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

Preminger, Arie and Storti, Giuseppe

Ben Gurion University of the Negev, Beer-Sheva, Israel., University of Salerno, Italy

17 January 2014

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

Arie Preminger † Giuseppe Storti ‡

October 4, 2014

Abstract

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 log-transform-based least squares estimator (LSE) for the GARCH (1,1) model. The asymptotic properties of the LSE are studied under very mild moment conditions for the errors. We establish the consistency, asymptotic normality at the standard convergence rate of $\sqrt{n}$ for our estimator. The finite sample properties are assessed by means of an extensive simulation study. Our results show that LSE is more accurate than the quasi-maximum likelihood estimator (QMLE) for heavy tailed errors. Finally, we provide some empirical evidence on two financial time series considering daily and high frequency returns. The results of the empirical analysis suggest that in some settings, depending on the specific measure of volatility adopted, the LSE can allow for more accurate predictions of volatility than the usual Gaussian QMLE.

JEL Classification: C13, C15, C22.

Keywords: GARCH (1,1), least squares estimation, consistency, asymptotic normality.

*The authors thank Christian M. Hafner, for helpful comments. 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).

†Ben Gurion University of the Negev, Beer-Sheva, Israel.
E-mail: `ariepr@bgu.ac.il`

‡Department of Economics and Statistics, Università di Salerno, 84084 Fisciano, Italy.
E-mail: `storti@unisa.it`

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 quasi-maximum likelihood estimator (QMLE) 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 (2009). 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 (Hall and Yao (2003), 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 $\sqrt{n}$. Straumann (2005) established similar results for a more general class of GARCH type models.

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 tailed errors. 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. The finite sample efficiency of the LSE is then assessed by means of a simulation study considering different error distributions as well as different persistence levels of the volatility process. The results suggest that the LSE can be more efficient than the Gaussian QMLE (GQMLE) in the following cases: i) in the presence of heavy tailed or skewed error distributions ii) when the volatility persistence is close to unity. It is important to note that both these features typically occur in the analysis of financial time series.

The paper also presents an empirical application to financial data whose aim is to evaluate the ability of the LSE to adequately reproduce the volatility dynamics of some commonly encountered classes of asset returns. To cover a wide range of features typically arising in financial applications, we consider two different datasets characterized by substantially different volatility patterns namely the daily log-returns on the S&P 500 stock market index and the 30 minutes log-returns on the US dollar/Swiss franc (USD/CHF) exchange rate. The results indicate that the LSE can produce more accurate predictions of volatility than the usual GQMLE. Further, in order to investigate if the LSE is able to adequately characterize the stochastic structure of the two datasets analyzed, we compare the theoretical autocorrelation functions of squared returns implied by the estimated volatility models to their sample counterparts. In both cases the results are compared with those yielded by the GQMLE.

The structure of the paper is as follows. In Section 2 we discuss the LSE and derive its asymptotic properties. In Section 3 we conduct a simulation study aiming at investigating the small sample properties of the estimator, while the results of an application of the proposed estimation approach to two financial time series are presented in Section 4. Section 5 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. The symbol $\underset{D}{\rightarrow}$ denotes converges in distribution. The symbol $\underset{a.s.}{\rightarrow}$ $(\underset{p}{\rightarrow})$ denotes convergence almost surely (in probability). $o_{a.s.}(1)$ denotes a series of random variables that converges to zero almost surely (a.s.).

2 Least squares estimation for the GARCH (1,1) model

The standard GARCH (1,1) model as proposed by Bollerslev (1986) is given by

$$y_t = \sqrt{h_{0t}}\varepsilon_t \tag{1}$$

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

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

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 history of the system. 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 assumptions below. By squaring the terms in (1) and taking the logarithm we obtain

$$z_t = \ln(h_{0t}) + \eta_t \tag{3}$$

where $z_t = \ln(y_t^2) - c_0$ and $\eta_t = \ln(\varepsilon_t^2) - c_0$ are zero mean iid random variables. This nonlinear regression can be estimated via a least squares estimation. Conditional on some initial positive value $\tilde{h}_1$(e.g. $\tilde{h}_1 = \omega$), the objective function is given by

$$\tilde{Q}_n(\theta) = \tfrac{1}{2n} \textstyle\sum_{t=1}^n \tilde{\ell}_t(\theta) = \tfrac{1}{2n} \sum_{t=1}^n (z_t - \ln \tilde{h}_t(\theta))^2 \tag{4}$$

where $\theta = (\omega, \alpha, \beta)'$ and $\tilde{h}_t(\theta)$ is defined recursively, for $t \geq 2$ by

$$\tilde{h}_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta \tilde{h}_{t-1} \tag{5}$$

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

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \tilde{Q}_n(\theta) \tag{6}$$

where $\Theta \subset (0, \infty) \times [0, \infty]^2$. It will also be convenient to work with $h_t(\theta)$ the unobserved conditional variance

$$h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta) \tag{7}$$

where h_1 is initialised from its stationary distribution. Note that $h_{0t} = h_t(\theta_0)$ and $\tilde{h}_{0t} = \tilde{h}_t(\theta_0)$. For the unobserved process we construct the following unobserved objective function

$$Q_n(\theta) = \tfrac{1}{2n} \textstyle\sum_{t=1}^n (z_t - \ln h_t(\theta))^2 = \tfrac{1}{2n} \sum_{t=1}^n \ell_t(\theta) \tag{8}$$

The primary difference between the two objective functions is that $Q_n(\theta)$ is computed as if we had a sample containing the infinite past observations. In practice, we can only use (4) for estimation. It will be shown that the choice of the initial values does not matter for the asymptotic properties of the LSE.

To show the strong consistency, the following assumptions will be made.

