2017/15

DP

Arie Preminger and Giuseppe Storti

Least squares estimation for Garch (1,1) model with heavy tailed errors

CORE
CENTER FOR
OPERATIONS RESEARCH
AND ECONOMETRICS

CORE
Voie du Roman Pays 34, L1.03.01
Tel (32 10) 47 43 04
Fax (32 10) 47 43 01
Email: immaq-library@uclouvain.be
https://uclouvain.be/en/research-institutes/immaq/core/discussion-papers.html

Least squares estimation for GARCH (1,1) model with heavy tailed errors

A. Preminger[†] and G. Storti[‡]

[†], *CORE, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.*
E-mail: ariepr@hotufi.net
[‡], *Università degli studi di Salerno, Via G. Paolo II, 132, 84084, Fisciano (SA), Italy.*
E-mail: storti@unisa.it
Received: February 2017

Summary GARCH (1,1) models are widely used for modelling processes with time varying volatility. These include financial time series, which can be particularly heavy tailed. In this paper, we propose a novel log-transform-based least squares approach to the estimation of GARCH(1,1) models. Within this approach the scale of the estimated volatility is dependent on an unknown tuning constant. By means of a backtesting exercise on both real and simulated data we show that knowledge of the tuning constant is not crucial for Value at Risk prediction. However, this does not apply to many other applications where correct identification of the volatility scale is required. In order to overcome this difficulty, we propose two alternative two-stage least squares estimators (LSE) and derive their asymptotic properties under very mild moment conditions for the errors. In particular, we establish the consistency and asymptotic normality at the standard convergence rate of $\sqrt{n}$ for our estimators. Their finite sample properties are assessed by means of an extensive simulation study.

Keywords: *GARCH (1,1), least squares estimation, heavy tails, consistency, asymptotic normality, two-step estimator.*

1. INTRODUCTION

In the last three decades there has been a large amount of theoretical and empirical research on modelling the conditional volatility of financial time series data. These time series, which appear to be uncorrelated, exhibit dependence in their squares, a notable example being the daily financial returns. The practical motivation lies in the increasing need to explain and to model risk and uncertainty usually associated with financial returns. One of the most successful approaches for modelling volatility makes use of the generalized autoregressive conditionally heteroskedasticity (GARCH) model, suggested by Bollerslev (1986), and its numerous extensions. Indeed, its simplicity and intuitive appeal make the GARCH model, especially the GARCH(1,1), a good starting point in many financial applications, see e.g. Hansen and Lunde (2005).

The main approach for the estimation of GARCH models is the Gaussian quasi-maximum likelihood estimator (GQMLE) approach where the estimates are obtained through maximization of a Gaussian likelihood function. Bollerslev and Wooldridge (1992) derived the asymptotic distribution of the QMLE under high level assumptions. When the errors have finite fourth moment the consistency and asymptotic normality of the QMLE for the GARCH(1,1) have been established by Lee and Hansen (1994) and Lumsdaine (1996). These results were extended to the case of GARCH (p,q) by Boussama (1998), Berkes et al. (2003) and Francq and Zakoïan (2004). However, empirical evidence indicates that for many financial time series, the distribution of errors is far from being Gaussian and it is usually heavy tailed (Mittnik and Rachev, 2000). Hall and Yao (2003) studied the QMLE for heavy tailed errors (without finite fourth moment). They showed that the asymptotic distribution may be non-Gaussian and the convergence rate is slower than $\sqrt{n}$. Mikosch, and Straumann (2006) establish similar results for a more general class of GARCH type models. Furthermore, even when the standardized errors have finite fourth moment, i.e., $E(\varepsilon_t^4) < \infty$, the divergence of the Gaussian likelihood from the true error density may result in substantial loss of efficiency, reflecting the cost of not knowing the true innovation distribution.

The non-regular behavior of the GQMLE in heavy tailed settings has motivated a long strand of research on the development of alternative quasi-likelihood estimators. Starting from Engle and Gonzalez-Rivera (1991), attention has initially been paid to semi-parametric procedures later moving to the development of estimators based on the maximization of non-Gaussian quasi likelihood functions. A common drawback to the early literature on non-Gaussian QMLE (NGQMLE) is that, unless the chosen parametric likelihood family contains the true likelihood, the resulting estimator is inconsistent, see e.g. Newey and Steigerwald (1997) and Berkes and Horwath (2004). Recently, Fan et al. (2014) have proposed a novel NGQMLE which is robust against density misspecification and more efficient than the GQMLE. They introduce a two-step estimation procedure in which a scale adjustment parameter is estimated via GQMLE in the first step, and then it is fed into a non-Gaussian likelihood in the second step. Related works include Lee and Lee (2009), where the likelihood is chosen from a parametric Gaussian mixture, and Francq et al. (2011), which construct a two-stage NGQMLE using a generalized Gaussian likelihood.

In this paper, we consider a log-transform-based least squares estimator (LSE) for the parameters of a GARCH(1,1) model. In order to establish our asymptotic theory, we impose mild moment conditions on the errors which account for the possibility of heavy tails. In addition, we require that the process satisfies the necessary and sufficient condition for strict stationarity as given by Nelson (1990), which allows for mildly explosive GARCH processes. We establish the consistency and asymptotic normality of the proposed LSE.

It is worth noting that the implementation of the proposed estimator requires knowledge of a scaling factor depending on the distribution of the underlying innovations. Otherwise, volatility will be known only up to some scaling factor. We show that, in some applications, such as VaR prediction, this issue is not relevant since knowledge of the correct scaling factor is not required. On an empirical ground, the results of an empirical application to four time series of financial returns provide strong evidence in this direction confirming that, although the correct scale of volatility is unknown, accurate VaR predictions can still be obtained using a misspecified and, possibly, biased LSE. To our knowledge, this issue has not previously been addressed in the financial econometrics literature. This makes the results of our VaR forecasting exercise an additional contribution of the paper, of interest for practitioners interested in predicting VaR in heavy tailed environments.

On the other hand, in order to deal with cases in which knowledge of the correct volatility scaling factor is required, we propose an alternative two-stage estimator (LSE_0). In the first stage, the model parameters are estimated up to an unknown scale factor by using our LS estimator with an arbitrarily chosen tuning constant. In the second stage, we recover the correct scale by imposing an identification restriction. In standard settings, where it is assumed that $E(\varepsilon_t^2) < \infty$, the most natural choice is $E(\varepsilon_t^2) = 1$, leading the h_t^2 to be the conditional variance of returns. However, in settings in which one is not willing to assume finite second moment for the standardized error ε_t, some other identification restriction should be imposed. Following Peng & Yao (2003), among others, we assume that the squared standardized errors have a unit median. Similarly, a two-step estimation approach was suggested by Chen et al. (2009). Under their setting no functional form is assumed for the volatility. In the first step the volatility is estimated non-parametrically up to some unknown scaling factor while, in the second step, by assuming a unit error variance (we assume unit median) this scaling factor is identified.

A further point to note is that the basic LSE can be easily shown to be equivalent to a QMLE based on a Gaussian approximation of the density of the log-transformed squared innovations where the mean of the distribution plays the role of a nuisance parameter. This analogy suggests a further two-stage variant of the LSE (LSE_Q) in which the tuning constant is pre-estimated by means of the GMQLE or some other consistent estimator. Namely, in a first step, the GQMLE is used to obtain a consistent estimate of the scaling parameter and, in a second step, the LSE is used to estimate the model parameters. Our approach is similar in spirit to the Non-Gaussian QMLE (NGQMLE) recently proposed by Fan et al. (2014) who also use the GQMLE to pre-estimate an unknown shape parameter appearing in the QL function. Consistency of the LSE_Q will crucially depend on the consistency of the first stage estimator of c_0. So, if the GQMLE is chosen for first stage estimation, we do not expect the LSE_Q estimator to be an alternative to the LSE_0 in cases in which the model innovations ε_t are characterized by heavy tails and, eventually, infinite second moment but we rather expect it to improve over the efficiency of the GQMLE in cases in which $E(\varepsilon_t^4) < \infty$.

The finite sample efficiency of the proposed two-stage LS estimators (LSE_0 and LSE_Q) has then been assessed by means of a simulation study considering different error distributions as well as different persistence levels of the volatility process. Their performances have been compared with those obtained by the classical GQMLE and the NGQMLE proposed by Fan et al. (2014). The results suggest that the LSE can be substantially more efficient than both the GQMLE and the NGQMLE in cases where the error distribution is characterized by substantial deviations from the Normality assumption and, in particular, it is heavy tailed.

The structure of the paper is as follows. In Section 2, we discuss the basic LSE when the volatility scaling factor is assumed known. Next, in section 3 we deal with the case in which the scaling factor is unknown. In particular, we propose and discuss two different two-stage variants of the standard LSE. The first (LSE_0) identifies the correct scaling factor by imposing some appropriately chosen identification restriction on the model while the second (LSE_Q) pre-estimates the scaling factor by means of some alternative consistent estimator such as the GQMLE. The asymptotic properties of the proposed estimators are derived in section 4 while, in section 5, we conduct a simulation study aiming at investigating the small sample properties of the proposed two-stage estimators: LSE_0 and LSE_Q. In section 6, we show that knowledge of the volatility scaling factor is not relevant for the estimation of Value at Risk (VaR) presenting the results of an application to the estimation of VaR for both real and simulated data. Finally, section 7 concludes. The mathematical proofs are presented in the Appendix.

We use the following notations throughout the paper. $|A| = (tr(A'A))^{1/2}$ denotes the Euclidian norm of a vector or a matrix and $||A||_r = (\mathrm{E}(|A|^r))^{1/r}$ denotes the L^r-norm of a random vector or matrix and let $\Rightarrow$ denote weak convergence with respect to the uniform metric. The symbol $\overset{d}{\to}$ denotes convergence in distribution. The symbol $\to_{a.s.}$ ($\to_p$) denotes convergence almost surely (or in probability). $o_{a.s.}(1)$ denotes a series of random variables that converges to zero almost surely (a.s.).

2. LEAST SQUARES ESTIMATION OF GARCH(1,1) MODELS: GENERAL FRAMEWORK

In this section we introduce the basic LSE for the GARCH(1,1) model as proposed by Bollerslev (1986). Our reference model is given by

$$y_t = \sqrt{h_{0t}}\varepsilon_t \tag{2.1}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (iid) random variables with $\mathrm{E}(\varepsilon_t) = 0$ and

$$h_{0t} = \omega_0 + \alpha_0 y_{t-1}^2 + \beta_0 h_{0t-1} \tag{2.2}$$

The process is described by an unknown parameter vector $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$. If $\mathrm{E}(\varepsilon_t^2) = 1$ then h_{0t} is the conditional variance of y_t given the past. However, without any moment conditions, $h_{0t}^{0.5}$ is the conditional scaling parameter of the observed process. Let $c_0 = \mathrm{E}(\ln \varepsilon_t^2)$ and assume that c_0 is finite, which is implied by our results below (namely, it is shown in Lemma 1(iii) that $c_0 < \infty$). By squaring the terms in (2.1) and taking the logarithm, we obtain

$$\ln y_t^2 - c_0 = \ln h_{0t} + \eta_t \tag{2.3}$$

where $(\ln y_t^2 - c_0)$ and $\eta_t = \ln \varepsilon_t^2 - c_0$ are zero mean iid random variables. This nonlinear regression can be estimated via least squares. Thus, the objective function is given by

$$\tilde{Q}_n(\theta) = \frac{1}{2n}\sum_{t=1}^{n} \tilde{\ell}_t(\theta) = \frac{1}{2n}\sum_{t=1}^{n} (\ln y_t^2 - c_0 - \ln \tilde{h}_t(\theta))^2 \tag{2.4}$$

where $\theta = (\omega, \alpha, \beta)'$ belongs to a parameter space $\Theta \in (0, \infty) \times [0, \infty)^2$; $\tilde{h}_t(\theta)$ is defined recursively, for $t \geq 2$ by

$$\tilde{h}_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta \tilde{h}_{t-1}(\theta) \tag{2.5}$$

with the initial condition $\tilde{h}_1(\theta) = \omega$ that, on practical grounds, mimics the situation in which the series starts in a very low volatility period virtually characterized by the absence of remarkable shocks. Note that, since our framework allows to deal with integrated GARCH processes, it would be unduly restrictive to initialize the recursion at the unconditional variance of returns $\omega(1 - \alpha - \beta)^{-1}$ that can be meaningfully calculated only for weakly stationary processes.

The LSE of θ is defined as any measurable solution $\hat{\theta}_n$ of

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \tilde{Q}_n(\theta). \tag{2.6}$$

It is trivial to show that the proposed LSE has a QMLE interpretation. Under this respect we note that the LSE estimator can be in general interpreted as the maximizer of an appropriately defined quasi-likelihood function for $\ln \varepsilon_t^2$. Namely, let $f(.|\theta, w)$ be the (unknown) density of $\ln \varepsilon_t^2$, conditional on information at time $t-1$ ($\mathcal{I}_{t-1}$), and w a vector of nuisance parameters. Let then $g(.)$ be some approximation of this unknown density that we use for building the quasi log-likelihood function we choose for estimation. A general class of QMLEs can be defined as:

$$\hat{\vartheta} = \arg\max_{\theta \in \Theta} \quad QL(\theta, w) \tag{2.7}$$

where $QL(\theta, w) = \sum_{t=1}^{T} \ln g(\ln(\varepsilon_t^2)|\theta, w)$. It is trivial to show that, the LSE estimator in (2.6) can be obtained as a special case of (2.7) by setting $g(.)$ equal to a Normal density with mean c_0 and standard deviation s_0. The quasi log-likelihood function is then given by

$$QL(\theta, c_0, s_0) = -\frac{T}{2} log(s_0^2) - \frac{T}{2} \ln(2\pi) - \sum_{t=1}^{T} \left(\frac{\ln \varepsilon_t^2 - c_0}{2s_0^2} \right),$$

where, in order to estimate θ, knowledge of c_0 is required while the same does not hold for s_0, since the first order conditions for the estimation of θ are linear in this parameter.

For deriving the asymptotic properties of $\hat{\theta}_n$, in the remainder it will be also convenient to work with $h_t(\theta)$ the unobserved conditional variance

$$h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta) = \omega + \alpha \sum_{k=0}^{\infty} \beta^k y_{t-k-1}^2. \tag{2.8}$$

The process $h_t(\theta)$ is the conditional variance model when the infinite past history of the data is observed. Also, note that $h_{0t} = h_t(\theta_0)$. For the unobserved process we construct the following unobserved objective function

$$Q_n(\theta) = \frac{1}{2n} \sum_{t=1}^{n} (\ln y_t^2 - c_0 - \ln h_t(\theta))^2 = \frac{1}{2n} \sum_{t=1}^{n} \ell_t(\theta) \tag{2.9}$$

The primary difference between the two objective functions is that $Q_n(\theta)$ is computed as if we had observed the random function $h_1(\theta)$. In practice, we can only use (2.5) for estimation. In section 4 it will be shown that the choice of the initial values does not matter for the asymptotic properties of the LSE.

A crucial issue in the implementation of the LSE in (2.6) is related to the identification of the tuning constant c_0. In this section, we have assumed that the scaling factor c_0 is known. This assumption simplifies the discussion and implies that the practitioner has some a-priori knowledge or can formulate some reasonable assumptions about the distribution of the errors [1].

It is also worth remarking that knowledge of the constant c_0 is not required in all financial applications, a notable

[1]For stochastic volatility models, an approach similar to ours was considered by Ruiz(1994) and Harvey et al.(1994), where it was assumed that the error term is Gaussian which implies that the scaling constant was set to -1.27. See also Francq and Zakoian (2006a)

example being given by Value at Risk (VaR) prediction. It is indeed easy to show that, independently from the chosen value of c_0, even a simple estimation approach based on an one step LSE can allow to obtain accurate VaR predictions. *To start let us rewrite the observation equation of the data generating process as*

$$y_t = h_{0t}^{0.5}\varepsilon_t = h_{0t}^{0.5}c_\nu \nu_t = \bar{h}_{0t}^{0.5}\nu_t$$

with c_ν being an arbitrarily chosen positive constant, $E(\nu_t^2) = 1$, $\bar{h}_{0t}^{0.5} = h_{0t}^{0.5}c_\nu$. The 1-step ahead VaR can then be either defined in terms of the model representation driven by ε_t or by ν_t

$$VaR_{t,p} = \bar{h}_{0,t+1}^{0.5}\nu^{(p)} = h_{0,t+1}^{0.5}\varepsilon^{(p)}$$

where $X^{(p)}$ denotes the order p quantile of the r.v. X. Said differently, it follows that VaR can be consistently predicted using either the correctly scaled volatility $h_{0,t+1}$ or its rescaled version $\bar{h}_{0,t+1}$. Changes in the scale of volatility will be compensated by changes in the scale of the errors. In order to empirically support our statement, in section 6 we will present the result of an application to VaR prediction for some time series of real and simulated returns. Nevertheless, there are many other situations in which knowledge of the correct scale of volatility is required such as, for example, when a GARCH model is fitted in the first stage of the estimation of a Dynamic Conditional Correlation (DCC) model (Engle, 2002). In the next section we present and discuss two different proposals for overcoming this problem and handling cases in which the constant c_0 is unknown.

3. TWO-STAGE LEAST SQUARES ESTIMATION OF GARCH(1,1) MODELS

Implementation of the LSE discussed in section 2 requires knowledge of the tuning constant c_0. On the other hand, if we treat c_0 as unknown, (α_0, ω_0) can be estimated[2] up to a scale parameter. More precisely, assuming that the chosen value of the tuning constant is equal to c_1 will transform the objective function in (2.4) into

$$Q_n(\theta_1) = \frac{1}{2n}\sum_{t=1}^{n}(\ln y_t^2 - c_1 - \ln h_t(\theta_1))^2 = \frac{1}{2n}\sum_{t=1}^{n}(\ln y_t^2 - c_0 - \ln(e^{c_1-c_0}h_t(\theta_1))^2 \quad (3.10)$$

where the relationship between θ_1 and $\theta = (\omega, \alpha, \beta)'$ is such that $\theta_1 = (e^{c_0-c_1}\omega, e^{c_0-c_1}\alpha, \beta)'$ and $h_t(\theta_1) = h_t(\theta)e^{c0-c1}$. It is evident that the correct scale of the volatility function can be recovered only if c_0 is known.

Direct estimation of c_0 along with the other parameters is indeed ruled out by an identification issue. *As pointed out by a referee this is even more immediately seen if, as in Fan et al. (2014), one considers the following reparameterization of the model*

$$y_t = \sigma h_t^* \epsilon_t^* \quad (3.11)$$

where σ is a scaling parameter such that $\sigma h_t^* = h_t$. Squaring both terms of (3.11) and taking logs one obtains

$$\begin{aligned}\ln y_t^2 &= \ln\sigma^2 + \ln(h_t^*) + \ln(\epsilon_t^*)^2 \\ &= \ln\sigma^2 + c_0 + \ln h_t^* + \eta_t \\ &= \mu + \ln h_t^* + \eta_t \quad (3.12)\end{aligned}$$

where $\mu = \ln\sigma^2 + c_0$ making evident that $\ln\sigma^2$ and c_0 cannot be separately identified unless the value of c_0 is pre-determined ex ante using prior information on the error distribution.

In this section we illustrate two different variants of the basic LSE that can be used to obtain consistent estimates of the elements of θ in cases in which the value of c_0 is not identified a priori. In this case, in-order to identify the model parameters, two different strategies can be pursued. Of these, the first aims at reaching identification by imposing some adequately chosen identification restriction. In particular we impose that the median of squared innovation is equal to one. Differently, the second is based on pre-estimation of the constant c_0 by means of some consistent preliminary estimator, such as the GQMLE.

It is worth noting that the two estimators should not be seen as competing alternatives. The first one should be

[2]The β parameter is invariant to rescaling of the error term.

preferred in cases where the empirical distribution of ε_t is well approximated by a model with non finite second moment. In this case the classical unit variance identification restriction becomes meaningless and should be replaced with some alternative constraint. On the other hand, in more regular settings, we expect researchers to be more willing to use the second estimator that allows for consistent estimation of c_0 in a simple and intuitive way. In cases in which the distribution of is very heavy tailed but still has finite second moment, the procedure could be further extended replacing the GQMLE used in the first stage with some non-Gaussian QMLE.

3.1. Two-stage least squares estimation under identification restrictions

In the first approach we consider, for dealing with the case in which c_0 is unknown, the model parameters are identified by imposing some suitable identification restrictions on ε_t^2. In the standard GARCH setting, $E(\varepsilon_t^2) = 1$ acts as an identification condition, and when it is removed, $(\omega_0,\ \alpha_0)$ are identified up to a scale factor[3]. Under our setting such an assumption may be too restrictive. In what follows, we will assume that the median of ε_t^2 instead of its variance equals one. This condition was also imposed by Peng and Yao (2003) and Pan et al. (2008). The estimated parameters of (3.14) are then used to fully identify the GARCH model in a two-step estimation procedure, described below.

