with the GARCH-AR specification (our new model). We begin by analyzing the effects on the estimated parameters in the conditional mean function if it is misspecified, i.e., if the researcher estimates an AR-GARCH model when the true data generating process is GARCH-AR. We derive the relative inconsistency theoretically and we quantify it under various distributional assumptions by conducting simulations. Secondly, we provide an empirical illustration showing that the GARCH-AR model provides a strong and a perhaps better alternative to the AR-GARCH model and the AR-GARCH-M as a representation of the change in the US 3 and 6 month treasury bill interest (spot) rates.

3.1 Effects of misspecification: A simple illustration

Consider the time series of interest y_t being generated according to the following GARCH(1,1)-AR(1) process (everything evaluated at the parameter vector)

$$y_t = \sigma_t \epsilon_t; \qquad \epsilon_t = \phi_0 \epsilon_{t-1} + \epsilon_{t-1}^2, \tag{8}$$

where v_t is white noise and $\sigma_t^2 = 1 + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$. Assume, for simplicity that a) the researchers primary interest is in estimating ϕ_0 and b) that she can observe the conditional variance function perfectly. However, the researcher "wrongly" assumes that the model is given by an AR(1)-GARCH(1,1), hence uses the representation in the mean given as

$$y_t = \phi y_{t-1} + \sigma_t v_t, \tag{9}$$

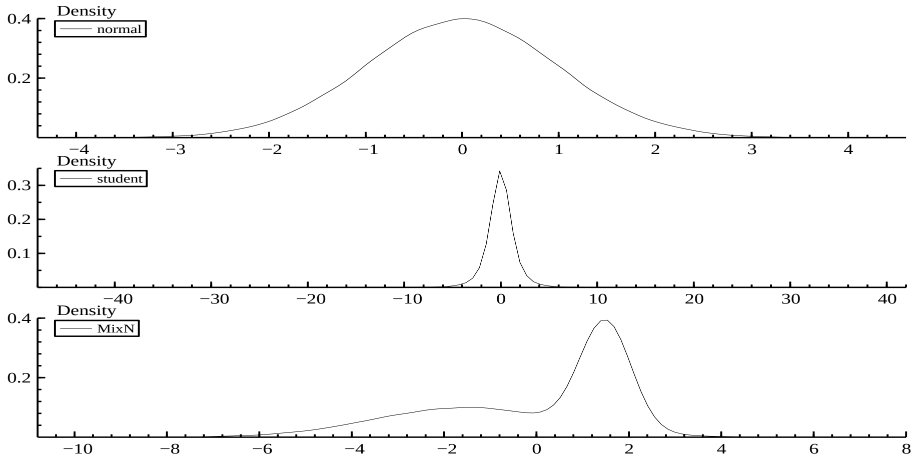

Figure 1: Alternative densities for v_t.

for estimation of the parameter ϕ_0. The main interest is to analyze the consequences of using the wrong model and to establish the asymptotic properties of the estimator of ϕ based on (9) given that σ_t is known and data is generated from (8). When σ_t is known, a proper estimator of ϕ_0 is the WLS (Weighted Least Squares) estimator $\left(\widehat{\phi}\right)$ given as

$$
\begin{aligned}
\widehat{\phi} &= \frac{\frac{1}{T}\sum \sigma_t^{-2} y_t y_{t-1}}{\frac{1}{T}\sum \sigma_t^{-2} y_{t-1}^2} \\
&= \phi_0 \frac{\frac{1}{T}\sum \left(\sigma_t^2/\sigma_{t-1}^2\right)^{1/2}\left(\sigma_t^{-2} y_{t-1}^2\right)}{\frac{1}{T}\sum \sigma_t^{-2} y_{t-1}^2} + o_p(1),
\end{aligned}
\tag{10}
$$

For simplicity assume that $y_t^2 \overset{a.s.}{\to} \infty$. Then from straightforward calculations we have $\sigma_t^{-2} y_{t-1}^2 \to 1/\alpha$, and $\sigma_t^2/\sigma_{t-1}^2 = 1/\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \beta$. By inserting in (10) it follows that

$$
\left|\frac{\left(\widehat{\phi} - \phi_0\right)}{\phi_0}\right| = \left|\left(\frac{1}{T}\sum \left(\alpha\epsilon_{t-1}^2 + \beta\right)^{1/2}\right) - 1\right|. \tag{11}
$$

We will denote $\left|\left(\widehat{\phi} - \phi_0\right)/\phi_0\right|$ as the measure of absolute relative inconsistency. From (11) it can be seen that estimating ϕ_0 based on (9) when (8) is the actual data generating process generally leads to relative inconsistency except in the trivial case when $\phi_0 = 0$. In general $\frac{1}{T}\sum \left(\alpha\epsilon_{t-1}^2 + \beta\right)^{1/2}$ can be far away from unity in particular for larger values of $|\phi_0|$ that will generate a large variance of ϵ_{t-1}^2.

To quantify the measure of absolute relative inconsistency under different distributional assumptions on v_t and for alternative values of $|\phi_0|$ we next conduct a small simulation study. We consider v_t being generated from three alternative densities depicted in Figure 1. As we expect the relative

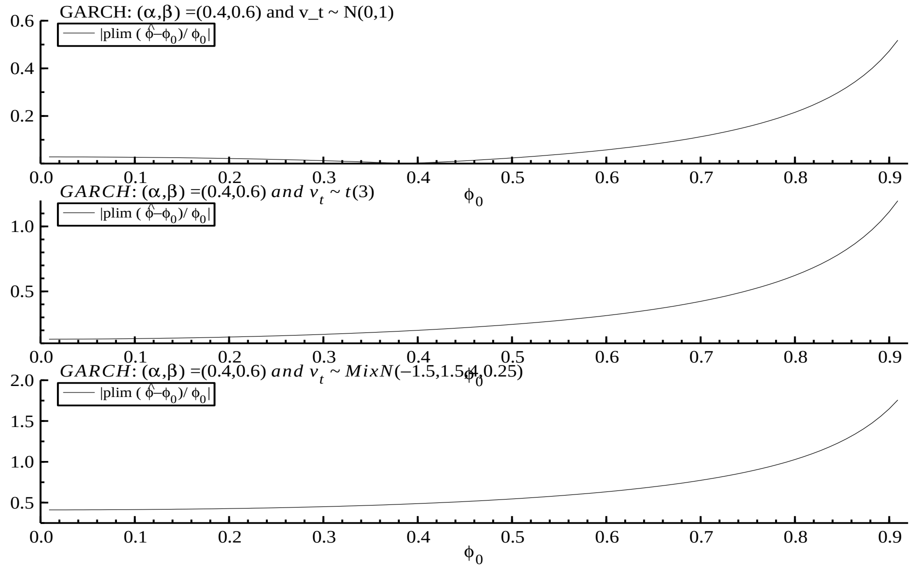

Figure 2: Measures of absolute relative inconsistency for alternative values of $|\phi_0|$ and alternative distributional assumptions.

inconsistency to grow with the variance of v_t, we include the standard Gaussian density as well as a student t-density with 3 degrees of freedom and a Gaussian mixture density.

In Figure 2 we have depicted the measure of absolute relative asymptotic inconsistency for alternative values of $|\phi_0|$ when $(\alpha, \beta) = (0.4, 0.6)$. Not surprisingly, the inconsistencies become more severe as the variance of ϵ_{t-1} increases. Note that $var(\epsilon_{t-1})$ increases with ϕ_0 and when the density of v_t goes from the normal to densities with fatter tails such as the student-t and Gaussian mixture. Such fat-tailed distributions are common in financial time series. In general the inconsistencies seem significant even for very small values of $|\phi_0|$ ranging from 5% in the Gaussian case to about 45% for the Gaussian mixture. This simple example stresses that careful model evaluation is needed to determine whether to model the dynamics in the innovation terms or in the conditional mean function, particularly when the density of the data exhibit fat-tail behavior.

3.2 Empirical illustration

Based on monthly data from January 1984 to March 2005 on the US 3 and 6 month treasury bill spot rates we compare the estimated AR-GARCH model and GARCH-AR models. We also consider the AR-GARCH-M specification. Following the theoretical framework by Cox, Ingersoll and Ross (1985), we consider the following general empirical model for the spot interest rates, denoted r_t,

Table 1: Estimation results for the GARCH-AR and AR-GARCH specification for the change in the 3 and 6 months US. t-bill spot rates. The sample period is January 1984 to March 2005.

	change in 3m t-bill		change in 6m t-bill	
	GARCH-AR	AR-GARCH	GARCH-AR	AR-GARCH
x_{t-1}	0.398***	0.388***	0.418***	0.456***
	(0.076)	(0.081)	(0.070)	(0.067)
x_{t-2}	-0.098	-0.089	-0.117*	-0.110
	(0.060)	(0.076)	(0.064)	(0.079)
x_{t-3}	0.165**	0.223**	0.138**	0.161***
	(0.068)	(0.093)	(0.063)	(0.057)
x_{t-6}	0.182***	0.188***	0.158***	0.163***
	(0.056)	(0.054)	(0.050)	(0.052)
x_{t-13}	-0.166***	-0.128**	-0.158***	-0.101*
	(0.056)	(0.054)	(0.057)	(0.053)
ω_0	0.006**	0.009	0.003***	0.010***
	(0.003)	(0.006)	(0.001)	(0.004)
y^2_{t-1}	0.153**	0.387**	0.166***	0.396**
	(0.069)	(0.193)	(0.038)	(0.183)
σ^2_{t-1}	0.576***	0.339	0.690***	0.356*
	(0.147)	(0.275)	(0.057)	(0.200)
Log likelihood	307.740	308.192	290.213	289.757
AIC	2.350	2.354	2.213	2.210

s.e. in parenthesis. p-values in brackets.

'*': significant on 10 percent level, double-sided (normal dist.).

'**': significant on 5 percent level, double-sided (normal dist.).

'***': significant on 1 percent level, double-sided (normal dist.).

given by

$$\begin{aligned}\Delta r_t &= \theta_0 + \theta_1 r_{t-1} + \theta_2 \log \sigma_t + \sum_{i=1}^{p_1} \widetilde{\phi}_i \Delta r_{t-1-i} + z_t, \\ z_t &= \sigma_t \epsilon_t, \\ \epsilon_t &= \sum_{i=1}^{p_2} \phi_i \epsilon_{t-1-i} + v_t, \\ v_t &\sim \text{i.i.d.}(0,1),\end{aligned}$$

where

$$\sigma_t^2 = \alpha_0 + \alpha z_{t-1}^2 + \beta \sigma_{t-1}^2,$$

for the AR(p_1)-GARCH (see e.g. Ling and McAleer (2003) for a general definition of AR-GARCH models) and

$$\sigma_t^2 = \alpha_0 + \alpha \Delta r_{t-1}^2 + \beta \sigma_{t-1}^2,$$

for the GARCH-AR(p_2) specification. Since the limiting properties of the estimated parameters in the GARCH-AR model have been derived under the assumption that $\theta_0 = \theta_1 = \theta_2 = 0$ and $\widetilde{\phi}_1 = \widetilde{\phi}_2 = ... = \widetilde{\phi}_{p_1} = 0$, we will estimate the model under this assumption and test the validity of the assumption using a series of robust misspecification tests proposed by Wooldridge (1991). Similarly, the limiting properties of the estimated parameters in the AR-GARCH is valid only if $\phi_1 = \phi_2 = ... = \phi_{p_2} = 0$ (as in Ling and McAleer (2003)), and again we estimate the model under this assumption and test its validity using the robust tests suggested by Wooldridge (1991).

