The limiting properties of the QMLE in a general class of asymmetric volatility models*

Christian M. Dahl[†]
CREATES, School of Economics and Management
University of Aarhus

Emma M. Iglesias[‡]
Department of Economics
Michigan State University

This version: July 2008.

Abstract

In this paper we analyze the limiting properties of the estimated parameters in a general class of asymmetric volatility models which are closely related to the traditional exponential GARCH model. The new representation has three main advantages over the traditional EGARCH: (1) It allows a much more flexible rep-

*We wish to thank Hira Koul, Timo Terasvirta, Tim Vogelsang and participants at a seminar at MSU, the Summer Econometrics Conference at University of Aarhus, the 2007 Canadian Econometrics Study Group and the 2007 Midwest Econometrics Group meetings for very helpful comments. The first author gratefully acknowledges the research support of CREATES (funded by the Danish National Research Foundation). The second author acknowledges support from the MSU Intramural Research Grants Program.

†Address: School of Economics and Management, Room 225, Building 1326, DK-8000 Aarhus C. Phone: +45 8942 1559. E-mail: cdahl@econ.au.dk.

‡Corresponding author. Address: Department of Economics, Michigan State University, 101 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA. Phone: 517-353-9916. E-mail: iglesia5@msu.edu.

resentation of the conditional variance function. (2) It is possible to provide a complete characterization of the asymptotic distribution of the QML estimator based on the new class of nonlinear volatility models, something which has proven very difficult even for the traditional EGARCH. (3) It can produce asymmetric news impact curves where, contrary to the traditional EGARCH, the resulting variances do not excessively exceed the ones associated with the standard GARCH model, irrespectively of the sign of an impact of moderate size. Furthermore, the new class of models considered can create a wide array of news impact curves which provide the researcher with a richer choice set relative to the traditional EGARCH. We also show in a Monte Carlo experiment the good finite sample performance of our asymptotic theoretical results and we compare them with those obtained from a parametric and the residual based bootstrap. Finally, we provide an empirical illustration.

Key words: Asymmetric volatility models; Asymmetric news impact curves; Quasi maximum likelihood estimation; Asymptotic Theory; Bootstrap.

JEL Codes: C12, C13, C15, C22, C51, C52, E43.

1 Introduction

Since the influential work of Engle (1981), a huge amount of literature on conditional heteroskedasticity has appeared in econometrics. ARCH (Autoregressive Conditional Heteroskedastic) models have obtained an enormous attention. A main shortcoming of the ARCH model of Engle (1981) and the generalized ARCH model by Bollerslev (1986) (these models have been studied in the pioneering papers by Lee and Hansen (1994) and Lumsdaine (1996)), is that they do not allow for asymmetries, i.e., the so-called leverage effects. The empirical relevance of leverage effects in particular has motivated an extensive search for asymmetric volatility models and many specifications have been suggested. The exponential GARCH (EGARCH) specification, by Nelson (1991) is probably one of the best known examples and has become a very important tool in volatility modelling. Surprisingly, however, very little is known about the statistical properties of the EGARCH model, particular estimation and inference theory seems to be lacking. He, Teräsvirta and Malmsten (2002) have analyzed the moment structure of first-order exponential GARCH models, but to the best of our knowledge there does not exist a formal proof establishing the limiting properties of maximum likelihood or quasi maximum likelihood (QML) estimators associated with this model, see, e.g., Mikosch and Straumann (2006) and Linton (2007).

This paper makes three main contributions. *First*, we propose a new exponential GARCH model that allows for a more flexible specification of the conditional variance function relative to the traditional EGARCH specification. *Secondly*, we provide a proof of the asymptotic normality of the QML estimator associated with the relative general class of specifications we propose. We illustrate - by a Monte Carlo simulation study - the finite sample performance of the asymptotic results and make comparisons with those obtained from parametric and residual based bootstraps. In some cases, we have evidence of the existence of possible refinements along the lines of Corradi and Iglesias (2008). *Thirdly*, we provide an empirical illustration, comparing the new EGARCH specification with the traditional EGARCH as well as the GARCH model. We show that the news impact curve of our new model is likely to offer a better fit to US interest rate data. Moreover, we show that contrary to the traditional EGARCH, the new model is able to generate news impacts that are highly asymmetric but do not exceed the news impacts generated by the GARCH and the EGARCH. In other words, the news impact curves generated by the new model are uniformly flatter than the news impact curves associated with the GARCH/EGARCH. We will argue that this, mainly due to stability issues related to the traditional EGARCH, is an important feature of the new EGARCH specification. Finally, we should note that since we allow for a very general class of conditional volatility functions, our approach can create many alternative news impact curves and they can all be tested in the framework suggested by Engle and Ng (1993).

The organization of the paper is as follows: Section 2 presents the new model and the first order validity of the QML method in this setting. Section 3 provides an empirical illustration, showing some advantages of our new specification. In addition, we provide simulation based evidence on the performance of the theoretically derived asymptotic approximation/distribution and comparisons to the parametric and the residual based bootstrap are made. Finally, Section 4 concludes. All the proofs are collected in the Appendix.

2 An EGARCH-type model with a general volatility function

We consider the following process, where y_t is the time series of interest, and its conditional variance, given by $\sigma_t^2(\theta)$, is taking the following representation

$$\begin{aligned} y_t &= \sigma_t(\theta)\,\upsilon_t, && (1) \\ \ln\sigma_t^2(\theta) &= \omega + \beta\ln\sigma_{t-1}^2(\theta) + \psi g(y_{t-1}) + \phi y_{t-1}, && (2) \end{aligned}$$

for $\theta = \left(\omega, \beta, \psi, \phi, \sigma_0^2(\theta)\right)'$. In what follows, the unobserved initial variance is parameterized as $\gamma = \sigma_0^2(\theta)$. In Lemma 1 in the Appendix, we will prove that this initial variance is asymptotically negligible (along the lines of Lemma 6 in Lumsdaine (1996)). The true parameter vector is defined as $\theta_0 = (\omega_0, \beta_0, \psi_0, \phi_0, \gamma_0)'$ and we let υ_t be a sequence of independent and identically distributed random variables with mean zero. We sometimes abbreviate $\sigma_t^2(\theta) = \sigma_t^2 > 0$, which is a $\mathcal{F}_{t-1}$-measurable function, where $\mathcal{F}_{t-1}$ is the sigma algebra generated by $\{\upsilon_{t-1}, \upsilon_{t-2}, ...\}$. Furthermore, we define $g(y_{t-1})$ as a measurable function of y_{t-1}, although the function could be generalized to include more lags. It should also be noted that the model includes as a special case the log-GARCH model of Geweke (1986) and Pantula (1986).

In what follows, we assume for simplicity that ω and γ are known constants. However, the extension to the case where ω is estimated follows straightforwardly (but tediously) using the same procedures as in the proofs in the Appendix (along the same lines as how the ARCH parameter is treated in Jensen and Rahbek (2004b, Section 3.5)). Moreover, the initial value of the conditional variance and the intercept in the conditional variance can be deduced from the unconditional variance.

Note that the process given by (1) - (2) has a structure that is very similar to the traditional exponential generalized autoregressive model (EGARCH) of Nelson (1991), where (2) is replaced by

$$\ln\sigma_t^2(\theta) = \omega + \beta\ln\sigma_{t-1}^2(\theta) + \psi g(\upsilon_{t-1}) + \phi\upsilon_{t-1}.$$

As Linton (2007) points out, "no results have yet been published for consistent and asymptotic normality of EGARCH from primitive conditions, although simulation evidence does suggest normality is a good approximation in large samples". Quoting also Mikosch and Straumann (2006), "the theoretical properties of the QMLE in EGARCH have not been studied in the literature", and "at the moment we cannot provide a proof of the asymptotic normality of the QMLE in the general EGARCH model".

In our "new" model, the main difference to the EGARCH of Nelson (1991) is that $g(\cdot)$ is a function of y_{t-1} instead of υ_{t-1}, and ϕ is associated y_{t-1} instead of υ_{t-1}. This feature is what makes the proof of the asymptotic normality in our "new" model different to the proof based on the traditional EGARCH, as dealing with the recursive nature of υ_{t-1} in the conditional variance function is avoided. However, since one can specify alternative choices of measurable g-functions, the specification is not limited to the EGARCH-type.

It should be emphasized that including y_{t-1} instead of υ_{t-1} in the conditional variance function is a commonly used approach. A leading example includes the double autoregressive model of Ling (2004), where

$$y_t + \phi y_{t-1} = \epsilon_t, \qquad \sigma_t^2 = \omega + \alpha y_{t-1}^2.$$

Similarly, Danielsson (1994) proposes a type of stochastic volatility model where asymmetries using y_{t-1} are introduced in the conditional variance equation. However, most examples can perhaps be found in the nonparametric literature, see, e.g. Linton and Mammen (2005) and the references therein, where the volatility function typically is specified as function of lags of observed data, i.e.,

$$\sigma_t = \sigma_t \left(y_{t-1}, ..., y_{t-p} \right).$$

Hence, specifying the process (1) - (2) and letting y_{t-1} enter the conditional volatility function is by no means a new idea. In particular, one can argue that we are merely extending the approach by Ling (2004) by introducing the existence of asymmetries and leverage effects in the conditional volatility.

2.1 Asymptotics

In order to establish the asymptotic results we will make the following set of assumptions which are all fairly common in the GARCH literature:

Assumptions A1-A5

A1 $\upsilon_t \sim i.i.d.\,(0,1)\,,$

A2 $|\beta| < 1, |\psi| < \infty$, and $|\phi| < \infty$,

A3 $\mathrm{E}\left(\left(1-\upsilon_t^2\right)^2\right) = \zeta < \infty$,

A4 $\sigma_0^2\,(\theta) = \gamma$ is a drawing from the stationary distribution,

A5 y_t is a strictly stationary and ergodic process with $\mathrm{E}\left[\left(|g\,(y_t)|\right)^4\right] < \infty$ and $\mathrm{E}\left[\left(|y_t|\right)^4\right] < \infty$.

Note that Assumptions A1-A5 allow for a very general class of conditional variance functions. Assumption A1 and A3 are restrictions on the innovation process and are both common. Assumption A4 restricts the initial value of σ_0^2 to be a drawing from a stationary distribution, but in the Appendix, in Lemma 1, we will show the asymptotic negligibility of treating it as a fixed and bounded constant. A1 could probably be replaced by the weaker martingale difference sequence assumption without altering the main results of the paper, but at the expense of adding to the complexity of the proofs. Therefore in this paper we are considering only strong-EGARCH type models. Assumption A2 and A4 restrict the unknown parameters to be defined on a compact parameters space. In the traditional EGARCH model, the existence of the rth moment in υ_t requires the existence of rth moment of y_t. Therefore, as it is widely used in the EGARCH literature as well as in the literature on more general asymmetric volatility models, we impose the moments restriction directly on y_t (and $g(y_t)$) as in A5. Carrasco and Chen (2002) show that Nelson's (1991) EGARCH process will be strictly stationary and geometrically ergodic under general conditions similar to Assumptions A1-A4. However, when replacing υ_{t-1} with y_{t-1} in the conditional variance specification it is no longer possible to use the methodology in Carrasco and Chen (2002) or the more general approach by Cline and Pu (1999) to establish strictly stationarity and ergodicity of y_t and/or σ_t^2. Establishing such properties of y_t and/or σ_t^2 within the current model setup will be very challenging but a topic for future research.

In general, A5 imposes restrictions on the parameter space. However, there are some special cases where we can provide the specific conditions for strict stationarity and ergodicity of y_t. For example, if we set $\phi = 0$ and $g(y_{t-1}) = \ln y_{t-1}^2$ in (1)-(2), we obtain

$$\begin{aligned} y_t &= \sigma_t(\theta)\upsilon_t, \\ \ln \sigma_t^2(\theta) &= \beta \ln \sigma_{t-1}^2(\theta) + \omega + \psi \ln y_{t-1}^2, \\ &= \beta \ln \sigma_{t-1}^2(\theta) + \omega + \psi\left(\ln \upsilon_{t-1}^2 + \ln \sigma_{t-1}^2(\theta)\right), \\ &= (\beta+\psi)\ln \sigma_{t-1}^2(\theta) + \left(\omega + \psi\left(\ln \upsilon_{t-1}^2\right)\right), \end{aligned}$$

where, following Carrasco and Chen (2002, Proposition 5, page 23), if $|\beta + \psi| < 1$ and there is an integer $s \geq 1$ such that $E\left|\omega + \psi \ln\left(\upsilon_t^2\right)\right|^s < \infty$, we have that $\sigma_t^2(\theta)$ is Markov geometrically ergodic. If σ_0^2 is initialized from the invariant measure, then y_t and $\sigma_t^2(\theta)$ are strictly stationary and $\beta-$mixing with exponential decay. Also, $E\left|\ln \sigma_t^2(\theta)\right|^s < \infty$. If moreover $s = 2$, then we have simultaneously that h_t is geometrically ergodic and $E\left|\ln \sigma_t^2(\theta)\right|^2 < \infty$.