In order to avoid confusion, we explicitly report the reparameterized model under the unit median restriction on the squared errors. This reads as

$$y_t = \sqrt{\hbar_{0t}}\varepsilon_t \tag{3.13}$$

where $\{\varepsilon_t\}$ is now redefined as a sequence of independent and identically distributed (iid) random variables with $\mathrm{E}(\varepsilon_t) = 0$, *$med(\varepsilon_t^2) = 1$ and*

$$\hbar_{0t} = \bar{\omega}_0 + \bar{\alpha}_0 y_{t-1}^2 + \bar{\beta}_0 h_{0t-1}. \tag{3.14}$$

So, letting $\bar{\theta}_0 = (\bar{\omega}_0, \bar{\alpha}_0, \bar{\beta}_0)'$, we have that $\hbar_{0t} = h_t(\bar{\theta}_0)$ as opposed to $h_{0t} = h_t(\theta_0)$. In order to estabilish the relationship between $\bar{\theta}_0$ and θ_0, we start noting that $med(\varepsilon_t^2) = 1 \Leftrightarrow med(\ln \varepsilon_t^2) = 0$. Then by i) squaring both terms in equation (3.13) ii) taking natural logs iii) adding and subtracting $c_0 = E(\ln \varepsilon_t^2)$ on the right hand side, we obtain

$$\ln y_t^2 = \ln \hbar_{0t} + \ln \varepsilon_t^2 = \ln(e^{c_0}\hbar_{0t}) + \ln \varepsilon_t^2 - c_0 = \ln h_{0t} + \eta_t, \tag{3.15}$$

where, by equation (2.3), it is evident that $e^{c_0}\hbar_{0t} = h_{0t} \equiv h_t(\theta_0)$ is characterized as the unique conditional volatility model for which $E(\eta_t) = 0$ and $med(\eta_t) = -c_0$. It is then easy to see that $\bar{\theta}_0 = A_0\theta_0$ where

$$A_0 = e^{med(\eta_t)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-med(\eta_t)} \end{pmatrix}$$

which allows to fully identify $\bar{\theta}_0$ from $(\theta_0, med(\eta_t))$. This motivates a two-step regression-like estimation procedure. First, estimate the following regression model by LS

$$\ln y_t^2 = \ln \hbar_{0t} + \ln \varepsilon_t^2 \tag{3.16}$$

and denote by $\hat{\theta}_n$ the resulting estimator. By the same argument in equation (3.10) this will be consistent for $e^{c_0}\hbar_{0t}$ and the resulting volatility estimates will be biased by a factor equal to e^{c_0}. Under the imposed identification restriction the scaling factor can be estimated from the LS residuals of the regression in (3.16) that, by (3.15), will be consistent for η_t. So, the second stage of the estimation procedure is implemented as follows. Given $\hat{\theta}_n$, calculate $\tilde{\eta}_t = \ln \tilde{\varepsilon}_t^2(\hat{\theta}_n)$ where $\tilde{\varepsilon}_t^2(\hat{\theta}_n) = \tilde{h}_t^{-1}(\hat{\theta}_n)y_t^2$. Then calculate the two-step estimator

$$\hat{\bar{\theta}}_n = A_n\hat{\theta}_n$$

where A_n is the sample counterpart of A_0 where $med(\tilde{\eta}_t)$ is used.

[3]In another case, this restriction is often relaxed at the cost of assuming that $\omega_0 = 1$, see e.g. Lee and Noh (2013)

Basically, in the second step, we obtain an estimator of $med(\tilde{\eta}_t)$, $\hat{m}_n$, through the minimization of the absolute deviation function

$$\hat{m}_n = \arg\min_{m\,\in\,\Re} \tilde{\tilde{Q}}_n(m) = \frac{1}{n}\sum\nolimits_{t=1}^{n} \tilde{l}_t(m, \hat{\theta}_n), \quad \tilde{l}_t(m, \theta) \ = |\ln \tilde{\varepsilon}_t^2(\theta) - m|. \tag{3.17}$$

In the reminder, for ease of reference, we will indicate $\hat{\hat{\theta}}_n$ as the LSE_0 estimator. Finally, it is interesting to note that the LSE_0 estimator discussed in this section does not appear to have a direct QML interpretation. However, each of the two estimation stages involved in the procedure can be separately related to the maximization of a different QL function. The first-stage estimator can be interpreted as the maximizer of a Gaussian QL estimator. In the second stage, the scaling-factor $\hat{m}_n$, which is used to adjust the bias of the first stage estimator, is implicitly defined as the maximizer of a Laplace quasi-likelihood function.

3.2. Two-stage least squares estimation with pre-estimated c_0

An alternative approach to dealing with cases in which c_0 is unknown, is to use the QMLE to pre-estimate it in a two-stage estimation procedure as described below:

1 Given $\hat{\theta}_n^{QML}$ the QML-estimator of the model in (2.1)-(3.14), we compute $\ln \hat{\varepsilon}_t^2(\hat{\theta}_n^{QML}) = \ln \tilde{h}_t^{-1}(\hat{\theta}_n^{QML}) y_t^2$

2 We calculate $\hat{c}_n = n^{-1}\sum_{t=1}^{n} \ln \hat{\varepsilon}_t^2(\hat{\theta}_n^{QML})$ and minimize the following objective function

$$\tilde{Q}_n(\theta, \hat{c}_n) = \tfrac{1}{2n}\sum\nolimits_{t=1}^{n} \tilde{\ell}_t(\theta, \hat{c}_n) = \tfrac{1}{2n}\sum\nolimits_{t=1}^{n} (\ln y_t^2 - \hat{c}_n - \ln h_t(\theta))^2$$

and

$$\hat{\theta}_n^{LS} = \arg\min_{\theta\in\Theta} \tilde{Q}_n(\theta, \hat{c}_n).$$

For ease of reference, throughout the text, we will denote this as the LSE_Q estimator. Again it is immediate to recognize that the LSE_Q admits a simple and appealing two stage QML intepretation. Within this setting, as pointed out by a Referee, the proposed two step estimation procedure, aimed at improving efficiency, resembles the one recently proposed by Fan et al. (2014). They consider a parametric family of quasi likelihood functions of the form $\frac{1}{\eta} f(\frac{\cdot}{\eta})$ where $\eta > 0$ is used to adjust the scale of the likelihood function. In the first step the GQMLE is used to calculate $\{\widehat{\varepsilon}_t^2\}$, the model standardized residuals and the "bias correction factor" η_f is estimated by

$$\hat{\eta}_f = \arg\max_{\eta} \frac{1}{n}\sum\nolimits_{t=1}^{n} \left(\ln \eta + \log f\left(\frac{\tilde{\varepsilon}_t}{\eta}\right)\right)$$

In the second step, a non-Gaussian quasi likelihood function, obtained with a plug in of $\hat{\eta}_f$, is maximized in order to obtain

$$\widehat{\theta}_n = \arg\max \frac{1}{n}\sum\nolimits_{t=1}^{n} \left(-\ln(\hat{\eta}_f h_t) + \ln f\left(\frac{y_t}{\hat{\eta}_f h_t}\right)\right)$$

where $f(\cdot)$ is related to a family of likelihoods defined in Fan el al. (2014) which include for example the generalized Gaussian or the Student t density function. $\widehat{\theta}_n$ is called the NGQMLE estimator.

A comparative analysis of both estimation approaches, NGQMLE and LSE_Q, allows to identify the main differences and similarities between the two estimators. In Fan et al. (2014) a quasi-likelihood function is defined considering a potentially Non-Gaussian parametric family of densities for ε_t discriminated by the scale parameter η. Differently, in the LSE_Q, the quasi likelihood function used for estimation is determined through the specification of a Gaussian density for $\ln varepsilon_t^2$, whose location parameter c_0 indirectly brings information on the distribution of ε_t. So, in our approach the location parameter of the density of the log-transformed squared errors c_0 plays the same role that is played by the scale parameter of the error density η in the NGQMLE. Also both c_0 and η need to be pre-estimated in order to identify the quasi-likelihood function maximized in the final stage of the estimation procedure.

This suggests that a generalization of the LSE_Q approach could be potentially achieved through replacement of the Gaussian quasi-likelihood for $\ln \varepsilon_t^2$ with a more general parametric family of distributions. It is important to note that such a generalization, although of great interest, goes beyond the scope of this work which is focused on the estimation of GARCH(1,1) models through least squares. Replacing the Gaussian quasi-log-likelihood with alternative specifications would indeed potentially imply moving the focus from least squares to alternative loss functions. So it has been currently left for further investigation in our future research work. Similarly, regarding the possibility of performing an analytical efficiency comparison of our LSE-Q estimator with Fan et al (2014)'s NGQMLE, for different distributions, sample sizes and volatility parameterizations, we find that this issue is far from being trivial and well beyond the scope of this paper.

4. ASYMPTOTIC PROPERTIES

In this section we derive the asymptotic properties of the basic LSE and its variants as illustrated in the previous sections. In particular, for each of them, we show consistency and asymptotic normality.

4.1. Basic LSE

To show the strong consistency of the LSE introduced in 2, the following assumptions will be made.

Assumptions

(A1) $\Theta \equiv \{\theta : \ 0 < \underline{\omega} \leq \omega \leq \bar{\omega}, \ 0 < \underline{\alpha} \leq \alpha \leq \bar{\alpha}, \ 0 \leq \underline{\beta} \leq \beta \leq \bar{\beta} < 1\}$, where $\theta_0 \in \Theta$.

(A2) $\gamma = E \ln(\alpha_0 \varepsilon_t^2 + \beta_0) < 0$

(A3)$\mathrm{E}|\varepsilon_t|^{2s} < \infty$ for some $s > 0$ and ε_t^2 is a non-degenerate iid random variable.

(A4) $\lim_{r \to 0} r^{-(1+\delta)} \Pr(\varepsilon_t^2 \leq r) < \infty$ for some $\delta > 0$.

Remark 4.1. *The first assumption allows for the possibility that the process is a pure ARCH process. Nelson (1990) showed that Assumption A2 is sufficient and necessary for strict stationarity of (2.1) and (3.14). Note that, by Jensen's inequality, Assumption A2 holds if $\alpha_0 + \beta_0 \leq 1$ and $\mathrm{E}(\varepsilon_t^2) = 1$. But the condition does not require that $\alpha_0 + \beta_0 \leq 1$ so that while strict stationarity still holds if $\gamma < 0$ and $(\alpha_0 + \beta_0) \geq 1$, weak stationarity fails due to infinite second moment.*

However, this conclusion does not necessarily hold if $E(\varepsilon_t) = \infty$. Nelson (1990) shows that when ε_t is standard Cauchy, $\gamma = 2E \ln(\beta_0^{0.5} + \alpha_0^{0.5})$, so that the set of parameter values which allows for strict stationarity is smaller than the set $\alpha_0 + \beta_0 < 1$. Assumption A3 is a mild moment condition which allows for heavy tailed errors and is needed to uniquely identify the model parameters. It resembles the one in Francq and Zakoian (2004). However, we do not assume that the squared standardized errors have a unit mean, since it is restrictive under our setting. When c_0 is unknown, an alternative identification restriction is discussed later on. Assumption A4 implies that the distribution of the error term is not concentrated around zero, and one sufficient condition is that the density of ε_t is bounded. A similar condition also appears in Berkes et al.(2003).

Remark 4.2. *The method underlying the proofs basically consists of two main stages. In the first stage, it is assumed that the process is initiated from its stationary distribution and we establish the finiteness of various moments of the first and second derivative of the objective function. This part is justified by the second stage in which we show that the choice of the initial values does not matter for the asymptotic properties of the estimator. The proofs appear in*

Appendix A. Our first result is given as follows.

Theorem 4.1. *Under Assumptions A1-A4,* $\hat{\theta}_n \underset{a.s.}{\rightarrow} \theta_0$

The next theorems establish the asymptotic normality of θ_n. For GQMLE the next result is obtained under the assumption that $\mathrm{E}(\varepsilon_t^4) < \infty$. For the LSE, we consider the additional assumption:

(A5) $\theta_0 \in \Theta^0$, where Θ^0 denotes the interior of Θ.

Remark 4.3. *Assumption A5 is needed to establish the asymptotic normality, otherwise when the parameters are on the boundary other methods should be used. For example, under the null hypothesis that $\alpha = 0$, the conditional volatility process is degenerate which implies that β is unidentifiable and the null value of α is on the boundary, so its distribution cannot be normal. Andrews (2001) and Francq and Zakoian (2007) study in detail the distribution of the QMLE in that case. This issue is beyond the scope of this paper.*

We can now derive the LSE asymptotic distribution.

Theorem 4.2. *Under Assumptions A1-A5,* $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Omega)$, *where* $\Omega = \kappa J^{-1}$, $J = \mathrm{E}\left(J_t J_t^{'}\right)$, $J_t = \frac{1}{h_{0t}} \frac{\partial h_{0t}}{\partial \theta}$ *and* $\kappa = \mathrm{E}(\eta_t^2)$.

Remark 4.4. *Let $\tilde{J}_t$ and $\tilde{\eta}_t^2$ be the sample counterparts of J_t and η_t^2 where $\hat{\theta}_n$ is used and the variance is conditional on some initial fixed value. Under Lemma 7, it is straightforward to show that $\hat{\Omega}_n = \hat{\kappa}_n \hat{J}_n^{-1}$ is a strongly consistent estimate of Ω, where $\hat{\kappa}_n = \frac{1}{n} \sum_{t=1}^{n} \tilde{\eta}_t$ and $\hat{J}_n = \frac{1}{n} \sum_{t=1}^{n} \tilde{J}_t \tilde{J}_t^{'}$.*

Remark 4.5. *An important use of the asymptotic normality shown in Theorem 4.2 is to construct a Wald statistic to test the null hypothesis,*

$$H_0 : R\theta_0 = r$$

where R is a given $k \times 3$ matrix and r is a given $k \times 1$ vector. This test statistic may be defined as

$$W_n = \left(R\hat{\theta}_n - r\right)' \left(R\hat{\Omega}_n R'\right)^{-1} \left(R\hat{\theta}_n - r\right) \tag{4.18}$$

and we reject H_0 for large values of W_n. The following theorem gives the limiting distribution of W_n under the null hypothesis.

Theorem 4.3. *Under Assumptions A1-A5,* $W_n \xrightarrow{d} \chi_k^2$,

Remark 4.6. *Other scale measures can be used instead of our objective function. Thus, instead of using the LSE one may use the L_q estimator in which the scale measure is based on the $q-th$ absolute moment ($q \geq 1$) of the fitted residuals. For example, for $q = 1$ the least absolute deviations estimator was proposed by Peng and Yao (2003). Another more general approach was taken by Mukherjee (2008) who proposed a general M-estimator which includes the Gaussian and the Exponential QMLE as special cases. However, in order to obtain the asymptotic normality* *for these estimators, it is required that the observed process is covariance stationary; this assumption may be too restrictive under our settings.*

Remark 4.7. *Our estimator can be treated as an alternative to the common GQMLE in cases where the error distribution does not have finite fourth moment. For example, we can consider the Cauchy distribution or the Student t distribution with ≤ 4 degrees of freedom.*

4.2. 2S-LSE under identification restrictions

This section derives the asymptotic properties of the two stage LSE discussed in section 3.1. Let $\lambda = (\theta', m)' \in \Theta \times \Re \equiv \Lambda$ and denote by $\hat{\lambda}_n = (\hat{\theta}'_n, \hat{m}_n)$ the estimates obtained in each step of the estimation procedure described above. Next, we establish that $\hat{\lambda}_n$ is strongly consistent to its true parameter value, $\lambda_0 = (\theta'_0, med(\eta_t))' \in \Lambda$ and asymptotically normal and since $\bar{\theta}_0 = A_0\theta_0$, similar results can be obtained for $\hat{\bar{\theta}}_n$, the two-step estimator. The proofs appear in Appendix B.

Denote by $f_{\ln\varepsilon_t^2}$ and $F_{\ln\varepsilon_t^2}$ the density function and the cumulative distribution function of $\ln\varepsilon_t^2$, respectively. In order to show that $\hat{m}_n \to_{a.s.} med(\eta_t)$, we assume that $f_{\ln\varepsilon_t^2}$ satisfies the following "local identification" condition.

(A6) $med(\varepsilon_t^2) = 1$ and $f_{\ln\varepsilon_t^2}$ is continuous and positive in some neighborhood of the origin.

Theorem 4.4. *Under Assumptions A1-A4 and A6 (i)*$\hat{\lambda}_n \to_{a.s.} \lambda_0$ *(ii)*$\hat{\bar{\theta}}_n \overset{a.s.}{\to} \bar{\theta}_0$.

The first order condition for minimization of (3.17) is

$$n^{-1}\sum_{t=1}^{n} \frac{\partial(\tilde{l}_t(med(\eta_t), \theta_0))}{\partial m} = -n^{-1}\sum_{t=1}^{n} sgn(\ln\tilde{\varepsilon}_t^2(\hat{\theta}_n) - \hat{m}_n) = 0 \qquad (4.19)$$

where $sgn(z) = 1 - 2I(z < 0)$ and $I(.)$ is the indicator function. Note that (4.19) is not differentiable due to the appearance of an indicator function. Thus the standard Taylor expansion is not applicable and the empirical process theory (see e.g. Andrews, 1994) is used to establish the asymptotic normality of $\hat{m}_n$ in the second stage of the estimation. Therefore, the following assumptions impose additional restrictions on the distribution of $\ln\varepsilon_t^2$, namely,

(A7) $\sup_{x\in\Re} f_{\ln\varepsilon_t^2}(x) = \bar{f} < \infty$.

The two step estimator asymptotic distribution is given in the following theorem.

Theorem 4.5. *Under Assumptions A1-A4 and A6-A7,*

(i) $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \overset{d}{\to} N(0, \Psi V \Psi')$, *(ii)* $\sqrt{n}(\hat{\bar{\theta}}_n - \bar{\theta}_0) \overset{d}{\to} N(0, \Xi V \Xi')$
where

$$\Psi = \begin{pmatrix} I_3 & 0 \\ \bar{J}' & f^{-1}_{\ln\varepsilon_t^2}(0) \end{pmatrix}, \quad \Xi = e^{med(\eta_t)} \times \begin{pmatrix} 1 & 0 & 0 & \omega_0 \\ 0 & 1 & 0 & \alpha_0 \\ 0 & 0 & e^{-med\eta_t} & 0 \end{pmatrix} \times \Psi \quad ,$$

$$V = \begin{pmatrix} \kappa J^{-1} & \bar{\kappa}\bar{J}'J^{-1} \\ \bar{\kappa}J^{-1}\bar{J} & 1 \end{pmatrix}, \quad \bar{J} = E(J_t), \ \bar{\kappa} = 2E(\eta_t I(\ln\varepsilon_t^2 \leq 0))$$

and I_3 is the identity matrix of size 3×3 *and κ and J are defined above.*

In general the elements Ξ and V can be consistently estimated by using their sample counterparts. As for the unknown density function,$f_{\ln \varepsilon_t^2}(0)$, we employ the following nonparametric estimator,

$$\hat{f}_{\ln \varepsilon_t^2}(0) = \frac{1}{na_n} \sum\nolimits_{t=1}^{n} H\left(-a_n^{-1}(\tilde{\eta}_t - \hat{m}_n)\right) \tag{4.20}$$

where a_n is a bandwidth and $H(\cdot)$ is a density kernel on $\Re$. In particular, we can use the result above to construct a Wald statistic to test the null hypothesis $R\bar{\theta}_0 = r$. A Wald test statistic is defined as

$$\bar{W}_n = \left(R\hat{\bar{\theta}}_n - r\right)' \left((R\hat{\Xi}_n)\hat{V}_n(R\hat{\Xi}_n)'\right)^{-1} \left(R\hat{\bar{\theta}}_n - r\right)$$

We reject the null hypothesis for large values of W_n. In the above expression,

$$\hat{V}_n = \begin{pmatrix} \hat{\kappa}_n \hat{J}_n^{-1} & \bar{\kappa}_n \hat{\bar{J}}_n' \\ \hat{\bar{\kappa}}_n \hat{\bar{J}}_n & 1 \end{pmatrix}, \hat{\Xi}_n = e^{\hat{m}_n} \times \begin{pmatrix} 1 & 0 & 0 & \hat{\omega}_n \\ 0 & 1 & 0 & \hat{\alpha}_n \\ 0 & 0 & e^{-\hat{m}_n} & 0 \end{pmatrix} \times \hat{\Psi}_n$$

where $\hat{\Psi}_n = \begin{pmatrix} I_3 & 0 \\ \hat{\bar{J}}_n' & \hat{f}^{-1}_{\ln \varepsilon_t^2}(0) \end{pmatrix}$, $\hat{\bar{\kappa}}_n = \frac{2}{n} \sum_{t=1}^{n} \tilde{\eta}_t I(\tilde{\eta}_t - \hat{m}_n \leq 0))$, $\hat{\bar{J}}_n = \frac{1}{n} \sum_{t=1}^{n} \hat{J}_t$, $\hat{\kappa}_n$ and $\hat{J}_n$ are defined in the previous section. The next Theorem gives the limiting distribution of $\bar{W}_n$ under the null hypothesis.

Theorem 4.6. *Suppose the conditions of Theorem 4.5 hold. If the kernel function $H(\cdot)$ and the bandwidth a_n satisfy the following assumptions:*

(i) $H(\cdot)$ is a Lipschitz function which satisfies that $\int H(u)du = 1$.

(ii) $a_n \to 0$ and $a_n^4 n \to \infty$ as $n \to \infty$.

Then $\bar{W}_n \xrightarrow{d} \chi_k^2$.

4.3. 2S-LSE with pre-estimated c_0

In this section we show consistency and asymptotic normality of the estimator proposed in section 3.2. To this end, we introduce the identification restriction

(A6'): $E(\varepsilon_t^2) = 1$.

Hence consistency is proved by the following theorem

Theorem 4.7. *Under Assumptions A1-A4 and A6' $\hat{\theta}_n^{LS} \xrightarrow{a.s.} \theta_0$.*

Similarly, in order to derive the asymptotic distribution of the proposed estimator, we need to introduce the additional assumption of finite fourth moment for the standardized errors.