The estimation results are reported in Table 1. To minimize the size of the table, let $x_{t-i-1} = \Delta r_{t-1-i}$ in relation to the AR-GARCH columns and let $x_{t-i-1} = \epsilon_{t-1-i}$ in the GARCH-AR columns. Similarly, let $y_{t-1}^2 = z_{t-1}^2$ and $y_{t-1}^2 = \Delta r_{t-1}^2$ in the AR-GARCH and GARCH-AR columns respectively. It should be noted that although the parameters entering the conditional mean function are practical identically, the estimated coefficients in the conditional variance equation differ substantially between the AR-GARCH and GARCH-AR specification. Note also, that the parameters in the conditional variance seem to be much more precisely estimated in the GARCH-AR model. We have also considered the AR-GARCH-M specification, but the risk premium effect through the model of Engle, Lilien and Robins (1987) seems not to be statistically significant in this framework. We show however, that there are risk premium effects but iterating with lags of the first difference of the treasury bill spot rates (what our model is able to generate). Various misspecification test are reported in Table 2. Although all the models pass the Ljung-Box tests, the AR-GARCH models could not be augmented with lags such that there was no evidence of neglected serial correlation according to Wooldridge 1. On the contrary there is no evidence of neglected serial correlation in the GARCH-AR models. Based on all models we cannot reject that $\theta_0 = \theta_1 = \theta_2 = 0$.

Table 2: Specification testing and forecast performance results for the GARCH-AR and AR-GARCH models for the change in the 3 and 6 months US. t-bill spot rates. The sample period is January 1984 to March 2005.

	change in 3m t-bill		change in 6m t-bill	
	GARCH-AR	AR-GARCH	GARCH-AR	AR-GARCH
LB[10]	7.892	15.293	8.278	9.084
	[0.639]	[0.122]	[0.602]	[0.524]
LB[20]	20.613	29.353	18.443	22.623
	[0.420]	[0.081]	[0.558]	[0.308]
LB[40]	42.600	48.637	40.049	44.269
	[0.360]	[0.164]	[0.468]	[0.296]
Wooldridge 1	3.740*	10.164***	3.182*	5.438**
	[0.053]	[0.001]	[0.074]	[0.020]
Wooldridge 2.1	1.469	0.007	1.146	0.429
	[0.225]	[0.934]	[0.284]	[0.512]
Wooldridge 2.2	.870	1.484	2.365	1.411
	[0.171]	[0.223]	[0.124]	[0.234]
Wooldridge 2.3	0.984	0.464	1.193	0.372
	[0.321]	[0.496]	[0.274]	[0.541]
Out-of-sample MSFE	0.024	0.026	0.023	0.024

LB[XX] is the Ljung-Box test for neglected serial dependence up to order XX.

Wooldridge 1 is Wooldridge's (1991) robust test for neglected serial dependence.

Wooldridge 2.1 is the robust test for omitted variable, see, e.g., Wooldridge's (1991)
For the AR-GARCH models the omitted variables is $\sigma_t \sigma_{t-1}^{-1} x_{t-1}$ while
it is x_{t-1} for the GARCH-AR models.

Wooldridge 2.2 is as 2.1, but where the omitted variable is the level of the interest rates.

Wooldridge 2.3 is as 2.1, but where the omitted variable is $log(\sigma_t^2)$

Models estimated for 1984m1-1999m12 and MSFE is computed using the out-of-sample period 2000m1-2005m3.

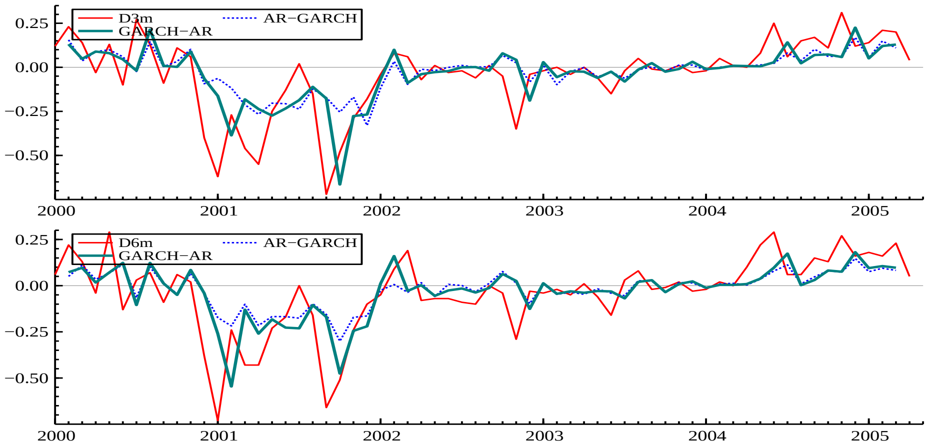

Figure 3: Out-of-sample forecasts

Finally, we look at the out-of sample forecast accuracy of the two models. We estimate the parameters for all models based on the sample from January 1984 to December 2000, and based on the remaining part of the sample we compute the mean squared forecast errors which in Table 2 is referred to as out-of-sample MSFE. It turns out that the GARCH-AR improves the forecast accuracy over the AR-GARCH model by about 5%. However, perhaps more importantly it can be seen from Figure 3 that the main difference between the models forecast seems to be in periods with "larger" variance or higher probability of more extreme observation. Note that the GARCH-AR model is able to generate these "extreme" forecast, something that the AR-GARCH seems not to be able to do in the two illustrations.

4 Conclusion

In this paper we introduce a new parametric volatility model that allows for serially dependent innovations in GARCH specifications. We first prove the asymptotic normality of the QML estimator in this setting, allowing for possible explosive and nonstationary behavior of the GARCH. We also show that this model is capable of generating an alternative measure of risk premium relative to the GARCH-M model of Engle, Lilien and Robins (1987). In particular the model generate iteration effects between functions of the conditional volatility and lags of the dependent variable of interest. Finally, we provide evidence of the usefulness of our approach in a Monte Carlo experiment and in a practical application. We first show the consequences of dynamically misspecifying the GARCH model when the actual data generating mechanism is an GARCH-AR specification. We provide evidence of the large inconsistencies that this type of misspecification can generate. Finally we show, using US interest rates, how the new model can improve the fit in-sample as well as out-of-sample relative to traditional GARCH type models.

Appendix 1

Proof of Lemma 1 The process is given by (note we only need ϵ_t to be ergodic and strictly stationary)

$$\sigma_t^2 = w_0 + \left(\alpha_0 \epsilon_{t-1}^2 + \beta_0\right) \sigma_{t-1}^2 = B_t + A_t \sigma_{t-1}^2,$$

and we define $A_t = \left(\alpha_0 \epsilon_{t-1}^2 + \beta_0\right)$ and $B_t = w_0$. Then, applying Theorem 1.1 of Bougerol and Picard (1992a, page 1715), we verify the conditions that $E\left(\log\left(\max\left\{\|1, A_0\|\right\}\right)\right) < \infty$, $E\left(\log\left(\max\left\{\|1, B_0\|\right\}\right)\right) < \infty$ (by assumption because $w_0, \alpha_0, \beta_0 > 0$), and also σ_t^2 is strictly stationary if the Lyapunov exponent τ

$$\tau = \inf\left\{E\left(\frac{1}{T+1}\log\|A_0 \cdots A_T\|\right)\right\} < 0.$$

In the case of one-dimensional recurrence equations

$$\frac{1}{T+1}E\left(\log|A_0 \cdots A_T|\right) = \frac{1}{T+1}\sum_{i=0}^{T} E\log|A_i| = E\log|A_0| < 0.$$

Therefore, σ_t^2 is strictly stationary if

$$E\log|A_0| = E\log\left(\alpha_0 \epsilon_t^2 + \beta_0\right) < 0.$$

This proves when σ_t^2 is strictly stationary. Since the pair $(y_t, \sigma_t^2)' = (\sigma_t \epsilon_t, \sigma_t^2)'$ is a fixed function of $(\sigma_t^2, \epsilon_t)'$ which is ergodic and strictly stationary, then it follows that if $(\sigma_t^2, \epsilon_t)'$ is strictly stationary, $(\sigma_t \epsilon_t, \sigma_t^2)'$ is also strictly stationary. ■

Proof of Lemma 2 By recursion (note again that we only need ϵ_t to be ergodic and strictly stationary)

$$\begin{aligned} \sigma_t^2 &= w_0 + \left(\alpha_0 \epsilon_{t-1}^2 + \beta_0\right) \sigma_{t-1}^2 \\ &= B_t + A_t \sigma_{t-1}^2 \\ &= A_t \cdots A_1 \sigma_0^2 + \sum_{i=0}^{t-1} A_t \cdots A_{t-i+1} B_{t-i} \\ &= \sigma_{1t}^2 + \sigma_{2t}^2, \end{aligned} \tag{12}$$

where $\sigma_{1t}^2 = A_t \cdots A_1 \sigma_0^2$ and $\sigma_{2t}^2 = \sum_{i=0}^{t-1} A_t \cdots A_{t-i+1} B_{t-i}$. Since σ_{2t}^2 is always positive (β_0 and w_0 are always positive), for Lemma 2, it suffices to show that $\frac{\log \sigma_{1t}^2}{t} \xrightarrow{a.s.} E\log\left(\alpha_0 \epsilon_t^2 + \beta_0\right) \geq 0$. After taking logarithms and dividing by t in the expression for σ_{1t}^2 in (12)

$$\frac{\log \sigma_{1t}^2}{t} = \frac{\sum_{i=0}^{t-1} \log\left(\alpha_0 \epsilon_{t-i}^2 + \beta_0\right) + \log \sigma_0^2}{t},$$

and by the strong law of large numbers for ergodic and strictly stationary processes, when $E \log A_t \geq 0$, and as $t \to \infty$, $\frac{\log \sigma_{1t}^2}{t} \xrightarrow{a.s.} E \log\left(\alpha_0 \epsilon_t^2 + \beta_0\right) \geq 0$, since $\frac{\log \sigma_0^2}{t} \xrightarrow{a.s.} 0$. ■

Proof of Theorem 1 In order to allow for nonstationarity in the GARCH along the lines of Jensen and Rahbek (2004a, 2004b), we first find the expressions for the first, second and third order derivatives (Lemmas 1 and 2 are used in our results). Later, Lemmas 3, 4 and 5 establish the Cramér type conditions. As in Jensen and Rahbek (2004a, 2004b) we also use the central limit theorem in Brown (1971). In order to make our results clear, we order the terms of the derivatives to find a similar structure as in Jensen and Rahbek (2004a, 2004b), in all those cases where this is possible. We also use Lemma 1 of Jensen and Rahbek (2004b) to prove uniqueness and the existence of the consistent and asymptotically Gaussian estimator. ■