Assumptions

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

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

Remark 1: The first assumption allows for the possibility that the process is a pure ARCH or even an iid. process. Nelson (1990) showed that Assumption A2 is sufficient and necessary for strict stationarity of (1) and (2). 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 $\alpha_0 + \beta_0 \leq 1$. Thus, we are allowing for the possibility of mildly explosive GARCH, in addition to integrated GARCH. However, this conclusion does not necessarily hold if ε_t has infinite second moment. 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. 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. This condition is necessary for both consistency and asymptotic normality. A similar condition also appears in Berkes et al.(2003). Assumptions A3 and A4 imply that z_t, η_t are finite a.s. and the scaling factor c_0 is finite (see Lemma 1(iii) in the Appendix for details).

Remark 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. Our first result is given as follows.

Theorem 1: Under Assumptions A1-A4, $\hat{\theta}_n \underset{a.s.}{\rightarrow} \theta_0$

The next theorems establish the asymptotic normality for our estimator. For GQMLE the former 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 denote the interior of Θ.

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

Remark 4: Let $\hat{J}_t$ and $\hat{\eta}_t^2$ 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 $\hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_t^2 \hat{J}_t$, is a strongly consistent estimate of Ω. Further, for the QMLE, it was shown that the covariance matrix estimate converges in probability to the true quantity (see e.g. Francq and Zakoian (2009)). It is worth nothing that the methods used in the Appendix can easily be applied to prove almost sure convergence to the true asymptotic covariance matrix also in the context of quasi-likelihood estimation.

Remark 5: An important use of the asymptotic normality shown in Theorem 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

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

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 3*:* Under Assumptions A1-A5, $W_n \underset{D}{\rightarrow} \chi_k^2$,

Remark 6: Other scale measures can be used as 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 (LADE) was proposed by Peng and Yao (2003). They showed that the LADE is locally asymptotically Gaussian with convergence rate $\sqrt{n}$ provided that the second moment of the error term is finite (see also Huang et al. (2008)). Another more general class of scale measures is the "regular scale about the origin", introduced by Sakata and White (2001), which allows for more robust estimation. The choice of a specific scale measure could be motivated by efficiency or robustness considerations. Further, the unique features of each estimation method should be considered before deriving its asymptotic properties for the GARCH case.

Remark 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.

Remark 8: When the fourth order moment is assumed to be finite, the GQMLE is $\sqrt{n}$ consistent for the true parameter values. However, in the presence of extreme non-normality, this estimator can fail to produce asymptotically efficient estimates. Hence, a two-step estimation procedure can be applied to gain efficiency. In the first step the GQMLE is used to obtain a consistent estimate of the scaling parameter and in the second step the LSE is used to estimate the model parameters. The issue of efficiency will be examined in the simulation study in the next section.

Remark 9: In our setting, we assume that the scaling factor c_0 is known. This assumption is standard[1]. It 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. Further, our empirical

[1]For stochastic volatility models, a similar approach 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 scaling constant was set to -1.27.

results, shown in the next section, clearly indicate that our findings are not sensitive to the choice of the scaling factor.

Remark 10: If we treat c_0 as unknown, (α_0, ω_0) can be estimated[2] up to a scale parameter. However, other GARCH estimation methods considered in the literature, R-estimation (Andrews (2012), M-estimation (Mukherjee (2008)), LAD-estimation (Peng and Yao (2003)), are also not used to directly estimate $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$. Instead, those methods are used to estimate $(\omega_0/d,\ \alpha_0/d,\ \beta_0)'$ where $d > 0$ is unknown when the error distribution is unknown. Another approach is to assume that ω_0 is known, see Linton el al. (2010).

Remark 11: Estimating θ_0 when c_0 is unknown is more complicated and requires modifying our estimation procedure. In what follows we describe in general, a possible estimation procedure for this case. However, investigating the asymptotic and empirical properties of the proposed estimator is left for future work. Note that from (1)-(2) and letting $\hbar_{0t} = h_{0t}/\omega_0 = 1 + (\alpha_0/\omega_0)y_{t-1}^2 + \beta_0\hbar_{0t-1}$, we have

$$\ln(y_t^2) = c_0 + \ln(\hbar_{0t}) + \zeta_t \tag{9}$$

where $c_0 = \mathrm{E}[\ln(\omega_0\varepsilon_t^2)]$ and $\{\zeta_t\}$ is a sequence of mean zero iid variables. As mentioned above, this nonlinear regression can be estimated via a least squares estimation. Thus, the unknown parameters $\psi_0 = (c_0,\ \alpha_0/\omega_0,\ \beta_0)$ are estimated by minimizing the following modified objective function

$$\tilde{Q}_n(\psi) = \tfrac{1}{2n}\textstyle\sum_{t=1}^{n}(\ln(y_t^2) - c - \ln\tilde{\hbar}_t(\theta))^2 \tag{10}$$

where $\psi = (c,\, \theta')'$, $\theta = (\alpha/\omega,\ \beta)'$ and $\hbar_t(\theta) = 1 + (\alpha/\omega)y_{t-1}^2 + \beta\hbar_{t-1}(\theta)$. In order to fully identify θ_0, we can use a standard two-step estimation procedure, see e.g. White (1994). In the first step, we apply the modified LSE to obtain a consistent estimate for the normalized series $\{y_t/\hbar_{0t}^{0.5}\}$ which should resemble $\sqrt{\omega_0}\varepsilon_t$ for large samples. In the second step, given the identified rescaled error distribution, θ_0 can be identified [3] via the maximum likelihood method (Rekkasa and Wong (2008); Francq and Zakoian (2013)).

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

[3]A simple way to identify the parameters would be to assume that $\mathrm{E}(\varepsilon_t^2) = 1$, which implies that the average of the squared rescaled errors converge to ω_0.

3 Simulation evidence

In this section, we investigate the finite sample properties of the LSE by means of a simulation study and compare the performance of the LSE with that of the GQMLE for a wide range of processes.