(A7'): $E(\varepsilon_t^4) < \infty$

The desired asymptotic normality result is then given in the following theorem

Theorem 4.8. *Under Assumptions A1-A5 and A6' A7',* $\sqrt{n}(\hat{\theta}_n^{LS} - \theta_0) \xrightarrow{d} N(0, \bar{\Xi}\bar{V}\bar{\Xi}')$ *where*

$$\bar{V} = \begin{pmatrix} \kappa & \varsigma \bar{J}' J^{-1} & \kappa \bar{J}' \\ \varsigma J^{-1} \bar{J} & \mu J^{-1} & \varsigma I_3 \\ \kappa \bar{J} & \varsigma I_3 & \kappa J \end{pmatrix}, \quad \bar{\Xi} = J^{-1} \begin{pmatrix} \bar{J} & I_3 \end{pmatrix} \begin{pmatrix} 1 & -\bar{J}' & 0 \\ 0 & 0 & I_3 \end{pmatrix} \quad ,$$

$$\varsigma = E(\eta_t \varepsilon_t^2), \;\; \mu = E(\varepsilon_t^4) - 1$$

In general the elements $\bar{\Xi}$ and $\bar{V}$ can be consistently estimated by using their sample counterparts. In particular, we can use the result above to construct a Wald statistic to test the null hypothesis $R\theta_0 = r$. The proposed test statistic is defined as

$$\bar{\bar{W}}_n = \left(R\hat{\bar{\theta}}_n - r\right)' \left((R\hat{\bar{\Xi}}_n)\hat{V}_n(R\hat{\bar{\Xi}}_n)'\right)^{-1} \left(R\hat{\bar{\theta}}_n - r\right)$$

We reject the null hypothesis for large values of $\bar{W}_n$. In the above expression,

$$\hat{\bar{V}}_n = \begin{pmatrix} \hat{\kappa}_n & \hat{\varsigma}_n \hat{\bar{J}}_n' \hat{J}_n^{-1} & \hat{\kappa}_n \hat{\bar{J}}_n' \\ \hat{\varsigma}_n \hat{J}_n^{-1} \hat{\bar{J}}_n & \hat{\mu}_n \hat{J}_n^{-1} & \hat{\varsigma}_n I_3 \\ \hat{\kappa}_n \hat{\bar{J}}_n & \hat{\varsigma}_n I_3 & \hat{\kappa}_n \hat{J}_n \end{pmatrix}, \quad \bar{\Xi} = \hat{J}_n^{-1} \begin{pmatrix} \hat{\bar{J}}_n & I_3 \end{pmatrix} \begin{pmatrix} 1 & -\hat{\bar{J}}_n' & 0 \\ 0 & 0 & I_3 \end{pmatrix}$$

where $\hat{c}_n = n^{-1}\sum_{t=1}^n \ln \hat{\varepsilon}_t^2$, $\hat{\varsigma}_n = n^{-1}\sum_{t=1}^n \hat{\eta}_t \hat{\varepsilon}_t^2$, $\hat{\mu}_n = n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t^4 - 1$, $\hat{\kappa}_n$ and $\hat{\eta}_t$, $\hat{J}_n$ are defined in the previous section. Note that $\hat{\varepsilon}_t$, $\hat{\eta}_t$ are residuals from the QML and LS estimation procedure, respectively. The next Theorem gives the limiting distribution of $\bar{\bar{W}}_n$ under the null hypothesis.

Theorem 4.9. *Under Assumptions A1-A4 and A6'-A7',* $\bar{\bar{W}}_n \xrightarrow{d} \chi_k^2$.

5. FINITE SAMPLE PROPERTIES

In order to compare the finite sample efficiency of the proposed LS estimators with that of their competitors, we have performed an extensive simulation study designed to reproduce a wide range of settings typically encountered in the analysis of financial time series. The set of estimators considered in the implementation of our simulation study includes the two stage LSEs introduced in section 3, LSE_0 and LSE_Q, the GQMLE and the Non-Gaussian QMLE (NGQMLE) proposed by Fan et al. (2014). Following these authors we have implemented the NGQMLE using a Student's t quasi-likelihood with 7 and 4 degrees of freedom, respectively. Fan et al. (2014) suggest the use of a t_4 quasi-likelihood function in order to account for situations characterized by heavy tailed error distributions where the hypothesis of infinite 4th order moment of the standardized errors is strongly supported by the data.

As Data Generating Processes (DGPs) we have considered different GARCH(1,1) structures characterized by

- Varying levels of persistence: integrated (I), high (H), medium (M) and low (L). The dynamic volatility parameters of the DGPs considered in our simulation study have been summarized in Table 3. In all cases, excluding the Integrated one, the variance of the DGP has been set equal to one.
- Different innovation distributions: standard Normal, t_5, t_3 and $t_{2.1}$. All the distributions considered have been rescaled to enforce the unit median identification constraint for the squared errors ε_t^2.

The NGQMLE has not been proven to be consistent under the assumption of an infinite second moment. So, we considered a Student's t distribution with 2.1 degrees of freedom within our simulation study in order to assess the efficiency of the NGQMLE for DGPs close to the theoretical bound of an infinite variance process.

From each of the above described DGPs we have generated 1000 pseudo-random time series of length $T \in \{500, 1000, 2000, 5000\}$ after discarding the first 500 observations taken as burn-in period. The efficiency of the estimators considered in our study has been evaluated in terms of the simulated Root Mean Squared Error (RMSE) computed as

$$RMSE_\theta(\hat{\theta}) = \sqrt{\sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2}$$

where θ and $\hat{\theta}_i$ are a generic unknown parameter to be estimated and its estimate obtained from the i-th simulated time series, respectively. In the case of Normal errors (Table 4) the GQMLE, as expected, is overall dominating the other estimators followed by the NGQMLE that, in most cases, outperforms the two LSEs considered. The only exceptions to this general framework arise in the Integrated case for $T \geq 1000$, where the LSE_Q turns out to be the most efficient in the estimation of the intercept ω_0. Also, for $T = 500$, the LSE_Q outperforms the NGQMLEs in the estimation of β_0 while, for $T = 1000$, the LSE_Q performs better than the NGQMLEs for all the parameters.

The results obtained in the t_5 (Table 7) and t_3 (Table 6) cases are very similar: the two LS estimators are outperforming GQMLE and NGQMLE with the LSE_0 slightly prevailing over the LSE_Q estimator. The NGQMLE is in general performing better than the GQMLE in the estimation of α_0 and β_0 but it is in most cases outperformed by the GQMLE in the estimation of the intercept ω_0. Also, it is worth noticing that, for both t_3 and t_5, the GQMLE is in some cases characterized by a non-regular behaviour since, in some cases, the simulated RMSE is increasing with the sample size.

Finally, in the $t_{2.1}$ case (Table 7), the LSE is still dominating its competitors with the LSE_0 estimator giving the best performance. The NGQMLE, still being less efficient than the LSE, remains consistent differently from the GQMLE for which the simulation results provide strong evidence of an inconsistent behaviour for all the parameters.

In conclusion, in line with previous findings our results confirm that, in the presence of a heavy tailed conditional distribution of returns, the NGQMLE allows substantial efficiency gains over the standard GQMLE. Nevertheless, the NGQMLE is in turn outperformed by both the LSEs considered with the LSE_0 slightly overperforming LSE_Q. Also, given our findings, the ranking of the estimators, appears to be insensitive to the persistence of the volatility process.

6. VAR ESTIMATION WITH UNKNOWN C_0

In section 2 it has been argued that, when using a one-stage LSE, the choice of the tuning constant c_0 is not expected to affect the quality of VaR forecasts generated from the fitted model. In order to provide empirical evidence supporting this statement, in this section we present the results of an empirical application to VaR forecasting. Namely, we consider returns on three different real series: a stock market index, the S&P500, an equity, the JP Morgan, and an exchange rate series, the USD/EUR (see table 7). In addition, to test the ability of our LS-GARCH estimator in tracking VaR in extreme settings, we consider a simulated series of 10000 observations generated from a GARCH(1,1) with Student's $t_{2.1}$ errors and parameters given by $\omega = 0.1$, $\alpha = 0.1$ and $\beta = 0.8$. For all the series we adopt a rolling window forecasting scheme for generating 1-period ahead prediction of VaR at three different coverage levels p=(0.05, 0.025, 0.01). The length of the estimation window has been set to 1000 observations and the model parameters are estimated every time a new observation becomes available. VaR predictions are generated according to a two-step procedure. First, a GARCH(1,1) model is fitted using the proposed LS-GARCH estimator with c_0 arbitrarily set equal to 0. Second, we generate the in-sample standardized residuals $\hat{\varepsilon}_t^*$ from the fitted model and compute the order p empirical quantile of the $\hat{\varepsilon}_t^*$ (t=1,. . . ,n) series. The predicted 1-step ahead VaR at time $t+1$ and coverage level p is then given by

$$V\hat{a}R_{t+1,p} = \sqrt{\hat{h}_{t+1}^*}\hat{\varepsilon}_{(p)}^*$$

where $\hat{\varepsilon}_{(p)}^*$ denotes the order p empirical quantile of $\hat{\varepsilon}_t^*$, which is the estimated standardized residuals obtained from fitting a LSE with $c_0 = 0$, and $\hat{h}_{t+1}^*$ is the (biased) estimate of the conditional variance at time $(t+1)$ obtained from fitting a GARCH(1,1) model. In order to assess the accuracy of VaR predictions we have performed two different tests: the likelihood ratio test for correct conditional coverage (LR_{cc}), proposed by Christoffersen (1998), and the Dynamic Quantile test (DQ), proposed by Engle and Manganelli (2004). The results have been reported in table 7. In all cases, the null hypothesis of an accurately predicted VaR series is accepted at any reasonable significance level.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we suggest using LSE for the estimation of a GARCH (1,1) model. The proposed estimator is based on the log transformation of the squared data. In the paper we consider three different variants of the proposed estimation method establishing the consistency and asymptotic normality for all of them. Our results have been obtained under mild regularity conditions that allow for heavy tailed error distributions that can be of particular interest in financial applications.

The basic LSE estimator has the limitation of being able to estimate volatility only up to an unknown scale factor, unless we fix the value of a tuning constant whose knowledge implicitly requires knowledge of the shape of the underlying conditional distribution of returns. However, in section 6, by means of an empirical application, we show that ignoring the tuning constant does not prevent the generation of accurate VaR predictions in heavy tailed settings. These results point out that the simple basic LSE can be an effective and easy-to-use tool for practitioners primarily interested in using the estimated GARCH model for predicting VaR or related risk measures, such as Expected Shortfall.

At the same time, in order to deal with cases in which full knowledge of the volatility process is required, this limitation can be easily overcome by considering two different two-stage variants of the basic LSE. Of these one is obtained at the price of imposing some suitable identification restriction on the innovations process. Namely, in the paper we focus on the case in which the median of squared innovations is assumed to be equal to one. However, different identification restrictions could be considered without altering the theoretical framework on which our results rely. This issue is currently left for future research. Alternatively, if one is willing to work under more restrictive moment assumptions, a different approach is to pre-estimate the tuning constant c_0 by using some other estimator, such as the standard QMLE or the NGQMLE.

The finite sample properties of the proposed 2-stage LSEs have been investigated via a simulation study, which shows that, in the presence of extreme non-normality, the estimator based on additional identification restrictions, LSE_0, can allow for substantial efficiency gains with respect to existing alternatives, such as the GQMLE and the NGQMLE. The simulation results also show that in most settings, the efficiency of the LSE based on a pre-estimated tuning constant, LSE_Q, can be very close to that of the LSE_0 estimator.

Finally, it should be noted that when working with heavy tailed returns series, such as ultra high frequency returns, an important issue is the robustness of the estimation procedure, since these data are typically characterized by a high fraction of very small returns, which, after the log transformation, can produce large negative values. Therefore, our estimator, which is based on the L_2 scale measure, may not be optimal in the presence of outlying observations. In order to overcome this problem, an estimator that employs a more robust scale measure such as the S-estimator can be used (see e.g. Sakata and White, 2001). In addition, our results can be extended to the GARCH (p,q) case as well as to other GARCH "type" models. The investigation of these issues is left for future work.

ACKNOWLEDGEMENTS

The authors thank Christian M. Hafner, Luc Bauwens and participants to the CFE 14 conference in Pisa for helpful comments. The authors would also like to thank two anonymous referees and the Associate Editor Andrew Patton for their valuable comments and suggestions that helped to significantly improve the quality of this paper. Giuseppe Storti gratefully acknowledges funding from the Italian Ministry of Education, University and Research (MIUR) through PRIN project "Forecasting economic and financial time series: understanding the complexity and modelling structural change" (code 2010J3LZEN).

APPENDIX

The Appendix contains the proofs of the asymptotic results presented in section 4. For reader's convenience it has been structured in three different parts denominated A, B and C. Namely, the proofs of theorems and lemmata relative to the one-stage LSE, LSE_0 and LSE_Q estimators have been presented in Appendix A, B and C, respectively.

Throughout the Appendix, we have adopted the following notational conventions.

- K will denote a generic positive number that may vary in different uses.
- In order to simplify the notation we define the new variables $z_t = \ln y_t^2 - c_0$ and $\xi_t = \ln \varepsilon_t^2$ and, in Appendix B, we introduce the shorthand notation $m_0 = med(\eta_t)$.
- Let $\tilde{h}_{0t} = \tilde{h}_t(\theta_0)$ and

$$\dot{h}_{it}(\theta) = \frac{\partial h_t(\theta)}{\partial \theta_i}, \ \ddot{h}_{ijt}(\theta) = \frac{\partial h_t(\theta)}{\partial \theta_i \partial \theta_j}, \ \dot{\tilde{h}}_{it}(\theta) = \frac{\partial \tilde{h}_t(\theta)}{\partial \theta_i}, \ \ddot{\tilde{h}}_{ijt}(\theta) = \frac{\partial \tilde{h}_t(\theta)}{\partial \theta_i \partial \theta_j}$$

- Let $\nabla \ell_t(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta}$, $\nabla \ell_{it}(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta_i}$ and $\nabla^2 \ell_t(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta \partial \theta'}$, $\nabla^2 \ell_{ijt}(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta_i \partial \theta_j}$ denote the first and second derivatives of $\ell_t(\theta)$ (and their elements), respectively.

Appendix A: proofs of theorems 4.1-4.3 and lemmata 1-7

Proof of Theorem 4.1: We use similar arguments as in Theorem 5.3.1 of Straumann (2005, p.101) showing strong consistency by contradiction. Suppose that $\hat{\theta}_n \not\to \theta_0$ a.s. so for some arbitrary $\gamma > 0$, the set $F = \{\omega \in \Omega | \lim_{n\to\infty} |\hat{\theta}_n(\omega) - \theta_0| \geq \gamma\}$ has a positive probability. Since the set $\mathrm{N} = \Theta \cap \{\theta : |\hat{\theta}_n(\omega) - \theta_0| \geq \gamma\}$ is compact, for each $\omega \in F$ we can find in the set N, a convergent subsequence $\hat{\theta}_{n(i)}(\omega) \to \tilde{\theta} \in \mathrm{N}$ as $i \to \infty$. By definition of the LSE

$$\begin{aligned} \liminf_{i\to\infty} \tfrac{1}{n(i)} \textstyle\sum_{t=1}^{n(i)} \tilde{\ell}_t(\theta_0) &\geq \liminf_{i\to\infty} \inf_{\theta\in\mathrm{N}} \tfrac{1}{n(i)} \textstyle\sum_{t=1}^{n(i)} \tilde{\ell}_t(\theta) \\ &= \liminf_{n\to\infty} \tfrac{1}{n(i)} \textstyle\sum_{t=1}^{n(i)} \tilde{\ell}_t F(\hat{\theta}_{n(i)}) \end{aligned} \tag{7.21}$$

From Lemma 5,

$$\liminf_{i\to\infty} \tfrac{1}{n(i)} \textstyle\sum_{t=1}^{n(i)} \ell_t(\theta_0) \ \geq \ \liminf_{i\to\infty} \tfrac{1}{n(i)} \textstyle\sum_{t=1}^{n(i)} \ell_t(\hat{\theta}_{n(i)}) \tag{7.22}$$

The inequality above and Lemmata 4(ii)-(iii) imply that with positive probability $\mathrm{E}\ell_t(\theta_0) \geq \mathrm{E}\inf_{\theta\in\mathrm{N}} \ell_t(\theta)$. This result contradicts Lemma 4(i) which states that in the limit $Q_n(\theta)$ is uniquely minimized at θ_0. Since $\gamma > 0$ is arbitrary, the strong consistency follows.

Proof of Theorem 4.2: By Theorem 4.1, $\bar{\theta}_n \to \theta_0$ a.s. so for n sufficiently large $\bar{\theta}_n \in \Theta^0$ a.s. and the results of Lemmas 6-7 can be applied. Using a mean-value expansion of $\tilde{Q}_n(\hat{\theta}_n) = \sum_{t=1}^n \tilde{\ell}_t(\hat{\theta}_n)$ around θ_0, we have

$$\begin{aligned} 0 &= n^{-0.5} \sum_{t=1}^n \nabla \tilde{\ell}_t(\hat{\theta}_n) \\ &= n^{-0.5} \sum_{t=1}^n \nabla \tilde{\ell}_t(\theta_0) + \left(\tfrac{1}{n} \textstyle\sum_{t=1}^n \nabla^2 \tilde{\ell}_t(\bar{\theta}_n)\right) \sqrt{n}(\hat{\theta}_n - \theta_0) \\ &= n^{-0.5} \sum_{t=1}^n \nabla \tilde{\ell}_t(\theta_0) \\ &+ \left[\left(\tfrac{1}{n} \textstyle\sum_{t=1}^n \nabla^2 \tilde{\ell}_t(\bar{\theta}_n) - \tfrac{1}{n} \sum_{t=1}^n \nabla^2 \ell_t(\bar{\theta}_n)\right) + \left(\tfrac{1}{n} \sum_{t=1}^n \nabla^2 \ell_t(\bar{\theta}_n) - J\right) + J\right] \sqrt{n}(\hat{\theta}_n - \theta_0) \end{aligned} \tag{7.23}$$

where $\bar{\theta}_n$ lies on the chord between $\hat{\theta}_n$ and θ_0. Lemma 6 and the asymptotic equivalence lemma (e.g. see White, 1994, p.172) imply that $n^{-0.5} \sum_{t=1}^n \nabla \ell_t(\theta_0) \overset{d}{\to} N(0, \kappa J)$ where κJ is a positive definite matrix. Next, Lemmas 7(i)-(ii) imply that the first and second terms, inside the square brackets in (7.23), converge a.s. to zero. Hence, to complete the proof it suffices to solve (7.23) and apply Slutsky's theorem.

Proof of Theorem 4.3: Let $R_t = \sup_{\theta \in \Theta} \left| \dot{h}_t(\theta) h_t^{-1}(\theta) \right|$ and note that by the mean-value theorem

$$\begin{aligned} |\eta_t - \tilde{\eta}_t| = \left| \ln \tilde{h}_t(\hat{\theta}_n) - \ln h_t(\theta_0) \right| &= \left| \ln(\tilde{h}_t(\hat{\theta}_n) h_t^{-1}(\hat{\theta}_n)) + \ln(h_t(\hat{\theta}_n) h_{0t}^{-1}) \right| \\ &\leq K \sup_{\theta \in \Theta} \left| \tilde{h}_t(\theta) - h_t(\theta) \right| + K \left| \hat{\theta}_n - \theta_0 \right| R_t \end{aligned}$$

using the same arguments as in Lemma 5, we can show that the first term after the first inequality converges a.s. to zero. Lemma 3(i) implies that R_t is almost surely finite and hence Theorem 4.1 implies that the second term after the first inequality is $o_{a.s.}(1)$. Therefore,

$$\tilde{\eta}_t = \eta_t + o_{a.s.}(1) \tag{7.24}$$

and by the ergodic theorem $\hat{\kappa}_n \to_{a.s.} \kappa$. Further, by combining Theorem 4.1, Lemmas 1(i), 3(i) and the similar argument as in the proof of Lemma 7(ii) we can show that $\hat{J}_n \to_{a.s.} J$. Hence, the desired result follows directly by applying 4.2 and Slutsky's theorem.

Lemma 1: Under Assumptions A1-A4, for some $p \in (0, 1)$

i) (y_t^2, h_{0t}) are strictly stationary and ergodic and $\mathrm{E}\left(h_{0t}^p\right) < \infty,\ \mathrm{E}\left(|y_t^2|^p\right) < \infty$.

ii) $\inf_{\theta \in \Theta} \ell_t(\theta),\ \ell_t(\theta),\ \nabla \ell_{it}(\theta)$ and$\nabla^2 \ell_{ijt}(\theta)$ are strictly stationary and ergodic.

iii)$\mathrm{E}\left(\ln \varepsilon_t^2\right) < \infty$ and $\mathrm{E}\left(\eta_t^2\right) < \infty$.

Proof:

i) Under Assumption A2, the result follows directly from (2.1)-(3.14) and Theorem 4 of Nelson (1990).

ii) From (2.8)-(2.9) and Theorem 2.7 of Stinchcombe and White (1992), we have that $\inf_{\theta \in \Theta} \ell_t(\theta)$ is measurable functions of y_{t-j} for all $j \geq 0$, and thus are strictly stationary and ergodic (see Stout, 1974, Theorem 3.5.8). The same result follows for $\ell_t(\theta)$ and its derivatives by Lemma 2(ii) of Lee and Hansen (1994).

iii) Let $w = \varepsilon_t^2,\ F(x) = \Pr(w \leq x)$ and $f(x)$ be the density function. By integration by parts

$$\int_0^1 [\ln w]^2 f(w)dw = [\ln 1]^2 F(2.1) - 2\int_r^1 \frac{\ln w}{w} F(w)dw - 2\int_0^r \frac{\ln w}{w} F(w)dw$$

The first integral on the RHS is bounded for any $r > 0$. Hence, by Assumption A4, when $r > 0$ is small enough, there exists some $\delta > 0$ such that the second integral is bounded by $K\int_0^r w^\delta \ln w dw$. This integral is finite for any $\delta > 0$. For $w \geq 1$, by Assumption 3, we get

$$\int_1^{+\infty} [\ln w]^2 f(w)dw \ < \ \int_1^{+\infty} w^s f(w)dw \ \leq \ \mathrm{E}|\varepsilon_t^2|^s < \infty.$$

Similarly we can show that $c_0 = E(\ln \varepsilon_t^2) < \infty$ and, since $\eta_t^2 = (\ln w)^2 - c_0^2$—, the desired result follows.

Lemma 2: Under Assumptions A1-A4, for some $p \in (0,1)$

i) $\left\| \sup_{\theta\in\Theta} \left| h_t(\theta) - \tilde{h}_t(\theta) \right| \right\|_p = O(\bar{\beta}^t)$ and $\mathrm{E}|\sup_{\theta\in\Theta} \tilde{h}_t(\theta)|^p \ < \ \infty.$

ii) $\left\| \sup_{\theta\in\Theta^0} \left| \dot{h}_{it}(\theta) - \dot{\tilde{h}}_{it}(\theta) \right| \right\|_p = O(\bar{\beta}^t)$ for all i.

iii) $\left\| \sup_{\theta\in\Theta^0} \left| \ddot{h}_{ijt}(\theta) - \ddot{\tilde{h}}_{ijt}(\theta) \right| \right\|_p = O(\bar{\beta}^t)$ for all i, j.