Within the scope of this paper we will, however, continue under the maintained assumption A5 that y_t is a strictly stationary and geometrically ergodic process. Assumption A5 is common in the GARCH literature, see, e.g., Kristensen and Linton (2006).

Finally, it should be noted, that given Assumptions A2 and A5 in particular, then also $g(y_t)$ and σ_t^2 are strictly stationary and ergodic processes, see, e.g., Theorem 3.35, page 42, in White (1984). Strict stationarity of σ_t^2 would actually "only" require Assumptions A1, A2, A4, strictly stationarity of y_t, $\mathrm{E}\left(\upsilon_t^2\right) < \infty$ and $\mathrm{var}\left(\psi g(y_t) + \phi y_t\right) < \infty$ as in He et al (2002). Note also that $g(y_t)$ needs a finite expectation, but it does not have to be equal to zero. To get a specification very similar to Nelson's (1991) EGARCH, one could in finite samples choose $g(x) = |x| - \overline{|x|}$, where $\overline{|x|}$ denotes the sample mean of $|x|$.

The quasi-likelihood function associated with (1) - (2) is given by

$$l_T(\theta) = \sum_{t=1}^{T} l_t(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln \sigma_t^2 + \frac{y_t^2}{\sigma_t^2}\right). \tag{3}$$

The following Theorem 1 establishes the asymptotic normality of the QML estimator which is the main theoretical result of the paper.

Theorem 1 Let Assumptions A1-A5 hold. With $(\omega, \gamma)'$ fixed at their true values, $(\omega_0, \gamma_0)'$, consider the model given by the quasi log-likelihood function as in (3). Then, there exists a fixed open neighborhood $U = U(\beta_0, \psi_0, \phi_0)$ of $(\beta_0, \psi_0, \phi_0)'$ such that with probability tending to one as $T \longrightarrow \infty$, $l_T(\beta, \psi, \phi)$ has a unique maximum point $\left(\widehat{\beta}, \widehat{\psi}, \widehat{\phi}\right)'$ in U. In addition, the QML estimator $\left(\widehat{\beta}, \widehat{\psi}, \widehat{\phi}\right)'$ is consistent and asymptotically normal,

$$\sqrt{T}\left[\left(\widehat{\beta}, \widehat{\psi}, \widehat{\phi}\right) - (\beta_0, \psi_0, \phi_0)\right]' \xrightarrow{d} N\left(0, \boldsymbol{\Omega}^{-1}\right),$$

where

$$\boldsymbol{\Omega} = \frac{1}{\zeta}\begin{pmatrix} \varpi_1^2 & \varpi_{12} & \varpi_{13} \\ \varpi_{12} & \varpi_2^2 & \varpi_{23} \\ \varpi_{13} & \varpi_{23} & \varpi_3^2 \end{pmatrix},$$

and the typical element of Ω is defined by Propositions 1 and 2 given in the Appendix.

Proof of Theorem 1 The proof of Theorem 1 is given in the Appendix

Theorem 1 provides the first order asymptotic theory for the QMLE applied to our model (1) - (2), when $(\omega, \gamma)'$ is fixed at their true values $(\omega_0, \gamma_0)'$. The extension to the case where ω is estimated is straightforward by using the same methodology as the one applied in the Appendix. In empirical applications, one would replace the unknown quantities in $\boldsymbol{\Omega}$ with the estimates obtained from the QML estimation procedure in order to obtain the estimated standard errors. In the following section, the finite sample performance of the asymptotic distribution given by Theorem 1 will be analyzed in detail.

3 Illustrations

In this section we compare and illustrate the empirical properties of the new EGARCH(1,1) specification. We are interested in analyzing how well the asymptotic approximation that we provide in our Theorem 1 performs in finite samples. We first provide a Monte Carlo study of a bootstrapped version of the QML estimator and discuss results obtained using first order the asymptotic theory provided in Theorem 1. Subsequently, we provide an empirical illustration based on one-month U.S. Treasury bill yields.

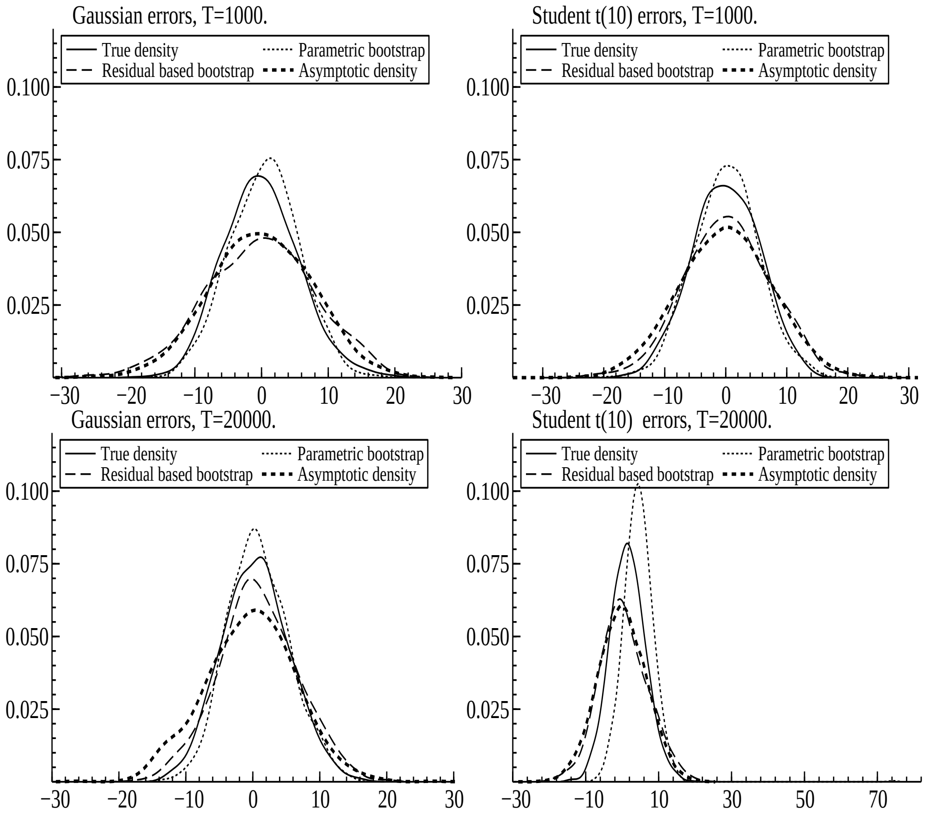
3.1 A Monte Carlo study in finite samples: bootstrap and first order asymptotic results

Hidalgo and Zaffaroni (2007) show that if the residual based bootstrap is applied to the QMLE in the context of an ARCH(∞) model, then first order validity is obtained. Gonçalves and White (2004) document the advantages of bootstrapping an objective function in the context of non-markovian models and provide the conditions for the block-bootstrap to be first order asymptotically valid. Corradi and Iglesias (2008) show the conditions under which, when bootstrapping a GARCH(1,1) model, one can obtain the refinements of the bootstrap procedure of Gonçalves and White (2004).

However, to the best of our knowledge, there are no results available in the literature with respect to applying the bootstrap to a GARCH-type model with leverage effect. In this section we provide simulated plots for the density function of the QMLE of ψ (the parameter directly related to the leverage effect) in (1)-(2). We provide four cases for the estimation of that density function: (1) first we show the true density function obtained by Monte Carlo (True density); (2) then we use our asymptotic theoretical results in Theorem 1 and we plug in the QMLE (Asymptotic density). (3) Later we apply a parametric bootstrap and we draw the bootstrap resamples assuming υ_t to be a $N(0,1)$ (Parametric bootstrap), and (4) finally we apply a residual based bootstrap along the lines of Hidalgo and Zaffaroni (2007, pages 842-843) by drawing from the empirical distribution of the residuals (Residual based bootstrap). The number of bootstrap resamples is 999 and we use 1000 replications. The results are available in Figure 1. We consider two cases: the case of υ_t being a $N(0,1)$ distribution (the first column of Figure 1) and also a t-distribution of 10 degrees of freedom (the second column of Figure 1). Moreover, we show the graphs for two sample sizes: 1000 (the first row of Figure 1) and 20000 (the second row of Figure 1). The data generating process (DGP) uses $(\omega, \beta, \psi, \phi) = (0.04, 0.6, -0.5, 0.1)$ and $g(y_{t-1}) = |y_{t-1}| - \overline{|y_{t-1}|}$.

When the true density of the innovation process is a normal distribution, our asymptotic density obtained from Theorem 1 is not very far away from the true density, even for a limited sample size of 1000 observations. Moreover, as expected (along the lines of the results of Corradi and Iglesias (2008) for the regular GARCH(1,1)), both the parametric

Figure 1: Alternative densities of $\sqrt{T}\left(\widehat{\psi}-\psi_0\right)$



and residual based bootstrap are closer to the true density, suggesting the possible existence of refinements in this context. When the innovation process follows a t-distribution with 10 degrees of freedom, then again as expected (as in Corradi and Iglesias (2008)), the parametric and the residual based bootstrap provide a worst approximation of the true density, and the gains of the bootstrap are smaller versus asymptotic theory.

Table 1: Alternative point estimates of s.e.$(T^{1/2}\hat{\psi})$ based on Gaussian errors

	Asymptotic	Par. bootstrap	Residual bootstrap	True (simulated)
T= 500	7.666	5.274	14.401	7.412
T= 1000	7.533	7.354	10.523	7.686
T= 2500	7.531	6.793	10.386	7.557
T= 5000	7.476	7.340	10.235	7.649
T=10000	7.501	7.391	10.409	7.575
T=15000	7.564	7.633	10.603	7.579
T=20000	7.604	7.639	10.809	7.560

To provide further evidence of the finite sample performance of our asymptotic approximation in Theorem 1, Table 1 shows, for different sample sizes (T) and when υ_t is simulated as a $N(0,1)$ distribution, alternative point estimates of the element of $\boldsymbol{\Omega}^{-1}$ related to the QMLE of ψ in our Theorem 1. We again show the results for the simulated true value (True (simulated)), when we use our asymptotic results in Theorem 1 (Asymptotic) and we plug in the QMLE, when we apply a parametric bootstrap (Par. bootstrap) and also the residual based bootstrap (Res. bootstrap). Our asymptotic approximation of the variance using Theorem 1 provides a very good approximation of the true value, and as expected, in this case the parametric bootstrap provides a much better approximation than the residual based bootstrap (similar results are reported in Corradi and Iglesias (2008) in the regular GARCH(1,1) model).

Table 2 provides the same results as Table 1, but now when υ_t is simulated as a t-distribution with 10 degrees of freedom. In this case again, our asymptotic result of Theorem 1 shows a very good approximation to the true values, and now as expected (along the lines of Corradi and Iglesias (2008) for the regular GARCH(1,1)), the parametric bootstrap offers a worst approximation than in Table 1 (since we are assuming that υ_t is a $N(0,1)$ distribution while it is truly a t-distribution with 10 degrees of freedom), and the residual based bootstrap provides better results.

From the limited simulations provided in this section we can conclude that, based on our DGP, our asymptotic results in Theorem 1 provide a very good approximation in

Table 2: Alternative point estimates of s.e.$(T^{1/2}\hat{\psi})$ based on Student t(10) errors

	Asymptotic	Par. bootstrap	Residual bootstrap	True (simulated)
T= 500	6.765	5.901	7.557	6.839
T= 1000	6.786	6.070	7.683	6.694
T= 2500	6.789	6.322	8.665	6.622
T= 5000	6.659	6.117	9.074	6.765
T=10000	6.719	6.031	8.933	6.511
T=15000	6.661	5.914	9.018	6.656
T=20000	6.659	5.910	9.151	6.676

finite samples. It should be noted that the traditional EGARCH model has a tendency to exhibit exploding behavior when simulating/bootstrapping the processes, particularly when the volatility process is parameterized such that it is stationary and ergodic but still highly persistent; see, e.g., the simulations performed by Andersen and Lund (1997). The approximation properties could therefore be affected in finite samples for other choices of DGP, particularly if more persistence in the volatility process is observed/imposed.