We note that for $\tilde{\theta}_n$, the GQMLE, $\sqrt{n}(\tilde{\theta}_n - \theta_0) \sim N(0, \kappa_N J^{-1})$ where $\kappa_N = \mathrm{E}(\varepsilon_t^4) - 1$. This relationship implies that the variability of the LSE relative to the GQMLE is captured by the efficiency ratio $\lambda = \kappa_N / \kappa$. The larger this quantity is, the more efficient the LSE is relative to the GQMLE. This relative efficiency depends on the distribution of the error term. The efficiency ratio for error distributions that have been used in the simulation study and have finite fourth moment are shown in Table 1. The results imply that the LSE can be substantially more efficient than the GQMLE when the distribution of the error term deviates from normality.

Table 1: : Efficiency of the LSE relative to GQMLE for different error distributions.

Distributions	κ	κ_N	λ
Normal	4.92	2	0.41
t_5	6.47	19.12	2.96
$\chi_1^2 - 1$	4.67	64.55	13.81

In the simulation study, in order to reflect a wide range of situations commonly encountered in practical financial modelling, we have considered different levels of persistence for the volatility model as well as different distributions for the errors. In particular, three different volatility parameterizations are used corresponding to three different levels of persistence in the volatility model: High (H), Medium (L) and Low (L). The selected volatility models have been summarized in Table 2. For each model in the table, the value of ω_0 in the volatility model was determined in order to constrain the variance of each of the DGP to be equal to 1.

Table 2: : Volatility models used for the simulation study.

	α_0	β_0
H	0.09	0.90
M	0.10	0.80
L	0.20	0.60

The error term was assumed to follow: standard normal, standardized

Student's t, with 3 and 5 degrees of freedom, and standardized χ^2_1. It is worth noting that $\mathrm{E}(\varepsilon_t^4) < \infty$ for all the distributions except for the t_3. In this case the asymptotic normality of the GQMLE is not expected to hold (Straumann (2005), p. 178).

Then, considering four different sample sizes,$T = 500, 1000, 2000, 5000$, a set of 1000 pseudo-random time series was simulated from each of the DGP's obtained matching the assumed error distributions with the volatility models summarized in Table 2. Next, a GARCH(1,1) model was fitted to each of the simulated series by using the GQMLE and the LSE, respectively. In particular, two different versions of the LSE have been used [4]. First, assuming knowledge of the underlying error distribution, the LSE was implemented using the correct scaling factor c_0. This can be easily approximated by simulating a very large sample[5] from the assumed distribution for error term. Then, $\tilde{c}_0$, a simulated approximation of c_0 can be obtained by taking the sample average of the natural logarithms of the squared simulated values. Furthermore, we also considered a two-stage LSE. In the first stage the GQMLE is used to obtain $\hat{c}_0$, a consistent estimate of the scaling factor. In the second stage the model is re-estimated by our method using the estimated scaling factor.

In order to assess the quality of the estimates, we have focused on the simulated values of bias and Mean Square Error (MSE). For the sake of brevity and ease of exposition, the results obtained for the two stage LSE have been omitted since they did not turn out to be significantly different from those obtained for the estimator based on the correct scaling factor ($\tilde{c}_0$). Also, to simplify the presentation of the results, we omit reporting the bias and MSE values for the constant term ω_0. However this set of results is available from the authors upon request.

A different situation appears for the High persistence GARCH model. In this case the GQMLE, differently from the LSE, is characterized by non-regular behaviour. Even in the case of normal errors, for large sample sizes, the value of the MSE is surprisingly higher than that registered for the LSE. This is probably due to the fact that the chosen DGP is very close to the border of the weak stationarity region. In the case of t_5 errors the LSE is by far more efficient than the QMLE if a sufficiently large sample size is considered ($T \geq 2000$). In the remaining cases the LSE is performing better than the QMLE, in terms of MSE, for all the sample sizes considered.

---

[4]The GQMLE was computed by using the MATLAB function fminunc to maximize the associated quasi likelihood function with respect to the unknown parameters. For the LSE, the relevant sum of squares was minimized using the MATLAB function lsqnonlin.

[5]In the simulation study a sample of length 10000 was used to approximate the scaling factor c_0.

It is interesting to note that, in general, the bias tends to be positive for the ARCH coefficient α while it is always negative for the GARCH coefficient β. This result is not surprising since it is in line with previous findings in the literature (see e.g. Straumann, 2005[6]). Furthermore, we must note that the overall behaviour observed in the cases of Low and Medium volatility persistence (see tables 3-6) is substantially different from that registered for the High persistence case (see tables 7-8). For the Low and Medium persistence models, in line with the results in Table 1, the GQMLE is performing substantially better than the LSE in the Gaussian case while, in non-Gaussian settings, the overall performance of the LSE model tends to improve over its competitor.

4 An application to financial data

In this section we present the results of an application of the proposed estimator to two time series of financial returns. First, we consider a time series of daily (percentage) log-returns on the S&P 500 index from January 5, 1971 to May 30, 2006 for a total of 8937 observations (Figure 1). Second, we consider a time series of 30 minutes returns on the USD/CHF exchange rate from April 1, 1996 to March 30, 2001 for a total of 62495 observations (Figure 2). In the latter case the data have been standardized in order to account for the presence of some observations exactly equal to zero. In order to remove any serial correlation structure, the S&P 500 series has been pre-filtered fitting an AR(2) model to the raw returns. Differently, the USD/CHF intraday exchange rate returns series has been pre-filtered in two steps: i) an AR(1) model has been fitted to the standardized returns to account for serial correlation ii) we have corrected for intraday seasonal patterns in volatility dividing the filtered returns by the corresponding seasonal factors. These have been calculated by simply averaging the squared returns in the various intraday intervals and taking square roots.