Proof:

i) By iterating (2.8) and using the fact $\alpha_0 y_{t-1-i}^2 \leq h_{0t}$, we get

$$\begin{aligned} h_t(\theta) &= \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta) \\ &= \sum_{i=0}^{t-1} (\omega + \alpha y_{t-1-i}^2)\beta^i + \beta^t h_1(\theta) \\ &= \sum_{i=0}^{\infty} (\omega + \alpha y_{t-1-i}^2)\beta^i \\ &= \frac{\omega}{1-\beta} + \alpha \sum_{i=0}^{\infty} \beta^i y_{t-1-i}^2 \\ &\leq \frac{\bar{\omega}}{1-\underline{\beta}} + \frac{\bar{\alpha}}{\alpha_0} \sum_{i=0}^{\infty} \bar{\beta}^i h_{0t}. \end{aligned} \tag{7.25}$$

Hence, the c_r inequality ($(a+b)^q \leq a^q + b^q$ for all $a, b > 0,\ q \in [0,1]$) and Lemma 1(i) imply that for some $p \in (0,1)$,

$$\mathrm{E}|\sup_{\theta\in\Theta} h_t(\theta)|^p \ \leq \ K + K\mathrm{E}(h_{0t}^p) \ < \ \infty. \tag{7.26}$$

Now, by iterating (2.5) we obtain

$$\tilde{h}_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta \tilde{h}_{t-1}(\theta) = \sum_{i=0}^{t-1} (\omega + \alpha y_{t-1-i}^2)\beta^i + \beta^t \tilde{h}_1(\theta). \tag{7.27}$$

Hence

$$\tilde{h}_t(\theta) - h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta \tilde{h}_{t-1}(\theta) = \beta^t(\tilde{h}_1(\theta) - h_1(\theta)) \tag{7.28}$$

and

$$\mathrm{E}\sup_{\theta\in\Theta^0} \left| h_t(\theta) - \tilde{h}_t(\theta) \right|^p \ \leq \ \beta^{pt}(\bar{\omega}^p + \mathrm{E}\sup_{\theta\in\Theta^0} |h_1(\theta)|^p) \leq K\bar{\beta}^{pt}. \tag{7.29}$$

Further, by Lemma 1(i) and the c_r inequality

$$\mathrm{E}(\bar{\omega}+\bar{\alpha}y_{t-1-i}^2)^p<\infty \tag{7.30}$$

and

$$\mathrm{E}\left(\sup_{\theta\in\Theta}\left|\tilde{h}_t(\theta)\right|^p\right)\leq\sum\nolimits_{i=0}^{t-1}\mathrm{E}(\bar{\omega}+\bar{\alpha}y_{t-1-i}^2)^p\bar{\beta}^{ip}+\bar{\beta}^{pt}\bar{\omega}^p<\infty.$$

ii) We start by showing that for some $p\in(0,1)$ and all i,

$$\mathrm{E}\left(\sup_{\theta\in\Theta^0}\left|\dot{h}_{it}(\theta)\right|^p\right)<\infty. \tag{7.31}$$

By (7.25) and the fact that $y_{t-1-i}^2\leq\alpha_0^{-1}h_{0t}$,

$$\frac{\partial h_t(\theta)}{\partial\omega}\;\leq\;\frac{1}{1-\underline{\beta}} \tag{7.32}$$

$$\frac{\partial h_t(\theta)}{\partial\alpha}=\sum\nolimits_{i=0}^{\infty}\beta^i y_{t-1-i}^2\;\leq\;\frac{1}{\alpha}\left[\sum\nolimits_{i=0}^{\infty}\alpha\beta^i y_{t-1-i}^2\right]\;\leq\;\frac{1}{\underline{\alpha}}h_t(\theta) \tag{7.33}$$

$$\begin{aligned}\frac{\partial h_t(\theta)}{\partial\beta}&=\sum\nolimits_{i=1}^{\infty}i\beta^i(\omega+\alpha y_{t-1-i}^2) \qquad (7.34)\\ &\leq\;\sum\nolimits_{i=1}^{\infty}i\beta^i\left(\omega+\frac{\alpha}{\alpha_0}h_{0t}\right)\;\leq\;\bar{\omega}\sum\nolimits_{i=1}^{\infty}i\bar{\beta}^i+\frac{\bar{\alpha}}{\alpha_0}\sum\nolimits_{i=0}^{\infty}\bar{\beta}^i h_{0t}.\end{aligned}$$

The term in (7.32) is bounded and admits moments of any order. As for (7.33)-(7.34), the result follows directly from the c_r inequality and Lemma 1(i). In view of (7.28), almost surely,

$$\sup_{\theta\in\Theta^0}\left|\dot{h}_{it}(\theta)-\dot{\tilde{h}}_{it}(\theta)\right|\;\leq\;t\bar{\beta}^{(t-1)}(\bar{\omega}+\sup_{\theta\in\Theta^0}h_1(\theta))+\bar{\beta}^t\sup_{\theta\in\Theta^0}|\dot{h}_{i1}(\theta)|\;\leq\;K\bar{\beta}^t$$

the desired result follows by (7.26), (7.31) and the c_r inequality.

iii) From (7.32)-(7.34) and direct calculations we get,

$$\frac{\partial^2 h_t}{\partial\omega^2}=\frac{\partial^2 h_t}{\partial\alpha^2}=\frac{\partial^2 h_t}{\partial\omega\partial\alpha}=0,\qquad\frac{\partial^2 h_t}{\partial\omega\partial\beta}\frac{1}{\beta}\leq\sum\nolimits_{i=1}^{\infty}i\bar{\beta}^i \tag{7.35}$$

which are bounded and admit moments of any order. We also find

$$\frac{\partial^2 h_t}{\partial\alpha\partial\beta}\;\leq\;\alpha\sum\nolimits_{i=1}^{\infty}i\beta^i y_{t-1-i}^2\;\leq\;\frac{\bar{\alpha}}{\alpha_0}\sum\nolimits_{i=1}^{\infty}i\beta^i h_{0t} \tag{7.36}$$

$$\frac{\partial^2 h_t}{\partial\beta^2}=\frac{1}{\beta}\sum\nolimits_{i=2}^{\infty}i(i-1)(\omega+\alpha y_{t-1-i}^2)\beta^i. \tag{7.37}$$

So, similar to Lemma 2(ii) we can show that for some $0<p<1$,

$$\mathrm{E}\left(\sup_{\theta\in\Theta^0}\left|\ddot{h}_{ijt}(\theta)\right|^p\right)<\infty \tag{7.38}$$

for all i, j. In view of (7.28), almost surely,

$$\begin{aligned}\sup_{\theta\in\Theta^0}\left|\ddot{h}_{ijt}(\theta)-\ddot{\tilde{h}}_{ijt}(\theta)\right| &\leq t(t-1)\bar{\beta}^{(t-2)}[\bar{\omega}+\sup_{\theta\in\Theta^0}h_1(\theta)]\\ &+t\bar{\beta}^{(t-1)}\left(\sup_{\theta\in\Theta^0}\left|\dot{h}_{j1}(\theta)\right|+\sup_{\theta\in\Theta^0}\left|\dot{h}_{i1}(\theta)\right|\right)\\ &+\bar{\beta}^t\sup_{\theta\in\Theta^0}|\ddot{h}_{ij1}(\theta)|\end{aligned}$$

and by (7.26), (7.31), (7.38) and the c_r inequality the desired result follows.

Lemma 3[4]: Under Assumptions A1-A4, for all $r \geq 1$

i) $\left\|\sup_{\theta\in\Theta^0} h_t^{-1}(\theta)\dot{h}_{it}(\theta)\right\|_r < \infty$ for all i.

ii) $\left\|\sup_{\theta\in\Theta^0} h_t^{-1}(\theta)\ddot{h}_{ijt}(\theta)\right\|_r < \infty$ for all i, j.

iii) $\left\|\sup_{\theta\in\Theta^0} \tilde{h}_t^{-1}(\theta)\dot{\tilde{h}}_{it}(\theta)\right\|_r < \infty$ for all i, and $\left\|\sup_{\theta\in\Theta^0} \tilde{h}_t^{-1}(\theta)\ddot{\tilde{h}}_{ijt}(\theta)\right\|_r < \infty$ for all i, j.

iv) $E|\sup_{\theta\in\Theta}\ln h_t(\theta)| < \infty$ and $E|\sup_{\theta\in\Theta}\ln \tilde{h}_t(\theta)| < \infty$.

Proof:

i) Eq. (7.32) and (7.33) imply that the derivatives of h_t with respect to ω and α (divided by h_t) are bounded and hence admit moments of any order. However, this is not true for the derivative with respect to β. From (7.25) we get $h_t(\theta) \geq \omega + (\omega + \alpha y_{t-1-i}^2)\beta^i$ for all $i \geq 1$. Using the fact that $x/(1+x) < x^{p/r}$ for all $x \geq 0$ and any $p \in (0,1)$,$r \geq 1$ (this idea of exploiting this inequality is due to Boussama, 2000), we get

$$\begin{aligned}\frac{\partial h_t}{\partial\beta}\frac{1}{h_t} &\leq \frac{1}{\beta}\sum\nolimits_{i=1}^{\infty} i\frac{(\omega+\alpha y_{t-1-i}^2)\beta^i}{\omega+(\omega+\alpha y_{t-1-i}^2)\beta^i}\\ &\leq \frac{1}{\beta}\sum\nolimits_{i=1}^{\infty} i\left[\frac{(\omega+\alpha y_{t-1-i}^2)\beta^i}{\omega}\right]^{p/r}\\ &\leq \frac{1}{\underline{\beta}\,\underline{\omega}^{p/r}}\sum\nolimits_{i=1}^{\infty} i\bar{\beta}^{ip/r}(\bar{\omega}+\bar{\alpha}y_{t-1-i}^2)^{p/r}.\end{aligned} \tag{7.39}$$

Therefore, by (7.30) and Minkowski's inequality we get

$$\left\|\sup_{\theta\in\Theta^0}\frac{\partial h_t}{\partial\beta}\frac{1}{h_t}\right\|_r \leq K\sum\nolimits_{i=1}^{\infty} i\bar{\beta}^{ip/r}\left[\mathrm{E}\left(\bar{\omega}+\bar{\alpha}y_{t-1-i}^2\right)^p\right]^{1/r} < \infty.$$

ii) From (7.35)-(7.37), we observe that the relevant second derivatives satisfy

$$\frac{\partial^2 h_t}{\partial\beta^2}\frac{1}{h_t} \leq \frac{1}{\beta}\sum\nolimits_{i=2}^{\infty} i(i-1)\frac{(\omega+\alpha y_{t-1-i}^2)\beta^i}{\omega+(\omega+\alpha y_{t-1-i}^2)\beta^i} \tag{7.40}$$

[4]Note that this lemma extends Lemma 4 of Lumsdaine (1996) and Lemmas 8 and 10 of Lee and Hansen (1994), since our results apply to moments of any order.

and

$$\frac{\partial^2 h_t}{\partial\alpha\partial\beta} \le \sum\nolimits_{i=1}^{\infty} i\beta^i \frac{(\omega + \alpha y_{t-1-i}^2)}{\omega + (\omega + \alpha y_{t-1-i}^2)\beta^i}$$

(the other derivatives are naturally bounded). Using the same arguments as in part (i) of the lemma the desired results follow.

iii) The proof is similar to part (i)-(ii) of the lemma, hence omitted.

iv) Let $\ln h_t^+ = \max(h_t, 0) \ge 0$ and $\ln h_t^- = \max(-h_t, 0) \ge 0$ where $\ln h_t = \ln h_t^+ - \ln h_t^-$. Similarly, define $\ln \tilde{h}_t^+$ and $\ln \tilde{h}_t^-$. From (2.5) and (2.8), for some $p \in (0, 1)$ we get that $\ln h_t^+ \le h_t^p$ and $\max(\ln h_t^-, \ln \tilde{h}_t^-) \le \max(0, -\ln\omega)$. Since $\tilde{h}_t - \omega \le h_t$, it easy to show that $\ln \tilde{h}_t^+ \le \max(|\ln\omega|,\ |\ln\bar{\omega}|) + \omega^{-1}h_t^p$. These results, (7.25) and the fact that $|\sup_{\theta\in\Theta} \ln h_t| \ \le\ \sup_{\theta\in\Theta} \ln h_t^+ + \sup_{\theta\in\Theta} \ln h_t^-$ imply the desired result.

Lemma 4: Under Assumptions A1-A4,

i) $\mathrm{E}(\ell_t(\theta_0)) \ \le\ \mathrm{E}(\ell_t(\theta))$ with equality if and only if $\theta = \theta_0$.

ii) For any compact set $\mathrm{N} \subseteq \Theta$,

$$\liminf_{n\to\infty} \inf_{\theta\in\mathrm{N}} \tfrac{1}{n} \sum\nolimits_{t=1}^{n} \ell_t(\theta) \ \ge\ \mathrm{E} \inf\nolimits_{\theta\in\mathrm{N}} \ell_t(\theta).$$

iii) $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \ell_t(\theta_0) = \mathrm{E}\ell_t(\theta_0)$.

Proof:

i) Note that

$$\begin{aligned} \mathrm{E}(\ell_t(\theta)) - \mathrm{E}(\ell_t(\theta_0)) &= \tfrac{1}{2}\mathrm{E}\left[(z_t - \ln h_t(\theta))^2 - \eta_t^2\right] \\ &= \tfrac{1}{2}\mathrm{E}\left[\ln(h_{0t} - \ln(h_t(\theta))\right]^2 + \mathrm{E}\left[\ln(h_{0t}/h_t(\theta))\right]\mathrm{E}(\eta_t) \\ &= \tfrac{1}{2}\mathrm{E}\left[\ln\left(h_t(\theta_0)/h_t(\theta)\right)\right]^2 - E[\ln(h_t(\theta))]E(\eta_t) \ \ge\ 0. \end{aligned} \tag{7.41}$$

Similar to lemma 3(iv), we can show that $E(\ln h_{0t}/h_t) \in \Re$. Hence the second term after the second equality equals zero. The term after this equality is zero only when $h_t(\theta_0) = h_t(\theta)$ and positive otherwise, because $(\ln(x))^2 \ge 0$ for all $x > 0$ with equality only when $x = 1$. Next, since, by Assumption A3, ε_t^2 is a non-degenerate random variable, it is trivial to show that $h_t(\theta) = h_t(\theta_0)$ implies $\theta = \theta_0$ and the desired result follows.

ii) For any compact set $\mathrm{N} \subseteq \Theta$ we have,

$$\liminf_{n\to\infty} \inf_{\theta\in\mathrm{N}} \tfrac{1}{n} \sum_{t=1}^{n} \ell_t(\theta) \ \ge\ \liminf\nolimits_{n\to\infty} \tfrac{1}{n} \sum_{t=1}^{n} \inf\nolimits_{\theta\in\mathrm{N}} \ell_t(\theta) \tag{7.42}$$

Further, note $\mathrm{E}\ell_t(\theta) < \infty$ is well defined and belongs to $\Re \cup \{+\infty\}$. Hence, by Lemma 1(ii), we can apply the ergodic theorem (see Billingsley, 1995, p. 284) to the stationary and ergodic sequence $\{\inf_{\theta\in\mathrm{N}} \ell_t(\theta)\}_t$ to obtain

$$\begin{aligned} \liminf_{n\to\infty} \inf_{\theta\in\mathrm{N}} \tfrac{1}{n} \sum\nolimits_{t=1}^{n} \ell_t(\theta) &\ge \liminf_{n\to\infty} \tfrac{1}{n} \sum\nolimits_{t=1}^{n} \inf\nolimits_{\theta\in\mathrm{N}} \ell_t(\theta) \\ &\ge \mathrm{E}\left(\inf_{\theta\in\mathrm{N}} \ell_t(\theta)\right). \end{aligned} \tag{7.43}$$

iii) Note that $\mathrm{E}\ell_t(\theta_0) = \mathrm{E}(\eta_t^2) < \infty$ by Lemma 1(iii). The desired result follows from Lemma 1(ii), and the ergodic theorem.

Lemma 5: Under Assumptions A1-A4,

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\tilde{\ell}_t(\theta)-\ell_t(\theta)\right)\right| \underset{a.s.}{\to} 0.$$

Proof:
Let $A_t(\theta)=\tilde{\ell}_t(\theta)-\ell_t(\theta)$. To prove this result, it suffices to check that $\mathrm{E}\sup_{\theta\in\Theta}|A_t(\theta)|^q$, is bounded by a summable sequence in t, for some $q\geq 0$. Indeed then (by Markov inequality) for all $\lambda>0$,

$$\sum_{t=1}^{\infty}\mathrm{P}(\sup_{\theta\in\Theta}|A_t(\theta)| \;>\; \lambda) \;\leq\; \sum_{t=1}^{\infty}\mathrm{E}\sup_{\theta\in\Theta}|A_t(\theta)|^q/\lambda^q<\infty \tag{7.44}$$

so that the Borel-Cantelli lemma implies that $\sup_{\theta\in\Theta}|A_t(\theta)|$ converges to zero a.s. This convergence and the Cesaro lemma imply the desired result. Now, since $\tilde{h}_t,\ h_t\geq\underline{\omega}>0$, an application of the mean-value theorem lead to

$$|\ln\tilde{h}_t(\theta)-\ln h_t(\theta)| \;\leq\; K|\tilde{h}_t(\theta)-h_t(\theta)|. \tag{7.45}$$

So, from (2.4), (2.9) and the c_r inequality, for some $p\in(0,1)$

$$\begin{aligned}
&\left\|\sup_{\theta\in\Theta}|\tilde{\ell}_t(\theta)-\ell_t(\theta)|\right\|_{p/2} \qquad (7.46)\\
\leq&\left\|\sup_{\theta\in\Theta}\left(\left|\ln\tilde{h}_t(\theta)-\ln h_t(\theta)\right|\cdot\left|2z_t+\ln\tilde{h}_t(\theta)+\ln h_t(\theta)\right|\right)\right\|_{p/2}\\
\leq&\left\|\sup_{\theta\in\Theta}\left|\tilde{h}_t(\theta)-h_t(\theta)\right|\right\|_{p}\cdot\left\|2\eta_t+\sup_{\theta\in\Theta}\ln\tilde{h}_t(\theta)+3\sup_{\theta\in\Theta}\ln h_t(\theta)\right\|_{p}=O(\bar{\beta}^t).
\end{aligned}$$

The second inequality holds by (7.45). The third inequality holds by Cauchy-Schwarz inequality. The last equality holds by Minkowski and Lyapunov inequalities, Lemmas 3(iv) and 2(i).

Lemma 6: Under Assumptions A1-A5,

i) $\left|n^{-1/2}\sum_{t=1}^{n}\left(\nabla\tilde{\ell}_t(\theta_0)-\nabla\ell_t(\theta_0)\right)\right| \to 0$ a.s.

ii) $n^{-1/2}\sum_{t=1}^{n}\nabla\ell_t(\theta_0)\overset{d}{\to}N\left(0,\kappa J\right)$ where J is positive definite and $\kappa=\mathrm{E}(\eta_t^2)>0$.

Proof:

i) We use the proof idea of Lemma 8 in Robinson and Zaffaroni (2006). Let $B_t=\nabla\ell_{it}(\theta_0)-\nabla\tilde{\ell}_{it}(\theta_0)$, the gradients of (2.4) and (2.9) are given by

$$\nabla\tilde{\ell}_{it}(\theta_0)=-(z_t-\ln\tilde{h}_{0t})\frac{\dot{\tilde{h}}_{0it}}{\tilde{h}_{0t}},\ \ \nabla\ell_{it}(\theta_0)=-(z_t-\ln h_{0t})\frac{\dot{h}_{0it}}{h_{0t}}=-\eta_t\frac{\dot{h}_{0it}}{h_{0t}} \tag{7.47}$$

where $\dot{h}_{0it}=\dot{h}_{it}(\theta_0),\ \dot{\tilde{h}}_{0it}=\dot{\tilde{h}}_{ijt}(\theta_0)$. Hence,

$$B_t=\nabla\tilde{\ell}_{it}(\theta_0)-\nabla\ell_{it}(\theta_0)=\eta_t\left(\frac{\dot{h}_{0it}}{h_{0t}}-\frac{\dot{\tilde{h}}_{0it}}{\tilde{h}_{0t}}\right)+\frac{\dot{\tilde{h}}_{0it}}{\tilde{h}_{0t}}\ln\left(\frac{\tilde{h}_{0t}}{h_{0t}}\right)$$

and

$$n^{-1/2}\sum\nolimits_{t=1}^{n} B_t \leq n^{-1/2}K\sum\nolimits_{t=1}^{n}\eta_t\left(\dot{h}_{0it}-\dot{\tilde{h}}_{0it}\right)+\frac{\dot{\tilde{h}}_{0it}}{\tilde{h}_{0t}}\left(h_{0it}-\tilde{h}_{0it}\right). \tag{7.48}$$

Next, by application of the c_r and Cauchy-Schwarz inequalities, we get that $\sum_{t=1}^{\infty}|B_t|$ has some finite $p>0$ moment and thus by Loeve (1977, p. 121) is a.s. finite. Further Lemma 2(i)-(ii) implies that a.s. $|B_t| \leq K\bar{\beta}^t$, $\forall t$. Hence, by Kronecker lemma (7.48) tends to zero a.s. as $n\to\infty$ and the desired result follows.

ii) From (7.47)

$$\mathrm{E}\left(\nabla\ell_{it}(\theta_0)\middle|\Im_{t-1}\right)=\frac{-\dot{h}_{0it}}{h_{0t}}\mathrm{E}\left(\eta_t\middle|\Im_{t-1}\right)=\frac{-\dot{h}_{0it}}{h_{0t}}\mathrm{E}\left(\eta_t\right)=0$$

where $\Im_t=\sigma(y_t,y_{t-1},\ldots)$ and

$$\|\nabla\ell_{it}(\theta_0)\nabla\ell_{jt}(\theta_0)\| \leq \mathrm{E}\left(\eta_t^2\right)\left\|\frac{\dot{h}_{0it}}{h_{0t}}\right\|_2\left\|\frac{\dot{h}_{0jt}}{h_{0t}}\right\|_2<\infty$$

by applying the Cauchy-Schwarz inequality and Lemmas 1(iii) and 3(i). Thus, we have shown that the second moment of each element of the gradient is finite hence $\mathrm{E}\left|\nabla\ell_t(\theta_0)\nabla\ell_t(\theta_0)'\right|<\infty$. These results and Lemma 1(ii) imply that$\{\nabla\ell_t(\theta_0),\Im_t\}$ is a stationary, ergodic and martingale difference sequence with finite variance

$$var(\nabla\ell_t(\theta_0))=\mathrm{E}(\eta_t^2)\mathrm{E}\left(\frac{1}{h_{0t}^2}\frac{\partial h_{0t}}{\partial\theta}\frac{\partial h_{0t}}{\partial\theta'}\right)=\kappa J$$

Next, by using similar arguments used in Lemma 5 in Lumsdaine (1996) we can show that J is a positive definite matrix $\kappa>0$ by Assumption A4. Thus, Theorem 23.1 of Billingsley (1968) and the Cramér-Wold device imply that $n^{-1/2}\sum_{t=1}^{n}\nabla^2\ell_t(\theta_0)\overset{d}{\to}N(0,\kappa J)$.