3.2 Empirical application

We use one-month U.S. Treasury bill yields (the average of bid and ask) which are collected from the risk-free rate file of the Center for Research and Security Prices (CRSP). The sample period is June 1964 to December 1989, for a total of 307 monthly observations, and corresponds to the sample period used by Chan, Karolyi, Longstaff and Sanders (1992) and Ball and Torous (1999)[1]. Ball and Torous (1999) show that Nelson's EGARCH model performs very well based on this data set (they show that the stochastic volatility model does not outperform the traditional EGARCH. See Ball and Torous (1999, pages 2348 and 2349)) and it would therefore be interesting to see how the new

[1]Standard augmented Dickey-Fuller (1979) and Phillips-Perron (1988) tests allow us not to reject the hypothesis of a unit root in the level of the series at 10% and 5% significant levels respectively. However, for the first difference, with both tests we reject the null of a unit root at the 1% significance level. Note that from A5, we need stationarity in our time series.

EGARCH specification compares. The empirical models under consideration are given as

$$\Delta r_t = k + \sum_{i=1}^{6} \delta_i \Delta r_{t-1-i} + \sigma_t \upsilon_t, \tag{4}$$

where $\upsilon_t \sim i.i.d(0,1)$ and σ_t evolves according to the following dynamic representations:

- Nelson's EGARCH(1,1), denoted EGARCH(Nelson)

  $$\ln \sigma_t^2 = \omega + \beta \ln \sigma_{t-1}^2 + \psi g(\upsilon_{t-1}) + \phi \upsilon_{t-1},$$

  where $g(\upsilon_{t-1}) = |\upsilon_{t-1}| - E\,|\upsilon_{t-1}|$.

- Dahl-Iglesias EGARCH(1,1), denoted EGARCH(DI)

  $$\ln \sigma_t^2 = \omega + \beta \ln \sigma_{t-1}^2 + \psi g(\Delta r_{t-1}) + \phi \Delta r_{t-1},$$

  where $g(\Delta r_{t-1}) = |\Delta r_{t-1}| - \overline{|\Delta r_{t-1}|}$ and $\overline{|\Delta r_{t-1}|}$ denotes the sample mean of $|\Delta r_{t-1}|$.

- GARCH(1,1)

  $$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \psi g(\sigma_{t-1}\upsilon_{t-1}),$$

  where $g(\sigma_{t-1}\upsilon_{t-1}) = \sigma_{t-1}^2 \upsilon_{t-1}^2$.

The estimation results are summarized in Table 3. Note first that Ball and Torous (1999) do not introduce an autoregressive process in the mean equation such as (4). However, they only check for possible neglected autocorrelation in the residuals up to lag 10. We carry out a much more detailed battery of tests where we check for possible autocorrelation up to lag 40; and in order to remove all neglected serial correlation up to that order, we need to include lags in the mean equation. Although the mean equation is of lesser interest here it should be noted that all lags of the dependent variable enters insignificantly in the EGARCH(DI) specification. The main implication of this is that the asymptotics derived in the previous section then can be expected to hold for the model. In addition, the estimates for the parameters in Table 3 fall inside the regions that are allowed by Assumption A2. Moreover, Assumption A5 can and has been checked by simulating a number of sequences of Δr_t based on the estimated model, similar to

Andersen and Lund (1997). If the simulated sequences appear to be mean and variance reverting, this gives support to the assumption. Regarding the conditional variance function note that the estimated parameters are of almost the same magnitude for the two EGARCH specifications and they are all significant. However, as we will see later the estimated conditional volatility for the two EGARCH models will evolve quite differently over time. Furthermore, the implied/estimated news impact curves also turn out not to be similar. As in Ball and Torous (1999) the estimator of ϕ in the EGARCH models is significantly positive. This implies that a positive impulse to the interest rate will have a larger impact on the volatility than a negative impulse. Impulses or news that affect interest rates upwards might be considered bad news so the estimated effects are qualitatively in line with the conventional wisdom that bad news has a larger effect than good news on volatility on the financial securities. A final important remark is that one of the main advantages of our model is that the news impact curve generated by our new model is uniformly flatter that the news impact curve associated with the traditional GARCH/EGARCH. In terms of estimated parameters, this advantage manifests itself through the estimation of ψ in Table 3: 0.386 for the new model instead of 0.525 for a standard EGARCH. Danielsson (1994) got exactly the same kind of result in the context of stochastic volatility: 0.06 instead of 0.229.

Next, we investigate whether the models are well specified. From rows 1 through 4 in Table 4 we see that we cannot reject that the standardized residuals satisfy the model assumption regarding their two first moments. All the estimated models pass the Ljung-Box (LB) tests for neglected serial dependence and neglected ARCH. We consider 3 different test for neglected serial dependence up to order 4, which are described in Wooldridge (1991): HE-LM-AR(4) is the standard LM test, see, e.g., Engle (1982), while "W Proc 3.1" and "W Robust" (described on page 16 and 21 in Wooldridge, 1991, respectively) are robust to misspecification of the conditional variance function. Following the robust specification testing methodology of Wooldridge (1991), we also test for the existence of omitted variables $\widehat{\sigma}_{t-1}$ and $\widehat{\sigma}_{t-2}$ since the so-called mean-effects as discussed in Engle et al (1987) are commonly found in interest rate series. From Table 4 we see that all the specifications pass all the above mentioned tests $-$ with one exception $-$ at the commonly used 5 per cent nominal significance level. Whereas at the

Table 3: Estimation results:

	EGARCH(Nelson)	EGARCH(DI)	GARCH
k	0.036***	0.040	0.009
	(0.008)	(0.031)	(0.029)
δ_1	-0.123***	-0.077	-0.185**
	(0.033)	(0.068)	(0.073)
δ_2	-0.010	0.009	0.019
	(0.011)	(0.064)	(0.068)
δ_3	0.010	-0.049	0.025
	(0.024)	(0.068)	(0.057)
δ_4	-0.011	-0.011	0.017
	(0.038)	(0.065)	(0.054)
δ_5	0.030	0.049	-0.017
	(0.030)	(0.063)	(0.050)
δ_6	-0.038	-0.049	-0.031
	(0.037)	(0.061)	(0.058)
ω	-0.030	-0.034*	0.062**
	(0.037)	(0.020)	(0.026)
ϕ	0.185***	0.310***	.
	(0.048)	(0.066)	.
ψ	0.525***	0.386***	0.505***
	(0.128)	(0.133)	(0.158)
β	0.928***	0.854***	0.479***
	(0.034)	(0.051)	(0.132)
logL	0.700	-4.781	-7.629

s.e. in parenthesis.

'*': 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.).

1% we reject that σ_{t-1} should be included in the new EGARCH specification, the term clearly cannot be omitted from neither Nelson's EGARCH nor the GARCH model.

As argued by Engle and Ng (1993) the news impact curve implied by volatility models is very important for portfolio selection and assets pricing. It also turns out that it is useful for model specification and by testing whether the news impact curve of a model offers a good fit to the data which can highlight the quality of the model. The final three specification tests in Table 4 are aimed at this target and are described in Engle and Ng (1993): Engle-Ng Spec. Test 1 and Engle-Ng Spec. Test 2 are standard LM test for whether $\{|\Delta r_{t-1}|, \Delta r_{t-1}\}$ and $\{|\upsilon_{t-1}|, \upsilon_{t-1}\}$ respectively are omitted from the volatility function/news impact curve (described on page 1758 in Engle and Ng, 1993). Finally, the Engle-Ng Joint Test is a standard LM test for whether $\{I(\epsilon_{t-1} < 0), I(\epsilon_{t-1} < 0)\epsilon_{t-1}, I(\epsilon_{t-1} \geq 0)\epsilon_{t-1}\}$ are omitted from the volatility function/ news impact curve, where $\epsilon_t = \sigma_t \upsilon_t$ and $I(\cdot)$ denotes the indicator function. From inspection of Table 4 we see that it is not possible to reject that news impact curve implied by the EGARCH(DI) adequately captures the observed features in the data. The Engle-Ng Spec. Test 1 indicates that $\{|\Delta r_{t-1}|, \Delta r_{t-1}\}$ should have been included in Nelson's EGARCH model and the GARCH model. On the other hand and based on Engle-Ng Spec. Test 2 there is not any strong evidence that $\{|\upsilon_{t-1}|, \upsilon_{t-1}\}$ should have been included in neither EGARCH(DI) nor the GARCH model. Most striking, however, is the results implied by the Engle-Ng Joint Test: The EGARCH(DI) clearly passes this diagnostic test. This, on the contrary, is not true for Nelson's EGARCH as well as the GARCH; both models clearly fail to capture the basic asymmetries in news impact curve implied by the data.

From inspection of Figure 2 we see how different the predictions from the three volatility models turn out to be. The estimated GARCH volatility increases primarily due to the large negative shocks in Sept. 1974, Apr. 1980, Oct. 1984, Oct. 1987 and Dec 1988. The estimated EGARCH volatility is to a lesser extent also affected by these events. This is in contrast to the estimated EGARCH(DI) volatility which is virtually unaffected by these specific shocks. The main increase in the EGARCH(DI) volatility occurs as a result of a series of large positive shocks from 1980 - 1982. Surprisingly, the size of this increase in volatility in this period is not predicted by neither the GARCH

Table 4: Model specification results:

	EGARCH(Nelson)	EGARCH(DI)	GARCH
$E_T(\hat{v}_t) = 0$	-9.7*10^{-5}	-0.013	0.0544
	[0.496]	[0.589]	[0.168]
$E_T(\hat{v}_t^2) = 1$	0.981	0.979	0.988
	[0.626]	[0.624]	[0.582]
LB[10,AR]	3.326	4.083	4.538
	[0.973]	[0.944]	[0.920]
LB[20,AR]	19.976	23.086	21.489
	[0.459]	[0.285]	[0.369]
LB[40,AR]	55.075*	53.686*	50.043
	[0.057]	[0.073]	[0.133]
LB[10,ARCH]	4.447	6.896	2.194
	[0.925]	[0.735]	[0.995]
LB[20,ARCH]	10.952	18.645	13.265
	[0.947]	[0.545]	[0.866]
LB[40,ARCH]	28.713	37.889	25.316
	[0.908]	[0.566]	[0.966]
W Robust	20.482*	20.021*	16.689
	[0.058]	[0.066]	[0.161]
W Proc 3.1	17.480	17.480	17.480
	[0.132]	[0.132]	[0.132]
HE-LM-AR(4)	14.240	17.413	15.457
	[0.285]	[0.134]	[0.217]
Omitted $\hat{\sigma}_{t-1}$	11.362***	5.350**	10.501***
	[0.001]	[0.020]	[0.001]
Omitted $\hat{\sigma}_{t-2}$	1.473	4.7*10^{-5}	2.383
	[0.224]	[0.994]	[0.122]
Engle-Ng Spec. Test 1	7.176**	·	8.009**
	[0.027]	·	[0.018]
Engle-Ng Spec. Test 2	·	4.189	5.555*
	·	[0.123]	[0.062]
Engle-Ng Joint Test	20.537***	3.334	16.599***
	[0.002]	[0.765]	[0.001]

p-values in brackets.