First order derivatives The first order derivatives are given by

$$\frac{\partial}{\partial z} l_T(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \left(\left(1 - \frac{\left(y_t - \rho \sigma_t(\theta) \sigma_{t-1}^{-1}(\theta) y_{t-1}\right)^2}{\sigma_t^2(\theta)} \right) \frac{\frac{\partial \sigma_t^2(\theta)}{\partial z}}{\sigma_t^2(\theta)} + \frac{\frac{\partial \left(y_t - \rho \sigma_t(\theta) \sigma_{t-1}^{-1}(\theta) y_{t-1}\right)^2}{\partial z}}{\sigma_t^2(\theta)} \right); \forall z = \alpha, \beta,$$

with

$$\begin{aligned}
\frac{\partial}{\partial \alpha} l_T(\theta) &= \sum_{t=1}^{T} s_{1t}(\theta), \\
\frac{\partial}{\partial \beta} l_T(\theta) &= \sum_{t=1}^{T} s_{2t}(\theta), \\
\frac{\partial}{\partial \rho} l_T(\theta) &= \sum_{t=1}^{T} s_{3t}(\theta) = \sum_{t=1}^{T} \frac{\left(y_t - \rho \sigma_t(\theta) \sigma_{t-1}^{-1}(\theta) y_{t-1}\right) \sigma_{t-1}^{-1}(\theta) y_{t-1}}{\sigma_t(\theta)},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\frac{\partial \sigma_t^2(\theta)}{\partial \alpha}}{\sigma_t^2(\theta)} &= \frac{\sum_{j=1}^{T} \beta^{j-1} y_{t-j}^2}{\sigma_t^2(\theta)}, \\
\frac{\frac{\partial \sigma_t^2(\theta)}{\partial \beta}}{\sigma_t^2(\theta)} &= \sum_{j=1}^{T} \beta^{j-1} \frac{\sigma_{t-j}^2(\theta)}{\sigma_t^2(\theta)}, \\
\frac{\partial \left(y_t - \rho \sigma_t(\theta) \sigma_{t-1}^{-1}(\theta) y_{t-1}\right)^2}{\partial z} &= \frac{-\left(y_t - \rho \sigma_t(\theta) \sigma_{t-1}^{-1}(\theta) y_{t-1}\right) \rho y_{t-1}}{\left(1 + \alpha y_{t-2}^2 + \beta \sigma_{t-2}^2(\theta)\right)^{3/2} \left(1 + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta)\right)^{1/2}} \\
&\quad \times \left(\left(1 + \alpha y_{t-2}^2 + \beta \sigma_{t-2}^2(\theta)\right) \frac{\partial \sigma_t^2(\theta)}{\partial z} - \left(1 + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta)\right) \frac{\partial \sigma_{t-1}^2(\theta)}{\partial z} \right),
\end{aligned}$$

for $\forall z = \alpha, \beta$. Then,

Lemma 3 *Let Assumptions 1 and 2 hold. Defining* $s_{it}(\theta_0) = s_{it}$, $\forall i = 1, 2, 3$ *associated with (5). Then*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{1t} \xrightarrow{d} N\left(0, \frac{\zeta}{4\alpha_0^2}\right),$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{2t} \xrightarrow{d} N\left(0, \frac{\zeta(1+\mu_1)\mu_2}{4\beta_0^2(1-\mu_1)(1-\mu_2)}\right),$$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{3t} \xrightarrow{d} N\left(0, \frac{1}{(1-\rho_0^2)}\right),$$

with $\mu_i = E\left(\beta_0 / \left(\alpha_0 \epsilon_t^2 + \beta_0\right)\right)^i, i = 1, 2$ *as* $T \longrightarrow \infty$.

Proof of Lemma 3 As in Jensen and Rahbek (2004b, Lemma 5 and its extension to the α parameter), using the law of iterated expectations and the properties of v_t, we have $E(s_{1t}/I_{t-1}) = E(s_{2t}/I_{t-1}) = E(s_{3t}/I_{t-1}) = 0$. Also, the proof of Lemma 3 requires that

$$E|s_{1t}| < \infty; \quad E|s_{2t}| < \infty; \quad E|s_{3t}| < \infty. \tag{13}$$

We prove now (13). For the first two scores, we have, evaluated at θ_0

$$\frac{1}{2}\left(1 - v_t^2\right) \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} + \frac{\rho_0}{2} \frac{v_t y_{t-1}(\left(w_0 + \alpha_0 y_{t-1}^2 + \beta_0 \sigma_{t-1}^2\right) \frac{\partial \sigma_{t-1}^2}{\partial z} - \left(w_0 + \alpha_0 y_{t-2}^2 + \beta_0 \sigma_{t-2}^2\right) \frac{\partial \sigma_t^2}{\partial z})}{\sigma_t^2 \sigma_{t-1}^3} =$$

$$\frac{1}{2}\left(1 - v_t^2\right) \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2} + \frac{\rho_0}{2} v_t \epsilon_{t-1} \left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2} - \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2}\right]; \qquad \forall z = \alpha, \beta,$$

where the first term follows from Lemma 5 in Jensen and Rahbek (2004b), and for the second term, note that

$$\left|\frac{\rho_0}{2} v_t \epsilon_{t-1} \left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial \alpha}}{\sigma_{t-1}^2} - \frac{\frac{\partial \sigma_t^2}{\partial \alpha}}{\sigma_t^2}\right]\right| \leq \left|\frac{\rho_0}{2} v_t \epsilon_{t-1}\right| \left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial \alpha}}{\sigma_{t-1}^2} - \frac{\frac{\partial \sigma_t^2}{\partial \alpha}}{\sigma_t^2}\right].$$

In addition

$$\begin{aligned} |v_t \epsilon_{t-1}| &= \left|\sum_{j=0}^{\infty} \rho_0^j v_t v_{t-1-j}\right| \\ &\leq \sum_{j=0}^{\infty} \left|\rho_0^j\right| \left|v_t v_{t-1-j}\right|, \end{aligned}$$

and from Holder's inequality

$$\begin{aligned} E\left|v_t v_{t-1-j}\right| &\leq \sqrt{E(v_t^2)}\sqrt{E(v_{t-1-j}^2)} \\ &= E(v_t^2) \\ &= 1. \end{aligned}$$

Finally, $E\left|v_t \epsilon_{t-1}\right| < \infty$ (as $\sum_{j=0}^{\infty}\left|\rho_0^j\right| < \infty$), and hence $E\left|s_{1t}\right| < \infty$ and $E\left|s_{2t}\right| < \infty$. For the third score, we have

$$s_{3t} = v_t \epsilon_{t-1},$$

and from the previous results for the first and second score, it follows directly that

$$E\left|s_{3t}\right| = E\left|v_t \epsilon_{t-1}\right| < \infty.$$

Besides

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} E\left(s_{1t}^2/I_{t-1}\right) &= \frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)^2 \\ &+\frac{1}{T}\sum_{t=1}^{T}\frac{\rho_0^2 y_{t-1}^2\left(\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z}-\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}\right)^2}{4\left(1+\alpha_0 y_{t-2}^2\right)^3\left(1+\alpha_0 y_{t-1}^2\right)^2} \\ &\xrightarrow{p} \frac{\zeta}{4\alpha_0^2}, \end{aligned}$$

using Lemmas 4-6 in Jensen and Rahbek (2004b) and its extension to the α parameter, since

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial\alpha}}{\sigma_t^2}\right)^2 \xrightarrow{p} \frac{\zeta}{4\alpha_0^2},$$

and

$$\frac{\rho_0^2}{4T}\sum_{t=1}^{T}\epsilon_{t-1}^2\left[\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right] \xrightarrow{p} 0.$$

For the second score and the outer product

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} E\left(s_{2t}^2/I_{t-1}\right) &\xrightarrow{p} \frac{\zeta\left(1+\mu_1\right)\mu_2}{4\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)}, \\ \frac{1}{T}\sum_{t=1}^{T} E\left(s_{1t}s_{2t}/I_{t-1}\right) &\xrightarrow{p} \frac{\zeta\mu_1}{4\alpha_0\beta_0\left(1-\mu_1\right)}, \end{aligned}$$

since

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}\left(\frac{\frac{\partial\sigma_t^2}{\partial\beta}}{\sigma_t^2}\right)^2 \xrightarrow{p} \frac{\zeta\left(1+\mu_1\right)\mu_2}{4\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)},$$

following Jensen and Rahbek (2004b), Lemmas 3, 4 and 5. For the last score

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} E\left(s_{3t}^2 / I_{t-1}\right) &= \frac{1}{T}\sum_{t=1}^{T}\frac{y_{t-1}^2}{\left(1+\alpha_0 y_{t-2}^2\right)} \\ &\xrightarrow{p} \frac{1}{\left(1-\rho_0^2\right)},\end{aligned}$$

under Assumption 1. Finally, we can derive a Lindeberg type condition as in Jensen and Rahbek (2004a), where we have

$$\begin{aligned}&\frac{1}{4}\left(\left(1-v_t^2\right)\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}+\rho_0\frac{v_t y_{t-1}(\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}-\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z})}{\sigma_t^2\sigma_{t-1}^3}\right)^2 \\ =\;& \frac{1}{4}[\left(1-v_t^2\right)^2\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)^2+\rho_0^2\frac{v_t^2 y_{t-1}^2(\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}-\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z})^2}{\sigma_t^4\sigma_{t-1}^6} \\ &+2\rho_0 v_t\left(1-v_t^2\right)\frac{\frac{\partial\sigma_t^2}{\partial z}y_{t-1}(\left(w_0+\alpha_0 y_{t-1}^2+\beta_0\sigma_{t-1}^2\right)\frac{\partial\sigma_{t-1}^2}{\partial z}-\left(w_0+\alpha_0 y_{t-2}^2+\beta_0\sigma_{t-2}^2\right)\frac{\partial\sigma_t^2}{\partial z})}{\sigma_t^4\sigma_{t-1}^3}] \\ =\;& \frac{1}{4}\left(1-v_t^2\right)^2\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)^2+\frac{1}{4}\rho_0^2\left(v_t\epsilon_{t-1}\right)^2\left(\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_t^2\sigma_{t-1}^4}-\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^4\sigma_{t-1}^2}\right) \\ &+\frac{1}{2}\rho_0\left(v_t\epsilon_{t-1}\right)\left(1-v_t^2\right)\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right)\left(\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}-\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}\right),\end{aligned}$$

and

$$s_{3t}^2 = v_t^2\epsilon_{t-1}^2,$$

with the following bounds for s_{1t}^2 (for s_{2t}^2, it would follow the same argument) and s_{3t}^2

$$s_{1t}^2 \leq \mu_{1t}^2 \equiv \frac{1}{4\alpha_0^2}\left(1-v_t^2\right)^2+\rho_0^2\left(v_t\epsilon_{t-1}\right)^2+\left|\rho_0\right|\left|\left(v_t\epsilon_{t-1}\right)\right|\left|\left(1-v_t^2\right)\right|, \tag{14}$$

$$s_{3t}^2 \leq \mu_{3t}^2 \equiv v_t^2\epsilon_{t-1}^2(1+\gamma_0) \text{ for } 0 \leq \gamma_0 < \infty\;. \tag{15}$$

Since v_t and ϵ_{t-1} are stationary and ergodic, then also any measurable mapping of v_t and ϵ_{t-1} will be stationary and ergodic, see, e.g., White (1984, Th. 3.35). Consequently, μ_{1t}^2 and μ_{3t}^2 are stationary and ergodic and it follows that

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left(s_{it}^2 I\left(\left|s_{it}\right|>\sqrt{T}\partial\right)\right) &\leq \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E\left(\mu_{it}^2 I\left(\left|\mu_{it}\right|>\sqrt{T}\partial\right)\right) \\ &= \lim_{T\to\infty}E\left(\mu_{i1}^2 I\left(\left|\mu_{i1}\right|>\sqrt{T}\partial\right)\right) \\ &\to 0.\end{aligned} \tag{16}$$

for $i = 1, 2, 3$. This establishes the Lindeberg type condition as in Jensen and Rahbek (2004a, 2004b). ■