Lemma 7: Under Assumptions A1-A5,

i) $\sup_{\theta\in\Theta^0}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\nabla^2\tilde{\ell}_t(\theta)-\nabla^2\ell_t(\theta)\right)\right| \to 0$ a.s.

ii) If $\tilde{\theta}_n \underset{a.s.}{\to} \theta_0$, then $\frac{1}{n}\sum_{t=1}^{n}\nabla^2\ell_t(\tilde{\theta}_n) \underset{a.s.}{\to} -J$.

Proof:

i) First, let $C_t(\theta) = \nabla^2\tilde{\ell}_{ijt}(\theta)-\nabla^2\ell_{ijt}(\theta)$. Using similar arguments as in Lemma 5, it suffices to check that $\mathrm{E}\sup_{\theta\in\Theta^0}|C_t(\theta)|^q$ is bounded by a summable sequence in t, for some $q\geq 0$. Second, given (2.4) and (2.9) the second derivatives are

$$\nabla^2\tilde{\ell}_{ijt}(\theta)=-(z_t-\ln\tilde{h}_t)\frac{\ddot{\tilde{h}}_{ijt}}{\tilde{h}_t}+(z_t-\ln\tilde{h}_t+1)\frac{\dot{\tilde{h}}_{it}\dot{\tilde{h}}_{jt}}{\tilde{h}_t^2} \tag{7.49}$$

and

$$\nabla^2\ell_{ijt}(\theta)=-(z_t-\ln h_t)\frac{\ddot{h}_{ijt}}{h_t}+(z_t-\ln h_t+1)\frac{\dot{h}_{it}\dot{h}_{jt}}{h_t^2}. \tag{7.50}$$

Third, note

$$\frac{\dot{h}_{it}\dot{\tilde{h}}_{jt}}{h_t^2}-\frac{\dot{\tilde{h}}_{it}\dot{\tilde{h}}_{jt}}{\tilde{h}_t^2} \leq K\left\{\frac{\dot{h}_{it}}{h_t}\left[\dot{\tilde{h}}_{it}-\dot{h}_{it}\right]+\frac{\dot{\tilde{h}}_{jt}}{\tilde{h}_t}\left[\dot{\tilde{h}}_{jt}-\dot{h}_{jt}\right]\right\}. \tag{7.51}$$

Finally, using (7.49)-(7.51) we obtain

$$
\begin{aligned}
\sup_{\theta\in\Theta^0} C_t(\theta) &\le \sup_{\theta\in\Theta^0}\left\{\left(1+\eta_t+\ln\left(\frac{h_{0t}}{h_t}\right)\right)\left(\frac{\ddot{h}_{ijt}}{h_t}-\frac{\ddot{\tilde{h}}_{ijt}}{\tilde{h}_t}\right)+\frac{\tilde{h}_{ijt}}{\tilde{h}_t}\ln\left(\frac{\tilde{h}_t}{h_t}\right)\right. \\
&+\left.\left(\frac{\dot{h}_{it}\dot{h}_{jt}}{h_t^2}-\frac{\dot{\tilde{h}}_{it}\dot{\tilde{h}}_{jt}}{\tilde{h}_t^2}\right)+\frac{\dot{\tilde{h}}_{it}\dot{\tilde{h}}_{jt}}{\tilde{h}_t^2}\ln\left(\frac{\tilde{h}_t}{h_t}\right)\right\} \\
&\le K\sup_{\theta\in\Theta^0}\left\{1+\eta_t+\ln\left(\frac{h_{0t}}{h_t}\right)\right\}\left\{\left(\frac{\tilde{h}_{ijt}}{h_t}+\frac{\dot{\tilde{h}}_{it}}{\tilde{h}_t}\frac{\dot{\tilde{h}}_{jt}}{\tilde{h}_t}\right)\left(h_t-\tilde{h}_t\right)\right. \\
&+\left.\frac{\dot{h}_{it}}{h_t}\left(\dot{\tilde{h}}_{it}-\dot{h}_{it}\right)+\frac{\dot{\tilde{h}}_{jt}}{\tilde{h}_t}\left(\dot{\tilde{h}}_{jt}-\dot{h}_{jt}\right)+\left(\ddot{h}_{ijt}-\ddot{\tilde{h}}_{ijt}\right)\right\}.
\end{aligned}
$$

By applying Holder, Minkowski and Lyapunov inequalities with Lemmas 1(iii), 2 and 3, we get for some $q\in(0,1)$ that $\mathrm{E}\sup_{\theta\in\Theta^0}|C_t(\theta)|^q=O(\bar{\beta}^t)$ and the desired result follows.

ii) From (7.50), $\mathrm{E}\left(\nabla^2\ell_t(\theta_0)\right)=J$ and

$$
\begin{aligned}
\mathrm{E}\sup_{\theta\in\Theta^0}\left|\nabla^2\ell_{ijt}(\theta)\right| &\le \mathrm{E}\sup_{\theta\in\Theta^0}\left|\left(1+\eta_t+\ln\left(\frac{h_{0t}}{h_t}\right)\right)\left(\frac{\ddot{h}_{ijt}}{h_t}+\frac{\dot{h}_{it}\dot{h}_{jt}}{h_t^2}\right)\right| \\
&\le \left\{1+||\eta_t||_2+\left\|\sup_{\theta\in\Theta^0}\ln\left(\frac{h_{0t}}{h_t}\right)\right\|_2\right\}\left\{\left\|\sup_{\theta\in\Theta^0}\frac{\ddot{h}_{ijt}}{h_t}\right\|_2\right. \\
&+\left.\left\|\sup_{\theta\in\Theta^0}\frac{\dot{h}_{it}}{h_t}\right\|_4\left\|\sup_{\theta\in\Theta^0}\frac{\dot{h}_{jt}}{h_t}\right\|_4\right\}<\infty.
\end{aligned}
$$

The second inequality holds by applying the Cauchy-Schwarz and Minkowski inequalities. The last inequality holds by Lemmas 1(iii) and 3. From the ergodic theorem (see Billingsley, 1995),

$$\sup_{\theta\in\Theta^0}\left|\tfrac{1}{n}\sum_{t=1}^n\nabla^2\ell_t(\theta)-\mathrm{E}\left(\nabla^2\ell_t(\theta)\right)\right| \underset{a.s.}{\to} 0.$$

Hence, given $\varepsilon>0$

$$\left|\tfrac{1}{n}\sum_{t=1}^n\nabla^2\ell_t(\tilde{\theta}_n)-\mathrm{E}\left(\nabla^2\ell_t(\tilde{\theta}_n)\right)\right| < \tfrac{1}{2}\varepsilon$$

a.s. for n sufficiently large. Since $\mathrm{E}\left(\nabla^2\ell_t(\theta)\right)$ is continues

$$\left|\mathrm{E}\left(\nabla^2\ell_{ijt}(\tilde{\theta}_n)\right)-\mathrm{E}\left(\nabla^2\ell_{ijt}(\theta_0)\right)\right| < \tfrac{1}{2}\varepsilon$$

a.s. for n sufficiently large since $\tilde{\theta}_n\to\theta_0$ a.s. and the desired result follows from an application of the triangle inequality, since ε is arbitrary.

Appendix B: proofs of theorems 4.4-4.6 and lemmata 8-10

The estimator $\hat{m}_n$ minimizes the objective function given in (3.17) and for the unobserved process we can construct similarly, the following unobserved objective function

$$\bar{Q}_n(m)=\frac{1}{n}\sum_{t=1}^n l_t(m,\hat{\theta}_n),\quad l_t(m,\theta) = |\ln\hat{\varepsilon}_t^2(\theta)-m|,\ \hat{\varepsilon}_t^2(\theta)=h_t^{-1}(\theta)y_t^2$$

and $\bar{Q}_n^0(m) = \frac{1}{n}\sum_{t=1}^n l_t(m,\theta_0)$. The first order conditions for minimization of $\bar{Q}_n(m)$ (3.17) are not differentiable due to the appearance of an indicator function. Thus, empirical process methods should be applied and further notations are needed. Let $\Lambda = [-M, M]$, for some positive $M < \infty$, and $\phi = (\theta', m)' \in \Theta \times \Lambda = \Phi$. Define $s_t(\phi) = I(\ln(h_t^{-1}(\theta)y_t^2) - m \leq 0)$ and $\tilde{s}_t(\phi) = I(\ln(\tilde{h}_t^{-1}(\theta)y_t^2) - m \leq 0)$. Their normalized summands are defined by $S_n(\phi) = n^{-0.5}\sum_{t=1}^n [s_t(\phi) - E(s_t(\phi))]$ and $\tilde{S}_n(\phi) = n^{-0.5}\sum_{t=1}^n [\tilde{s}_t(\phi) - E(\tilde{s}_t(\phi))]$, respectively. Also let $S(\phi)$ denote a mean zero Gaussian process with covariance kernel $K(\phi_i, \phi_j)$. This means that for any $\{\phi_1, \ldots, \phi_k\} \in \Phi^k$, $\{S(\phi_1), \ldots, S(\phi_k)\}$ is multivariate normal with mean zero and covariances $E(S(\phi_i)S(\phi_j)') = K(\phi_i, \phi_j)$. In order to establish the asymptotic normality of the two-stage estimator, we basically need to show that the empirical process $\tilde{S}_n(\phi)$ converges over Φ to the Gaussian process $S(\phi)$ as well as other results proven in the Lemmata below.

Proof of Theorem 4.4: Let $R_t = \sup_{\theta\in\Theta}\left|\dot{h}_t(\theta)h_t^{-1}(\theta)\right|$ and we first show that

$$\sup_{m \in \Lambda} |\bar{Q}_n^0(m) - \bar{Q}_n(m)| = o_{a.s}(1). \tag{7.52}$$

We have,

$$\begin{aligned}\sup_{m \in \Lambda} |\bar{Q}_n^0(m) - \bar{Q}_n(m)| &\leq n^{-1}\sum_{t=1}^n \sup_{m \in \Lambda} |l_t(m,\theta_0) - l_t(m,\hat{\theta}_n)| \leq n^{-1}\sum_{t=1}^n \left|\ln h_t(\theta_0) - \ln h_t(\hat{\theta}_n)\right| \\ &\leq n^{-1}\sum_{t=1}^n \sup_{\theta\in\Theta}\left|\dot{h}_t(\theta)h_t^{-1}(\theta)\right| \cdot |\hat{\theta}_n - \theta_0| = n^{-1}\sum_{t=1}^n R_t|\hat{\theta}_n - \theta_0|.\end{aligned}$$

The second inequality uses the fact that the absolute function is Lipschitz continuous. Lemma 3(i) implies that R_t is almost surely finite. For t fixed, the strong consistency of $\hat{\theta}_n$ implies that $R_t|\hat{\theta}_n - \theta_0| = o_{a.s.}(1)$ and the desired result follows from Cesaro lemma.

Second, we show that

$$\sup_{m\in\Lambda} |\bar{Q}_n(m) - \tilde{\bar{Q}}_n(m)| = o_{a.s}(1). \tag{7.53}$$

Since,

$$\begin{aligned}\sup_{m \in \Lambda} |\bar{Q}_n(m) - \tilde{\bar{Q}}_n(m)| &\leq n^{-1}\sum_{t=1}^n \sup_{m \in\Lambda} |l_t(m,\hat{\theta}_n) - \tilde{l}_t(m,\hat{\theta}_n)| \\ &\leq n^{-1}\sum_{t=1}^n \sup_{\theta\in\Theta} |\ln h_t(\theta) - \ln \tilde{h}_t(\theta)| \leq Kn^{-1}\sum_{t=1}^n \sup_{\theta\in\Theta}\left|h_t(\theta) - \tilde{h}_t(\theta)\right|\end{aligned}$$

(7.53) follows from Lemma 2(i), Markov inequality and Cesaro lemma. By the triangle inequality (7.52)-(7.53) imply that $\sup_{m\in\Lambda} |\bar{Q}_n^0(m) - \tilde{\bar{Q}}_n(m)| = o_{a.s}(2.1)$. This result and the ergodic theorem imply

$$P\left(\lim_{n\to\infty} \sup_{m \in \Lambda} |\tilde{\bar{Q}}_n(m) - \bar{Q}_\infty^0(m)| = 0\right) = 1 \tag{7.54}$$

where $\bar{Q}_\infty^0(m) = E|\eta_t - m|$. Let $\gamma > 0$ and define the compact set $D = \{m \in \Lambda : |m - m_0| \geq \gamma\}$. Assumption A6 ensures that $\bar{Q}_\infty^0(m) > \bar{Q}_\infty^0(m_0)$ when $m \in D$, hence $\min_{m\in D}(\bar{Q}_\infty^0(m_0) - \bar{Q}_\infty^0(m)) < 0$ which implies by (7.54) that there exists N such that for all $n > N$, $\min_{m\in D}(\tilde{\bar{Q}}_n(m_0) - \tilde{\bar{Q}}_n(m)) < 0$ with probability one. Since $\tilde{\bar{Q}}_n(m_0) - \tilde{\bar{Q}}_n(\hat{m}_n) \geq 0$ by definition, we deduce that $\hat{m}_n \notin D$ which leads that $\hat{m}_n \to_{a.s.} m_0$ since $\gamma > 0$ is arbitrary. Next, by the convexity of the objective function and by using similar arguments to those in Lemma A of Newey and Powell (1987), we can show that for any m outside Λ, $\tilde{\bar{Q}}_n(m) > \tilde{\bar{Q}}_n(\hat{m}_n)$ with probability 1, implying that $\hat{m}_n$ obtains minimum value over $\Re$. This result and Theorem 4.1 imply part (i). When further combined with Slutsky's theorem imply part (ii).

Proof of Theorem 4.5: The first order condition for minimization of (3.17) is given by (4.19) which is not everywhere differentiable. Hence, we will establish the asymptotic normality by taking a mean-value expansion of $n^{-1}\sum_{t=1}^{n} Esgn(\ln\tilde{\varepsilon}_t^2(\theta_0) - m_0)$ around $\hat{m}_n$ which yields

$$\begin{aligned} o_P(1) &= \frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n} \frac{\partial E(\tilde{l}_t(m_0,\theta_0))}{\partial m} \\ &= \frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n} \frac{\partial E(\tilde{l}_t(\hat{m}_n,\theta_0))}{\partial m} + \frac{1}{n}\sum\nolimits_{t=1}^{n} \frac{\partial^2 E(\tilde{l}_t(\bar{m}_n,\theta_0))}{\partial m^2}\sqrt{n}(\hat{m}_n - m_0) \end{aligned} \tag{7.55}$$

where the first equality holds by Lemma 8(iii) and $\bar{m}_n$ lies on the line joining $\hat{m}_n$ and m_0. Similar arguments and a mean-value expansion of the first term on (7.55) around $\hat{\theta}_n$ yield

$$\begin{aligned} o_P(2.1) &= \frac{1}{\sqrt{n}}\sum\nolimits_{t=1}^{n} \frac{\partial E(\tilde{l}_t(\hat{m}_n,\hat{\theta}_n))}{\partial m} + \frac{1}{n}\sum\nolimits_{t=1}^{n} \frac{\partial E(\tilde{l}_t(\hat{m}_n,\bar{\theta}_n))}{\partial m\partial\theta'}\sqrt{n}(\hat{\theta}_n - \theta_0) \\ &+ \frac{1}{n}\sum\nolimits_{t=1}^{n} \frac{\partial^2 E(\tilde{l}_t(\bar{m}_n,\theta_0))}{\partial m^2}\sqrt{n}(\hat{m}_n - m_0) \\ &= -2S_n(\phi_0) + \left(f_\xi(0) + o_{a.s.}\right)\sqrt{n}(\hat{m}_n - m_0) + \left(-f_\xi(0)\bar{J}' + o_{a.s}(1)\right)\sqrt{n}(\hat{\theta}_n - \theta_0) \end{aligned} \tag{7.56}$$

where $\bar{\theta}_n$ lies on the chord between $\hat{\theta}_n$ and θ_0. The first, second and third terms, after the second equality, result from Lemmata 9(i)-(iii), respectively. Next, using Theorem 4.2 and the Lindeberg-Lévy CLT, we obtain that

$$\begin{aligned} \begin{pmatrix} 2S_n(\phi_0) \\ \sqrt{n}(\hat{\theta}_n - \theta_0) \end{pmatrix} &= \frac{1}{\sqrt{n}}\begin{pmatrix} J^{-1}\sum_{t=1}^{n}\eta_t\dot{h}_{0t}h_{0t}^{-1} \\ \sum_{t=1}^{n} 2(I(\xi_t \le 0) - 0.5) \end{pmatrix} + o_P(2.1) \xrightarrow{d} \begin{pmatrix} Z_\theta \\ Z_m \end{pmatrix} \\ &:= Z \sim N\left\{0,\, V := \begin{pmatrix} \kappa J^{-1} & \bar{\kappa}\bar{J}'J^{-1} \\ \bar{\kappa}J^{-1}\bar{J} & 1 \end{pmatrix}\right\}. \end{aligned}$$

Hence, to show part (i)

$$\sqrt{n}(\hat{m}_n - m_0) = f_\xi^{-1}(0)\left(2S_n(\phi_0) + f_\xi(0)\bar{J}'\sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1)\right) \xrightarrow{d} \left(Z_m/f_\xi(0) + \bar{J}'Z_\theta\right)$$

and

$$\begin{aligned} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ m_0 - \hat{m}_n \end{pmatrix} &= \sqrt{n}(\hat{\phi}_n - \phi_0) \xrightarrow{d} \begin{pmatrix} Z_\theta \\ Z_m/f_\xi(0) + \bar{J}'Z_\theta \end{pmatrix} \\ &= \begin{pmatrix} I_3 & 0 \\ \bar{J}' & f_\xi^{-1}(0) \end{pmatrix}\begin{pmatrix} Z_\theta \\ Z_m \end{pmatrix} \sim N(0, \Psi V\Psi'). \end{aligned} \tag{7.57}$$

As for part (ii) applying (7.57), Theorems 1 and 4 to a mean-value expansion of $\hat{\hat{\theta}}_n = A_n\hat{\theta}_n$ around(θ_0, m_0) yield that

$$\sqrt{n}\begin{pmatrix} \bar{\omega}_0 - \hat{\hat{\omega}}_n \\ \bar{\alpha}_0 - \hat{\hat{\alpha}}_n \\ \bar{\beta}_0 - \hat{\hat{\beta}}_n \end{pmatrix} = e^{m_0} \times \begin{pmatrix} 1 & 0 & 0 & \omega_0 \\ 0 & 1 & 0 & \alpha_0 \\ 0 & 0 & e^{-m_0} & 0 \end{pmatrix}\sqrt{n}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ m_0 - \hat{m}_n \end{pmatrix} + o_P(2.1) \xrightarrow{d} \Xi Z.$$

Thus, $\sqrt{n}(\bar{\theta}_0 - \hat{\hat{\theta}}) \xrightarrow{d} N(0, \Xi V\Xi')$.

Lemma 8: Under Assumptions A1-A3 and A5-A7

i) $\sup_{\phi\in\Phi}\left|S_n(\phi) - \tilde{S}_n(\phi)\right| = o_p(1)$.

ii) $\tilde{S}_n \Rightarrow S$.

iii) $n^{-0.5}\sum_{t=1}^{n}\frac{\partial E(\tilde{l}_t(m_0,\theta_0))}{\partial m} = o_P(1)$.