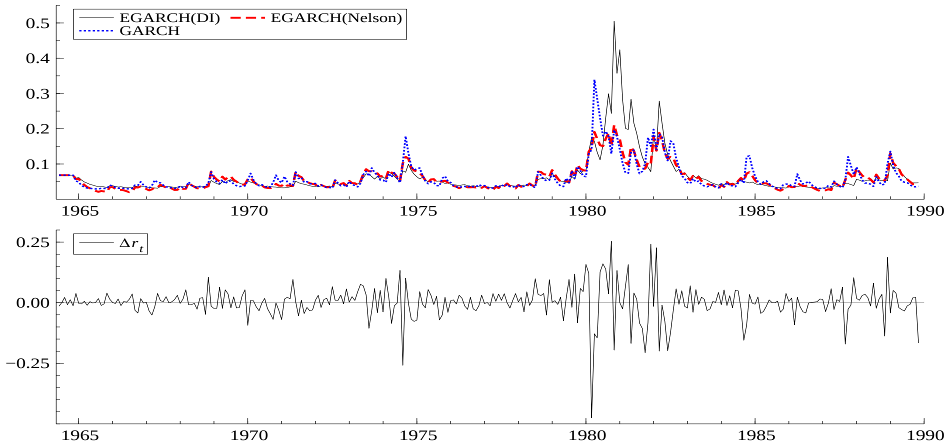
LB[XX,AR] is the Ljung-Box test for neglected serial dependence up to order XX

LB[XX,ARCH] is the Ljung-Box test for neglected ARCH up to order XX

Engle-Ng Spec. Test 1 uses $|\Delta r_{t-1}|$ and Δr_{t-1} under the alternative

Engle-Ng Spec. Test 2 uses $|v_{t-1}|$ and v_{t-1} under the alternative

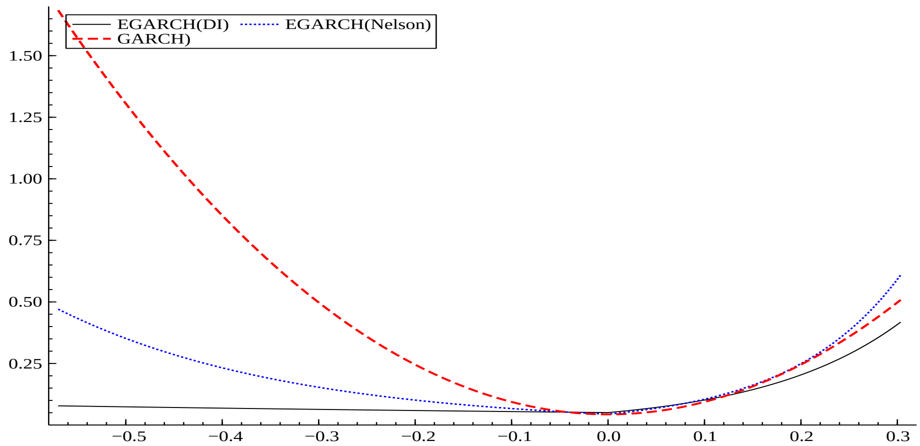
Figure 2: The t-bill rates (Δr_t) and estimated volatility functions



nor the EGARCH model. Strong recessions in 1980 and 1982 implied that the Federal Reserve kept monetary policy tight as the only anti-inflationary instrument available in this period (see e.g. Cline (1989)).

The estimated news impact curves depicted in Figure 3 confirm why the estimated EGARCH(DI) volatility does not change much due to the large negative shocks referred to previously. In particular, only relatively large positive shock will affect the EGARCH(DI) volatility but the effect of such news will be uniformly smaller than their effect on the estimated GARCH-volatility and EGARCH-volatility. We believe this is a very nice feature of the model, since one of the main arguments against the EGARCH specification is that it often predicts too much volatility in case of bad news and as a result tends to become “unstable”, see e.g. Engle and Ng (1993) and the simulation studies in Andersen and Lund (1997) and Ball and Torous (1999). Note also, that in Figure 3 we only show the shape of the NIC of the new EGARCH model where $g(\Delta r_{t-1}) = |\Delta r_{t-1}| - \overline{|\Delta r_{t-1}|}$. But this is only one possible NIC that our new model can generate since we can allow for different $g(\Delta r_{t-1})$ functions. This gives the researcher a lot of flexibility to generate alternative NIC's if data cannot support the EGARCH type specification.

Figure 3: The estimated news impact curves



4 Conclusions

In this paper we propose a new asymmetric volatility model along the lines of the traditional exponential GARCH model. Our new model has the advantage over the traditional EGARCH that allows a much more flexible function in the conditional variance. We prove the asymptotic normality of the QML estimator in this setting. A Monte Carlo simulation study shows that if we apply different types of bootstrapping procedure to the QML estimator in our model, we obtain very similar results compared to first order asymptotic theory. When the true innovation process follows a normal distribution the bootstrap based results indicate the possible existence of refinements versus first order asymptotic theory along the lines of the results in Corradi and Iglesias (2008) for the regular GARCH(1,1) model. Finally, we show how the new EGARCH model can generate asymmetric news impact curves without the disadvantage of significantly increasing the volatility in relation to the traditional GARCH model. Note also, that since we allow for a general class of functions in the conditional variance, the model can create many different shapes of the news impact curves. An empirical application completes the paper.

REFERENCES

Andersen, T. G. and J. Lund (1997), Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate, *Journal of Econometrics* 77, 343-377.

Ball, C. A. and W. N. Torous (1999), The Stochastic Volatility of Short-Term Interest Rates: Some International Evidence, *The Journal of Finance* 54, 6, 2339-2359.

Bierens, H. (2004), *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge University Press.

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

Carrasco, M. and X. Chen, (2002), Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory* 18, 17-39.

Chan, K. C., A. Karolyi, F. Longstaff and A. Sanders (1992), An Empirical Comparison of Alternative Models of the Short-term Interest Rate, *The Journal of Finance* 48,1209-1227.

Cline, W. R. (1989), Macroeconomic Influences on Trade Policy, *American Economic Review, papers and proceedings* 79, 2, 123-127.

Cline, D.B. and H.H. Pu (1999), Geometric ergodicity of nonlinear time series, *Statistica Sinica* 9, 1103-1118.

Corradi, V. and E. M. Iglesias (2008), Bootstrap Refinements for QML Estimators of the GARCH(1,1) Parameters, *Journal of Econometrics* 144, 500-510.

Danielsson, J. (1994), Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood, *Journal of Econometrics* 64, 375-400.

Dickey, D. A. and W. A. Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, vol. 74, 427-431.

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 V. K. Ng (1993), Measuring and Testing the Impact of News on Volatility, *Journal of Finance* 48, 5, 1749-1778.

Geweke, J. (1986), Comment on "Modeling the persistence of conditional variances", by R. F. Engle and T. Bollerslev. *Econometric Reviews* 5, 57-61.

Gonçalves, S. and H. White (2004), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics* 119, 199-219.

He, C., T. Teräsvirta and H. Malmsten (2002), Moment Structure of a Family of First-Order Exponential GARCH Models, *Econometric Theory* 18, 868-885.

Hidalgo, F.J., and P. Zaffaroni (2007), A Goodness of Fit Test for ARCH(∞), *Journal of Econometrics,* 141, 835-875.

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.

Kristensen, D. and O. Linton (2006), A Closed-form Estimator for the GARCH(1,1) Model, *Econometric Theory* 22, 323-337.

Lee, S. W. 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. (2007), ARCH Models, *The New Palgrave Dictionary of Economics*, 2nd Edition, forthcoming.

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

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

Mikosch, T. and D. Straumann (2006), Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach, *The Annals of Statistics* 34, 5, 2449-2495.

Nelson, D. B. (1991), Conditional Heteroskedasticity in Asset Returns: a New Approach, *Econometrica* 59, 347-370.

Pantula, S. G. (1986), Comment on "Modeling the persistence of conditional variances," by R. F. Engle and T. Bollerslev, *Econometric Reviews* 5, 71-74.

Phillips, P. C. B. and P. Perron (1988), Testing for a Unit Root in Time Series Regression, *Biometrika* 75, 335-346.

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.

Appendix

All the main results and proofs of this paper are summarized in this Appendix. In order to cut down the size of this paper we have excluded detailed derivations of the 2nd and 3rd order derivatives of the quasi likelihood functions. These derivations can be obtained from the corresponding author upon request. To establish the main theoretical result given by Theorem 1 we first state and prove five very useful lemmas. In addition, for easy reference, all the first, second and third order derivatives of the log-likelihood function given by (3) have been provided. These derivatives are summarized in Result 1, Result 2 and Result 3 respectively. Before starting with the first order derivatives, we show the asymptotic negligibility of the initial value σ_0^2 in Lemma 1.

Lemma 1 Let Assumptions A1-A5 hold. In (2) we assume that the initial value σ_0^2 is a constant, chosen arbitrarily. Define now

$$\begin{aligned} {}_u l_T(\theta) &= \sum_{t=1}^{T} {}_u l_t(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left({}_u \ln\sigma_t^2 + \frac{y_t^2}{{}_u\sigma_t^2}\right), \\ {}_u \ln\sigma_t^2(\theta) &= \omega_{i=0}^{t-1}\beta^i + \phi\sum_{i=1}^{t-1}\beta^i y_{t-i-1} + \psi\sum_{i=1}^{t-1}\beta^i g(y_{t-i-1}) + \beta_u^t \ln\sigma_0^2(\theta), \end{aligned}$$

where in ${}_u \ln\sigma_t^2(\theta)$, the initial value $\ln\sigma_0^2(\theta)$ is drawn from the stationary distribution. Then, if the true parameter vector $\theta_0 \in \Theta$ is in the interior or Θ,

$$\sup_{\theta\in\Theta}\left|T^{-1/2}\left[{}_u l_T(\theta) - l_T(\theta)\right]\right| \xrightarrow{p} 0,$$

$$\sup_{\theta\in\Theta^+}\left|T^{1/2}\left[T^{-1}\sum_{t=1}^{T}\frac{\partial_u^2 l_t(\theta)}{\partial\theta\partial\theta'} - T^{-1}\sum_{t=1}^{T}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]\right| \xrightarrow{p} 0,$$

$$\sup_{\theta\in\Theta^+}\left|T^{1/2}\left[T^{-1}\sum_{t=1}^{T}\frac{\partial_u l_t(\theta)}{\partial\theta}\frac{\partial_u l_t(\theta)}{\partial\theta'} - T^{-1}\sum_{t=1}^{T}\frac{\partial l_t(\theta)}{\partial\theta}\frac{\partial l_t(\theta)}{\partial\theta'}\right]\right| \xrightarrow{p} 0.$$

Proof of Lemma 1 The proof follows along the lines of Lumsdaine (1996, Lemma 6, page 587), replacing $\sigma_t^2(\theta)$ by the corresponding $\ln\sigma_t^2(\theta)$.

Result 1: The first order derivatives of the loglikelihood function

$$\begin{aligned}
\frac{\partial}{\partial\beta} l_T(\theta) &\equiv \sum_{t=1}^{T} s_{1t}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{y_t^2}{\sigma_t^2}\right)\left(\ln\sigma_{t-1}^2 + \beta\frac{\partial\sigma_{t-1}^2/\partial\beta}{\sigma_{t-1}^2}\right),\\
\frac{\partial}{\partial\psi} l_T(\theta) &\equiv \sum_{t=1}^{T} s_{2t}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{y_t^2}{\sigma_t^2}\right)\left(\beta\frac{\partial\sigma_{t-1}^2/\partial\psi}{\sigma_{t-1}^2} + g(y_{t-1})\right),\\
\frac{\partial}{\partial\phi} l_T(\theta) &\equiv \sum_{t=1}^{T} s_{3t}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left(1-\frac{y_t^2}{\sigma_t^2}\right)\left(\beta\frac{\partial\sigma_{t-1}^2/\partial\phi}{\sigma_{t-1}^2} + y_{t-1}\right).
\end{aligned}$$

In particular,

$$s_{1t}(\theta_0) = -\frac{1}{2}\left(1-\upsilon_t^2\right)\ln\sigma_{t-1}^2 - \frac{1}{2}\left(1-\upsilon_t^2\right)\beta_0\left(\frac{\partial\sigma_{t-1}^2/\partial\beta}{\sigma_{t-1}^2}|_{\theta_0}\right), \tag{5}$$

$$s_{2t}(\theta_0) = -\frac{1}{2}\left(\sum_{i=1}^{t}\beta_0^{(i-1)}\left(1-\upsilon_t^2\right)g(y_{t-i})\right), \tag{6}$$

$$s_{3t}(\theta_0) = -\frac{1}{2}\left(\sum_{i=1}^{t}\beta_0^{(i-1)}\left(1-\upsilon_t^2\right)y_{t-i}\right). \tag{7}$$

Lemma 2 Let Assumptions A1-A5 hold and define the sequence $I_{t-1} = \{y_{t-1}, y_{t-2}, ...\}$ to be sub-σ-algebras of I. Then $\{s_{it}(\theta_0), I_{t-1}\}$ for $i = 1, 2, 3$, are martingale difference sequences.