The performance of the LSE in reproducing the volatility of returns has been compared with that of the classical GQMLE. To evaluate the sensitivity of the LSE to different choices of the scaling factor, we consider estimating c_0 under different distributional assumptions for the error series: a standardized t_5, a standard normal and a Cauchy random variable with location and scale parameters equal to 0 and 1, respectively. In order to assess the relative performance of the estimators considered, we use the squared returns as a

[6]Note that the model considered by Straumann (2005) is slightly different from the GARCH(1,1) we consider since it includes an additional parameter which accounts for the presence of leverage effects.

Table 3: : Simulated bias (x 100) over 1000 pseudo-random replicates for the Low persistence volatility process with $\omega_0 = 0.20$, $\alpha_0 = 0.20$, $\beta_0 = 0.60$.

	GQMLE		LSE ($\tilde{c}_0$)	
Error Distribution	α	β	α	β
T=500				
Normal	0.1520	-2.8439	1.6772	-7.8704
t_5	0.6734	-4.1621	1.3123	-6.1606
t_3	4.7444	-6.5784	2.2778	-6.2961
χ_1^2	1.3163	-5.7422	1.3898	-4.1527
T=1000				
Normal	0.0341	-1.3305	0.7641	-3.2013
t_5	0.2991	-2.5013	0.6677	-1.6520
t_3	2.1055	-4.3241	1.0071	-2.0541
χ_1^2	1.6215	-3.6119	0.4695	-0.9259
T=2000				
Normal	0.1050	-0.5921	0.5958	-1.9491
t_5	0.0945	-1.3977	0.4892	-1.3129
t_3	1.2401	-2.5547	0.5408	-1.3392
χ_1^2	0.6412	-1.9104	0.4328	-0.5639
T=5000				
Normal	0.0132	-0.4075	0.2277	-0.8376
t_5	0.0775	-0.8912	0.2779	-0.8582
t_3	1.5100	-1.2480	0.2603	-0.3487
χ_1^2	0.7095	-1.2712	0.4327	-0.6224

Table 4: : Simulated Mean Square Error (x 100) over 1000 pseudo-random replicates for the Low persistence volatility process with $\omega_0 = 0.20$, $\alpha_0 = 0.20$, $\beta_0 = 0.60$.

	GQMLE		LSE ($\tilde{c}_0$)	
Error Distribution	α	β	α	β
T=500				
Normal	0.3758	1.8616	0.9682	5.3519
t_5	1.2075	3.3092	1.0069	4.6324
t_3	20.9969	5.8684	1.3078	4.4953
χ_1^2	3.1462	5.6772	0.8740	3.0159
T=1000				
Normal	0.1770	0.7460	0.4717	2.1371
t_5	0.5562	1.8335	0.4830	1.8687
t_3	4.1587	4.0998	0.6168	1.9349
χ_1^2	1.7442	3.2239	0.4064	1.0868
T=2000				
Normal	0.0907	0.3216	0.2242	1.0220
t_5	0.2837	0.8548	0.2332	0.8009
t_3	2.4501	2.8793	0.3100	0.8990
χ_1^2	0.6925	1.7321	0.1893	0.4337
T=5000				
Normal	0.0350	0.1270	0.0866	0.3420
t_5	0.1122	0.3376	0.0982	0.3065
t_3	2.3541	1.6481	0.1176	0.2966
χ_1^2	0.2847	0.6921	0.0804	0.1763

Table 5: : Simulated bias (x 100) over 1000 pseudo-random replicates for the Medium persistence volatility process with $\omega_0 = 0.10$,$\alpha_0 = 0.10$, $\beta_0 = 0.80$.

	GQMLE		LSE ($\tilde{c}_0$)	
Error Distribution	α	β	α	β
T=500				
Normal	0.0819	-4.3154	2.0268	-18.9382
t_5	0.9619	-6.2946	2.2417	-14.2595
t_3	4.0729	-11.2138	2.5491	-13.5762
χ^2_1	2.4365	-9.4596	1.5899	-7.5517
T=1000				
Normal	0.0669	-1.8689	1.2244	-10.0284
t_5	0.4380	-3.1592	1.1640	-7.0427
t_3	1.6529	-6.5714	1.2531	-5.2478
χ^2_1	1.1753	-4.5431	0.6193	-2.4494
T=2000				
Normal	0.0489	-0.9953	0.7294	-3.3948
t_5	0.3451	-1.6849	0.5889	-2.0423
t_3	2.2618	-2.9461	0.5049	-1.7278
χ^2_1	0.7206	-2.3926	0.1515	-0.6626
T=5000				
Normal	0.0058	-0.3597	0.3141	-1.2743
t_5	0.1450	-0.6466	0.2254	-0.8282
t_3	1.4510	-2.0021	0.2324	-0.5920
χ^2_1	0.2104	-0.7452	0.1319	-0.3485

Table 6: : Simulated Mean Square Error (x 100) over 1000 pseudo-random replicates for the Medium persistence volatility process with $\omega_0 = 0.10$,$\alpha_0 = 0.10$, $\beta_0 = 0.80$.

	GQMLE		LSE ($\tilde{c}_0$)	
Error Distribution	α	β	α	β
T=500				
Normal	0.2022	1.8818	0.5225	11.4787
t_5	0.5860	2.8612	0.5130	7.9819
t_3	7.4474	5.5943	0.7750	7.9700
χ_1^2	1.9106	4.8282	0.3737	4.1705
T=1000				
Normal	0.0814	0.6311	0.2662	5.4332
t_5	0.3151	1.3267	0.2441	3.6226
t_3	1.8726	2.9159	0.2570	2.5763
χ_1^2	0.5817	2.1179	0.1246	1.0478
T=2000				
Normal	0.0393	0.2548	0.1130	1.3540
t_5	0.1139	0.5641	0.1061	0.8534
t_3	2.6353	1.6702	0.1128	0.6794
χ_1^2	0.2732	0.9862	0.0564	0.2240
T=5000				
Normal	0.0163	0.0831	0.0419	0.2913
t_5	0.0477	0.1737	0.0353	0.1521
t_3	1.0281	0.9039	0.0410	0.1307
χ_1^2	0.0844	0.2784	0.0213	0.0707

Table 7: : Simulated bias (x 100) over 1000 pseudo-random replicates for the High persistence volatility process with $\omega_0 = 0.01$,$\alpha_0 = 0.09$, $\beta_0 = 0.90$.