Proof:

i) The indicator $I(z) = I(z \le 0)$ can be approximated by regular sequence[5] $\{\mathrm{J}_n(z)\}_{n\ge 1}$ cf. Lighthill (1958). Let $\{N_n\}$ be a sequence of finite positive numbers, the rate to be chosen below. Define $J_n(z) = \int_{-\infty}^{\infty} I(\tau)G(N_n(\tau - z))N_n e^{-\tau^2 N_n^{-2}}d\tau$ where $G(\varpi) = e^{-1/(1-\varpi^2)}\Big/\int_{-1}^{1} e^{-1/(1-w^2)}dw$ if $|\varpi| < 1$ and $G(\varpi) = 0$ if $|\varpi| \ge 1$. The function $\mathrm{J}_n(z)$ is uniformly bounded in z, and continuous and differentiable. $I(z)$ is differentiable, except at zero, with derivative $\delta(z)$, the Dirac delta function, hence $\delta(z)$ has regular sequence $D_n(z) = (N_n/\pi)^{0.5}\exp(-N_n z^2)$ see Lighthill (1958). Define $e_t(a) = \ln(y_t^2) - m - \ln a$ hence by definition $I(\ln(h_t^{-1}y_t^2) - m) = I(\mathrm{e}_t(h_t))$. Since the rate $N_n \to \infty$ can be made to be as fast as we choose, it can be set to ensure

$$\begin{aligned}
&\sup_{\phi\in\Phi} n^{-0.5}\left|\sum\nolimits_{t=1}^{n} s_t(\phi) - \tilde{s}_t(\phi)\right| \qquad (7.58)\\
&= \sup_{\phi\in\Phi} n^{-0.5}\left|\sum\nolimits_{t=1}^{n} I(\ln(h_t^{-1}(\theta)y_t^2) - m) - I(\ln(\tilde{h}_t^{-1}(\theta)y_t^2) - m)\right|\\
&= \sup_{\phi\in\Phi} n^{-0.5}\sum\nolimits_{t=1}^{n}\left|I(e_t(h_t)) - I(e_t(\tilde{h}_t))\right|\\
&\le n^{-0.5}\sum\nolimits_{t=1}^{n}\sup_{\phi\in\Phi}\left|J_n(e_t(h_t)) - J_n(e_t(\tilde{h}_t))\right| + o_P(1).
\end{aligned}$$

Therefore, by the mean value theorem, for some h_t^* that satisfies $|h_t^* - \tilde{h}_t| \le |h_t - \tilde{h}_t|$, it follows for some $r > 0$

$$\begin{aligned}
&\sup_{\phi\in\Phi} n^{-0.5}\sum\nolimits_{t=1}^{n} E\left|\mathrm{J}_n(\mathrm{e}_t(h_t)) - \mathrm{J}_n(\mathrm{e}_t(\tilde{h}_t)\right|^{r/2}\\
&\le Kn^{-0.5}\sup_{\phi\in\Phi}\sum\nolimits_{t=1}^{n} E\left|\mathrm{D}_n(\mathrm{e}_t(h_t^*))\left(h_t(\theta) - \tilde{h}_t(\theta)\right)\right|^{r/2}\\
&\le Kn^{-0.5}\left(\sum\nolimits_{t=1}^{n} E\sup_{\phi\in\Phi}\left|\mathrm{D}_n(\mathrm{e}_t(h_t(\theta)))\right|^r\right)^{1/2}\left(\sum\nolimits_{t=1}^{n} E\sup_{\phi\in\Phi}\left|h_t(\theta) - \tilde{h}_t(\theta)\right|^r\right)^{1/2}.
\end{aligned}$$

The last inequality follows from the Cauchy–Schwarz inequality. Distribution continuity implies that $\mathrm{e}_t(h_t^*) \ne 0$ a.s for any $\theta \in \Theta$. Hence, by construction the rate at which $\sup_{\theta\in\Theta}|\mathrm{D}_n(\mathrm{e}_t(h_t(\theta)))| \to_p 0$ can be made so fast by the choice of $\{N_n\}$ that $E\sup_{\phi\in\Phi}|\mathrm{D}_n(\cdot)|^r \to 0$ as fast as we choose for some small $r > 0$ by the dominated convergence theorem. Therefore by Lemma 2(i), we find that the RHS converges to zero as $n \to \infty$ which implies by the Markov's inequality that the first summand after the last inequality in (7.58) converges to zero in probability. Next by (3.15), for some $p \in (0, 1)$,

[5] A similar approach was used by Philips (1995) and Hill (2013) to approximate the indicator function.

$$\sup_{\phi\in\Phi} n^{-0.5}\sum\nolimits_{t=1}^{n}|E(s_t(\phi))-E(\tilde{s}_t(\phi))|^p \tag{7.59}$$

$$\begin{aligned}
&= \sup_{\phi\in\Phi} n^{-0.5}\sum\nolimits_{t=1}^{n}\left|E[I(\ln(h_t^{-1}(\theta)h_{0t})-m<-\eta_t)-I(\ln(\tilde{h}_t^{-1}(\theta)h_{0t})-m<-\eta_t)]\right|^p\\
&= \sup_{\phi\in\Phi} n^{-0.5}\sum\nolimits_{t=1}^{n}E\left(I(\ln(\tilde{h}_t(\theta)h_{0t}^{-1})<\eta_t-m<\ln(h_t(\theta)h_{0t}^{-1}))\right)^p\\
&= \sup_{\phi\in\Phi} n^{-0.5}\sum\nolimits_{t=1}^{n}E\left(E\left(I(\ln(\tilde{h}_t^{-1}(\theta)h_{0t})+c_0+m<\xi_t<\ln(h_t^{-1}(\theta)h_{0t})+c_0+m)\right)\Big|\Im_{t-1}\right)^p\\
&= \sup_{\phi\in\Phi} n^{-0.5}\sum\nolimits_{t=1}^{n}E\left(\left|\int_{\ln(h_{0t}^{-1}\tilde{h}_t(\theta))+c_0+m}^{\ln(h_{0t}^{-1}h_t(\theta))+c_0+m}f_\xi(z)dz\right|\right)^p\\
&\le Kn^{-0.5}\sum\nolimits_{t=1}^{n}\sup_{\theta\in\Theta}E|h_t(\theta)-\tilde{h}_t(\theta)|^p\\
&\le Kn^{-0.5}\sum\nolimits_{t=1}^{\infty}\bar{\beta}^{pt}\to 0.
\end{aligned}$$

The first inequality follows from Assumption A7 and the last inequality follows from Lemma 2(i). The desired result follows from (7.58), (7.59) and the triangle inequality.

ii) We start by showing the weak convergence of S_n to $S(\phi)$ a mean zero Gaussian process. Note that for each $\theta\in\Theta$, $h_t(\theta)$ is a GARCH (1,1) process which under Assumptions A2 and A3 is regular mixing with exponential rate by Theorem 4.3 of Francq and Zakoïan (2006b). Hence for each ϕ *in* Φ, $s_t(\phi)=I(\ln(h_t^{-1}(\theta)y_t^2)-m)$ is a strictly stationary, regular mixing process which admits moments of any order, to which the pointwise central limit theorem applies. Therefore, by the Cramer-Wold device, we can establish the finite dimensional distributional convergence of $(S_n(\phi_1),\ldots,S_n(\phi_N))$ to a multivariate normal distribution with zero mean and covariance kernel $K(\phi_j,\phi_k)$. To show that $S_n\Rightarrow S$, we further need to establish the stochastic equicontinuity. Since $s_t(\phi)$ is strictly stationary, regular mixing with exponential rate, we can appeal to Theorem 4.1, Application 1 of Doukhan et al. (1994). Thus, it suffices to show that the log of L^{2+r} bracketing number is integrable for some $r>0$. Because, $\Phi\subset\Re^4$ we can find a set $\{\phi_1,\ldots,\phi_N\}\in\Phi^N$ and a positive constant $G<\infty$ so that for all ϕ there is some $\phi_k\in\Phi^N$ satisfying $|\phi_k-\phi|\ \le GN^{-0.25}$ where $\lambda=1/(2+r)$. Using Assumption A7 and (3.15), we get for all $\phi\in\Phi$

$$\left\|I(\ln(h_t^{-1}(\theta)y_t^2)-m)-I(\ln(h_t^{-1}(\theta_k)y_t^2)-m_k)\right\|_{2+r} \tag{7.60}$$

$$\begin{aligned}
&=\left\|I(\ln(h_t^{-1}(\theta_k)h_{0t})-m_k<-\eta_t)-I(\ln(h_t^{-1}(\theta)h_{0t})-m<-\eta_t)\right\|_{2+r}\\
&=\left\|I(m-\ln(h_t^{-1}(\theta)h_{0t})<\eta_t<m_k-\ln(h_t^{-1}(\theta_k)h_{0t}))\right\|_{2+r}\\
&=E\left(E\left(I(m_k+c_0-\ln(h_t^{-1}(\theta_k)h_{0t})<\xi_t<m+c_0-\ln(h_t^{-1}(\theta)h_{0t})\right)\Big|\Im_{t-1}\right)^{1/(2+r)}\\
&=E\left(\left|\int_{m_k+c_0-\ln(h_{0t}^{-1}h_t(\theta_k))}^{m+c_0-\ln(h_{0t}^{-1}h_t(\theta))}f_\xi(z)dz\right|\right)^\lambda\le\bar{f}^\lambda E\left(|\ln(h_t(\theta_k)/h_t(\theta))+m-m_k|\right)^\lambda\\
&\le\bar{f}^\lambda|m-m_k|^\lambda+\bar{f}^\lambda E\left(\sup_{\theta\in\Theta}\left|h_t^{-1}(\theta)\dot{h}_t(\theta)\right|^\lambda\right)|\theta-\theta_k|^\lambda\le K|\phi-\phi_k|^\lambda\ =\ KG^\lambda N^{-0.25\lambda}
\end{aligned}$$

The fourth equality follows by the independence of ϵ_t and the second inequality follows by the mean-value theorem and the c_r inequality since $\lambda<1$. Now, by setting $N(\delta)=G^4(K\delta)^{4/\lambda}$ for some $\delta>0$ we have that $\left\|I(\ln(h_t^{-1}(\theta)y_t^2)-m)-I(\ln(h_t^{-1}(\theta_k)y_t^2)-m_k)\right\|_{2+r}\le\delta$. Thus, $N(\delta)$ satisfies the definition of bracketing number.

Since we can easily verify that $\sqrt{\log(N(\delta))}$ is integrable over the unit interval, the conditions of Doukhan et al. (1994) are met, establishing that S_n is stochastically equicontinuous. Thus, for any $\varepsilon > 0$ and $\delta > 0$ there exists $\eta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{\phi_1,\phi_2\in\Phi,|\phi_1-\phi_2|\leq\eta} |S_n(\phi_1) - S_n(\phi_2)| > \varepsilon\right) < \delta \tag{7.61}$$

Hence $S_n \Rightarrow S$ and the desired result follows by Lemma 8(i) and the asymptotic equivalence lemma (White, 1994, p.172).

iii) Note that

$$n^{-0.5}\sum\nolimits_{t=1}^{n} \frac{\partial E(\tilde{l}_t(m_0,\theta_0))}{\partial m} = n^{-0.5}\sum\nolimits_{t=1}^{n}\left(\frac{\partial E(\tilde{l}_t(m_0,\theta_0))}{\partial m} - \frac{\partial E(l_t(m_0,\theta_0))}{\partial m}\right) + \frac{\partial E(l_t(m_0,\theta_0))}{\partial m}$$

$$= 2n^{-0.5}\sum\nolimits_{t=1}^{n} E\left(\tilde{s}_t(\phi_0) - s_t(\phi_0)\right) = o_P(2.1)$$

The second summand after the first equality is zero by Assumption A6 and the last equality follows by Lemma 8(i) and the bounded convergence theorem.

Lemma 9: Under Assumptions A1-A3 and A5-A7

i) -$\sqrt{n}\frac{\partial E(\tilde{l}_t(\hat{m}_n,\hat{\theta}_n))}{\partial m} = 2S_n(\phi_0) + o_P(2.1) = n^{-0.5}\sum_{t=1}^{n} 2(I(\xi_t \leq 0) - 0.5) + o_P(1)$.

ii) If $\tilde{m}_n \to_{a.s.} m_0$ then $n^{-1}\sum_{t=1}^{n} \frac{\partial E(\tilde{l}_t(\tilde{m}_n,\theta_0))}{\partial m^2} = f_\xi(0) + o_{a.s.}(1)$.

iii) If $\tilde{m}_n \to_{a.s.} m_0$ and $\tilde{\theta}_n \to_{a.s.} \theta_0$ then $n^{-1}\sum_{t=1}^{n} \frac{\partial E(\tilde{l}_t(\tilde{m}_n,\tilde{\theta}_n))}{\partial m\partial\theta'} = f_\xi(0)\bar{J}' + o_{a.s}(1)$.

Proof:

i) By (4.19)

$$\begin{aligned}
-n^{-0.5}\sum\nolimits_{t=1}^{n} \frac{E\partial(\tilde{l}_t(\hat{m}_n,\hat{\theta}_n))}{\partial m} &= n^{-0.5}\sum\nolimits_{t=1}^{n}\left(\frac{\partial\tilde{l}_t(\hat{m}_n,\hat{\theta}_n)}{\partial m} - \frac{E\partial(\tilde{l}_t(\hat{m}_n,\hat{\theta}_n))}{\partial m}\right) \\
&= n^{-0.5}\sum\nolimits_{t=1}^{n}\left(sgn(\ln\tilde{\varepsilon}_t^2(\hat{\theta}_n) - \hat{m}_n) - Esgn(\ln\tilde{\varepsilon}_t^2(\hat{\theta}_n) - \hat{m}_n)\right) \\
&= 2\left(\tilde{S}_n(\hat{\phi}_n) - S_n(\hat{\phi}_n) + S_n(\hat{\phi}_n) - S_n(\phi_0)\right) + 2S_n(\phi_0).
\end{aligned} \tag{7.62}$$

We have to show that the first term on the RHS of (7.62) is $o_P(1)$. This follows because given any $\varepsilon > 0$ and $\delta > 0$, there exists a $\eta > 0$ such that

$$\begin{aligned}
&\limsup_{n\to\infty} P\left(\left|\tilde{S}_n(\hat{\phi}_n) - S_n(\hat{\phi}_n) + S_n(\hat{\phi}_n) - S_n(\phi_0)\right| > 2\varepsilon\right) \\
&\quad\leq\ \overline{\lim}_{n\to\infty} P\left(\left|\tilde{S}_n(\hat{\phi}_n) - S_n(\hat{\phi}_n)\right| > \varepsilon \cup \left|S_n(\hat{\phi}_n) - S_n(\phi_0)\right| > \varepsilon\right) \\
&\quad\leq \overline{\lim}_{n\to\infty} P\left(\left|\tilde{S}_n(\hat{\phi}_n) - S_n(\hat{\phi}_n)\right| > \varepsilon\right) + \overline{\lim}_{n\to\infty} P\left(\left|S_n(\hat{\phi}_n) - S_n(\phi_0)\right| > \varepsilon\right) \\
&\quad\leq \overline{\lim}_{n\to\infty} P\left(\left|\tilde{S}_n(\hat{\phi}_n) - S_n(\hat{\phi})\right| > \varepsilon\right) + \overline{\lim}_{n\to\infty} P\left(\left|S_n(\hat{\phi}_n) - S_n(\phi_0)\right| > \varepsilon \cup |\hat{\phi}_n - \phi_0|\ \leq \eta\right) \\
&\quad+ \overline{\lim}_{n\to\infty} P\left(|\hat{\phi}_n - \phi_0|\ > \eta\right) \leq \overline{\lim}_{n\to\infty} P\left(\sup_{\phi\in\Phi,|\phi-\phi_0|\leq\eta} |S_n(\phi) - S_n(\phi_0)| > \varepsilon\right) < \delta
\end{aligned}$$

where the fourth inequality follows from Theorem 4.1, Lemma 8(i) and the fifth inequality uses (7.61) hence,

$$
\begin{aligned}
-n^{-0.5}\sum\nolimits_{t=1}^{n}\frac{E\partial(\tilde{l}_t(\tilde{m}_n,\hat{\theta}_n))}{\partial m} &= 2S_n(\phi_0)+o_P(1)\\
&= 2n^{-0.5}\sum\nolimits_{t=1}^{n}(I(\eta_t-m_0\le 0)-E(I(\eta_t-m_0\le 0)))+o_P(1)\\
&= n^{-0.5}\sum\nolimits_{t=1}^{n}2(I(\xi_t\le 0)-0.5)+o_P(1).
\end{aligned}
$$

ii) Let $Z_t=\ln\tilde{h}_{0t}h_{0t}^{-1}$ and $B(\tau)$ be a closed ball of radius $\tau>0$ centered at zero. In what follows, we use the fact that the first derivative of $|u|$ is $2I(u\le 0)-1$, the second derivative is two times the Dirac delta function given by $2\delta(u)$ and that $\int\delta(u-z)f_\xi(u)du=f_\xi(z)$for all $z\in\Re$. Now, Assumption A3 allows us to change the order of differentiation and integration (see Davidson, 1994, p.141), and since ξ_t is independent of $\Im_{t-1}$, we have

$$
E\left(\frac{\partial(l_t(m_0,\theta_0))}{\partial m^2}\bigg|\Im_{t-1}\right)=E\left(\delta(\ln y_t^2-\ln h_{0t}-m_0)\big|\Im_{t-1}\right)=\int\delta(\xi_t)f_\xi(z)dz=f_\xi(0) \tag{7.63}
$$

Next, by combining (7.28), (7.45) and the Borel-Cantelli lemma we deduce that $Z_t\to_{a.s.}0$ as $t\to\infty$. Hence by Theorem 4.4 for all $B(\tau)$ there exists some $n_1(\tau)$ such that for all $t,n>n_1(\tau)$ almost surely $\tilde{m}_n-m_0\in B(\tau/2)$ and $Z_t\in B(\tau/2)$. Therefore, similar to (7.63), for all $t,n>n_1(\tau)$

$$
\begin{aligned}
\left|E\left(\frac{\partial(\tilde{l}_t(\tilde{m}_n,\theta_0))}{\partial m^2}\bigg|\Im_{t-1}\right)\right| &\le \sup_{m-m_0\in B(\tau/2)}\left|E\left(\frac{\partial(\tilde{l}_t(m,\theta_0))}{\partial m^2}\bigg|\Im_{t-1}\right)\right| \qquad (7.64)\\
&= \sup_{m-m_0\in B(\tau/2)}\left|\int\delta(\xi_t-((m-m_0)+Z_t))f_\xi(z)dz\right|\\
&\le \sup_{z\in B(\tau)}\left|\int\delta(\xi_t-z)f_\xi(z)dz\right| = \sup_{z\in B(\tau)}f_\xi(z)
\end{aligned}
$$

Because f_ξ is continuous on $B(\bar{\tau})$ for some $\bar{\tau}>0$ by Assumption A6, given $\varepsilon>0$ there exists $0<\tau(\epsilon)<\bar{\tau}$ such that $\sup_{z\in B(\tau(\varepsilon))}|f_\xi(z)-f_\xi(0)|<\varepsilon$. Therefore, given (7.63)-(7.64) for all $t,n>n_1(\tau(\varepsilon))$ almost surely

$$
\left|E\left(\frac{\partial(\tilde{l}_t(\tilde{m}_n,\theta_0))}{\partial m^2}\right)-\frac{E\partial(l_t(m_0,\theta_0))}{\partial m^2}\right| \le \sup_{z\in B(\tau(\varepsilon))}|f_\xi(z)-f_\xi(0)|<\varepsilon
$$

and for $n>\max\left(n_1(\tau(\varepsilon)),\,2\bar{f}n_1(\tau(\varepsilon))/\varepsilon\right)$ we have that

$$
\begin{aligned}
&\left|n^{-1}\sum\nolimits_{t=1}^{n}\frac{E\partial(\tilde{l}_t(\tilde{m}_n,\theta_0))}{\partial m^2}-\frac{E\partial(l_t(m_0,\theta_0))}{\partial m^2}\right|\\
&\le\left|n^{-1}\sum\nolimits_{t=1}^{n_1(\varepsilon)}+\sum\nolimits_{t=n_1(\varepsilon)+1}^{n}\left|\frac{E\partial(\tilde{l}_t(\tilde{m}_n,\theta_0))}{\partial m^2}-\frac{E\partial(l_t(m_0,\theta_0))}{\partial m^2}\right|\right|<2\varepsilon
\end{aligned}
$$

and the desired result follows since $\varepsilon>0$ is arbitrary.

iii) We note that

$$
\frac{\partial E(l_t(m_0,\theta_0))}{\partial m\partial\theta'}=E\left(\delta(\xi_t)\dot{h}_{0t}h_{0t}^{-1}\right)=f_\xi(0)\bar{J}'
$$

and the rest of the proof resembles part (ii) of this lemma, hence is omitted.

Lemma 10: Under Assumptions A1-A3 and A5-A7

i) $\hat{\bar{\kappa}}_n \to_p \bar{\kappa}$.

ii) $\hat{f}_\xi(0) \to_p f_\xi(0)$.

iii) $\hat{\Xi}_n \hat{V}_n \hat{\Xi}_n \to_p \Xi V \Xi$.

Proof:

i) In equation (7.60) by replacing θ, θ_k with $\hat{\theta}_n$, θ_0, respectively and using Theorems 1 and 4, we can show $\|I(\hat{\eta}_t - \hat{m}_n) - I(\eta_t - m_0)\|_2 = o(1)$ which implies by the Markov inequality that $I(\tilde{\eta}_t - \hat{m}_n) - I(\eta_t - m_0) \to_p 0$. Using this result, (7.24) and some direct calculations, we get,

$$\begin{aligned}\tilde{\eta}_t I(\tilde{\eta}_t - \hat{m}_n) - \eta_t I(\xi_t) &= \tilde{\eta}_t I(\tilde{\eta}_t - \hat{m}_n) - \eta_t I(\eta_t - m_0) \\ &= (\tilde{\eta}_t - \eta_t)\left(I(\tilde{\eta}_t - \hat{m}_n) - I(\eta_t - m_0)\right) + I(\eta_t - m_0)(\tilde{\eta}_t - \eta_t) \\ &+ \eta_t\left(I(\tilde{\eta}_t - \hat{m}_n) - I(\eta_t - m_0)\right) \\ &\leq 3(\tilde{\eta}_t - \eta_t) + \eta_t\left(I(\tilde{\eta}_t - \hat{m}_n) - I(\eta_t - m_0)\right) = o_P(1).\end{aligned} \tag{7.65}$$

The ergodic theorem applied to $n^{-1}\sum_{t=1}^n \eta_t I(\xi_t)$ and (7.65) indicate that $\hat{\bar{\kappa}}_n \to_p \bar{\kappa}$.

ii) Let $\hat{f}_\xi^*(0) = \frac{1}{na_n}\sum_{t=1}^n H\left(-a_n^{-1}(\ln \hat{\varepsilon}_t^2(\hat{\theta}_n) - \hat{m}_n)\right)$ and $\hat{f}_\xi^\diamondsuit(0) = \frac{1}{na_n}\sum_{t=1}^n H\left(-a_n^{-1}(\eta_t - m_0)\right)$.
By condition (i),

$$\begin{aligned}E\left(\left|\hat{f}_\xi^*(0) - \hat{f}_\xi(0)\right|\right) &\leq \frac{K}{na_n^2} E\left(\sum_{t=1}^n \left|\ln \tilde{\varepsilon}_t^2(\hat{\theta}_n) - \ln \hat{\varepsilon}_t^2(\hat{\theta}_n)\right|\right) \\ &= \frac{K}{na_n^2} E\left(\sum_{t=1}^n \left|\ln \tilde{h}_t(\hat{\theta}_n) - \ln h_t(\hat{\theta}_n)\right|\right) \\ &= \frac{K}{na_n^2} E\sum_{t=1}^n \ln\left(1 + \frac{\tilde{h}_t(\hat{\theta}_n) - h_t(\hat{\theta}_n)}{h_t(\hat{\theta}_n)}\right) \\ &\leq \frac{K}{na_n^2} E\left(\sum_{t=1}^n \left|\tilde{h}_t(\hat{\theta}_n) - h_t(\hat{\theta}_n)\right|\right) \\ &\leq \frac{K}{na_n^2}\sum_{t=1}^n E\sup_{\theta\in\Theta}\left|\tilde{h}_t(\theta) - h_t(\theta)\right| = O(n^{-1}a_n^{-2})O(2.1) = o(1).\end{aligned} \tag{7.66}$$

The third inequality follows since $\ln(1+x) \leq x$, the second equality results from Lemma 2(i), and condition (ii) implies the last equality. Therefore, by the Markov inequality $\hat{f}_\xi^*(0) \to_p \hat{f}_\xi(0)$.