Second order derivatives We start now with the second order derivatives. They are given by (σ_t should be evaluated at θ in all the expressions, but for reasons of space, we have omitted that)

$$\begin{aligned}
\frac{\partial^2}{\partial z_1 \partial z_2} l_T(\theta) &= \frac{1}{2} \sum_{t=1}^{T} \left(\left(1 - \frac{2\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2(\theta)} \right) \frac{\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_2}}{\sigma_t^4(\theta)} + \left(\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2(\theta)} - 1 \right) \frac{\frac{\partial^2\sigma_t^2}{\partial z_1 \partial z_2}}{\sigma_t^2(\theta)} \right) \\
&+ \frac{1}{2} \sum_{t=1}^{T} \left(-\frac{\frac{\partial^2\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1 \partial z_2}}{\sigma_t^2} + \frac{\left(\frac{\partial \sigma_t^2}{\partial z_1} \frac{\partial\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2} + \frac{\partial \sigma_t^2}{\partial z_2} \frac{\partial\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1} \right)}{\sigma_t^4} \right), \\
\frac{\partial^2}{\partial \rho^2} l_T(\theta) &= -\sum_{t=1}^{T} \sigma_{t-1}^{-2} y_{t-1}^2, \\
\frac{\partial^2}{\partial z \partial \rho} l_T(\theta) &= \sum_{t=1}^{T} \left(\frac{\frac{\partial\left[\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z}}{\sigma_t^2} - \frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_{t-1}^{-1}y_{t-1}\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^3} \right),
\end{aligned}$$

where

$$\begin{aligned}
\frac{\frac{\partial^2\sigma_t^2}{\partial \alpha^2}}{\sigma_t^2} &= 2\frac{\sum_{j=1}^{T}(j-1)\beta^{j-2}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^2\sigma_t^2}{\partial \beta^2}}{\sigma_t^2} &= 2\sum_{j=1}^{t}(j-1)\beta^{j-2}\frac{\sigma_{t-j}^2}{\sigma_t^2}, \\
\frac{\partial\left[\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z} &= \frac{y_t y_{t-1}}{2\sigma_t\sigma_{t-1}} \left(\frac{\partial \sigma_t^2}{\partial z} - \frac{\sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2} \right) - \rho y_{t-1}^2 \left(\frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_{t-1}^2} - \frac{\sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^4} \right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1 \partial z_2} &= \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\left(-\rho y_{t-1}\right) \left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2}}{\sigma_{t-1}^3 \sigma_t} \right] \\
&+ \left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\left(\rho y_{t-1}\right) \left[\frac{\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1} \right)\left(3\frac{\partial \sigma_{t-1}^2}{\partial z_2}\sigma_t + \sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2}\sigma_t^{-1} \right)}{2\sigma_{t-1}^5 \sigma_t^2} \right] \\
&+ \rho^2 y_{t-1}^2 \frac{\left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1} \right]\left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2} \right]}{2\sigma_{t-1}^6 \sigma_t^2},
\end{aligned}$$

for $\forall z, z_1, z_2 = \alpha, \beta$. Then

Lemma 4 *Under Assumptions 1 and 2, with the expressions of the second order derivatives evaluated at θ_0*

(a) $\frac{1}{T}\left(-\frac{\partial^2}{\partial \alpha^2} l_T(\theta) \mid_{\theta=\theta_0} \right) \xrightarrow{p} \frac{1}{2\alpha_0^2} > 0,$

(b) $\frac{1}{T}\left(-\frac{\partial^2}{\partial \beta^2} l_T(\theta) \mid_{\theta=\theta_0} \right) \xrightarrow{p} \frac{(1+\mu_1)\mu_2}{2\beta_0^2(1-\mu_1)(1-\mu_2)} > 0,$

(c) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\beta}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}\frac{\mu_1}{2\alpha_0\beta_0(1-\mu_1)}$,
(d) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\rho}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}0$,
(e) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\rho}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}0$,
(f) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\rho^2}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}\frac{1}{(1-\rho_0^2)}>0$,
as $T\longrightarrow\infty$.

Proof of Lemma 4 For expressions (a), (b) and (c)

$$-\frac{1}{2T}\sum_{t=1}^{T}\left(\left(1-\frac{2\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4}+\left(\frac{\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}-1\right)\frac{\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_t^2}\right)$$
$$\xrightarrow{p}\frac{1}{2\alpha_0^2},\text{ with } z_1=z_2=\alpha,$$
$$\xrightarrow{p}\frac{\left(1+\mu_1\right)\mu_2}{2\beta_0^2\left(1-\mu_1\right)\left(1-\mu_2\right)},\text{ with } z_1=z_2=\ \beta,$$
$$\xrightarrow{p}\frac{\mu_1}{2\alpha_0\beta_0\left(1-\mu_1\right)},\text{ with } z_1=\alpha\text{ and } z_2=\beta,$$

because of Lemma 6 in Jensen and Rahbek (2004b), and its extension to α and the cross products of α and β. Also

$$-\frac{1}{2T}\sum_{t=1}^{T}\left(-\frac{\frac{\partial^2\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2}}{\sigma_t^2}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1}\right)}{\sigma_t^4}\right)\xrightarrow{p}0;\quad\forall z_1,z_2=\alpha,\beta,$$

by using again Lemmas 3 and 4 in Jensen and Rahbek (2004b) and the same results as in Lemma 3 for our score. An expression of the most complicated term in the previous expression is

$$\frac{\rho_0^2}{4T}\sum_{t=1}^{T}\frac{y_{t-1}^2}{\sigma_{t-1}^2}\left[\frac{\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^4}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^2\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}}{\sigma_{t-1}^4}\right]\xrightarrow{p}0;\quad\forall z_1,z_2=\alpha,\beta.$$

For expressions (d) and (e)

$$-\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\frac{\partial\left[\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right]}{\partial z}}{\sigma_t^2}-\frac{\left(y_t-\rho_0\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_{t-1}^{-1}y_{t-1}\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^3}\right)\xrightarrow{p}0;\quad\forall z=\alpha,\beta,$$

since

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\rho_0y_{t-1}^2}{\sigma_{t-1}^2}\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right)-\frac{y_ty_{t-1}}{2\sigma_t\sigma_{t-1}}\left(\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t^2}-\frac{\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2}\right)\xrightarrow{p}0;\quad\forall z=\alpha,\beta,$$

and by a simple application again of Lemmas 3 and 4 in Jensen and Rahbek (2004b). Finally, expression (f)

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\sigma_{t-1}^{-2}y_{t-1}^{2} &= \frac{1}{T}\sum_{t=1}^{T}\epsilon_{t-1}^{2}\\ &\xrightarrow{p} \frac{1}{\left(1-\rho_{0}^{2}\right)},\end{aligned}$$

by Assumption 1. ■

Third order derivatives The third order derivatives are given by (σ_t should be evaluated at θ in all the expressions, but for reasons of space, we have omitted that)

$$\begin{aligned}
\frac{\partial^{3}}{\partial z_{1}^{2}\partial z_{2}}l_{T}\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\sigma_{t}^{2}}\right)\frac{\frac{\partial^{3}\sigma_{t}^{2}}{\partial z_{1}^{2}\partial z_{2}}}{\sigma_{t}^{2}}\\
&\quad -\sum_{t=1}^{T}\left(1-\frac{3\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\sigma_{t}^{2}}\right)\frac{\left(\frac{\partial\sigma_{t}^{2}}{\partial z_{1}}\right)^{2}\frac{\partial\sigma_{t}^{2}}{\partial z_{2}}}{\sigma_{t}^{6}}\\
&\quad +\sum_{t=1}^{T}\frac{\frac{1}{2}\left(\frac{\partial^{2}\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{1}^{2}}\frac{\partial\sigma_{t}^{2}}{\partial z_{2}}+\frac{\partial^{2}\sigma_{t}^{2}}{\partial z_{1}^{2}}\frac{\partial\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{2}}\right)}{\sigma_{t}^{4}}\\
&\quad +\sum_{t=1}^{T}\frac{\left(\frac{\partial^{2}\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{1}\partial z_{2}}\frac{\partial\sigma_{t}^{2}}{\partial z_{1}}+\frac{\partial^{2}\sigma_{t}^{2}}{\partial z_{1}\partial z_{2}}\frac{\partial\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{1}}\right)}{\sigma_{t}^{4}}\\
&\quad -\sum_{t=1}^{T}\left(2\frac{\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\sigma_{t}^{2}}-1\right)\frac{\left(\frac{\partial^{2}\sigma_{t}^{2}}{\partial z_{1}\partial z_{2}}\frac{\partial\sigma_{t}^{2}}{\partial z_{1}}+\frac{1}{2}\frac{\partial^{2}\sigma_{t}^{2}}{\partial z_{1}^{2}}\frac{\partial\sigma_{t}^{2}}{\partial z_{2}}\right)}{\sigma_{t}^{4}}-\frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial^{3}\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{1}^{2}\partial z_{2}}}{\sigma_{t}^{2}}\\
&\quad -\sum_{t=1}^{T}\frac{\left(\frac{\partial\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{2}}\left(\frac{\partial\sigma_{t}^{2}}{\partial z_{1}}\right)^{2}+2\frac{\partial\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{1}}\frac{\partial\sigma_{t}^{2}}{\partial z_{1}}\frac{\partial\sigma_{t}^{2}}{\partial z_{2}}\right)}{\sigma_{t}^{6}},\\
\frac{\partial^{3}}{\partial\rho^{3}}l_{T}\left(\theta\right) &= 0,\\
\frac{\partial^{3}}{\partial z_{1}\partial z_{2}\partial\rho}l_{T}\left(\theta\right) &= \sum_{t=1}^{T}\frac{2\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\frac{\partial\sigma_{t}^{2}}{\partial z_{1}}\frac{\partial\sigma_{t}^{2}}{\partial z_{2}}}{\sigma_{t}^{6}}+\sum_{t=1}^{T}\frac{\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\frac{\partial^{2}\sigma_{t}^{2}}{\partial z_{1}\partial z_{2}}}{\sigma_{t}^{4}}\\
&\quad +\frac{1}{2}\sum_{t=1}^{T}\left(-\frac{\frac{\partial^{3}\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{1}\partial z_{2}\partial\rho}}{\sigma_{t}^{2}}+\frac{\left(\frac{\partial\sigma_{t}^{2}}{\partial z_{1}}\frac{\partial^{2}\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{2}\partial\rho}+\frac{\partial\sigma_{t}^{2}}{\partial z_{2}}\frac{\partial^{2}\left(y_{t}-\rho\sigma_{t}\sigma_{t-1}^{-1}y_{t-1}\right)^{2}}{\partial z_{1}\partial\rho}\right)}{\sigma_{t}^{4}}\right),
\end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial^3}{\partial z \partial \rho^2} l_T(\theta) &= \sum_{t=1}^{T} \frac{y_{t-1}^2 \frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^4}, \\
\frac{\frac{\partial^3 \sigma_t^2}{\partial \alpha^3}}{\sigma_t^2} &= 3 \frac{\sum_{j=1}^{T} (j-1)(j-2) \beta^{j-3} y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^3 \sigma_t^2}{\partial \beta^3}}{\sigma_t^2} &= 3 \sum_{j=1}^{t} (j-1)(j-2) \beta^{j-3} \frac{\sigma_{t-j}^2}{\sigma_t^2},
\end{aligned}
$$