Proof of Lemma 2 Note that for each t a) $s_{it}(\theta_0)$ is measurable I_t, and b) $I_{t-1} \subset I_t$. It is also quite trivial to see that c) $\Pr(\mathrm{E}(s_{it}(\theta_0)|I_{t-1}) = 0) = 1$. To complete the proof of Lemma 2 we need to verify that d) $\mathrm{E}(|s_{it}(\theta_0)|) < \infty$ for $i = 1, 2, 3$, see, e.g., Definition 7.4, page 191 in Bierens (2004). We begin with the case where $i = 1$ by first noticing that

$$|s_{1t}(\theta_0)| \leq |\frac{1}{2}\left(1-\upsilon_t^2\right)\ln\sigma_{t-1}^2| + \left|\frac{1}{2}\left(1-\upsilon_t^2\right)\beta_0\left(\frac{\partial\sigma_{t-1}^2/\partial\beta}{\sigma_{t-1}^2}|_{\theta_0}\right)\right|, \tag{8}$$

where

$$\begin{aligned}
\frac{\partial\sigma_{t-1}^2/\partial\beta}{\sigma_{t-1}^2}|_{\theta_0} &= (t-1)\beta_0^{t-2}\ln\sigma_0^2 + \frac{(t-1)\beta_0^{t-2}}{(\beta_0-1)} + \frac{1-\beta_0^{t-1}}{\beta_0^2-2\beta_0+1}\\
&\quad +\phi_0\sum_{i=1}^{t-1}(i-1)\beta_0^{i-2}y_{t-i-1} + \psi_0\sum_{i=1}^{t-1}(i-1)\beta_0^{i-2}g(y_{t-i-1}).
\end{aligned}$$

Substituting in the expression for $\ln\sigma_{t-1}^2$, the first term in on the right hand side of (8) can be bounded as

$$\begin{aligned}
\left|\left(1-\upsilon_t^2\right)\ln\sigma_{t-1}^2\right| &\leq \left|\frac{\beta_0^t-\beta_0}{\beta_0^2-\beta_0}\left(1-\upsilon_t^2\right)\right| + \left|\left(1-\upsilon_t^2\right)\beta_0^{t-1}\ln\sigma_0^2\right|\\
&\quad + |\psi_0|\sum_{i=1}^{t-1}|\beta_0|^{i-1}\left|\left(1-\upsilon_t^2\right)g(y_{t-i-1})\right| + |\phi_0|\sum_{i=1}^{t-1}|\beta_0|^{i-1}\left|\left(1-\upsilon_t^2\right)y_{t-i-1}\right|,
\end{aligned}$$

and by using that $\mathrm{E}|XY| \leq \sqrt{\mathrm{E}\, X^2\, \mathrm{E}\, Y^2}$ we obtain

$$
\begin{aligned}
E\left|\left(1-v_t^2\right)\ln\sigma_{t-1}^2\right| &\leq \left(\frac{|\beta_0|^t-|\beta_0|}{|\beta_0|^2-|\beta_0|}+\left|\beta_0^{t-1}\ln\sigma_0^2\right|\right)\mathrm{E}\left|1-v_t^2\right| \\
&\quad+|\psi_0|\frac{|\beta_0|^t-|\beta_0|}{|\beta_0|^2-|\beta_0|}\sqrt{\mathrm{E}\left(\left(1-v_t^2\right)^2\right)}\sqrt{\mathrm{E}\left(g\left(y_{t-i-1}\right)^2\right)} \\
&\quad+|\phi_0|\frac{|\beta_0|^t-|\beta_0|}{|\beta_0|^2-|\beta_0|}\sqrt{\mathrm{E}\left(\left(1-v_t^2\right)^2\right)}\sqrt{\mathrm{E}\left(y_{t-i-1}^2\right)}.
\end{aligned}
$$

Consequently $\mathrm{E}\left|\left(1-v_t^2\right)\ln\sigma_{t-1}^2\right| < \infty$ by Assumptions A1-A5 uniformly in t. Next, we turn to the second term in (8). First, notice that for all finite $s = 0, \pm 1, \pm 2, \ldots$ and $T < \infty$ it holds that

$$
\sup_{t<T} t\,|\beta|^{t+s} < \infty, \tag{9}
$$

$$
\lim_{t\to\infty} t\,|\beta|^{t+s} = \lim_{t\to\infty}\frac{t}{1/|\beta|^{t+s}} = \frac{1}{\lim_{t\to\infty}-\frac{\ln|\beta|}{|\beta|^{s+t}}} = \frac{\lim_{t\to\infty}|\beta|^{s+t}}{-\ln|\beta|} = 0. \tag{10}
$$

Then,

$$
\begin{aligned}
\left|\left(1-v_t^2\right)\frac{\partial\sigma_{t-1}^2/\partial\beta}{\sigma_{t-1}^2}\right| &\leq \left|\left(1-v_t^2\right)(t-1)\,\beta_0^{t-2}\ln\sigma_0^2\right|+\left|\left(1-v_t^2\right)\frac{(t-1)}{\beta_0-1}\beta_0^{t-2}\right| \\
&\quad+\left|\left(1-v_t^2\right)\frac{1-\beta_0^{t-1}}{\beta_0^2-2\beta_0+1}\right|+|\phi_0|\sum_{i=1}^{t-1}(i-1)\,|\beta_0|^{i-2}\left|\left(1-v_t^2\right)y_{t-i-1}\right| \\
&\quad+|\psi_0|\sum_{i=1}^{t-1}(i-1)\,|\beta_0|^{i-2}\left|\left(1-v_t^2\right)g\left(y_{t-i-1}\right)\right|.
\end{aligned}
$$

Given Assumptions A1-A5 and the results in (9) and (10) it can now be established that

$$
\begin{aligned}
\sup_{t>0}\mathrm{E}\left|\left(1-v_t^2\right)(t-1)\,\beta_0^{t-2}\ln\sigma_0^2\right| &= \sup_{t>0}\mathrm{E}\left(1-v_t^2\right)(t-1)\,|\beta_0|^{t-2}\left|\ln\sigma_0^2\right| < \infty, \\
\sup_{t>0}\mathrm{E}\left|\left(1-v_t^2\right)\frac{(t-1)}{\beta_0-1}\beta_0^{t-2}\right| &= \sup_{t>0}\mathrm{E}\left|\left(1-v_t^2\right)\right|(t-1)\,|\beta_0|^{t-2}\left|(\beta_0-1)^{-1}\right| < \infty, \\
\sup_{t>0}\mathrm{E}\left|\left(1-v_t^2\right)\frac{1-\beta_0^{t-1}}{\beta_0^2-2\beta_0+1}\right| &= \sup_{t>0}\mathrm{E}\left|\left(1-v_t^2\right)\right|\left|\left(1-\beta_0^{t-1}\right)\right|\left|\left(\beta_0^2-2\beta_0+1\right)^{-1}\right| < \infty, \\
\sup_{t>0}\mathrm{E}\left(\sum_{i=1}^{t-1}(i-1)\,|\beta_0|^{i-2}\left|\left(1-v_t^2\right)y_{t-i-1}\right|\right) &\leq \sup_{t>0}\frac{|\beta_0|-2\,|\beta_0|^t-|\beta_0|^t-t\,|\beta_0|^{t-1}+|\beta_0|^{t-1}}{|\beta_0|-2\,|\beta_0|^2+|\beta_0|^3}\times \\
&\quad\sqrt{\mathrm{E}\left(\left(1-v_t^2\right)^2\right)}\sqrt{\mathrm{E}\left(y_{t-i-1}^2\right)} < \infty, \\
\sup_{t>0}\mathrm{E}\left(\sum_{i=1}^{t-1}(i-1)\,|\beta_0|^{i-2}\left|\left(1-v_t^2\right)g\left(y_{t-i-1}\right)\right|\right) &\leq \sup_{t>0}\frac{|\beta_0|-2\,|\beta_0|^t-|\beta_0|^t-t\,|\beta_0|^{t-1}+|\beta_0|^{t-1}}{|\beta_0|-2\,|\beta_0|^2+|\beta_0|^3}\times \\
&\quad\sqrt{\mathrm{E}\left(\left(1-v_t^2\right)^2\right)}\sqrt{\mathrm{E}\left(g\left(y_{t-i-1}\right)^2\right)} < \infty,
\end{aligned}
$$

Note that the term $\frac{|\beta_0|-2|\beta_0|^t-|\beta_0|^t-t|\beta_0|^{t-1}+|\beta_0|^{t-1}}{|\beta_0|-2|\beta_0|^2+|\beta_0|^3}$ attains maximum at $t = \frac{1}{-\ln|\beta_0|+|\beta_0|\ln|\beta_0|}\left(-|\beta_0|+|\beta_0|\ln|\beta_0|+1\right)$. Consequently $\mathrm{E}\left|\left(1-v_t^2\right)\frac{\partial\sigma_{t-1}^2/\partial\beta}{\sigma_{t-1}^2}\right| < \infty$ uniformly on t and it can be concluded that $\mathrm{E}\left|s_{1t}\left(\theta_0\right)\right| < \infty$ for all t.

Under Assumptions A1-A5, condition d) in the cases where $i = 2, 3$ follows easily as

$$\begin{aligned}
\mathrm{E}\left|s_{2t}\left(\theta_0\right)\right| &\leq \frac{1}{2}\left(\sum_{i=1}^{t}\left|\beta_0\right|^{i-1}\mathrm{E}\left|\left(1-v_t^2\right)g\left(y_{t-i}\right)\right|\right), \\
&\leq \frac{1}{2}\frac{1}{\left|\beta_0\right|-1}\left(\left|\beta_0\right|^t-1\right)\sqrt{\mathrm{E}\left(\left(1-v_t^2\right)^2\right)}\sqrt{\mathrm{E}\left(g\left(y_t\right)^2\right)} < \infty,
\end{aligned}$$

uniformly in t, i.e. $\sup_t \mathrm{E}\left|s_{2t}\left(\theta_0\right)\right| < \infty$ and

$$\begin{aligned}
\mathrm{E}\left|s_{3t}\left(\theta_0\right)\right| &\leq \frac{1}{2}\left(\sum_{i=1}^{t}\left|\beta_0\right|^{i-1}\mathrm{E}\left|\left(1-v_t^2\right)y_{t-i}\right|\right), \\
&\leq \frac{1}{2}\frac{1}{\beta_0-1}\left(\left|\beta_0\right|^t-1\right)\sqrt{\mathrm{E}\left(\left(1-v_t^2\right)^2\right)}\sqrt{\mathrm{E}\left(y_t^2\right)} < \infty,
\end{aligned}$$

also uniformly in t. This completes the proof of Lemma 2.

Lemma 3 Define the processes

$$\begin{aligned}
u_{1t}\left(\theta_0\right) &= \sum_{i=1}^{t}\beta_0^{(i-1)}\ln\sigma_{t-i}^2\left(\theta_0\right), \\
u_{2t}\left(\theta_0\right) &= \sum_{i=1}^{t}\beta_0^{(i-1)}g\left(y_{t-i}\right), \\
u_{3t}\left(\theta_0\right) &= \sum_{i=1}^{t}\beta_0^{(i-1)}y_{t-i}.
\end{aligned}$$

Given Assumptions A1-A5, then

$$\mathrm{E}\left(\left|u_{it}\left(\theta_0\right)\right|^p\right) \leq M_{i,p} < \infty,$$

for $p = 1, 2, 3$ and $i = 1, 2, 3$.