	GQMLE		LSE ($\tilde{c}_0$)	
Error Distribution	α	β	α	β
T=500				
Normal	0.9561	-2.3596	1.1684	-8.7027
t_5	2.6349	-4.3835	1.5388	-5.8657
t_3	5.9734	-7.9647	2.0560	-7.3836
χ_1^2	5.1557	-7.4965	1.3561	-3.6250
T=1000				
Normal	1.3435	-1.8156	0.5509	-2.1848
t_5	2.4439	-2.7361	0.6940	-1.7094
t_3	4.4315	-4.8574	0.6724	-1.8669
χ_1^2	5.3084	-3.6914	0.4448	-0.7459
T=2000				
Normal	1.4091	-1.3727	0.2109	-0.5496
t_5	2.0595	-1.6655	0.1693	-0.4252
t_3	3.1710	-3.0927	0.2558	-0.4894
χ_1^2	3.7968	-2.6326	0.2052	-0.2945
T=5000				
Normal	1.6350	-1.5051	0.1300	-0.2902
t_5	1.9663	-1.2381	0.0633	-0.1555
t_3	2.9075	-2.0648	0.1768	-0.2310
χ_1^2	2.4037	-1.0823	0.1242	-0.1656

Table 8: : Simulated Mean Square Error (x 100) over 1000 pseudo-random replicates for the High persistence volatility process with $\omega_0 = 0.01$,$\alpha_0 = 0.09$, $\beta_0 = 0.90$.

	GQMLE		LSE ($\tilde{c}_0$)	
Error Distribution	α	β	α	β
T=500				
Normal	0.4972	0.4333	0.2831	5.2575
t_5	1.0339	0.9597	0.3238	2.9953
t_3	18.5386	2.8339	0.5464	4.1824
χ_1^2	4.2910	2.4702	0.2873	1.8381
T=1000				
Normal	0.3060	0.1981	0.1177	0.7986
t_5	0.7952	0.4527	0.1193	0.5341
t_3	10.1927	1.4221	0.1322	0.6742
χ_1^2	9.7875	0.7489	0.0708	0.1113
T=2000				
Normal	0.2468	0.1437	0.0504	0.0639
t_5	0.6320	0.2137	0.0418	0.0453
t_3	2.7973	0.5827	0.0487	0.1306
χ_1^2	3.4428	0.6729	0.0263	0.0240
T=5000				
Normal	0.2346	0.1520	0.0167	0.0203
t_5	0.9340	0.2495	0.0150	0.0156
t_3	2.5498	0.5194	0.0193	0.0180
χ_1^2	9.7692	0.2160	0.0103	0.0096

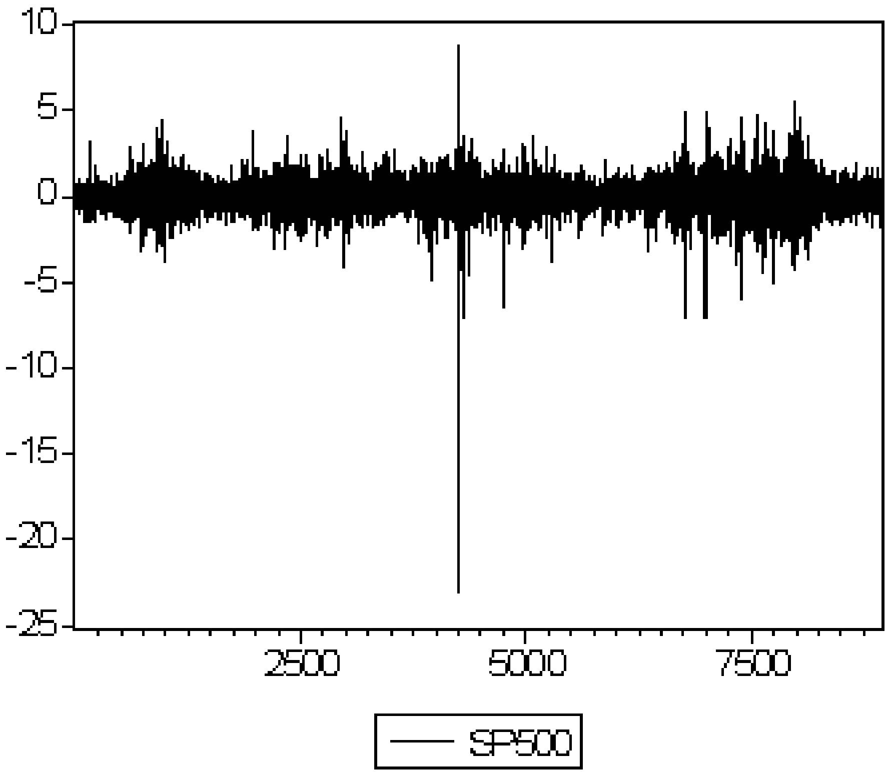

Figure 1: S&P 500 daily returns from 5.01.1971 to 30.05.2006.