for $\forall z, z_1, z_2 = \alpha, \beta$ where

$$
\begin{aligned}
\frac{\partial^3 \left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right)^2}{\partial z_1^2 \partial z_2} = {} & -\rho y_{t-1} \left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right) \lbrack \frac{\left(\frac{\partial \sigma_{t-1}^2}{\partial z_2} \frac{\partial^2 \sigma_t^2}{\partial z_1^2} + \frac{\partial^3 \sigma_{t-1}^2}{\partial z_1^2 \partial z_2} \sigma_{t-1}^2 - \frac{\partial \sigma_t^2}{\partial z_2} \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1^2} - \sigma_t^2 \frac{\partial^3 \sigma_{t-1}^2}{\partial z_1^2 \partial z_2} \right)}{\sigma_{t-1}^3 \sigma_t} \\
& - \frac{\left(\sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1^2} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1^2} \right) \left(3 \sigma_{t-1} \sigma_t \frac{\partial \sigma_{t-1}^2}{\partial z_2} + \sigma_t^{-1} \sigma_{t-1}^3 \frac{\partial \sigma_t^2}{\partial z_2} \right)}{2 \sigma_{t-1}^6 \sigma_t^2} \\
& - \frac{\left(\frac{\partial \sigma_{t-1}^2}{\partial z_1} \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} + \frac{\partial \sigma_t^2}{\partial z_1} \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2} \right)}{\sigma_{t-1}^4 \sigma_t} \\
& + \frac{\left(\frac{\partial \sigma_t^2}{\partial z_1} \frac{\partial \sigma_{t-1}^2}{\partial z_1} \right) \left(2 \sigma_{t-1}^2 \sigma_t \frac{\partial \sigma_{t-1}^2}{\partial z_2} + \frac{1}{2} \sigma_t^{-1} \sigma_{t-1}^4 \frac{\partial \sigma_t^2}{\partial z_2} \right)}{\sigma_{t-1}^8 \sigma_t^2} - \frac{\left(\frac{\partial \sigma_t^2}{\partial z_1} \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} \right)}{\sigma_{t-1}^2 \sigma_t^3} + \frac{3 \left(\frac{\partial \sigma_{t-1}^2}{\partial z_1} \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2} \right)}{\sigma_{t-1}^6 \sigma_t^{-1}} \\
& + \frac{\left(\frac{\partial \sigma_t^2}{\partial z_1} \right)^2 \left(\frac{\sigma_t^3}{2} \frac{\partial \sigma_{t-1}^2}{\partial z_2} + \sigma_t \sigma_{t-1}^2 \frac{3}{4} \frac{\partial \sigma_t^2}{\partial z_2} \right)}{\sigma_{t-1}^4 \sigma_t^6} \\
& - \frac{\frac{3}{2} \left(\frac{\partial \sigma_{t-1}^2}{\partial z_1} \right)^2 \left(3 \sigma_{t-1}^4 \sigma_t^{-1} \frac{\partial \sigma_{t-1}^2}{\partial z_2} - \sigma_t^{-3} \sigma_{t-1}^6 \frac{1}{2} \frac{\partial \sigma_t^2}{\partial z_2} \right)}{\sigma_{t-1}^{12} \sigma_t^{-2}} \rbrack \\
& - \frac{\rho^2 y_{t-1}^2}{2} \lbrack \left[\frac{\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2} \right)}{\sigma_{t-1}^3 \sigma_t} \right] \left[\frac{\left(2 \frac{\partial \sigma_t^2}{\partial z_1} \frac{\partial \sigma_{t-1}^2}{\partial z_1} \right)}{\sigma_{t-1}^4 \sigma_t} + \frac{\left(\frac{\partial \sigma_t^2}{\partial z_1} \right)^2}{\sigma_{t-1}^2 \sigma_t^3} - \frac{3 \left(\frac{\partial \sigma_{t-1}^2}{\partial z_1} \right)^2}{\sigma_{t-1}^6 \sigma_t^{-1}} \right] \\
& - \frac{2 \left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1} \right) \left(\frac{\partial \sigma_t^2}{\partial z_1} \frac{\partial \sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2} \frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2} \right)}{\sigma_{t-1}^6 \sigma_t^2} \\
& + \frac{\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1} \right) \left(3 \sigma_{t-1}^4 \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^6 \frac{\partial \sigma_t^2}{\partial z_2} \right)}{\sigma_{t-1}^{12} \sigma_t^4} \\
& - \frac{\left(\sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1^2} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1^2} \right) \left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2} \right)}{\sigma_{t-1}^6 \sigma_t^2} \rbrack,
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial^2\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z\partial\rho} &= \left[\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2\right]\left[\frac{\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^3}-\frac{\frac{\partial\sigma_t^2}{\partial z}}{\sigma_t\sigma_{t-1}}\right], \\
\frac{\partial^3\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\partial z_1\partial z_2\partial\rho} &= \left[\frac{\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right] \\
&\quad\times\left[\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2-\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}\right] \\
&\quad+\left[\frac{\left(\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}\sigma_t+\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^6\sigma_t^2}\right] \\
&\quad\times\left[\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2-\left(y_t-\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)y_{t-1}\right] \\
&\quad+2\rho y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2},
\end{aligned}$$

again for $\forall z, z_1, z_2 = \alpha, \beta$. Following Jensen and Rahbek (2004b), if we denote θ_0 the true parameter value $(\theta_0 = (w_0, \alpha_0, \beta_0, \gamma_0, \rho_0)')$, we introduce lower and upper values for each parameter in θ_0

$$w_L < w_0 < w_U; \quad \alpha_L < \alpha_0 < \alpha_U,$$

$$\beta_L < \beta_0 < \beta_U; \quad \gamma_L < \gamma_0 < \gamma_U; \quad \rho_L < \rho_0 < \rho_U,$$

and we define the neighborhood $N(\theta_0)$ around θ_0 as

$$(17) \qquad N(\theta_0) = \left\{\theta \backslash w_L \leq w \leq w_U, \alpha_L \leq \alpha \leq \alpha_U, \beta_L \leq \beta \leq \beta_U, \gamma_L < \gamma < \gamma_U \text{ and } \rho_L \leq \rho \leq \rho_U\right\}.$$

Then we establish now that the individual terms of each third order derivative are uniformly bounded

Lemma 5 *Under Assumptions 1 and 2, there exists a neighborhood $N(\theta_0)$ given in (17) for which*

$$(a) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\alpha^3}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{1t}; \qquad (b) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\beta^3}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{2t};$$

$$(c) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\rho^3}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{3t}; \qquad (d) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\beta}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{4t};$$

$$(e) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\rho}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{5t}; \qquad (f) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\beta^2\partial\rho}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{6t};$$

$$(g) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta^2}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{7t}; \qquad (h) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\rho^2}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{8t};$$

$$(i) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\beta\partial\rho^2}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{9t} \qquad (j) \quad \sup_{\theta\in N(\theta_0)}\left|\frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta\partial\rho}l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^T w_{10t},$$

where $w_{1t}, ..., w_{9t}$ and w_{10t} are stationary and have finite moments, $Ew_{it} = M_i < \infty, \forall i = 1, ..., 10$. Furthermore $\frac{1}{T}\sum_{t=1}^T w_{it} \xrightarrow{a.s.} M_i, \forall i = 1, ..., 10$.

Proof of Lemma 5 In this proof we change the notation slightly. We first define σ_t^2 when the conditional variance is evaluated at θ, and $\sigma_t^2(\theta_0)$ when it is evaluated at the true parameter. By definition

$$\frac{\left(y_t - \rho\sigma_t(\theta_0)\,\sigma_{t-1}^{-1}(\theta_0)\,y_{t-1}\right)^2}{\sigma_t^2(\theta_0)} = v_t^2.$$

For expressions (a), (b), (d) and (g), then $w_{1t}(\theta), w_{2t}(\theta), w_{4t}(\theta)$ and $w_{7t}(\theta)$ are given by

$$-\frac{1}{2}\sum_{t=1}^{T}\left(1 - \frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^3\sigma_t^2}{\partial z_1^2\partial z_2}}{\sigma_t^2} - \sum_{t=1}^{T}\left(1 - \frac{3\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6}$$
$$-\sum_{t=1}^{T}\left(2\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2} - 1\right)\frac{\left(\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1} + \frac{1}{2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^4},$$

with z_1 and z_2 being the corresponding α and β that are needed. If in all the three previous ratios we replace $\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\sigma_t^2}$ by

$$\frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t - \rho\sigma_t(\theta_0)\,\sigma_{t-1}^{-1}(\theta_0)\,y_{t-1}\right)^2}v_t^2,$$

then, using Lemmas 3 and 9 of Jensen and Rahbek (2004b) as in their Lemma 10, and knowing that

$$\frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \leq R_1 < \infty; \quad \frac{\sigma_t^2}{\sigma_t^2(\theta_0)} \leq R_2 < \infty\ ,$$

therefore

$$\begin{aligned}
\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t - \rho_0\sigma_t(\theta_0)\,\sigma_{t-1}^{-1}(\theta_0)\,y_{t-1}\right)^2}v_t^2 &= \frac{\left(y_t^2 + \rho^2\sigma_t^2\sigma_{t-1}^{-2}\sigma_{t-1}^2(\theta_0)\,\epsilon_{t-1}^2 - 2\rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}y_t\right)}{\sigma_t^2(\theta_0)} \\
&\leq \epsilon_t^2 + \rho^2\frac{\sigma_t^2}{\sigma_t^2(\theta_0)}\frac{\sigma_{t-1}^2(\theta_0)}{\sigma_{t-1}^2}\epsilon_{t-1}^2 + \frac{\sigma_t^2}{\sigma_t^2(\theta_0)}\sqrt{\frac{\sigma_{t-1}^2(\theta_0)}{\sigma_{t-1}^2}}2\left|\rho\right|\left|\epsilon_{t-1}\epsilon_t\right| \\
&\leq \epsilon_t^2 + \rho^2R_1R_2\epsilon_{t-1}^2 + 2\left|\rho\right|R_2\sqrt{R_1}\left|\epsilon_{t-1}\epsilon_t\right|,
\end{aligned}$$

and, for example, for the second of the previous ratios

$$\begin{aligned}
D_1 &\equiv \left(3\frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t - \rho\sigma_t(\theta_0)\,\sigma_{t-1}^{-1}(\theta_0)\,y_{t-1}\right)^2}v_t^2 - 1\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6} \\
&\leq \left(3\frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\frac{\left(y_t - \rho\sigma_t\sigma_{t-1}^{-1}y_{t-1}\right)^2}{\left(y_t - \rho\sigma_t(\theta_0)\,\sigma_{t-1}^{-1}(\theta_0)\,y_{t-1}\right)^2}v_t^2 + 1\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6} \\
&\leq 3A\left(R_1\epsilon_t^2 + \max\left(\rho_L^2, \rho_U^2\right)R_1^2R_2\epsilon_{t-1}^2 + 2\max\left(\left|\rho_L\right|, \left|\rho_U\right|\right)R_2R_1\left|\epsilon_{t-1}\epsilon_t\right| + 1/3\right) \\
&\equiv w_{1t}^*,
\end{aligned}$$

where A is the lower bound that is obtained in Lemma 9 of Jensen and Rahbek (2004b). Finally $E\left(w_{it}^{*}\right) < \infty$, $\forall i = 1, 2, 4, 7$ as $E(\epsilon_t^2) < \infty$, $E(\epsilon_{t-1}^2) < \infty$ and $E(|\epsilon_{t-1}\epsilon_t|) < \infty$ provided that $|\rho_0| < 1$ and $E(v_t^2) < \infty$. For the remaining ratios (and using the results of Lemma 4 with the same type of ratios that have been already analyzed), it is enough to deal with the extra expression

$$\frac{\partial^3 \left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right)^2}{\partial z_1^2 \partial z_2}.$$