Proof of Lemma 3 Consider first $u_{2t}\left(\theta_0\right)$ and define

$$d_{2t}\left(\theta_0\right) = \sum_{i=1}^{t}\left|\beta_0\right|^{i-1}\left|g\left(y_{t-i}\right)\right|.$$

¿From the triangle inequality it follows that

$$\left|u_{2t}\left(\theta_0\right)\right| \leq d_{2t}\left(\theta_0\right),$$

for all t. Furthermore, since $\mathrm{E}\left|g\left(y_{t-i}\right)\right| < \infty$ (by A5) it follows that

$$\begin{aligned}
\mathrm{E}\left(d_{2t}\left(\theta_0\right)\right) &= \frac{1}{\left|\beta_0\right|-1}\left(\left|\beta_0\right|^t-1\right)\mathrm{E}\left|g\left(y_{t-i}\right)\right|, \\
&< \infty \text{ for all } t.
\end{aligned}$$

Consequently,

$$\mathrm{E}\left(\left|u_{2t}\left(\theta_0\right)\right|\right) < \infty.$$

Secondly, notice that

$$
\begin{aligned}
\mathrm{E}\left(\left|u_{2t}\left(\theta_0\right)\right|^2\right) &= \mathrm{E}\left(\left|\left(\sum_{i=1}^{t}\beta_0^{(i-1)}g\left(y_{t-i}\right)\right)\right|^2\right), \\
&\leq \mathrm{E}\left(\sum_{i=1}^{t}\sum_{j=1}^{t}\left|\beta_0\right|^{(j+i-2)}\left|g\left(y_{t-i}\right)g\left(y_{t-j}\right)\right|\right), \\
&\leq \frac{\left|\beta_0\right|-2\left|\beta_0\right|^{t+1}+\left|\beta_0\right|^{2t+1}}{1+\left|\beta_0\right|\left(\left|\beta_0\right|-2\right)}\mathrm{E}\left(\left|g\left(y_t\right)\right|^2\right), \\
&< \infty,
\end{aligned}
$$

since $\mathrm{E}\left(\left|g\left(y_t\right)\right|^2\right)$ exists by Assumption A5. Higher order moments of $u_{2t}\left(\theta_0\right)$ exist to the extent that the higher order moments of $g\left(y_t\right)$ exist. Using similar techniques it is easy to show that the moments of u_{3t} exist to the degree that the moments of y_t exist, i.e., up to order 4 by Assumption 5. Notice also that $u_{2t}\left(\theta_0\right)$ and $u_{3t}\left(\theta_0\right)$ are strictly stationary. Next, note that we can write

$$
\begin{aligned}
\left|u_{1t}\left(\theta_0\right)\right| &= \left|\sum_{i=1}^{t}\beta_0^{(i-1)}\ln\sigma_{t-i}^2\left(\theta_0\right)\right|, \\
&\leq \sum_{i=1}^{t}\left|\beta_0\right|^{(i-1)}\left(\sum_{j=0}^{t-i-1}\left|\beta_0\right|^j+\left|\beta_0\right|^{t-i}\left|\ln\sigma_0^2\right|+\left|\psi_0\right|\sum_{j=1}^{t-i}\left|\beta_0\right|^{j-1}\left|g\left(y_{t-i-j}\right)\right|+\left|\phi_0\right|\sum_{j=1}^{t-i}\left|\beta_0\right|^{j-1}\left|y_{t-i-j}\right|\right), \\
&= \sum_{i=1}^{t}\sum_{j=0}^{t-i-1}\left|\beta_0\right|^{(j+i-1)}+\sum_{i=1}^{t}\left|\beta_0\right|^{(t+2*i-1)}\left|\ln\sigma_0^2\right| \\
&\quad +\left|\psi_0\right|\sum_{i=1}^{t}\sum_{j=1}^{t-i}\left|\beta_0\right|^{(j+i-2)}\left|g\left(y_{t-i-j}\right)\right|+\left|\phi_0\right|\sum_{i=1}^{t}\sum_{j=1}^{t-i}\left|\beta_0\right|^{(j+i-2)}\left|y_{t-i-j}\right|, \\
&= t\frac{\left|\beta_0\right|^t}{\left|\beta_0\right|^2-\left|\beta_0\right|}+\frac{1-\left|\beta_0\right|^t}{\left|\beta_0\right|^2-2\left|\beta_0\right|+1}+\frac{1}{\left|\beta_0\right|^2-1}\left(\left|\beta_0\right|^{3t+1}-\left|\beta_0\right|^{t+1}\right) \\
&\quad +\left|\psi_0\right|\sum_{i=1}^{t}\sum_{j=1}^{t-i}\left|\beta_0\right|^{(j+i-2)}\left|g\left(y_{t-i-j}\right)\right|+\left|\phi_0\right|\sum_{i=1}^{t}\sum_{j=1}^{t-i}\left|\beta_0\right|^{(j+i-2)}\left|y_{t-i-j}\right|, \\
&\equiv d_{1t}\left(\theta_0\right),
\end{aligned}
$$

and it is easy to see that the moments of $d_{1t}\left(\theta_0\right)$ will exists to the degree that the moments of $\left|g\left(y_{t-i-j}\right)\right|$) and $\left|y_{t-i-j}\right|$ are bounded (as assumed in A5). Consequently, as $\left|u_{1t}\left(\theta_0\right)\right| \leq d_{1t}\left(\theta_0\right)$ uniformly on t, the corresponding moments of $\left|u_{1t}\left(\theta_0\right)\right|$ will exist. This completes the proof of Lemma 3.

Lemma 4 Let Assumptions A1-A5 hold. Then

$$
\frac{1}{T}\sum_{t=1}^{T}s_{it}^2 \overset{a.s.}{\to} \frac{\zeta}{4}\varpi_i^2, \tag{11}
$$

where $i = 1, 2, 3$ as $T \to \infty$.

Proof of Lemma 4 Define $z_{it}\left(\theta_0\right) = u_{it}^2\left(\theta_0\right)/T$ and rewrite (11) as

$$
\frac{1}{4T}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right)^2\left(\sum_{i=1}^{t}\beta_0^{(i-1)}\ln\sigma_{t-i}^2\left(\theta_0\right)\right)^2 = \frac{1}{4}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right)^2 z_{1t}\left(\theta_0\right).
$$

Note that by Lemma 3

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{4}\sum_{t=1}^{T}\mathrm{E}\left(\left(1-\upsilon_t^2\right)^2 z_{1t}\left(\theta_0\right)\right) &\leq \lim_{T\to\infty}\frac{\zeta}{4}\sum_{t=1}^{T}\frac{M_{1,2}}{T}, \\
&= \frac{\zeta}{4}M_{1,2}<\infty,
\end{aligned}$$

where $M_{i,p}$ is defined in Lemma 3. From Proposition 3.52 page 48 in White (1984) it then follows that

$$\frac{1}{4}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right)^2 z_{1t}\left(\theta_0\right)\overset{a.s.}{\to}\frac{\zeta}{4}\varpi_1^2.$$

Similarly, by Lemma 3

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{4}\sum_{t=1}^{T}\mathrm{E}\left(\left(1-\upsilon_t^2\right)^2 z_{2t}\left(\theta_0\right)\right) &\leq \frac{\zeta}{4}M_{2,2}, \\
\lim_{T\to\infty}\frac{1}{4}\sum_{t=1}^{T}\mathrm{E}\left(\left(1-\upsilon_t^2\right)^2 z_{3t}\left(\theta_0\right)\right) &\leq \frac{\zeta}{4}M_{3,2},
\end{aligned}$$

implying that

$$\begin{aligned}
\frac{1}{4}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right)^2 z_{2t}\left(\theta_0\right)&\overset{a.s.}{\to}\frac{\zeta}{4}\varpi_2^2, \\
\frac{1}{4}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right)^2 z_{3t}\left(\theta_0\right)&\overset{a.s.}{\to}\frac{\zeta}{4}\varpi_3^2.
\end{aligned}$$

This completes the proof of Lemma 4.

Lemma 5 Let Assumptions A1-A5 hold and define $\delta>0$. Then

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{it}^2\left(\theta_0\right)1\left\{\left|s_{it}\left(\theta_0\right)\right|>\delta\sqrt{T}\right\}\right)\to 0,$$

for all $\delta>0$ for $i=1,2,3$.

Proof of Lemma 5 Consider first the following moment

$$\begin{aligned}
\mathrm{E}\left(s_{it}^2\left(\theta_0\right)1\left\{\left|s_{it}\left(\theta_0\right)\right|>\delta\sqrt{T}\right\}\right) &= \mathrm{E}\left(\frac{1}{4}\left(1-\upsilon_t^2\right)^2 u_{it}^2\left(\theta_0\right)1\left\{\left|s_{it}\left(\theta_0\right)\right|>\delta\sqrt{T}\right\}\right), \\
&\leq \mathrm{E}\left(\frac{1}{4}\left(1-\upsilon_t^2\right)^2 u_{it}^2\left(\theta_0\right)\right), \\
&= \frac{\zeta}{4}\mathrm{E}\left(u_{it}^2\left(\theta_0\right)\right), \\
&< \infty. \qquad (12)
\end{aligned}$$

uniformly on i and t. The first inequality follows since

$$\frac{1}{4}\left(1-\upsilon_t^2\right)^2 u_{it}^2\left(\theta_0\right)1\left\{\left|s_{it}\left(\theta_0\right)\right|>\delta\sqrt{T}\right\}\leq\frac{1}{4}\left(1-\upsilon_t^2\right)^2 u_{it}^2\left(\theta_0\right),$$

and the last inequality is due to Lemma 3. As shown in Lemma 3, $u_{it}\left(\theta_0\right)$ is strictly stationary implying that $s_{it}\left(\theta_0\right)$ will share the same property, see, e.g., Theorem 3.35, page 42 in White (1984). The implication of this property is that there exists an integer $S>0$ such that for all $t,j>S$ then

$$\mathrm{E}\left(s_{it}^2\left(\theta_0\right)1\left\{\left|s_{it}\left(\theta_0\right)\right|>\delta\sqrt{T}\right\}\right)=\mathrm{E}\left(s_{ij}^2\left(\theta_0\right)1\left\{\left|s_{ij}\left(\theta_0\right)\right|>\delta\sqrt{T}\right\}\right).$$

Hence,

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{it}^{2}(\theta_0)\,1\left\{|s_{it}(\theta_0)|>\delta\sqrt{T}\right\}\right) &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{S}\mathrm{E}\left(s_{it}^{2}(\theta_0)\,1\left\{|s_{it}(\theta_0)|>\delta\sqrt{T}\right\}\right)\\
&\quad+\lim_{T\to\infty}\frac{1}{T}\sum_{t=S+1}^{T}\mathrm{E}\left(s_{it}^{2}(\theta_0)\,1\left\{|s_{it}(\theta_0)|>\delta\sqrt{T}\right\}\right),\\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{S}\mathrm{E}\left(s_{it}^{2}(\theta_0)\,1\left\{|s_{it}(\theta_0)|>\delta\sqrt{T}\right\}\right)\\
&\quad+\lim_{T\to\infty}\left(\frac{T-(S+1)}{T}\right)\mathrm{E}\left(s_{iS+1}^{2}(\theta_0)\,1\left\{|s_{iS+1}(\theta_0)|>\delta\sqrt{T}\right\}\right),\\
&\le \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{S}\frac{\zeta}{4}\mathrm{E}\left(u_{it}^{2}(\theta_0)\right)\\
&\quad+\lim_{T\to\infty}\mathrm{E}\left(s_{iS+1}^{2}(\theta_0)\,1\left\{|s_{iS+1}(\theta_0)|>\delta\sqrt{T}\right\}\right),\\
&\le \lim_{T\to\infty}\frac{S\zeta}{T4}M_{i,2}+\mathrm{E}\left(s_{iS+1}^{2}(\theta_0)\lim_{T\to\infty}1\left\{|s_{iS+1}(\theta_0)|>\delta\sqrt{T}\right\}\right),\\
&\to 0,
\end{aligned}$$

since $\lim_{T\to\infty}1\left\{|s_{iS+1}(\theta_0)|>\delta\sqrt{T}\right\}=0$ for all $i=1,2,3$, uniformly in δ and $\lim_{T\to\infty}\frac{S\zeta}{T4}M_{i,2}=0$ (since S,ζ and $M_{i,2}$ are all bounded). Note that the dominated convergence theorem (implying that $\lim\mathrm{E}(x)=\mathrm{E}(\lim x)$) is used in the second inequality which is feasible due to the result in (12). Also $\max_{m\ge T}1\left\{|s_{iS+1}(\theta_0)|>\delta\sqrt{m}\right\}$ converges to zero in probability as $T\to\infty$. This completes the proof of Lemma 5.

Lemma 6 Given Assumptions A1-A5 then

$$\sup_{T\ge1}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{it}^{2}(\theta_0)\right)<\infty. \tag{13}$$

Proof of Lemma 6 The result of Lemma 3 implies that for all i and t

$$\begin{aligned}
\mathrm{E}\left(s_{it}^{2}(\theta_0)\right) &= \frac{\zeta}{4}\mathrm{E}\left(u_{it}^{2}(\theta_0)\right),\\
&\le \frac{\zeta}{4}M_{i,2},\\
&< \infty.
\end{aligned}$$

Then

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left(s_{it}^{2}(\theta_0)\right) &\le \frac{1}{T}\sum_{t=1}^{T}\frac{\zeta}{4}M_{i,2},\\
&\le \frac{\zeta}{4}M_{i,2},
\end{aligned}$$

and the result in (13) follows which completes the proof.