Next, it follows by condition (i) and the mean-value theorem for some $\bar{\theta}_n \in (\hat{\theta}_n, \theta_0)$,

$$\begin{aligned} \hat{f}^*_\xi(0) - \hat{f}^\diamondsuit_\xi(0) &\leq \frac{K}{na_n^2} \sum\nolimits_{t=1}^n \left| \ln \hat{\varepsilon}_t^2(\hat{\theta}_n) - \eta_t \right| + \frac{K}{\sqrt{n}a_n^2} \left(\left| \sqrt{n}(\hat{m}_n - m_0) \right| \right) \\ &= \frac{K}{na_n^2} E\left(\sum\nolimits_{t=1}^n \left| \ln h_t(\hat{\theta}_n) - \ln h_t(\theta_0) \right| \right) + O(n^{-0.5} a_n^{-2}) O_P(1) \\ &\leq \frac{K}{na_n^2} \sum\nolimits_{t=1}^n \left| \frac{\dot{h}_t(\bar{\theta}_n)}{h_t(\bar{\theta}_n)} (\hat{\theta}_n - \theta_0) \right| + o_P(1) \\ &\leq \frac{K}{\sqrt{n}a_n^2} \left| \sqrt{n}(\hat{\theta}_n - \theta_0) \right| \cdot \frac{1}{n} \left(\sum\nolimits_{t=1}^n \sup_{\theta \in \Theta} \left| \frac{\dot{h}_t(\theta)}{h_t(\theta)} \right| \right) + o_P(1) \\ &= o(1) O_P(1) O_{a.s}(1) + o_P(1) = o_p(1). \end{aligned} \tag{7.67}$$

In the first summand, after the second equality, the first term follows from condition (ii), the second term by Theorem 4.2 and the third term by Lemma 3(i) and the ergodic theorem.

Now, since by the triangle inequality,

$$|\hat{f}_\xi(0) - f_\xi(0)| \leq |\hat{f}_\xi(0) - \hat{f}^*_\xi(0)| + |\hat{f}^*_\xi(0) - \hat{f}^\diamondsuit_\xi(0)| + |\hat{f}^\diamondsuit_\xi(0) - f_\xi(0)|$$

and the third term is $o_P(1)$ by conditions (i)-(ii) and Devroye et al. (1979), the desired result follows by (7.66) and (7.67).

iii) By using similar arguments as in the proof of Lemma 7(ii) we can show that $\hat{\bar{J}}_n \to_{a.s.} \bar{J}$. Further, it is established in the proof of theorem 4.3 that $\hat{J}_n^{-1} \to_{a.s.} J^{-1}$ and $\hat{\kappa}_n \to_{a.s.} \kappa$. Hence, the desired result follows directly by applying Lemmas 10(i)-(ii), Theorem 4.1 and Slutsky's theorem.

Proof of Theorem 4.6: The desired result follows directly by Lemma 10(iii), Theorem 4.5 and Slutsky's theorem.

Appendix C: proofs of theorems 4.7-4.9 and lemmata 11-13

Let $\bar{B}(\tau) = B(\tau) \cap [\beta, \bar{\beta}]$ where $B(\tau)$ is a closed ball of radius $\tau > 0$ centered at β_0. Let $\tilde{\theta}_n$ be a strongly consistent estimate for θ_0 and $\hat{\varepsilon}_t^2 = \tilde{h}_t^{-1}(\tilde{\theta}_n) y_t^2$, $\hat{\eta}_t = \ln \hat{\varepsilon}_t^2 - \hat{c}_n$. The estimates $\hat{c}_n$, $\hat{\varsigma}_n$, $\hat{\mu}_n$, $\hat{\kappa}_n$, $\hat{\bar{J}}_n$ in Lemmas 12-13 are calculated based on $\hat{\varepsilon}_t^2$, $\hat{\eta}_t$. Further, let $\Delta\theta = \theta - \theta_0$, $\hat{h}_t = h_t(\hat{\theta}_n)$, $\hat{\tilde{h}}_t = \tilde{h}_t(\hat{\theta}_n)$ and when the mean-value theorem is applied, let $\bar{\theta}_n$ be the point on the line joining $\tilde{\theta}_n$ and θ_0

Lemma 11: Under Assumptions A1-A3, $\left\| \sup_{\theta \in \bar{B}(\tau)} h_t^{-1}(\theta) h_{0t} \right\|_r < \infty$, for some $\tau > 0$ and all $r \geq 1$.

Proof: By (7.24) that $x/(1+x) < x^{p/r}$for all $x > 0$ and any $p \in (0,1)$, $r \geq 1$, we get

$$\begin{aligned} \sup_{\theta \in \bar{B}(\tau)} \frac{h_{0t}}{h_t(\theta)} &\leq \frac{\omega_0}{\omega(1-\beta_0)} + \sup_{\theta \in \bar{B}(\tau)} \frac{\alpha_0 \sum_{i=0}^\infty \beta_0^i y_{t-i-1}^2}{\omega + \alpha \sum_{i=0}^\infty \beta_2^i y_{t-i-1}^2} \\ &\leq K + \alpha_0 \sum\nolimits_{i=0}^\infty \sup_{\theta \in \bar{B}(\tau)} \left(\frac{\beta_0}{\beta} \right)^i \frac{\omega^{-1} \alpha \beta^i y_{t-i-1}^2}{1 + \omega^{-1} \alpha \beta^i y_{t-i-1}^2} \\ &\leq K + K \sum\nolimits_{i=0}^\infty (1+\tau)^i \, \bar{\beta}^{ip/r} |y_{t-i-1}^2|^{p/r}. \end{aligned} \tag{7.68}$$

Therefore, by Lemma 1(i), setting $\tau < (\bar{\beta}^{-p/r} - 1)$ and using the c_r and Minkowski inequalities we get

$$\left\| \sup_{\theta \in \bar{B}(\tau)} h_t^{-1}(\theta) h_{0t} \right\|_r \leq K + K \sum_{i=1}^{\infty} (1+\tau)^i \bar{\beta}^{ip/r} \left(E|y_{t-i-1}^2|^p \right)^{1/r} < \infty$$

Lemma 12: Under Assumptions A1-A4 and A6':
 i) $\hat{c}_n \to_{a.s.} c_0$.
 ii) $\hat{J}_n^{-1} \to_p J^{-1}$.
 iii) $\hat{\bar{J}}_n \to_p \bar{J}$.

Proof: Since $\tilde{\theta}_n \to_{a.s.} \theta_0$ for any $\tau > 0$ there exist a sufficiently large n such that $\tilde{\theta}_n \in \bar{B}(\tau)$ almost surely. Hence, in what follows, we assume that, we can apply Lemma 11 for any $r \geq 1$.

i) Define the function $G_t(\theta) = h_t^{-1}(\theta)\dot{h}_t(\theta) - h_{0t}^{-1}\dot{h}_{0t}$ and for the sake of brevity, $h_t(\tilde{\theta}_n), \tilde{h}_t(\tilde{\theta}_n)$ are abbreviated to $h_t, \tilde{h}_t$. First, $EG_t(\theta_0) = 0$ by the law of iterated expectation, $G_t(\theta)$ is stationary, ergodic and $E\sup_{\theta\in\Phi}|G_t(\theta)| < \infty$ by Lemmas 1(i) and 3(i). So, by using these results and the same kind of arguments as in the proof of Lemma 7(ii) we get,

$$n^{-1}\sum_{t=1}^{n} G_t(\tilde{\theta}_n) \to_{a.s.} 0. \tag{7.69}$$

Using again Lemmas 1(i), 3(i) we get by the ergodic theorem

$$n^{-1}\sum_{t=1}^{n} h_{0t}^{-1}\dot{h}_{0t} \to_{a.s.} \bar{J}. \tag{7.70}$$

Second, let $A_t = \sup_{\theta\in\Theta}(h_t - \tilde{h}_t)$ and note that$E|t^{0.5}A_t|^p = O(t^{0.5p}\bar{\beta}^{pt})$ for some $p \in (0,1)$ by Lemma 2(i), hence, the Borel-Cantelli lemma implies that $t^{-0.5}A_t$ is almost surely summable and by Kronecker's lemma

$$n^{-0.5}\sum_{t=1}^{n} \sup_{\theta\in\Theta}(h_t - \tilde{h}) \to_{a.s.} 0. \tag{7.71}$$

Third, note

$$\begin{aligned}\hat{c}_n = n^{-1}\sum_{t=1}^{n} \ln \hat{\tilde{h}}_t^{-1} y_t^2 &= n^{-1}\sum_{t=1}^{n} \ln h_{0t}^{-1} y_t^2 - n^{-1}\sum_{t=1}^{n} (\ln \hat{h}_t - \ln h_{0t}) \\ -n^{-1}\sum_{t=1}^{n} (\ln \hat{\tilde{h}}_t - \ln \hat{h}_t) &= n^{-1}\sum_{t=1}^{n} \ln \varepsilon_t^2 - n^{-1}\sum_{t=1}^{n} \Delta\tilde{\theta}_n' h_{0t}^{-1}\dot{h}_{0t} \\ -n^{-1}\sum_{t=1}^{n} \Delta\tilde{\theta}_n' G_t(\bar{\theta}_n) &+ n^{-1}\sum_{t=1}^{n} \ln \hat{h}_t \hat{\tilde{h}}_t^{-1}\end{aligned} \tag{7.72}$$

where the mean-value theorem is applied to the second term after the second equality. From (7.69)-(7.72) we get

$$\hat{c}_n - c_0 = n^{-1}\sum_{t=1}^{n} (\ln \varepsilon_t^2 - c_0) - \Delta\bar{\theta}_n'(\bar{J} + o_{a.s.}(1)) + o_{a.s.}(n^{-0.5}). \tag{7.73}$$

The desired result follows by applying Kolmogorov's SLLN and Theorem 4.1 for the first and second summands of (7.73), respectively.

ii) By using the same kind of arguments used to show (7.69)-(7.70) we can show that $\hat{J}_n \to_p J$, so the desired result follows by Slutsky's theorem.

iii) The result follows immediately by replacing $\hat{\theta}_n$ by $\tilde{\theta}_n$ in the Theorem 4.3 and using the same arguments.

Lemma 13: Under Assumptions A1-A4 and A7':
i) $\hat{\varsigma}_n \to_p \varsigma$.
ii) $\hat{\mu}_n \to_p \mu$.

iii) $\hat{\kappa}_n \to_p \kappa$.
iv) $\hat{\bar{\Xi}}_n \hat{\bar{V}}_n \hat{\bar{\Xi}}_n' \to_p \bar{\Xi}\bar{V}\bar{\Xi}'$.

Proof:

i) we observe that

$$\hat{\varepsilon}_t^2 - \varepsilon_t^2 = \hat{\hat{h}}_t^{-1} y_t^2 - h_{0t}^{-1} y_t^2 = (h_{0t}\hat{h}_t^{-1} - 1)\varepsilon_t^2 + \hat{\hat{h}}_t^{-1}\hat{h}_t^{-1}(\hat{h}_t - \hat{\hat{h}}_t) y_t^2 \tag{7.74}$$
$$= \Delta\tilde{\theta}_n' h_t^{-1}(\bar{\theta}_n)\dot{h}_t(\bar{\theta}_n) h_t^{-1} h_{0t}\varepsilon_t^2 + \hat{\hat{h}}_t^{-1}(\hat{h}_t - \hat{\hat{h}}_t) h_{0t}\hat{h}_t^{-1}\varepsilon_t^2 = A_{nt}^1 + A_{nt}^2$$

and

$$\hat{\eta}_t - \eta_t = \ln \hat{\hat{h}}_t^{-1} y_t^2 - \ln \varepsilon_t^2 - (\hat{c}_n - c_0) = -\ln \hat{h}_t h_{0t}^{-1} + \ln \hat{\hat{h}}_t^{-1}\hat{h}_t + o_{a.s.}(1) \tag{7.75}$$
$$= -\Delta\tilde{\theta}_n' h_t^{-1}(\bar{\theta}_n)\dot{h}_t(\bar{\theta}_n)\hat{h}_t^{-1}\varepsilon_t^2 + K(\hat{\hat{h}}_t - \hat{h}_t) + o_{a.s.}(1) = -A_{nt}^1 + A_{nt}^3 + o_{a.s.}(1)$$

where $K < \omega^{-1}$and the mean-value theorem is applied to the first term after the second equality in (7.74) and the equality on (7.75), respectively. By Assumption A6', Lemmas 1(iii), 3(i), 11 and Cauchy-Schwarz inequality, we can show that for some $\tau > 0$

$$E \sup_{\theta \in \bar{B}(\tau)} |A_{nt}^1 \varepsilon_t^2| \leq ||\Delta\tilde{\theta}_n||_2|| \sup_{\theta \in \bar{B}(\tau)} h_t^{-1}\dot{h}_t||_4|| \sup_{\theta \in \bar{B}(\tau)} h_t^{-1} h_{0t}||_4 = o(1)$$

hence by Markov inequality the first term after the fourth equality in (67) is $o_P(1)$. As to the second term, by using again Assumption A6', Lemma 11, Cauchy-Schwarz inequality with similar arguments used in Lemma 5, we can show that it converges almost surely to zero. So the terms after the last equality converge to zero in probability. The second equality in (68) follows by 12(iii) and using similar arguments, we can establish the same result for (7.75) as well. Next, note that

$$\hat{\varsigma}_n - \varsigma = n^{-1}\sum\nolimits_{t=1}^n \hat{\eta}_t\hat{\varepsilon}_t^2 - n^{-1}\sum\nolimits_{t=1}^n \eta_t\varepsilon_t^2 + o_{a.s.}(1) = n^{-1}\sum\nolimits_{t=1}^n (\hat{\eta}_t - \eta_t)(\hat{\varepsilon}_t^2 - \varepsilon_t^2) \tag{7.76}$$
$$+ n^{-1}\sum\nolimits_{t=1}^n \eta_t(\hat{\varepsilon}_t^2 - \varepsilon_t^2) + n^{-1}\sum\nolimits_{t=1}^n (\hat{\eta}_t - \eta_t)\varepsilon_t^2 + o_{a.s.}(1).$$

The first equality follows by Lemma 1(iii), Assumption A7' and Kolmogorov's SLLN. Since the expectation of $\eta_t^2$and ε_t^2 are finite, they are almost surely finite as well (Loeve 1977, p. 121), so by (7.73)-(7.75) all the terms after the second equality converge to zero in probability which implies the desired result.

ii) From (7.74), we get

$$\hat{\mu}_n - \mu = \hat{\mu}_n - n^{-1}\sum\nolimits_{i=1}^n (\varepsilon_t^4 - 1) = \hat{\mu}_n - \mu + o_{a.s.}(1) = n^{-1}\sum\nolimits_{i=1}^n (\hat{\varepsilon}_t^4 - \varepsilon_t^4) + o_{a.s.}(1)$$
$$= n^{-1}\sum\nolimits_{i=1}^n (\hat{\varepsilon}_t^2 - \varepsilon_t^2)(\hat{\varepsilon}_t^2 + \varepsilon_t^2) + o_{a.s.}(1) = n^{-1}\sum\nolimits_{i=1}^n o_p(1)(1 + 2\varepsilon_t^2) + o_{a.s.}(1) = o_p(1)$$

and by using similar arguments used in to part (i) of the lemma the desired result follows.

iii) The proof is similar to parts (i)-(ii) of the lemma, hence omitted.

iv) By using the results of Lemma 12 parts (i)-(ii) of the lemma with Slutsky's theorem we get the desired result.

Proof of Theorem 4.7: Let $\hat{z}_t(\hat{c}_n) = \ln y_t^2 - \hat{c}_n$ and

$$n^{-1}\sum\nolimits_{t=1}^n \sup_{\theta \in \Theta} \left| \tilde{\ell}_t(\theta, \hat{c}_n) - \tilde{\ell}_t(\theta, c_0) \right|$$

$$= n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|\left(\hat{z}_t(\hat{c}_n)-\hat{z}_t(c_0)\right)\left(\hat{z}_t(\hat{c}_n)+\hat{z}_t(c_0)-2\ln h_t(\theta)\right)\right|$$

$$\leq |\hat{c}_n - c_0|\left|\left(2\sup_{\theta\in\Theta}\ln h_t(\theta)+c_0+o_{a.s.}(1)-2\ln y_t^2\right)\right|$$

$$\leq o_{a.s.}(1)\left(1+n^{-1}\sum_{i=1}^{n}(\sup_{\theta\in\Theta}|\ln h_t(\theta)|+|\ln\varepsilon_t^2|\right)=o_{a.s.}(1)O_{a.s.}(1)=o_{a.s.}(1).$$

The first term in the product after the last equality results from Lemma 12(iii). The second term in the product converges almost surely by Lemmas 1(ii)-(iii) and 3(iv). Thus,

$$\sup_{\theta\in\Theta}|\tilde{Q}_n(\theta,\hat{c}_n)-Q_n(\theta)|=o_{a.s.}(1)$$

and by using similar arguments used in Theorem 4.1, the desired result follows.

Proof of Theorem 4.8: Let $\nabla\tilde{\ell}(\theta,c)=\partial\tilde{\ell}(\theta,c)/\partial\theta$ and $\nabla_c\tilde{\ell}(\theta,c)=\partial\nabla\tilde{\ell}(\theta,c)/\partial c$. First, note that

$$\tfrac{1}{n}\sum_{t=1}^{n}\nabla_c\tilde{\ell}_t(\theta,c)=\tfrac{1}{n}\sum_{t=1}^{n}\dot{\tilde{h}}_t\tilde{h}_t^{-1}=\tfrac{1}{n}\sum_{t=1}^{n}\dot{h}_t h_t^{-1}+\tilde{h}_t^{-1}(\dot{\tilde{h}}_t-\dot{h}_t)+\tilde{h}_t^{-1}h_t^{-1}\dot{h}_t(h_t-\tilde{h}_t)$$

and Lemmas 1(i), 2(i), 2(ii), 3(i), the ergodic theorem and Theorem 4.1 imply

$$\tfrac{1}{n}\sum_{t=1}^{n}\nabla_c\tilde{\ell}_t(\hat{\theta}_n^{\mathrm{LS}},\bar{c}_n)=\bar{J}+o_{a.s.}(1). \tag{7.77}$$

Second by Assumption A7' and similar arguments used in Lemma 6(ii), we can show

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\sum_{t=1}^{n}\eta_t\\ J^{-1}\sum_{t=1}^{n}(\varepsilon_t^2-1)\dot{h}_{0t}h_{0t}^{-1}\\ \sum_{t=1}^{n}\eta_t\dot{h}_{0t}h_{0t}^{-1}\end{pmatrix}\overset{d}{\to}\begin{pmatrix}Z_c\\ Z_\theta^{QML}\\ Z_\theta^{LS}\end{pmatrix}=Z\sim N\left\{0,\,\bar{V}\right\} \tag{7.78}$$

where

$$\bar{V}:=\begin{pmatrix}\kappa & \varsigma\bar{J}'J^{-1} & \kappa\bar{J}'\\ \varsigma J^{-1}\bar{J} & \mu J^{-1} & \varsigma I_3\\ \kappa\bar{J} & \varsigma I_3 & \kappa J\end{pmatrix}.$$

Recall that under Assumptions A1-A4 and A7', $\sqrt{n}(\hat{\theta}_n^{QML}-\theta_0)=J^{-1}\sum_{t=1}^{n}(\varepsilon_t^2-1)\dot{h}_{0t}h_{0t}^{-1}+o_p(1)$, (see e.g. Francq and Zakoian (2004)) , hence by (7.78) and setting $\hat{\theta}_n^{QML}=\tilde{\theta}_n$ in (7.73) we get

$$\sqrt{n}(\hat{c}_n-c_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\ln\varepsilon_t^2-c_0\right)-\bar{J}'\sqrt{n}(\hat{\theta}_n^{QML}-\theta_0)+o_P(1)\to_D Z_c-\bar{J}'Z_\theta^{QML}. \tag{7.79}$$

Third, note that$\nabla\tilde{\ell}(\theta,c_0)=\nabla\tilde{\ell}_t(\theta)$ and $\nabla^2\tilde{\ell}(\theta,c_0)=\nabla^2\tilde{\ell}_t(\theta)$. Next, by using a mean value expansion of the gradient of $\tilde{Q}_n(\theta,c)$ with respect to θ around $\hat{\theta}_n^{LS}$ and $\hat{c}_n$ respectively, yields

$$\begin{aligned}0&=n^{-0.5}\sum_{t=1}^{n}\nabla\tilde{\ell}(\hat{\theta}_n^{\mathrm{LS}},\hat{c}_n) && (7.80)\\ &=n^{-0.5}\sum_{t=1}^{n}\nabla\tilde{\ell}(\hat{\theta}_n^{\mathrm{LS}},c_0)+\left(\tfrac{1}{n}\nabla_c\tilde{\ell}(\hat{\theta}_n^{\mathrm{LS}},\bar{c}_n)\right)\sqrt{n}(\hat{c}_n-c_0)\\ &=n^{-0.5}\sum_{t=1}^{n}\nabla\tilde{\ell}(\theta_0,c_0)+\tfrac{1}{n}\sum_{t=1}^{n}\nabla^2\tilde{\ell}(\bar{\theta}_n^{\mathrm{LS}},c_0)\sqrt{n}(\hat{\theta}_n^{\mathrm{LS}}-\theta_0)+\left[\bar{J}+o_{a.s.}(1)\right]\sqrt{n}(\hat{c}_n-c_0)\\ &=-n^{-0.5}\sum_{t=1}^{n}\eta_t\dot{h}_{0t}h_{0t}^{-1}+\left[J+o_{a.s.}(1)\right]\sqrt{n}(\hat{\theta}_n^{\mathrm{LS}}-\theta_0)+\bar{J}\sqrt{n}(\hat{c}_n-c_0)+o_p(1) && (7.81)\end{aligned}$$

where $\bar{\theta}_n^{\mathrm{LS}}$ and $\bar{c}_n$ are points on the lines joining $\hat{\theta}_n^{\mathrm{LS}}$ with θ_0 and $\hat{c}_n$ with c_0, respectively. The third and fourth equalities follow from (7.45), (7.77) and arguments already given in Theorem 4.2 above. Hence by (7.78)-(7.80),

$$\sqrt{n}(\hat{\theta}_n^{\text{LS}} - \theta_0) = -J^{-1} \begin{pmatrix} \bar{J} & -I_3 \end{pmatrix} \begin{pmatrix} \sqrt{n}(\hat{c}_n - c_0) \\ n^{-0.5} \sum_{t=1}^{n} \eta_t \dot{h}_{0t} h_{0t}^{-1} \end{pmatrix} + o_p(1)$$

$$= -J^{-1} \begin{pmatrix} \bar{J} & -I_3 \end{pmatrix} \begin{pmatrix} 1 & -\bar{J}' & 0 \\ 0 & 0 & I_3 \end{pmatrix} \begin{pmatrix} Z_c \\ Z_\theta^{QML} \\ Z_\theta^{LS} \end{pmatrix} \xrightarrow{d} \bar{\Xi} Z$$

which implies the desired result.