The most complicated terms of the previous expression are the terms

$$\begin{aligned}
&-\frac{\rho^2 y_{t-1}^2}{2\sigma_{t-1}^2}[\left[\frac{\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}\sigma_t^3}\right] \left[\frac{\left(\beta + \alpha\epsilon_{t-1}^2\right)\left(2\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^2 \sigma_t \left(\sigma_t^2 - 1\right)} + \frac{\left(\beta + \alpha\epsilon_{t-1}^2\right)\left(\frac{\partial \sigma_t^2}{\partial z_1}\right)^2}{\left(\sigma_t^2 - 1\right)\sigma_t^3} - \frac{3\left(\frac{\partial \sigma_{t-1}^2}{\partial z_1}\right)^2}{\sigma_{t-1}^6 \sigma_t^{-1}}\right] \\
&-\frac{2\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1}\right)\left(\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2}\right)}{\sigma_{t-1}^6 \sigma_t^4} \\
&+\frac{\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1}\right)\left(3\sigma_{t-1}^4 \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2} + \sigma_{t-1}^6 \frac{\partial \sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^6} \\
&-\frac{\left(\sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1^2} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1^2}\right)\left(\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^6 \sigma_t^4}],
\end{aligned}$$

for $\forall z_1, z_2 = \alpha, \beta$, where we have re-arranged the terms and used the fact that $\sigma_{t-1}^2 = \frac{\left(\sigma_t^2 - 1\right)}{\left(\beta + \alpha\epsilon_{t-1}^2\right)}$. Applying lemma 9 of Jensen and Rahbek (2004b) directly again, we get that the previous expectation is bounded. The proof for expression (c) is trivial. For the proof of expressions (e) (f) and (j), we need to consider the extra ratios

$$\frac{\frac{\partial^3 \left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right)^2}{\partial z_1 \partial z_2 \partial \rho}}{\sigma_t^2}, \text{ and } \frac{\frac{\partial^2 \left(y_t - \rho \sigma_t \sigma_{t-1}^{-1} y_{t-1}\right)^2}{\partial z \partial \rho}}{\sigma_t^2}.$$

For the first term, we use the law of iterated expectations and the fact that

$$\left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2}}{\sigma_{t-1}^2 \sigma_t^2}\right]\left[\frac{\rho y_{t-1}^2}{\sigma_{t-1}^2}\right],$$

and

$$\left[\frac{\left(\sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1}\right)\left(3\sigma_{t-1}\frac{\partial \sigma_{t-1}^2}{\partial z_2}\sigma_t + \sigma_{t-1}^3 \frac{\partial \sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^5 \sigma_t^3}\right]\left[\frac{\rho y_{t-1}^2}{\sigma_{t-1}^2}\right],$$

as well as

$$\frac{2\rho y_{t-1}^2}{\sigma_{t-1}^2}\frac{\left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^4 \sigma_t^4},$$

for $\forall z_1, z_2 = \alpha, \beta$ are all bounded. For the second term we need

$$\left[\frac{\rho y_{t-1}^2}{\sigma_{t-1}^2}\right]\left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2} - \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^2}\right],$$

which is also bounded. For the proof of expressions (h) and (i), we have

$$w_{it}\left(\theta\right) = \epsilon_{t-1}^2 \frac{\frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^2};\ \ \forall z = \alpha, \beta;\ i = 8, 9.$$

and due to Assumption 1, $E\left(w_{it}\right) < \infty$, $i = 8, 9$. ■

Appendix 2

Proof of Theorem 3 We begin again with the first order derivatives (Lemma 6). Later, we move to the second and third order derivatives (Lemmas 7 and 8). Brown (1971) provides the type of central limit theorem we need. Lemmas 1 and 2 also hold for the AR(p), and they are used in our proofs. ■

First order derivatives The first order derivatives are given by ($\forall z = \alpha, \beta$)

$$\frac{\partial}{\partial z} l_T\left(\theta\right) = -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1 - \frac{\left(\sigma_t\left(\theta\right) v_t\right)^2}{\sigma_t^2\left(\theta\right)}\right)\frac{\frac{\partial \sigma_t^2(\theta)}{\partial z}}{\sigma_t^2\left(\theta\right)} + \frac{\frac{\partial\left(y_t - \sum_{i=1}^{p}\rho_i \sigma_t(\theta)\sigma_{t-i}^{-1}(\theta) y_{t-i}\right)^2}{\partial z}}{\sigma_t^2\left(\theta\right)}\right),$$

with

$$\begin{aligned} \frac{\partial}{\partial \alpha} l_T\left(\theta\right) &= \sum_{t=1}^{T} s_{1t}\left(\theta\right), \\ \frac{\partial}{\partial \beta} l_T\left(\theta\right) &= \sum_{t=1}^{T} s_{2t}\left(\theta\right), \end{aligned}$$

and

$$\frac{\partial}{\partial \rho_i} l_T\left(\theta\right) = \sum_{t=1}^{T} s_{(2+i)t}\left(\theta\right) = \sum_{t=1}^{T}\frac{\left(\sigma_t\left(\theta\right) v_t\right)\sigma_{t-i}^{-1}\left(\theta\right) y_{t-i}}{\sigma_t\left(\theta\right)}; \forall i = 1, ..., p,$$

where

$$\begin{aligned} \frac{\frac{\partial \sigma_t^2(\theta)}{\partial \alpha}}{\sigma_t^2\left(\theta\right)} &= \frac{\sum_{j=1}^{T}\beta^{j-1} y_{t-j}^2}{\sigma_t^2\left(\theta\right)}, \\ \frac{\frac{\partial \sigma_t^2(\theta)}{\partial \beta}}{\sigma_t^2\left(\theta\right)} &= \sum_{j=1}^{T}\beta^{j-1}\frac{\sigma_{t-j}^2\left(\theta\right)}{\sigma_t^2\left(\theta\right)}, \end{aligned}$$

$$\frac{\partial \left(y_t - \sum_{i=1}^{p} \rho_i \sigma_t \left(\theta\right) \sigma_{t-i}^{-1} \left(\theta\right) y_{t-i}\right)^2}{\partial z} = -\left(\sigma_t \left(\theta\right) v_t\right) \left[\sum_{i=1}^{p} \frac{\left(\sigma_{t-i}^2 \left(\theta\right) \frac{\partial \sigma_t^2(\theta)}{\partial z} - \sigma_t^2 \left(\theta\right) \frac{\partial \sigma_{t-i}^2(\theta)}{\partial z}\right) \rho_i y_{t-i}}{\sigma_{t-i}^3 \left(\theta\right) \sigma_t \left(\theta\right)}\right],$$

Then,

Lemma 6 *Let Assumptions 2 and 3 hold and write the scores associated with (7) as* $s_{lt}\left(\theta_0\right) = s_{lt}$. *Then*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{1t} \xrightarrow{d} N\left(0, \frac{\zeta}{4\alpha_0^2}\right),$$
$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{2t} \xrightarrow{d} N\left(0, \frac{\zeta \left(1+\mu_1\right) \mu_2}{4\beta_0^2 \left(1-\mu_1\right)\left(1-\mu_2\right)}\right),$$
$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{(2+i)t} \xrightarrow{d} N\left(0, \frac{1}{\left(1-\rho_{i0}^2\right)}\right),$$

for $\forall i = 1, ..., p,$ *with* $\mu_j = E\left(\beta_0 / \left(\alpha_0 \epsilon_t^2 + \beta_0\right)\right)^j, \quad j = 1, 2$ *as* $T \longrightarrow \infty$.

Proof of Lemma 6 The proof follows the same type of argument as Lemma 3 and the proof in Jensen and Rahbek (2004b). Note that Lemma 6 is the same as Lemma 3, but instead of for $i = 1$ (the first lag) for $i = 1, ..., p$. Using the law of iterated expectations and the properties of v_t, $E\left(s_{1t}/I_{t-1}\right) = E\left(s_{2t}/I_{t-1}\right) = E\left(s_{(2+i)t}/I_{t-1}\right) = 0$. Also, by the same argument as in Lemma 3

$$E\left|s_{1t}\right| < \infty; \qquad E\left|s_{2t}\right| < \infty; \qquad E\left|s_{(2+i)t}\right| < \infty; \ \forall i = 1, ..., p.$$

Besides

$$\frac{1}{T} \sum_{t=1}^{T} E\left(s_{1t}^2 / I_{t-1}\right) \xrightarrow{p} \frac{\zeta}{4\alpha_0^2},$$
$$\frac{1}{T} \sum_{t=1}^{T} E\left(s_{2t}^2 / I_{t-1}\right) \xrightarrow{p} \frac{\zeta \left(1+\mu_1\right) \mu_2}{4\beta_0^2 \left(1-\mu_1\right)\left(1-\mu_2\right)},$$
$$\frac{1}{T} \sum_{t=1}^{T} E\left(s_{1t} s_{2t} / I_{t-1}\right) \xrightarrow{p} \frac{\zeta \mu_1}{4\alpha_0 \beta_0 \left(1-\mu_1\right)},$$
$$\frac{1}{T} \sum_{t=1}^{T} E\left(s_{(2+i)t}^2 / I_{t-1}\right) \xrightarrow{p} \frac{1}{\left(1-\rho_{i0}^2\right)}.$$

■

Second order derivatives The second order derivatives are given by (σ_t should be evaluated at θ in all the expressions, but for reasons of space, we have omitted that)

$$
\begin{aligned}
\frac{\partial^2}{\partial z_1 \partial z_2} l_T(\theta) &= \frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{2(\sigma_t v_t)^2}{\sigma_t^2}\right)\frac{\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial \sigma_t^2}{\partial z_2}}{\sigma_t^4}+\left(\frac{(\sigma_t v_t)^2}{\sigma_t^2}-1\right)\frac{\frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2}}{\sigma_t^2}-\frac{\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1\partial z_2}}{\sigma_t^2}\right) \\
&+\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\frac{\partial \sigma_t^2}{\partial z_1}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_2}+\frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1}}{\sigma_t^4}\right), \\
\frac{\partial^2}{\partial \rho_i \partial \rho_j} l_T(\theta) &= -\sum_{t=1}^{T}\sigma_{t-i}^{-1}y_{t-i}\sigma_{t-j}^{-1}y_{t-j}, \\
\frac{\partial^2}{\partial z \partial \rho_i} l_T(\theta) &= \sum_{t=1}^{T}\frac{\frac{\partial\left[\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)\sigma_t\sigma_{t-i}^{-1}y_{t-i}\right]}{\partial z}}{\sigma_t^2}-\sum_{t=1}^{T}\frac{(\sigma_t v_t)\,\sigma_{t-i}^{-1}y_{t-i}\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t^3},
\end{aligned}
$$