Proposition 1 Let Assumptions A1-A5 hold and let the scores be as defined in (5)-(7). Then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{it}\left(\theta_0\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{\zeta}{4}\varpi_i^2\right),$$

for $T \rightarrow \infty$ and $i = 1, 2, 3$, where ϖ_i^2 is defined as in Lemma 4.

Proof of Proposition 1 By Lemma 2 $s_{it}\left(\theta_0\right)$, for all i, is a martingale difference sequence. Furthermore, the results of Lemmas 4, 5 and 6 above corresponds exactly to the conditions a), b), and c) in Theorem 7.10 in Bierens (2004) respectively. The result of Proposition 1 therefore follows immediately.

Result 2: The second order derivatives of the loglikelihood function

$$\begin{aligned}
\frac{\partial^2}{\partial\beta^2} l_T\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{y_t^2}{\sigma_t^2}\right)\left(\sum_{i=1}^{t}\left(\left(i-1\right)\beta^{(i-2)}\ln\sigma_{t-i}^2+\beta^{(i-1)}\ln\sigma_{t-i-1}^2\right)\right)+\frac{y_t^2}{\sigma_t^2}\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)^2\right), \\
\frac{\partial^2}{\partial\psi^2} l_T\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\frac{y_t^2}{\sigma_t^2}\left(\sum_{i=1}^{t}\beta^{(i-1)}g\left(y_{t-i}\right)\right)^2, \\
\frac{\partial^2}{\partial\phi^2} l_T\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\frac{y_t^2}{\sigma_t^2}\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)^2, \\
\frac{\partial^2}{\partial\beta\partial\psi} l_T\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{y_t^2}{\sigma_t^2}\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}g\left(y_{t-i}\right)\right)+\frac{y_t^2}{\sigma_t^2}\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}g\left(y_{t-i}\right)\right)\right), \\
\frac{\partial^2}{\partial\beta\partial\phi} l_T\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\left(\left(1-\frac{y_t^2}{\sigma_t^2}\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)+\frac{y_t^2}{\sigma_t^2}\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)\right), \\
\frac{\partial^2}{\partial\psi\partial\phi} l_T\left(\theta\right) &= -\frac{1}{2}\sum_{t=1}^{T}\frac{y_t^2}{\sigma_t^2}\left(\sum_{i=1}^{t}\beta^{(i-1)}g\left(y_{t-i}\right)\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right).
\end{aligned}$$

Proposition 2 Let Assumptions A1-A5 hold. Then

(a) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta^2}l_T\left(\theta\right)|_{\theta=\theta_0}\right) \xrightarrow{p} \frac{1}{2}\varpi_1^2 > 0,$

(b) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\psi^2}l_T\left(\theta\right)|_{\theta=\theta_0}\right) \xrightarrow{p} \frac{1}{2}\varpi_2^2 > 0,$

(c) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\phi^2}l_T\left(\theta\right)|_{\theta=\theta_0}\right) \xrightarrow{p} \frac{1}{2}\varpi_3^2 > 0,$

(d) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\psi}l_T\left(\theta\right)|_{\theta=\theta_0}\right) \xrightarrow{p} \frac{1}{2}\varpi_{12},$

(e) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\beta\partial\phi}l_T\left(\theta\right)|_{\theta=\theta_0}\right) \xrightarrow{p} \frac{1}{2}\varpi_{13},$

(f) $\frac{1}{T}\left(-\frac{\partial^2}{\partial\psi\partial\phi}l_T\left(\theta\right)|_{\theta=\theta_0}\right) \xrightarrow{p} \frac{1}{2}\varpi_{23},$

as $T \longrightarrow \infty$.

Proof of Proposition 2 Define $z_{it}(\theta_0)$ and σ_i^2 for $i = 1, 2, 3$, as in the proof of Lemma 4. Further, consider

$$u_{4t}(\theta_0) \equiv \sum_{i=1}^{t} \left((i-1)\, \beta_0^{(i-2)} \ln \sigma_{t-i}^2 (\theta_0) + \beta_0^{(i-1)} \ln \sigma_{t-i-1}^2 (\theta_0) \right).$$

Then,

$$\begin{aligned} \frac{1}{T}\left(-\frac{\partial^2}{\partial \beta^2} l_T(\theta)\mid_{\theta=\theta_0}\right) &= \frac{1}{2T}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right) u_{4t}(\theta_0) + \frac{1}{2}\sum_{t=1}^{T} \upsilon_t^2 z_{1t}(\theta_0), \\ &\xrightarrow{p} \frac{1}{2}\varpi_1^2, \end{aligned}$$

as T$\rightarrow \infty$ since, by the ergodicity theorem,

$$\frac{1}{2T}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right) u_{4t}(\theta_0) \xrightarrow{p} \frac{1}{2}\mathrm{E}\left(\left(1-\upsilon_t^2\right)\right)\mathrm{E}\left(u_{4t}(\theta_0)\right),$$

as υ_t is iid, $\mathrm{E}\left(\left(1-\upsilon_t^2\right)\right) = 0$ and $\mathrm{E}\left(u_{4t}(\theta_0)\right) < \infty$. This completes the proof of (a). Similarly, it follows that

$$\begin{aligned} \frac{1}{T}\left(-\frac{\partial^2}{\partial \psi^2} l_T(\theta)\mid_{\theta=\theta_0}\right) &= \frac{1}{2}\sum_{t=1}^{T} \upsilon_t^2 z_{2t}(\theta_0), \\ &\xrightarrow{p} \frac{1}{2}\varpi_2^2, \\ \frac{1}{T}\left(-\frac{\partial^2}{\partial \phi^2} l_T(\theta)\mid_{\theta=\theta_0}\right) &= \frac{1}{2}\sum_{t=1}^{T} \upsilon_t^2 z_{3t}(\theta_0), \\ &\xrightarrow{p} \frac{1}{2}\varpi_3^2, \end{aligned}$$

as $T \rightarrow \infty$ hereby proving (b) and (c).

Consider next the cross derivatives in (d) - (f) and let $\mathrm{E}\left(u_{it}(\theta_0)\, u_{jt}(\theta_0)\right) = \sigma_{ij}$ for $i \neq j$ where $u_{it}(\theta_0)$ is defined as in Lemma 3. By applying the ergodicity theorem and the results on the existence of moments of $u_{it}(\theta_0)$ in Lemma 3, then

$$\begin{aligned} \frac{1}{T}\left(-\frac{\partial^2}{\partial \beta \partial \psi} l_T(\theta)\mid_{\theta=\theta_0}\right) &= \frac{1}{2T}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right) u_{2t}(\theta_0) + \frac{1}{2T}\sum_{t=1}^{T} \upsilon_t^2 u_{1t}(\theta_0)\, u_{2t}(\theta_0), \\ &\xrightarrow{p} \frac{1}{2}\varpi_{12}, \\ \frac{1}{T}\left(-\frac{\partial^2}{\partial \beta \partial \phi} l_T(\theta)\mid_{\theta=\theta_0}\right) &= \frac{1}{2T}\sum_{t=1}^{T}\left(1-\upsilon_t^2\right) u_{3t}(\theta_0) + \frac{1}{2T}\sum_{t=1}^{T} \upsilon_t^2 u_{1t}(\theta_0)\, u_{3t}(\theta_0), \\ &\xrightarrow{p} \frac{1}{2}\varpi_{13}. \\ \frac{1}{T}\left(-\frac{\partial^2}{\partial \psi \partial \phi} l_T(\theta)\mid_{\theta=\theta_0}\right) &= \frac{1}{2T}\sum_{t=1}^{T} \upsilon_t^2 u_{2t}(\theta_0)\, u_{3t}(\theta_0), \\ &\xrightarrow{p} \frac{1}{2}\varpi_{23}. \end{aligned}$$

as $T \rightarrow \infty$. This completes the proof of Proposition 2.

Result 3: The third order derivatives

$$
\begin{aligned}
\frac{\partial^3 l_T(\theta)}{(\partial\beta)^3} &= -\frac{1}{2}\sum_{t=1}^{T}\left(1-y_t^2\sigma_t^{-2}\right)\left(\sum_{i=1}^{t}(i-1)(i-2)\beta^{(i-3)}\ln\sigma_{t-i}^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}\left(1-y_t^2\sigma_t^{-2}\right)\left(\sum_{i=1}^{t}\sum_{j=1}^{t-i}(i+j-2)\beta^{(i+j-3)}\ln\sigma_{t-i-j}^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}\left(1-y_t^2\sigma_t^{-2}\right)\left(\sum_{i=1}^{t}\sum_{j=1}^{t-i}\sum_{n=1}^{t-i-n}\beta^{(i+j+n-3)}\ln\sigma_{t-i-j-n}^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)^3 \\
&\quad +\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)\left(\left(\sum_{i=1}^{t}(i-1)\beta^{(i-2)}\ln\sigma_{t-i}^2\right)+\left(\sum_{i=1}^{t}\sum_{j=1}^{t-i}\beta^{(i+j-2)}\ln\sigma_{t-i-j}^2\right)\right), \\
\frac{\partial^3 l_T(\theta)}{(\partial\psi)^3} &= -\frac{1}{2}\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right)^3, \\
\frac{\partial^3 l_T(\theta)}{(\partial\phi)^3} &= -\frac{1}{2}\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)^3, \\
\frac{\partial^3 l_T(\theta)}{(\partial\beta)^2\partial\psi} &= -\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}(i-1)\beta^{i-2}g(y_{t-i})\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right)\left(\sum_{i=1}^{t}(i-1)\beta^{(i-2)}\ln\sigma_{t-i}^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right)\left(\left(\sum_{i=1}^{t}\sum_{j=1}^{t-i}\beta^{(i+j-2)}\ln\sigma_{t-i-j}^2\right)-\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}\left(1-y_t^2\sigma_t^{-2}\right)\left(\sum_{i=1}^{t}(i-1)(i-2)\beta^{i-3}g(y_{t-i})\right), \\
\frac{\partial^3 l_T(\theta)}{(\partial\psi)^2\partial\beta} &= -\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}(i-1)\beta^{i-2}g(y_{t-i})\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right) \\
&\quad +\frac{1}{2}\sum_{t=1}^{T}y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right)^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial^3 l_T(\theta)}{(\partial\beta)^2\partial\phi} &= -\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}(i-1)\beta^{i-2}y_{t-i}\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)\left(\sum_{i=1}^{t}(i-1)\beta^{(i-2)}\ln\sigma_{t-i}^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\left(\sum_{i=1}^{t}\sum_{j=1}^{t-i}\beta^{(i+j-2)}\ln\sigma_{t-i-j}^2\right)-\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)^2\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T}\left(1-y_t^2\sigma_t^{-2}\right)\left(\sum_{i=1}^{t}(i-1)(i-2)\beta^{i-3}y_{t-i}\right) \\
\frac{\partial^3 l_T(\theta)}{\partial\beta(\partial\phi)^2} &= -\frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(2\left(\sum_{i=1}^{t}(i-1)\beta^{i-2}y_{t-i}\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)\right) \\
&\quad +\frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)^2\right), \\
\frac{\partial^3 l_T(\theta)}{(\partial\psi)^2\partial\phi} &= \frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right)^2\right), \\
\frac{\partial^2 l_T(\theta)}{\partial\psi(\partial\phi)^2} &= \frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)^2\right), \\
\frac{\partial^3 l_T(\theta)}{\partial\psi\partial\phi\partial\beta} &= -\frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\left(\sum_{i=1}^{t}(i-1)\beta^{i-2}y_{t-i}\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right)\right) \\
&\quad -\frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)\left(\sum_{i=1}^{t}(i-1)\beta^{i-2}g(y_{t-i})\right)\right) \\
&\quad +\frac{1}{2}\sum_{t=1}^{T} y_t^2\sigma_t^{-2}\left(\sum_{i=1}^{t}\beta^{(i-1)}\ln\sigma_{t-i}^2\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)\left(\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})\right),
\end{aligned}
$$

Definition 1 Denote $\theta_0 = (\beta_0, \psi_0, \phi_0, \gamma_0)'$. Define the lower and upper values for each parameter in θ_0 as

$$
\begin{gathered}
\beta_L < \beta_0 < \beta_U; \quad \psi_L < \psi_0 < \psi_U, \\
\phi_L < \phi_0 < \phi_U; \quad \gamma_L < \gamma_0 < \gamma_U,
\end{gathered}
$$