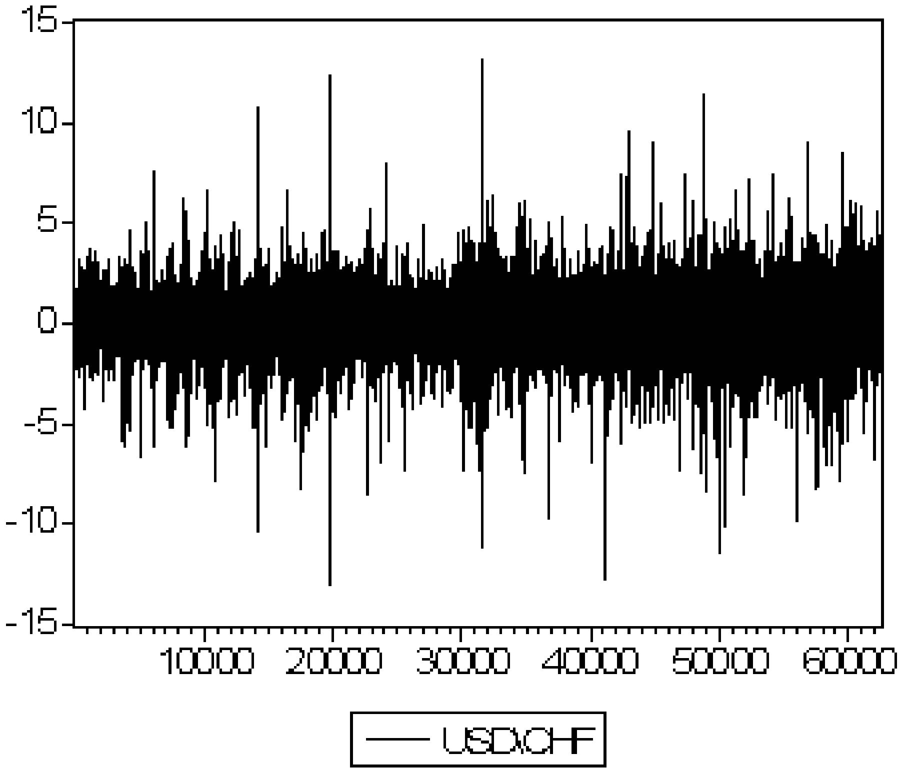

Figure 2: 30 minutes returns on the USD/CHF exchange rate from 1.04.1996 to 30.03.2001.

proxy of volatility and then refer to the following well-known loss functions: the Mean Square Error (MSE), the QLIKE, the Mean Absolute Error (MAE) and its equivalent formulation in terms of standard deviations (MAE-SD). A discussion of these loss functions and their properties can be found in Patton (2011). For MSE and QLIKE, the expected loss is minimized if the volatility estimate used to compute the loss function coincides with the true conditional variance. Differently, for MAE and MAE-SD, optimality is achieved in correspondence of the true conditional median of the squared returns.

The volatility of each of the two series, S&P 500 and USD/CHF exchange rate returns, has been modelled as a GARCH(1,1) whose parameters have been estimated by QML and by the LSE (Table 9). For the S&P 500, the estimates of the ARCH coefficient α obtained by the LSE are substantially lower than that yielded by the GQMLE while the opposite applies to the GARCH parameterβ. Furthermore, it is interesting to analyze the behaviour of the different estimators under the four loss functions considered (Table 8). For the MSE, all the estimators yield very similar performances. The only exception is given by the LSE constructed under the assumption of Cauchy errors which is characterized by a value of the MSE much higher than was observed for its competitors.

A different picture arises if we consider the QLIKE criterion. For the daily S&P 500 returns series, except for the Cauchy case, the performance of LSE is quite close to that of the GQMLE. The gap substantially increases in the case of the 30 minutes USD/CHF exchange rate returns. For the other two loss functions considered, MAE and MAE-SD, and for both datasets, the LSE is always outperforming the QMLE. The LSE performance is optimized if we estimate the scaling constant c_0 under the assumption of Cauchy errors with location and scale parameters equal to 0 and 1, respectively. However, in general, it is worth noting that the performance of the LSE appears to be quite robust to the choice of the scaling factor c_0.

The message we get from these results is that, if one is interested in the conditional variance of returns as a measure of volatility, no clear advantage derives from using the LSE instead of the usual GQMLE. Differently, if the focus is on an alternative measure of volatility, such as the conditional median of squared returns, the use of the LSE can potentially allow for substantial accuracy gains.

Finally, in order to evaluate the ability of the different estimators to correctly reproduce volatility persistence, we have compared the sample autocorrelation of squared returns with the autocorrelation function implied by each of the estimated models (Figure 1 and Figure 2). For this exercise, however, we haven't considered the LSE obtained under the assumption

Table 9: GARCH(1,1) parameter estimates under different estimators (* x 10^{-4}). Key to table: LS-D is the Least Squares estimator under distribution D (N=Normal, C=Cauchy, t_5= Student's with 5 df.)

	S&P 500			*USD/CHF*		
	ω	α	β	ω	α	β
QML	0.0007*	0.0658	0.9271	0.0448	0.0832	0.8752
LS-N	0.0036	0.0395	0.9486	0.0615*	0.1293	0.8302
LS-t_5	0.0030	0.0322	0.9478	0.0499*	0.1030	0.8312
LS-C	0.0013	0.0131	0.9386	0.0208*	0.0374	0.8259

Table 10: Evaluation of volatility estimates for the daily S&P 500 and 30 min. USD/CHF returns by means of different loss functions: MAE, MSE and MSE-LOG. Key to table: LS-D is the Least Squares estimator under distribution D (N=Normal, C=Cauchy, t_5= Student's with 5 df.)

	S&P 500				*USD/CHF*			
	MSE	*QLIKE*	*MAE*	*MAE-SD*	*MSE*	*QLIKE*	*MAE*	*MAE-SD*
QML	34.63	0.72	1.06	0.53	14.69	0.90	1.19	0.60
LS-N	34.53	0.74	0.99	0.49	14.77	1.45	1.08	0.53
LS-t_5	34.56	0.82	0.93	0.46	14.76	1.73	1.02	0.50
LS-C	35.41	2.60	0.89	0.44	15.58	5.50	0.95	0.47

of Cauchy errors since in this case the autocorrelation function of squared returns cannot be defined. Also, for the USD/CHF exchange rate returns series, the set of model coefficients estimated using LSE violate the condition for the existence of a finite fourth moment which is

$$(3\alpha^2 + 2\alpha\beta + \beta^2) < 1$$

For this reason, it has been necessary to approximate the corresponding autocorrelation function by means of the formula proposed by Ding and Granger (1996) for conditionally Gaussian GARCH(1,1) models

$$\rho(k) = (\alpha + \beta)^{k-1} \left(\alpha + \frac{\beta}{3} \right), \qquad k \geq 1$$

where $\rho(k)$ is the lag k autocorrelation function of a squared GARCH(1,1) process. For the daily S&P 500 returns series, it is evident how the LSE is interpolating the decay of the sample autocorrelation function of squared returns much better than the QML approach. Differently, for the 30 minutes USD/CHF exchange rate returns series, the autocorrelation patterns implied by the t_5-LSE and QMLE result quite close while the normal LSE drastically overestimates the value of the autocorrelation function of squared returns.