for $\forall i, j = 1, ..., p$,and $\forall z, z_1, z_2 = \alpha, \beta$, where

$$
\begin{aligned}
\frac{\frac{\partial^2 \sigma_t^2}{\partial \alpha^2}}{\sigma_t^2} &= 2\frac{\sum_{j=1}^{T}(j-1)\,\beta^{j-2}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^2 \sigma_t^2}{\partial \beta^2}}{\sigma_t^2} &= 2\sum_{j=1}^{t}(j-1)\,\beta^{j-2}\frac{\sigma_{t-j}^2}{\sigma_t^2}, \\
\frac{\partial^2(\sigma_t v_t)^2}{\partial z_1 \partial z_2} &= -(\sigma_t v_t)\left[(\rho_1 y_{t-1})\left(\frac{\frac{\partial \sigma_{t-1}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}}{\sigma_{t-1}^3\sigma_t}\right)\right] \\
&-...-(\sigma_t v_t)\left[(\rho_p y_{t-p})\left(\frac{\frac{\partial \sigma_{t-p}^2}{\partial z_2}\frac{\partial \sigma_t^2}{\partial z_1}+\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial \sigma_t^2}{\partial z_2}\frac{\partial \sigma_{t-p}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}}{\sigma_{t-p}^3\sigma_t}\right)\right] \\
&+(\sigma_t v_t)\left[(\rho_1 y_{t-1})\frac{\left(\sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial \sigma_{t-1}^2}{\partial z_1}\right)\left(3\frac{\partial \sigma_{t-1}^2}{\partial z_2}\sigma_t+\sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-1}^5\sigma_t^2}\right] \\
&+...+(\sigma_t v_t)\left[(\rho_p y_{t-p})\frac{\left(\sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial \sigma_{t-p}^2}{\partial z_1}\right)\left(3\frac{\partial \sigma_{t-p}^2}{\partial z_2}\sigma_t+\sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_2}\sigma_t^{-1}\right)}{2\sigma_{t-p}^5\sigma_t^2}\right] \\
&+\rho_1^2y_{t-1}^2\frac{\left[\sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial \sigma_{t-1}^2}{\partial z_1}\right]\left[\sigma_{t-1}^2\frac{\partial \sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial \sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6\sigma_t^2} \\
&+...+\rho_p^2y_{t-p}^2\frac{\left[\sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial \sigma_{t-p}^2}{\partial z_1}\right]\left[\sigma_{t-p}^2\frac{\partial \sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial \sigma_{t-p}^2}{\partial z_2}\right]}{2\sigma_{t-p}^6\sigma_t^2},
\end{aligned}
$$

and

$$\frac{\partial\left[\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t\left(\theta\right)\sigma_{t-i}^{-1}\left(\theta\right)y_{t-i}\right)\sigma_t\sigma_{t-i}^{-1}y_{t-i}\right]}{\partial z}=-\frac{y_ty_{t-i}}{2}\left[\sigma_t\sigma_{t-i}^{-3}\frac{\partial\sigma_{t-i}^2}{\partial z}-\sigma_t^{-1}\sigma_{t-i}^{-1}\frac{\partial\sigma_t^2}{\partial z}\right]$$
$$-\rho_1y_{t-1}y_{t-i}\left[\sigma_{t-1}^{-1}\sigma_{t-i}^{-1}\frac{\partial\sigma_t^2}{\partial z}-\frac{\sigma_t^2}{2}\left[\sigma_{t-1}^3\sigma_{t-i}^{-1}\frac{\partial\sigma_{t-1}^2}{\partial z}+\sigma_{t-1}^{-1}\sigma_{t-i}^{-3}\frac{\partial\sigma_{t-i}^2}{\partial z}\right]\right]$$
$$-\ldots-\rho_py_{t-p}y_{t-i}\left[\sigma_{t-p}^{-1}\sigma_{t-i}^{-1}\frac{\partial\sigma_t^2}{\partial z}-\frac{\sigma_t^2}{2}\left[\sigma_{t-p}^3\sigma_{t-i}^{-1}\frac{\partial\sigma_{t-p}^2}{\partial z}+\sigma_{t-p}^{-1}\sigma_{t-i}^{-3}\frac{\partial\sigma_{t-i}^2}{\partial z}\right]\right].$$

Then

Lemma 7 *Under Assumptions 2 and 3, with the expressions of the second order derivatives evaluated at* θ_0

(a) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha^2}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}\frac{1}{2\alpha_0^2}>0,$

(b) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta^2}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}\frac{(1+\mu_1)\mu_2}{2\beta_0^2(1-\mu_1)(1-\mu_2)}>0,$

(c) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\beta}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}\frac{\mu_1}{2\alpha_0\beta_0(1-\mu_1)},$

(d) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\alpha\partial\rho_i}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}0,\ \forall i=1,...,p,$

(e) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\rho_i}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}0,\ \forall i=1,...,p,$

(f) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\rho_i^2}l_T\left(\theta\right)|_{\theta=\theta_0}\right)\xrightarrow{p}\frac{1}{\left(1-\rho_{i0}^2\right)}>0,\ \forall i=1,...,p,$

with $\mu_j=E\left(\beta_0/\left(\alpha_0\epsilon_t^2+\beta_0\right)\right)^j,\quad j=1,2$ *as* $T\longrightarrow\infty.$

Proof of Lemma 7 The proof follows the same type of arguments as in Lemma 4. Note that Lemma 7 is the same as Lemma 4, but instead of for $i=1$ (the first lag) for $i=1,...,p$. All the results of the proof of Lemma 4 apply here directly. ■

Third order derivatives The third order derivatives are given by (σ_t should be evaluated at θ in all the expressions, but for reasons of space, we have omitted that)

$$\begin{aligned}
\frac{\partial^3}{\partial z_1^2\partial z_2}l_T\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{\left(\sigma_t v_t\right)^2}{\sigma_t^2}\right)\frac{\frac{\partial^3\sigma_t^2}{\partial z_1^2\partial z_2}}{\sigma_t^2}-\sum_{t=1}^{T}\left(2\frac{\left(\sigma_t v_t\right)^2}{\sigma_t^2}-1\right)\frac{\left(\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}+\frac{1}{2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^4} \\
&\quad -\sum_{t=1}^{T}\left(1-\frac{3\left(\sigma_t v_t\right)^2}{\sigma_t^2}\right)\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_t^6}-\frac{1}{2}\sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1^2\partial z_2}}{\sigma_t^2} \\
&\quad +\sum_{t=1}^{T}\left(\frac{\left(\frac{1}{2}\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1^2}\frac{\partial\sigma_t^2}{\partial z_2}+\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1\partial z_2}\frac{\partial\sigma_t^2}{\partial z_1}\right)}{\sigma_t^4}\right) \\
&\quad +\sum_{t=1}^{T}\left(\frac{\left(\frac{1}{2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_2}+\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1}\right)}{\sigma_t^4}\right) \\
&\quad -\sum_{t=1}^{T}\frac{\left(\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_2}\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2+2\frac{\partial\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_t^6}, \\
\frac{\partial^3}{\partial z_1\partial z_2\partial\rho_i}l_T\left(\theta\right) &= \sum_{t=1}^{T}\frac{2\left(\sigma_t v_t\right)\sigma_t y_{t-i}\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_t^2}{\partial z_2}}{\sigma_{t-i}\sigma_t^6}+\sum_{t=1}^{T}\frac{\left(\sigma_t v_t\right)\sigma_t y_{t-i}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}}{\sigma_{t-i}\sigma_t^4}-\sum_{t=1}^{T}\frac{\frac{\partial^3\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1\partial z_2\partial\rho_i}}{2\sigma_t^2} \\
&\quad +\frac{1}{2}\sum_{t=1}^{T}\left(\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_2\partial\rho_i}+\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t(\theta)\sigma_{t-i}^{-1}(\theta)y_{t-i}\right)^2}{\partial z_1\partial\rho_i}\right)}{\sigma_t^4}\right), \\
\frac{\partial^3}{\partial z\partial\rho_i\partial\rho_j}l_T\left(\theta\right) &= \sum_{t=1}^{T}\frac{y_{t-i}y_{t-j}\left(\sigma_{t-i}\frac{\partial\sigma_{t-j}}{\partial z}+\sigma_{t-j}\frac{\partial\sigma_{t-i}}{\partial z}\right)}{\sigma_{t-i}^2\sigma_{t-j}^2}, \\
\frac{\partial^3}{\partial\rho_i\partial\rho_j\partial\rho_k}l_T\left(\theta\right) &= 0,
\end{aligned}$$

for $\forall z_1, z_2=\alpha,\beta$ and $\forall i,j,k=1,...,p$ where

$$\begin{aligned}
\frac{\frac{\partial^3\sigma_t^2}{\partial\alpha^3}}{\sigma_t^2} &= 3\frac{\sum_{j=1}^{T}\left(j-1\right)\left(j-2\right)\beta^{j-3}y_{t-j}^2}{\sigma_t^2}, \\
\frac{\frac{\partial^3\sigma_t^2}{\partial\beta^3}}{\sigma_t^2} &= 3\sum_{j=1}^{t}\left(j-1\right)\left(j-2\right)\beta^{j-3}\frac{\sigma_{t-j}^2}{\sigma_t^2},
\end{aligned}$$

and

$$\frac{\partial^3\left(y_t-\sum_{i=1}^{p}\rho_i\sigma_t\left(\theta\right)\sigma_{t-i}^{-1}\left(\theta\right)y_{t-i}\right)^2}{\partial z_1^2\partial z_2}=-\rho_1 y_{t-1}\left(\sigma_t v_t\right)[\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}+\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\sigma_{t-1}^2-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^3\sigma_{t-1}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t}$$

$$-\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(3\sigma_{t-1}\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_t^{-1}\sigma_{t-1}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-1}^6\sigma_t^2}-\frac{\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t}$$

$$+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(2\sigma_{t-1}^2\sigma_t\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\frac{1}{2}\sigma_t^{-1}\sigma_{t-1}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^8\sigma_t^2}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^2\sigma_t^3}+\frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^{-1}}$$

$$+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_t\sigma_{t-1}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^4\sigma_t^6}-\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2\left(3\sigma_{t-1}^4\sigma_t^{-1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}-\sigma_t^{-3}\sigma_{t-1}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^{-2}}]$$

$$-\ldots-\rho_p y_{t-p}\left(\sigma_t v_t\right)[\frac{\left(\frac{\partial\sigma_{t-p}^2}{\partial z_2}\frac{\partial^2\sigma_t^2}{\partial z_1^2}+\frac{\partial^3\sigma_{t-p}^2}{\partial z_1^2\partial z_2}\sigma_{t-p}^2-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^3\sigma_{t-p}^2}{\partial z_1^2\partial z_2}\right)}{\sigma_{t-p}^3\sigma_t}$$

$$-\frac{\left(\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}\right)\left(3\sigma_{t-p}\sigma_t\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_t^{-1}\sigma_{t-p}^3\frac{\partial\sigma_t^2}{\partial z_2}\right)}{2\sigma_{t-p}^6\sigma_t^2}-\frac{\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}+\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^4\sigma_t}$$

$$+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(2\sigma_{t-p}^2\sigma_t\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\frac{1}{2}\sigma_t^{-1}\sigma_{t-p}^4\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^8\sigma_t^2}-\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^2\sigma_t^3}+\frac{3\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-p}^6\sigma_t^{-1}}$$

$$+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2\left(\frac{\sigma_t^3}{2}\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_t\sigma_{t-p}^2\frac{3}{4}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^4\sigma_t^6}-\frac{\frac{3}{2}\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)^2\left(3\sigma_{t-p}^4\sigma_t^{-1}\frac{\partial\sigma_{t-p}^2}{\partial z_2}-\sigma_t^{-3}\sigma_{t-p}^6\frac{1}{2}\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^{12}\sigma_t^{-2}}]$$

$$-\frac{\rho_1^2y_{t-1}^2}{2}[\left[\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)}{\sigma_{t-1}^4\sigma_t}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-1}^2\sigma_t^3}-\frac{3\left(\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)^2}{\sigma_{t-1}^6\sigma_t^{-1}}\right]$$

$$-\frac{2\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-1}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}$$

$$+\frac{\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_1}\right)\left(3\sigma_{t-1}^4\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}+\sigma_{t-1}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-1}^{12}\sigma_t^4}-\frac{\left(\sigma_{t-1}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-1}^2}{\partial z_1^2}\right)\left(\sigma_{t-1}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-1}^2}{\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}]$$