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

$$
N(\theta_0) = \{\theta \backslash \beta_L \le \beta \le \beta_U, \psi_L \le \psi \le \psi_U, \phi_L \le \phi \le \phi_U, \text{ and } \gamma_L < \gamma < \gamma_U\}. \tag{14}
$$

Proposition 3 Under Assumptions A1-A5, there exists a neighborhood $N(\theta_0)$ given in (14) for which

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

$$
\begin{array}{llll}
(c) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\phi^3}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{3t}; & (d) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\beta^2\partial\psi}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{4t};\\
(e) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\beta^2\partial\phi}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{5t}; & (f) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\psi^2\partial\phi}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{6t};\\
(g) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\beta\partial\psi^2}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{7t}; & (h) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\beta\partial\phi^2}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{8t};\\
(i) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\psi\partial\phi^2}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{9t} & (j) & \sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3}{\partial\beta\partial\psi\partial\phi}l_T(\theta)\right|\le\frac{1}{T}\sum_{t=1}^{T}w_{10t},
\end{array}
$$

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

Proof of Proposition 3 We define in this Proposition σ_t^2 to be the conditional variance evaluated at θ and $\sigma_t^2(\theta_0)$ when evaluated at the true parameter. Then, following Jensen and Rahbek (2004a, 2004b), by definition

$$
\frac{y_t^2}{\sigma_t^2(\theta_0)}=v_t^2.
$$

Therefore

$$
\begin{aligned}
\frac{\sigma_t^2(\theta_0)}{\sigma_t^2} &= \frac{\exp\left(\sum_{i=0}^{t-1}\beta_0^i+\beta_0^t\ln\sigma_0^2+\psi_0\sum_{i=1}^{t}\beta_0^{(i-1)}g(y_{t-i})+\phi_0\sum_{i=1}^{t}\beta_0^{(i-1)}y_{t-i}\right)}{\exp\left(\sum_{i=0}^{t-1}\beta^i+\beta^t\ln\sigma_0^2+\psi\sum_{i=1}^{t}\beta^{(i-1)}g(y_{t-i})+\phi\sum_{i=1}^{t}\beta^{(i-1)}y_{t-i}\right)}\\
&=\exp\left(\ln\sigma_0^2(\beta_0-\beta)^t\right)\prod_{i=0}^{t-1}\exp\left((\beta_0-\beta)^i\right)\prod_{i=1}^{t}\frac{\exp\left(\psi_0\beta_0^{(i-1)}g(y_{t-i})\right)}{\exp\left(\psi\beta^{(i-1)}g(y_{t-i})\right)}\\
&\quad\times\prod_{i=1}^{t}\frac{\exp\left(\phi_0\beta_0^{(i-1)}y_{t-i}\right)}{\exp\left(\phi\beta^{(i-1)}y_{t-i}\right)}\le B_1,
\end{aligned}
$$

where for $\beta_0\le\beta\le\beta_U,\psi_0\le\psi\le\psi_U,\phi_0\le\phi\le\phi_U$, and $\gamma_0<\gamma<\gamma_U$ we have that

$$
\begin{aligned}
B_1 &= \exp\left(\left|\ln\sigma_0^2(\beta_0-\beta_U)\right|^t\right)\prod_{i=0}^{t-1}\exp\left(\left|(\beta_0-\beta_U)^i\right|\right)\prod_{i=1}^{t}1/\exp\left(\left|\left(\psi_U\beta_U^{(i-1)}-\psi_0\beta_0^{(i-1)}\right)y_{t-i}\right|\right)\\
&\quad\times\prod_{i=1}^{t}1/\exp\left(\left|\left(\phi_U\beta_U^{(i-1)}-\phi_0\beta_0^{(i-1)}\right)y_{t-i}\right|\right),
\end{aligned}
$$

and for $\beta_L\le\beta\le\beta_0,\psi_L\le\psi\le\psi_0,\phi_L\le\phi\le\phi_0$, and $\gamma_L<\gamma<\gamma_0$ we get

$$
\begin{aligned}
B_1 &= \exp\left(\left|\ln\sigma_0^2(\beta_L-\beta_0)^t\right|\right)\prod_{i=0}^{t-1}\exp\left(\left|(\beta_L-\beta_0)^i\right|\right)\prod_{i=1}^{t}1/\exp\left(\left|\left(\psi_0\beta_0^{(i-1)}-\psi_L\beta_L^{(i-1)}\right)y_{t-i}\right|\right)\\
&\quad\times\prod_{i=1}^{t}1/\exp\left(\left|\left(\phi_0\beta_0^{(i-1)}-\phi_L\beta_L^{(i-1)}\right)y_{t-i}\right|\right),
\end{aligned}
$$

and where for example we have obtained the inequality of the term $\frac{\exp\left(\phi_0\beta_0^{(i-1)}y_{t-i}\right)}{\exp\left(\phi_U\beta_U^{(i-1)}y_{t-i}\right)}\le\frac{1}{\exp\left(\left|\left(\phi_U\beta_U^{(i-1)}-\phi_0\beta_0^{(i-1)}\right)y_{t-i}\right|\right)}$ along the same lines of expression (23) in Jensen and Rahbek (2004b). Then, under assumptions A1-A5,

$B_1 < \infty$. Therefore expression $\frac{\sigma_t^2(\theta_0)}{\sigma_t^2}$ is uniformly bounded in a region around the true value of the parameter space. We start now with the proof of (a). We can define, due to the expression of the third order derivatives, that

$$\left|\frac{\partial^3}{\partial \beta^3} l_T(\theta)\right| \leq \frac{1}{T}\sum_{t=1}^{T} w_{1t}(\theta),$$

with

$$\begin{aligned}
w_{1t}(\theta) &= \frac{1}{2}\left(1+\upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\right)\left(\left|\sum_{i=1}^{t}(i-1)(i-2)\beta^{(i-3)} \ln \sigma_{t-i}^2\right|\right) \\
&+\frac{1}{2}\left(1+\upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\right)\left(\left|\sum_{i=1}^{t}\sum_{j=1}^{t-i}(i+j-2)\beta^{(i+j-3)} \ln \sigma_{t-i-j}^2\right|\right) \\
&+\frac{1}{2}\left(1+\upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\right)\left(\left|\sum_{i=1}^{t}\sum_{j=1}^{t-i}\sum_{n=1}^{t-i-n}\beta^{(i+j+n-3)} \ln \sigma_{t-i-j-n}^2\right|\right) \\
&+\frac{1}{2}\sum_{t=1}^{T}\upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\left(\left|\sum_{i=1}^{t}\beta^{(i-1)} \ln \sigma_{t-i}^2\right|\right)^3 \\
&+\sum_{t=1}^{T}\upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2}\left(\left|\sum_{i=1}^{t}\beta^{(i-1)} \ln \sigma_{t-i}^2\right|\right)\left(\left(\left|\sum_{i=1}^{t}(i-1)\beta^{(i-2)} \ln \sigma_{t-i}^2\right|\right)+\left(\left|\sum_{i=1}^{t}\sum_{j=1}^{t-i}\beta^{(i+j-2)} \ln \sigma_{t-i-j}^2\right|\right)\right),
\end{aligned}$$

where we prove now that each of the quantities in $w_{1t}(\theta)$ are bounded by functions that have any desired moments, and $E(w_{1t}) = M_1 < \infty$ under assumptions A1-A5. The proof involves, the same as before, to use the expressions given in the proof of Proposition 2. For example from Lemma 3

$$\begin{aligned}
& \sup_{\theta\in N(\theta_0)}\left|\sum_{i=1}^{t}\beta^{(i-1)} \ln \sigma_{t-i}^2(\theta)\right|, \\
\leq\ & \sup_{\theta\in N(\theta_0)} c[t\frac{|\beta|^t}{|\beta|^2-|\beta|}+\frac{1-|\beta|^t}{|\beta|^2-2|\beta|+1}+\frac{1}{|\beta|^2-1}\left(|\beta|^{3t+1}-|\beta|^{t+1}\right) \\
& +|\psi|\sum_{i=1}^{t}\sum_{j=1}^{t-i}|\beta|^{(j+i-2)}|g(y_{t-i-j})|+|\phi|\sum_{i=1}^{t}\sum_{j=1}^{t-i}|\beta|^{(j+i-2)}|y_{t-i-j}|] \\
\equiv\ & A_{1t},
\end{aligned}$$

where moments of A_{1t} will exists to the degree that the moments of $|g(y_{t-i-j})|)$ and $|y_{t-i-j}|$ are bounded (as assumed in A5). Consequently, as $\sup_{\theta\in N(\theta_0)}\left|\sum_{i=1}^{t}\beta^{(i-1)} \ln \sigma_{t-i}^2(\theta)\right| \leq A_{1t}$ uniformly on t, the corresponding moments of $\sup_{\theta\in N(\theta_0)}\left|\sum_{i=1}^{t}\beta^{(i-1)} \ln \sigma_{t-i}^2(\theta)\right|$ will exist to the degree that the moments of $|g(y_{t-i-j})|)$ and $|y_{t-i-j}|$ are bounded. We apply the same inequality to each of the terms in $w_{1t}(\theta)$ and we find the dominating function w_{1t} where $\sup_{\theta\in N(\theta_0)} w_{1t}(\theta) \leq w_{1t}$. Finally, the convergence $\frac{1}{T}\sum_{t=1}^{T} w_{1t} \overset{a.s.}{\to} M_1$ follows by the ergodic theorem. In relation to (b), we have that

$$w_{2t}(\theta) = \upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{2\sigma_t^2}\left(\left|\sum_{i=1}^{t}\beta^{(i-1)} g(y_{t-i})\right|\right)^3.$$

After we find the dominating function given in Lemma 3, $\mathrm{E}(w_{2t}) = M_2 < \infty$ under assumptions A1-A5, and where $\mathrm{E}\left[(|g(y_{t-i})|)^3\right] < \infty$. We use again the bounds given in Definition 1 following Jensen and

Rahbek (2004b). (c) follows exactly the same as (b) but where we need to impose $\mathrm{E}\left[(|y_{t-i}|)^3\right] < \infty$, which is true by the assumptions.(d) needs

$$\begin{aligned} w_{4t}(\theta) &= \upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \left(\left| \sum_{i=1}^{t} (i-1)\beta^{i-2} g(y_{t-i}) \right| \right) \left(\left| \sum_{i=1}^{t} \beta^{(i-1)} \ln \sigma_{t-i}^2 \right| \right) \\ &+ \frac{1}{2} \upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \left(\left| \sum_{i=1}^{t} \beta^{(i-1)} g(y_{t-i}) \right| \right) \left(\left| \sum_{i=1}^{t} (i-1)\beta^{(i-2)} \ln \sigma_{t-i}^2 \right| \right) \\ &+ \frac{1}{2} \upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \left(\left| \sum_{i=1}^{t} \beta^{(i-1)} g(y_{t-i}) \right| \right) \left(\left(\left| \sum_{i=1}^{t} \sum_{j=1}^{t-i} \beta^{(i+j-2)} \ln \sigma_{t-i-j}^2 \right| \right) - \left(\left| \sum_{i=1}^{t} \beta^{(i-1)} \ln \sigma_{t-i}^2 \right| \right)^2 \right) \\ &+ \frac{1}{2} \left(1 + \upsilon_t^2 \frac{\sigma_t^2(\theta_0)}{\sigma_t^2} \right) \left(\left| \sum_{i=1}^{t} (i-1)(i-2)\beta^{i-3} g(y_{t-i}) \right| \right), \end{aligned}$$

where again all the quantities are bounded by functions with the desired moments under assumptions A1-A5 and the dominating functions are given in Lemma 3 and the in the proof of Proposition 2. Expression (e) follows directly from the proof of (d). (f) follows from (b), but under the assumption that $\mathrm{E}\left[|y_{t-i}| (|g(y_{t-i})|)^2\right] < \infty$. Expressions (g) and (h) follow as well directly from the assumptions and the previous line of proof. (i) requires $E\left[(|y_{t-i}|)^2 (|g(y_{t-i})|)\right] < \infty$, and finally, (j) follows directly from the assumptions and the same proof structure. The proof is completed with the convergence $\frac{1}{T}\sum_{t=1}^{T} w_{it} \xrightarrow{a.s.} M_i, \forall i = 1, ..., 10$, that follows for expressions (a)-(j) from the ergodic theorem.

Proof of Theorem 1 Given the conditions provided by Propositions 1 - 3 the results of Theorem 1 follow straightforwardly from Lemma 1, page 1206 in Jensen and Rahbek (2004b).