5 Conclusions and future work

In this paper, we suggest using LSE for the estimation of a GARCH (1,1) model. The estimator is based on the log transformation of the squared data. We establish the consistency and asymptotic normality of the proposed estimator. 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. Its finite sample properties have been investigated via a simulation study, which shows that, in the presence of extreme non-normality, the proposed LSE can allow for some efficiency gains with respect to the QMLE. We also provide empirical evidence that applying the LSE can yield better volatility forecasts than the standard QMLE. Our estimates also fit quite well the autocorrelation function of the squared returns.

When working with 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

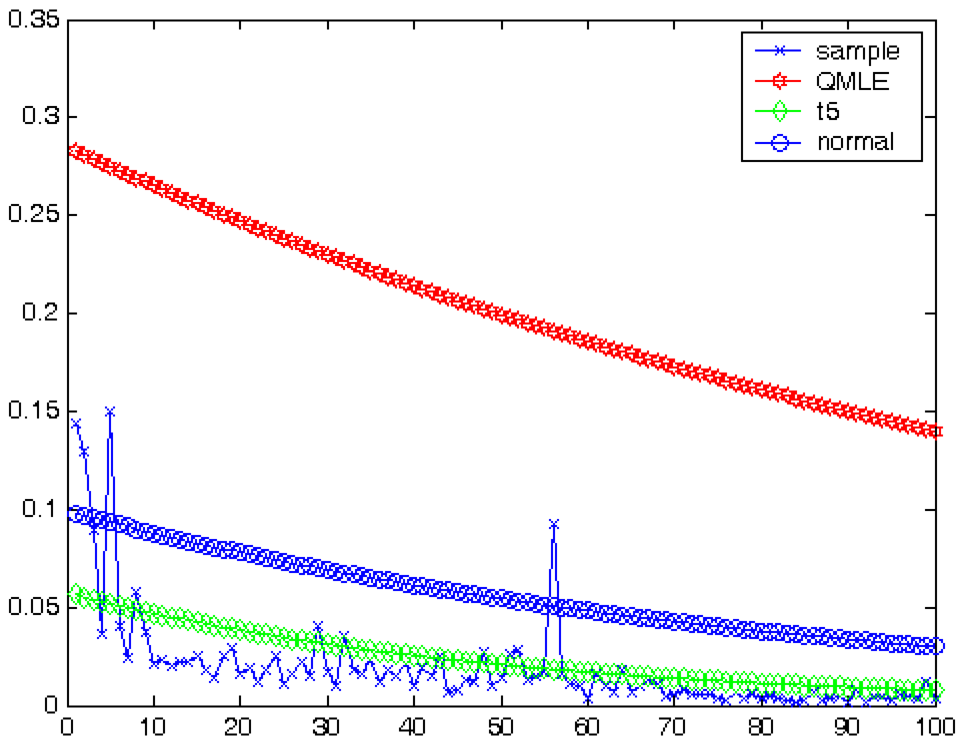

Figure 3: Implied autocorrelation function of squared returns versus sample autocorrelations for the S&P500 series (lags from 1 to 100) : QML and alternative LSE.

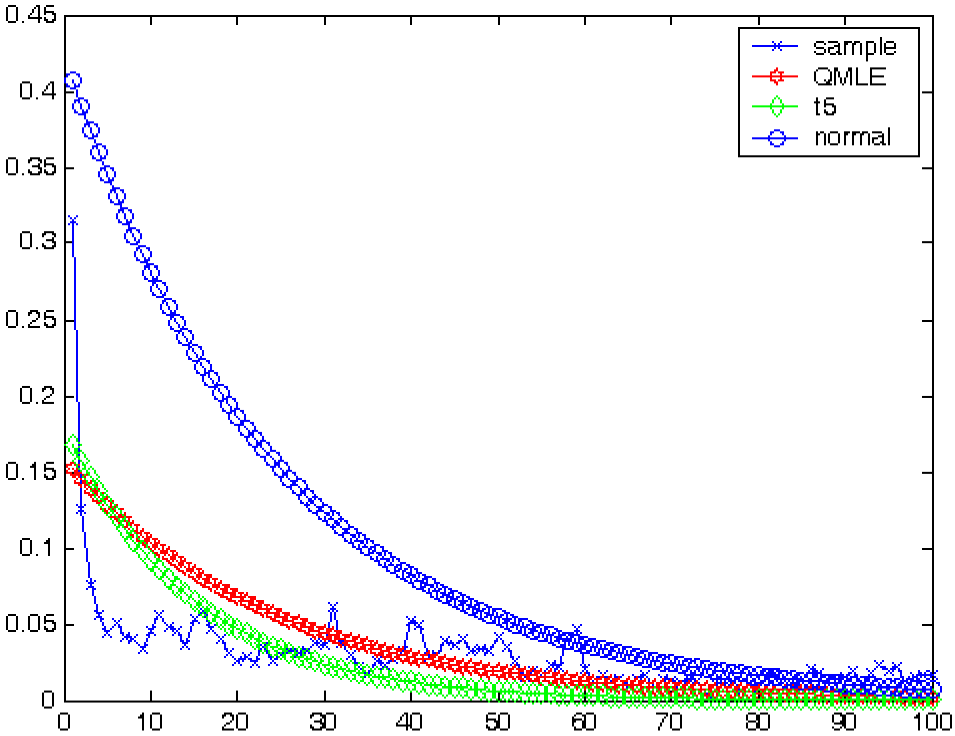

Figure 4: Implied autocorrelation function of squared returns versus sample autocorrelations for the USD/CHF series (lags from 1 to 100) : QML and alternative LSE.

employs a more robust scale measure such as the S-estimator can be used. 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.