Proof of Theorem 4.9: the desired result follows directly by Lemma 13(vi), Theorem 4.8 and Slutsky's theorem.

REFERENCES

Andrews B. (2012). Rank based estimation for GARCH processes. *Econometric Theory*, 28(5), 1037-1064.

Andrews D.W.K. (1994). Empirical process methods in econometrics, in *Handbook of Econometrics*, Vol. 4, Ch.37, 2247-2294.

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

Berkes I. Horvath L. and P.S. Kokoszka (2003). GARCH processes: structure and estimation. *Bernoulli*, 9, 201-227.

Berkes I. and L. Horvath (2003). The rate of consistency of the quasi-maximum likelihood estimator. *Statistics and Probability Letters*, 61, 133-143.

Billingsley P. (1968) *Convergence of Probability Measures.* New York, John Wiley.

Billingsley P. (1995). *Probability and Measure.* New York, John Wiley.

Bollerslev T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

Bollerslev T. and J.M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11, 143-172.

Boussama, F. (1998) Ergodicity, Mixing and Estimation in GARCH Models, Ph.D. thesis, University of Paris.

Boussama F. (2000). Normalité asymptotique de l'estimateur du pesudo-maximum de vraisemblance d'un mode'le GARCH, *C.R.Acad Sci. Paris*, 331, 81-84.

Chen L.H, M-H. Cheng and L. Peng (2009). Conditional variance estimation in heteroscedastic regression models. *Journal of Statistical Planning and Inference*, 130, 236-245.

Christoffersen, P. (1998). Evaluating Interval Forecasts, *International Economic Review*, 39, 841–862.

Dacorogna M.M., Gencay R., Muller U., Olsen R.B. and O.V. Pictet (2001). *An Introduction to High Frequency Finance.* San Diego, CA: Academic Press.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press,

Devroye, L.P. and T.J. Wagner (1979) The L_1 convergence of kernel density estimates, *The Annals of Statistics*, 7(5), 1136–1139.

Ding, Z. and C. W. J. Granger (1996). Modelling volatility persistence of speculative returns: a new approach, *Journal of Econometrics*, 73, 185–215.

Doukhan, P., Massart, P. and E. Rio (1995). Invariance principle for absolutely regular empirical processes. *Annales de I'Institute H. Poincare*, 31, 393-427.

Engle, R. F. (2002). Dynamic Conditional Correlation, *Journal of Business & Economic Statistics*,20(3), 339–350.

Engle, R.F. and G. Gonzalez-Rivera (1991). Semiparametric ARCH Models, *Journal of Business and Economic Statistics*, 9(4), 345-359.

Engle, R. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles, *Journal of Business and Economic Statistics*, 22(4), 367–381.

Fan, J., L. Qi and D. Xiu (2014). Quasi Maximum Likelihood estimation of GARCH models with heavy-tailed Likelihoods, *Journal of Business and Economic Statistics*, 32(2), 178–191.

Francq, C. and J-M. Zakoian (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, *Bernoulli*, 10(4), 605–637.

Francq C. and J-M Zakoian (2006a). Linear-representation based estimation of stochastic volatility models, *Scandinavian Journal of Statistics*, 33(4), 785–806.

Francq, C. and J-M. Zakoian (2006b). Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process, *Econometric Theory*, 22(5), 815-834.

Francq C. and J-M. Zakoian (2007). Quasi-likelihood inference in GARCH processes when some coefficients are equal to zero. *Stochastic Processes and their Applications*, 117(9), 1265-1284.

Francq, C., Lepage, G. and Zakoian J.-M. (2011). Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics*, 165(1), 246-257.

Hall P. and Q. Yao (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, 71, 285-317.

Hansen P.R and A.Lunde (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873-889.

Harvey A.C., Ruiz E. and N.G. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61, 247-264.

Hill, J.B. (2013) Least tail-trimmed squares for infinite variance autoregressions, *Journal of Time Series*, 34, 168-186.

Huang D., Wang H. and Q. Yao (2008). Estimating GARCH models: when to use what? *Econometrics Journal*, 11, 27–38

Lee S.W. and B.E. Hansen (1994). Asymptotic theory for the GARCH(1,1) quasi- maximum likelihood estimator. *Econometric Theory*, 10, 29-52.

Lee, T. and Lee S. (2009). Normal mixture quasi-maximum likelihood estimator for GARCH. models. *Scandinavian Journal of Statistics*, 36 157-170.

Lee, S. and Noh, J. (2013) Quantile regression estimator for GARCH. *Scandinavian Journal of Statistics*, 40(1), 2-20.

Lighthill, M.J. (1958) Introduction to Fourier Analysis and Generalized Functions, Cambridge Univ. Press, Cambridge.

Linton O, Pan J. and H Wang (2010). Estimation for A non-stationary semi-strong GARCH(1,1) models with heavy-tailed errors. *Econometric theory*, 26(2.1), 1-28.

Loeve M. (1977). *Probability Theory 1*, New York, Springer.

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, 575-596.

Mikosch, T., Straumann, D.(2006). Stable limits of martingale transforms with application to the estimation of GARCH parameters. *The Annals of Statistics*, 34(1), 493–522.

Mittnik S. and S.T. Rachev (2000) *Stable Paretian Models in Finance*. New York, John-Wiley.

Mukherjee K. (2008). M-estimation in GARCH models. *Econometric Theory*, 24(2.6), 1530–1553.

Nelson D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(2.3), 318-334.

Newey, W. K.and Powell, J. L. (1987) Asymmetric Least Squares Estimation and Testing, *Econometrica*, 55, 819–847.

Newey, W.K. and Steigerwald, D. G. (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models, *Econometrica*, 65, 587–599.

Pan, J., Wang, H. and Tong, H. (2008) Estimation and tests for power-transformed and threshold GARCH models, *Journal of Econometrics*, 142, 352–378.

Patton A. (2011) Volatility forecast evaluation and comparison using imperfect volatility proxies. *Journal of Econometrics*, 160, 246–256.

Peng, L. and Q. Yao (2003). Least absolute deviation estimation for ARCH and GARCH models, *Biometrika*, 90, 967-975.

Philips, P.C.B. (1995) Robust nonstationary regression, *Econometric Theory*, 11, 912-951.

Robinson P.M. and P. Zaffaroni (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models. *Annals of Statistics*, 34, 1049–1074.

Ruiz E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63, 289–306.

Sakata S. and H. White (2001). S-estimation of nonlinear regression models with dependent and heterogeneous observations. *Journal of Econometrics*, 103, 5-72.

Stinchcombe M. B. and H. White (1992). Some measurability results for extrema of random functions over random sets, *Review of Economic Studies*, 59(3), 495-514.

Stout, W. F. (1974). *Almost Sure Convergence*, Academic Press, New York.

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

White H. (1994). *Estimation, Inference and Specification Analysis*. New York, Cambridge University Press.

series	starts	ends	n. obs.
S&P500	3/1/1990	15/3/2015	6349
JPM	3/1/1990	15/3/2015	6349
USD/EUR	2/1/2002	20/3/2015	3448

Table 1. Observation periods and time series lengths for the S&P500, JP Morgan and USD/EUR exchange rate series.

	S&P500			JPM		
p	$\hat{p}$	LR_{cc}	DQ	$\hat{p}$	LR_{cc}	DQ
0.05	0.0550	0.2188	0.2549	0.0525	0.2727	0.2965
0.025	0.0288	0.1679	0.2228	0.0258	0.7090	0.2917
0.01	0.0125	0.0870	0.0779	0.0097	0.2669	0.1070
	USD/EUR			$t_{2.1}$		
p	$\hat{p}$	LR_{cc}	DQ	$\hat{p}$	LR_{cc}	DQ
0.05	0.0503	0.5310	0.6761	0.0500	0.4514	0.5772
.025	0.0278	0.5093	0.6073	0.0252	0.9419	0.6694
0.01	0.0110	0.6521	0.7227	0.0104	0.6035	0.3165

Table 2. VaR prediction for S&P 500, JP Morgan, USD/EUR exchange rate and a simulated series from a GARCH(1,1) with $t_{2.1}$ errors: empirical coverages ($\hat{p}$), p-values of the likelihood ratio conditional coverage (LR_{cc}) and dynamic quantile (DQ) tests.

	ω_0	α_0	β_0
I	0.1	0.1	0.90
H	0.01	0.09	0.90
M	0.1	0.10	0.80
L	0.2	0.20	0.60

Table 3. Volatility models used for the simulation study: integrated (I), high persistence (H), medium persistence (M), low persistence (L).

	I			H			M			L		
	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0
T=500												
GQMLE	151.890	**1.379**	**3.238**	**1.289**	**1.451**	**5.629**	**5.148**	**2.093**	**14.168**	**4.628**	**2.916**	**13.615**
$NGQMLE_7$	404.723	1.727	5.946	8.783	2.170	29.044	14.232	2.289	32.491	6.101	3.043	16.787
$NGQMLE_4$	411.031	1.832	6.129	7.606	1.804	24.029	14.186	2.357	33.091	6.095	3.123	16.730
LSE_Q	**19.770**	2.106	5.327	4.348	2.287	15.896	11.049	3.424	28.053	7.416	4.265	21.001
LSE_0	40.052	2.377	6.798	4.685	2.272	14.885	11.952	3.429	30.113	7.442	4.255	21.116
T=1000												
GQMLE	14.097	**0.897**	**1.860**	**0.407**	0.884	**2.488**	**3.333**	**1.360**	**9.082**	**2.947**	**2.042**	**8.737**
$NGQMLE_7$	356.278	1.386	5.040	0.471	**0.861**	2.578	7.551	1.449	17.975	3.272	2.052	9.453
$NGQMLE_4$	372.110	1.367	5.188	0.475	0.891	2.579	7.304	1.510	17.432	3.686	2.132	10.321
LSE_Q	**11.425**	1.299	3.129	0.803	1.280	4.413	6.802	2.095	17.616	5.026	3.088	14.644
LSE_0	18.033	1.422	3.424	0.816	1.331	4.672	7.562	2.085	19.658	5.074	3.065	14.670
T=2000												
GQMLE	**4.431**	**0.645**	**1.238**	**0.215**	**0.647**	**1.513**	**1.609**	**0.966**	**4.879**	**1.944**	**1.494**	**6.043**
$NGQMLE_7$	5.170	0.724	1.490	0.242	0.664	1.675	2.664	0.979	7.027	2.048	1.523	6.292
$NGQMLE_4$	5.503	0.762	1.573	0.264	0.692	1.839	2.566	1.018	7.078	2.150	1.574	6.590
LSE_Q	5.993	0.911	1.995	0.376	0.981	2.680	3.075	1.482	8.845	3.260	2.225	10.097
LSE_0	6.561	0.944	2.108	0.369	0.988	2.678	3.311	1.457	9.150	3.309	2.226	10.199
T=5000												
GQMLE	**1.631**	**0.432**	**0.779**	**0.111**	**0.413**	**0.915**	**0.914**	**0.614**	**2.849**	**1.239**	**0.974**	**3.783**
$NGQMLE_7$	1.706	0.472	0.920	0.119	0.620	1.195	0.956	0.626	3.007	1.299	0.992	3.993
$NGQMLE_4$	1.774	0.494	0.969	0.123	0.431	1.005	0.995	0.651	3.126	1.358	1.027	4.180
LSE_Q	2.378	0.607	1.213	0.175	0.592	1.433	1.486	0.940	4.660	1.909	1.409	6.034
LSE_0	2.391	0.613	1.241	0.175	0.594	1.438	1.472	0.937	4.615	1.911	1.409	6.029

Table 4. Simulated RMSEs under standardized Normal errors.

	I			H			M			L		
	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0
T=500												
GQMLE	44.222	4.122	63.203	0.973	3.832	18.416	4.319	4.011	23.469	4.897	6.279	28.187
$NGQMLE_7$	28.828	2.672	44.229	3.518	2.575	58.148	11.562	2.632	53.244	6.017	3.018	28.349
$NGQMLE_4$	25.320	2.455	40.011	3.245	2.358	53.857	10.957	2.346	50.480	5.527	2.712	26.291
LSE_Q	7.709	1.193	12.451	0.887	1.121	15.404	4.370	1.500	20.973	3.141	2.107	16.283
LSE_0	**7.183**	**0.998**	**11.611**	**0.762**	**1.039**	**13.530**	**4.035**	**1.476**	**19.490**	**3.048**	**2.106**	**15.938**
T=1000												
GQMLE	4.851	2.664	9.547	0.703	2.825	14.076	3.465	2.400	18.446	3.898	3.583	21.408
$NGQMLE_7$	21.690	1.629	32.949	2.751	1.616	45.848	9.110	1.687	42.012	3.442	1.945	17.012
$NGQMLE_4$	18.421	1.433	28.187	2.301	1.439	38.648	8.489	1.523	39.105	3.158	1.741	15.562
LSE_Q	2.439	0.660	4.153	0.184	0.660	3.569	1.892	0.923	9.561	1.855	**1.374**	9.781
LSE_0	**2.276**	**0.645**	**3.896**	**0.153**	**0.642**	**3.066**	**1.607**	**0.912**	**8.368**	**1.822**	**1.374**	**9.660**
T=2000												
GQMLE	33.517	2.519	46.928	0.229	1.102	4.949	2.562	1.873	13.701	2.812	2.495	15.299
$NGQMLE_7$	10.228	0.969	15.758	2.512	1.112	42.136	6.186	1.080	28.649	1.929	1.334	9.909
$NGQMLE_4$	8.083	0.912	13.033	1.523	0.880	25.735	6.065	0.964	28.079	1.718	1.191	8.820
LSE_Q	0.964	0.466	1.930	0.096	0.456	2.076	1.024	0.620	5.335	1.177	**0.941**	6.290
LSE_0	**0.805**	**0.441**	**1.667**	**0.088**	**0.452**	**1.947**	**0.972**	**0.615**	**5.109**	**1.168**	**0.941**	**6.258**
T=5000												
GQMLE	7.174	1.149	11.096	0.126	0.717	2.948	1.420	1.253	7.735	1.772	1.896	9.737
$NGQMLE_7$	1.639	0.469	2.873	2.168	0.757	36.651	2.422	0.572	11.433	1.094	0.860	5.820
$NGQMLE_4$	2.082	0.423	3.418	0.497	0.439	8.542	2.410	0.521	11.363	0.977	0.774	5.213
LSE_Q	0.447	**0.279**	0.972	0.049	0.281	1.128	0.538	**0.387**	2.892	0.744	0.613	4.005
LSE_0	**0.432**	**0.279**	**0.950**	**0.047**	**0.279**	**1.099**	**0.530**	0.389	**2.864**	**0.740**	**0.612**	**3.985**

Table 5. Simulated RMSEs under standardized t_5 errors.

	I			H			M			L		
	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0
T=500												
GQMLE	21.567	6.773	73.118	0.966	15.114	42.281	3.955	16.093	41.789	3.320	15.521	34.965
NGQMLE_7	17.678	2.178	60.837	1.874	2.188	66.747	7.111	2.387	60.811	4.309	2.545	36.561
NGQMLE_4	17.620	1.872	60.820	1.700	1.879	60.855	6.664	1.942	57.000	3.737	2.108	31.910
LSE_Q	7.272	1.108	25.606	0.674	1.045	24.758	2.932	1.191	25.735	1.896	1.499	17.038
LSE_0	**5.511**	**0.869**	**19.593**	**0.578**	**0.956**	**21.569**	**2.582**	**1.155**	**22.928**	**1.802**	**1.489**	**16.391**
T=1000												
GQMLE	7.583	9.614	31.286	0.807	13.194	37.982	3.571	11.285	36.244	3.026	11.270	31.056
NGQMLE_7	13.933	1.699	48.283	1.632	3.841	59.134	6.654	1.418	56.590	3.540	1.747	30.613
NGQMLE_4	13.365	1.196	46.184	1.487	1.272	53.773	5.696	1.218	48.643	2.587	1.432	22.639
LSE_Q	2.641	0.543	9.558	0.277	0.470	10.136	1.513	0.729	13.458	1.198	1.026	11.187
LSE_0	**2.168**	**0.510**	**7.999**	**0.258**	**0.440**	**9.242**	**1.277**	**0.714**	**11.548**	**1.155**	**1.024**	**10.880**
T=2000												
GQMLE	18.893	3.925	63.356	0.675	7.129	27.926	2.866	6.547	27.744	2.615	6.127	25.620
NGQMLE_7	10.205	0.828	35.302	1.481	0.951	53.057	6.116	0.947	52.061	2.862	1.184	24.780
NGQMLE_4	11.432	0.767	39.863	1.122	0.691	40.381	4.449	0.751	37.939	1.530	0.961	13.569
LSE_Q	0.711	0.372	2.866	0.055	0.329	2.311	0.618	0.458	5.713	0.776	0.716	7.291
LSE_0	**0.437**	**0.313**	**1.842**	**0.047**	**0.317**	**2.042**	**0.578**	**0.457**	**5.410**	**0.765**	**0.715**	**7.212**
T=5000												
GQMLE	6.717	31.916	42.771	0.382	4.125	17.214	2.010	3.589	19.686	1.953	4.715	19.611
NGQMLE_7	3.022	0.343	10.445	1.413	0.817	51.030	5.771	0.636	49.502	2.315	0.738	20.068
NGQMLE_4	2.521	0.294	8.785	0.719	0.359	26.077	2.643	0.426	22.655	0.621	0.605	5.870
LSE_Q	0.237	0.195	1.025	0.029	0.195	1.257	0.323	0.280	3.069	**0.442**	**0.451**	4.217
LSE_0	**0.226**	**0.193**	**0.990**	**0.027**	**0.192**	**1.199**	**0.312**	**0.278**	**2.983**	**0.442**	**0.451**	**4.216**

Table 6. Simulated RMSEs under standardized t_3 errors.

	I			H			M			L		
	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0	ω_0	α_0	β_0
T=500												
GQMLE	**1.051**	56.344	71.515	**0.103**	55.251	70.074	**0.451**	54.067	62.249	**0.473**	54.325	49.396
NGQMLE_7	2.420	2.204	74.789	0.233	2.577	72.837	1.020	1.933	64.660	0.715	2.060	45.431
NGQMLE_4	2.455	1.565	75.731	0.237	1.573	73.654	1.036	1.577	65.617	0.727	1.586	46.077
LSE_Q	1.827	1.103	56.426	0.184	1.073	57.109	0.780	1.144	**49.465**	0.562	1.153	**35.737**
LSE_0	1.711	**1.099**	**52.954**	0.173	**1.071**	**53.894**	0.735	**1.141**	46.668	0.563	**1.151**	35.818
T=1000												
GQMLE	**1.041**	60.312	75.623	**0.110**	60.147	75.930	**0.502**	58.959	67.406	**0.466**	57.691	50.861
NGQMLE_7	2.297	1.909	70.997	0.222	1.310	69.225	0.976	1.324	61.600	0.672	1.559	42.526
NGQMLE_4	2.317	1.595	71.579	0.224	1.064	69.770	0.990	0.880	62.355	0.679	1.127	42.881
LSE_Q	1.636	**0.648**	50.583	0.170	0.701	52.897	0.728	0.627	45.882	0.518	0.790	32.790
LSE_0	1.488	1.141	**46.123**	0.156	**0.697**	**48.486**	0.679	**0.618**	**42.886**	0.513	**0.785**	**32.523**
T=2000												
GQMLE	1.143	61.608	78.657	0.115	62.496	79.816	**0.502**	60.361	69.320	0.412	60.615	51.134
NGQMLE_7	2.018	0.902	62.282	0.201	0.737	62.359	0.884	1.129	55.809	0.579	0.780	36.552
NGQMLE_4	2.028	0.645	62.530	0.201	0.620	62.355	0.890	0.576	56.083	0.576	0.587	36.309
LSE_Q	1.269	0.416	39.150	0.128	0.374	39.814	0.585	0.391	36.950	0.387	0.410	24.469
LSE_0	**1.110**	**0.407**	**34.297**	**0.113**	**0.366**	**35.124**	**0.533**	**0.387**	**33.666**	**0.369**	0.409	**23.393**
T=5000												
GQMLE	1.085	62.962	80.026	0.115	61.688	79.555	0.463	62.586	70.915	0.383	62.748	52.175
NGQMLE_7	1.519	0.441	47.090	0.152	0.386	47.062	0.687	0.360	43.261	0.418	0.402	26.423
NGQMLE_4	1.495	0.347	46.326	0.150	0.315	46.602	0.683	0.288	42.980	0.405	0.310	25.581
LSE_Q	0.580	0.171	18.009	0.064	0.133	19.713	0.343	0.176	21.624	0.224	**0.212**	14.231
LSE_0	**0.492**	**0.164**	**15.310**	**0.053**	**0.125**	**16.469**	**0.298**	**0.173**	**18.837**	**0.216**	**0.212**	**13.683**

Table 7. Simulated RMSEs under standardized $t_{2.1}$ errors.