$$-\ldots-\frac{\rho_p^2y_{t-p}^2}{2}[\left[\frac{\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right)}{\sigma_{t-p}^3\sigma_t}\right]\left[\frac{\left(2\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)}{\sigma_{t-p}^4\sigma_t}+\frac{\left(\frac{\partial\sigma_t^2}{\partial z_1}\right)^2}{\sigma_{t-p}^2\sigma_t^3}-\frac{3\left(\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)^2}{\sigma_{t-p}^6\sigma_t^{-1}}\right]$$

$$-\frac{2\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(\frac{\partial\sigma_t^2}{\partial z_1}\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1\partial z_2}-\frac{\partial\sigma_t^2}{\partial z_2}\frac{\partial\sigma_{t-p}^2}{\partial z_1}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1\partial z_2}\right)}{\sigma_{t-1}^6\sigma_t^2}$$

$$+\frac{\left(\sigma_{t-p1}^2\frac{\partial\sigma_t^2}{\partial z_1}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_1}\right)\left(3\sigma_{t-p}^4\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}+\sigma_{t-p}^6\frac{\partial\sigma_t^2}{\partial z_2}\right)}{\sigma_{t-p}^{12}\sigma_t^4}+\frac{\left(\sigma_{t-p}^2\frac{\partial^2\sigma_t^2}{\partial z_1^2}-\sigma_t^2\frac{\partial^2\sigma_{t-p}^2}{\partial z_1^2}\right)\left(\sigma_{t-p}^2\frac{\partial\sigma_t^2}{\partial z_2}-\sigma_t^2\frac{\partial\sigma_{t-p}^2}{\partial z_2}\right)}{\sigma_{t-p}^6\sigma_t^2}],$$

for $\forall z_1, z_2 = \alpha, \beta$. Furthermore,

$$
\begin{aligned}
\frac{\partial^2 \left(y_t - \sum_{i=1}^{p} \rho_i \sigma_t(\theta) \sigma_{t-i}^{-1}(\theta) y_{t-i}\right)^2}{\partial z \partial \rho} &= \left[(\sigma_t v_t) y_{t-1} - \rho_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1}^2\right] \left[\frac{\sigma_t \frac{\partial \sigma_{t-1}^2}{\partial z}}{\sigma_{t-1}^3} - \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t \sigma_{t-1}}\right] \\
&\quad +... + \left[(\sigma_t v_t) y_{t-p} - \rho_p \sigma_t \sigma_{t-p}^{-1} y_{t-p}^2\right] \left[\frac{\sigma_t \frac{\partial \sigma_{t-p}^2}{\partial z}}{\sigma_{t-p}^3} - \frac{\frac{\partial \sigma_t^2}{\partial z}}{\sigma_t \sigma_{t-p}}\right] \\
\frac{\partial^3 (\sigma_t v_t)^2}{\partial z_1 \partial z_2 \partial \rho} &= \left[\frac{\frac{\partial \sigma_{t-1}^2}{\partial z_2} \frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-1}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2} \frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-1}^2}{\partial z_1 \partial z_2}}{\sigma_{t-1}^3 \sigma_t}\right] \\
&\quad \times \left[\rho_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1}^2 - (\sigma_t v_t) y_{t-1}\right] \\
&\quad +... + \left[\frac{\frac{\partial \sigma_{t-p}^2}{\partial z_2} \frac{\partial \sigma_t^2}{\partial z_1} + \sigma_{t-p}^2 \frac{\partial^2 \sigma_t^2}{\partial z_1 \partial z_2} - \frac{\partial \sigma_t^2}{\partial z_2} \frac{\partial \sigma_{t-p}^2}{\partial z_1} - \sigma_t^2 \frac{\partial^2 \sigma_{t-p}^2}{\partial z_1 \partial z_2}}{\sigma_{t-p}^3 \sigma_t}\right] \\
&\quad \times \left[\rho_p \sigma_t \sigma_{t-p}^{-1} y_{t-p}^2 - (\sigma_t v_t) y_{t-p}\right] \\
&\quad + \left[\frac{\left(\sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1} - \sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1}\right) \left(3\sigma_{t-1} \frac{\partial \sigma_{t-1}^2}{\partial z_2} \sigma_t + \sigma_{t-1}^3 \frac{\partial \sigma_t^2}{\partial z_2} \sigma_t^{-1}\right)}{2\sigma_{t-1}^6 \sigma_t^2}\right] \\
&\quad \times \left[\rho_1 \sigma_t \sigma_{t-1}^{-1} y_{t-1}^2 - (\sigma_t v_t) y_{t-1}\right] \\
&\quad +... + \left[\frac{\left(\sigma_t^2 \frac{\partial \sigma_{t-p}^2}{\partial z_1} - \sigma_{t-p}^2 \frac{\partial \sigma_t^2}{\partial z_1}\right) \left(3\sigma_{t-p} \frac{\partial \sigma_{t-p}^2}{\partial z_2} \sigma_t + \sigma_{t-p}^3 \frac{\partial \sigma_t^2}{\partial z_2} \sigma_t^{-1}\right)}{2\sigma_{t-p}^6 \sigma_t^2}\right] \\
&\quad \times \left[\rho_p \sigma_t \sigma_{t-p}^{-1} y_{t-p}^2 - (\sigma_t v_t) y_{t-p}\right] \\
&\quad + 2\rho_1 y_{t-1}^2 \frac{\left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_1}\right] \left[\sigma_{t-1}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-1}^2}{\partial z_2}\right]}{2\sigma_{t-1}^6 \sigma_t^2} \\
&\quad +... + 2\rho_p y_{t-p}^2 \frac{\left[\sigma_{t-p}^2 \frac{\partial \sigma_t^2}{\partial z_1} - \sigma_t^2 \frac{\partial \sigma_{t-p}^2}{\partial z_1}\right] \left[\sigma_{t-p}^2 \frac{\partial \sigma_t^2}{\partial z_2} - \sigma_t^2 \frac{\partial \sigma_{t-p}^2}{\partial z_2}\right]}{2\sigma_{t-p}^6 \sigma_t^2},
\end{aligned}
$$

for $\forall z, z_1, z_2 = \alpha, \beta$. Following again Jensen and Rahbek (2004b), if we denote θ_0 the true parameter value $(\theta_0 = (w_0, \alpha_0, \beta_0, \gamma_0, \rho_{10}, ..., \rho_{p0})')$, we introduce lower and upper values for each parameter in θ_0

$$
\begin{gathered}
w_L < w_0 < w_U; \quad \alpha_L < \alpha_0 < \alpha_U; \quad \beta_L < \beta_0 < \beta_U, \\
\gamma_L < \gamma_0 < \gamma_U; \quad \rho_L < \rho_{10} < \rho_U; ...; \rho_L < \rho_{p0} < \rho_U,
\end{gathered}
$$

and we define the neighborhood $N(\theta_0)$ around θ_0 as

(18)

$$N(\theta_0) = \{\theta \backslash w_L \leq w \leq w_U, \alpha_L \leq \alpha \leq \alpha_U, \beta_L \leq \beta \leq \beta_U, \gamma_L < \gamma < \gamma_U, \rho_L < \rho_1 < \rho_U, \rho_L \leq \rho_p \leq \rho_U\},$$

Then we can establish that the individual terms of each third order derivative are uniformly bounded. In particular,

Lemma 8 *Under Assumptions 2 and 3, there exists a neighborhood $N(\theta_0)$ given in (18) for which*

$$(a) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{\partial^3}{\partial\alpha^3} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{1t}; \qquad (b) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{\partial^3}{\partial\beta^3} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{2t};$$

$$(c) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\rho_i\partial\rho_j\partial\rho_k} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{3t}; \qquad (d) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\beta} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{4t};$$

$$(e) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\alpha^2\partial\rho_i} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{5t}; \qquad (f) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\beta^2\partial\rho_i} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{6t};$$

$$(g) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta^2} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{7t}; \qquad (h) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\rho_i\partial\rho_j} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{8t};$$

$$(i) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\beta\partial\partial\rho_i\partial\rho_j} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{9t}; \qquad (j) \quad \sup_{\theta\in N(\theta_0)} \left| \frac{1}{T}\frac{\partial^3}{\partial\alpha\partial\beta\partial\rho_i} l_T(\theta) \right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{10t},$$

with $i, j, k = 1, ...p$, where $w_{1t}, ..., w_{9t}$ and w_{10t} are stationary and have finite moments, $Ew_{lt} = M_l < \infty, \forall l = 1, ..., 10$. Furthermore $\frac{1}{T}\sum_{t=1}^{T} w_{lt} \xrightarrow{a.s.} M_l, \forall l = 1, ..., 10$.

Proof of Lemma 8 The proof follows the same argument as Lemma 5. Note that Lemma 8 is the same as Lemma 5, but instead of for $i = 1$ (the first lag), for $i = 1, ..., p$. All the results of the proof of Lemma 5 apply here directly. ■

REFERENCES

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* 31, 307-327.

Bougerol P. and N. Picard (1992a), Strict Stationarity of Generalized Autoregressive Processes, *The Annals of Probability* 20, 4, 1714-1730.

Bougerol P. and N. Picard (1992b), Stationarity of GARCH Processes and of some non-negative time series, *Journal of Econometrics* 52, 115-127.

Brown, B. M. (1971), Martingale Central Limit Theorems, *Annals of Mathematical Statistics* 42, 59-66.

Cox, J., J. Ingersoll and S. Ross (1985), A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.

Dahl, C. M., and M. Levine (2005), Nonparametric estimation of volatility models with serially dependent innovations. *Unpublished manuscript* (available at http://www.mgmt.purdue.edu /faculty/dahlc), Purdue University.

Drost, F. C. and T. E. Nijman (1993), Temporal Aggregation of GARCH Processes, *Econometrica* 61, 909-927.

Engle, R. F. (1982), Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, 987-1007.

Engle, R. F., D. M. Lilien and R. P. Robins (1987), Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model, *Econometrica* 55, 2, 391-407.

Engle, R. F. and K. F. Kroner (1995), Multivariate Simultaneous Generalised ARCH, *Econometric Theory* 11, 122-150.

Galbraith, R. F. and J. I. Galbraith (1974), On the Inverses of Some Patterned Matrices Arising in the Theory of Stationary Time Series, *Journal of Applied Probability* 11, 63-71.

Hamilton, J. D. (1994), *Time Series Analysis.* Princeton University Press, Princeton.

Jensen, S. T. and A. Rahbek (2004a), Asymptotic Normality of the QML Estimator of ARCH in the Nonstationary Case, *Econometrica*, 72, 2, 641-646.

Jensen, S. T. and A. Rahbek (2004b), Asymptotic Inference for Nonstationary GARCH, *Econometric Theory* 20, 6, 1203-1226.

Lee, S. and B. Hansen (1994), Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, *Econometric Theory* 10, 29-52.

Ling, S. and McAleer, M. (2003), Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280-310.

Ling, S. (2004), Estimation and Testing Stationarity for Double-Autoregressive Models, *Journal of the Royal Statistical Society Series B* 66, 1, 63-78.

Linton, O. and E. Mammen (2005), Estimating Semiparametric ARCH(∞) Models by Kernel Smoothing Methods, *Econometrica* 73, 3, 771-836.

Lumsdaine, R. L. (1996), Consistency and Asymptotic Normality of the Quasi Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica* 64, 3, 575-596.

Meddahi, N. and E. Renault (2004), Temporal Aggregation of Volatility Models, *Journal of Econometrics* 119, 355-379.

Nelson, D. B. (1990), Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory* 6, 318-334.

White, H. (1984), *Asymptotic Theory for Econometricians.* New York: Academic Press.

Wooldridge, J. (1991), On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances, *Journal of Econometrics* 47, 5-46.