Appendix

Throughout the Appendix, K will denote a generic positive number that may vary in different uses. To simplify the notation we set

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

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.

5.1 A. Proofs of theorems

Proof of Theorem 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 compact set $F=\{\omega\in\Omega|\limsup_{n\to\infty}||\hat{\theta}_n-\theta_0||\ \geq\gamma,\ \ \hat{\theta}_n\in\Theta\}$ has a positive probability. Since the set $\mathrm{N}=\Theta\cap\{\theta:|\hat{\theta}_n-\theta_0|\ \geq\gamma\}$ is compact, there exists a non-null subset $\bar{F}\subset F$ such that for every$\omega\in\bar{F}$, one can find inN, a convergent subsequence$\hat{\theta}_{n_i}(\omega)\to\theta\in\mathrm{N}$. By definition of the LSE

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

From Lemma 5,

$$\liminf_{n\to\infty}\tfrac{1}{n_i}\textstyle\sum_{t=1}^{n_i}\ell_t(\theta_0)\ \geq\ \liminf_{n\to\infty}\tfrac{1}{n_i}\textstyle\sum_{t=1}^{n_i}\ell_t(\hat{\theta}_{n_i}) \tag{11}$$

The inequality above and Lemmas 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 2: By Theorem 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

$$
\begin{aligned}
0 &= n^{-0.5} \sum_{t=1}^{n} \nabla \tilde{\ell}_t(\hat{\theta}_n) \qquad (12) \\
&= n^{-0.5} \sum_{t=1}^{n} \nabla \tilde{\ell}_t(\theta_0) + \left(\tfrac{1}{n} \sum_{t=1}^{n} \nabla^2 \tilde{\ell}_t(\bar{\theta}_n) \right) \sqrt{n}(\hat{\theta}_n - \theta_0) \\
&= n^{-0.5} \sum_{t=1}^{n} \nabla \tilde{\ell}_t(\theta_0) \\
&+ \left[\left(\tfrac{1}{n} \sum_{t=1}^{n} \nabla^2 \tilde{\ell}_t(\bar{\theta}_n) - \tfrac{1}{n} \sum_{t=1}^{n} \nabla^2 \ell_t(\bar{\theta}_n) \right) + \left(\tfrac{1}{n} \sum_{t=1}^{n} \nabla^2 \ell_t(\bar{\theta}_n) + J \right) - J \right] \\
&\quad \sqrt{n}(\hat{\theta}_n - \theta_0)
\end{aligned}
$$

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 $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \partial \tilde{\ell}_t(\theta_0) \big/ \partial \theta \underset{D}{\rightarrow} N(0, H)$ where $H = \kappa J$ and J is a positive definite matrix. Next, Lemmas 7(i)-(ii) imply that the first and second terms, inside the square brackets in (12), converge a.s. to zero. Hence, to complete the proof it suffices to solve (12) and apply Slutsky's theorem.

Proof of Theorem 3: The result follows immediately from Theorems 1-2 and Lemma 7.

B. Lemmata

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

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

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) $\mathrm{E}\left(\eta_t^2\right) < \infty$

Proof:

i) Under Assumption A2, the result follows directly from (1)-(2) and Theorem 4 of Nelson (1990).

ii) From (7)-(8) 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, since $\eta_t = w - c_0$, the result follows if $\int_0^{+\infty} [\ln(w)]^2 f(w)dw < \infty$. By integration by parts

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

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 $w \geq 1$ we get $\int_1^{+\infty} [\ln(w)]^2 f(w)dw < \int_1^{+\infty} w^{2s} f(w)dw \leq \mathrm{E}|\varepsilon_t|^{2s}$, since $\ln(w) < w^{s/2}$ for any $s > 0$, and the desired result follows by Assumption A3.

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

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

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

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

Proof: i) By iterating (7) and using the fact $\alpha_0 y_{t-1-i}^2 \leq h_{0t}$, we get

$$\begin{aligned} h_t(\theta) &= \omega + \alpha y_{t-1}^2 + \beta h_{t-1}(\theta) \qquad (13) \\ &= \sum\nolimits_{i=0}^{t-1} (\omega + \alpha y_{t-1-i}^2)\beta^i + \beta^t h_1(\theta) \\ &= \sum\nolimits_{i=0}^{\infty} (\omega + \alpha y_{t-1-i}^2)\beta^i \\ &= \frac{\omega}{1-\beta} + \alpha \sum\nolimits_{i=0}^{\infty} \beta^i y_{t-1-i}^2 \\ &\leq \frac{\bar{\omega}}{1-\underline{\beta}} + \frac{\bar{\alpha}}{\alpha_0} \sum\nolimits_{i=0}^{\infty} \bar{\beta}^i h_{0t} \end{aligned}$$

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 $p \in (0,1)$,

$$\mathrm{E}|\sup_{\theta\in\Theta} h_t(\theta)|^p \ \leq\ K + K\mathrm{E}h_{0t}^p \ <\ \infty \tag{14}$$

Now, without loss of generality, set $\tilde{h}_1 = 0.5(\bar{\omega} + \omega)$, by iterating (5) we obtain

$$\tilde{h}_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta\tilde{h}_{t-1}(\theta) = \sum\nolimits_{i=0}^{t-1}(\omega + \alpha y_{t-1-i}^2)\beta^i + \beta^t\tilde{h}_1 \tag{15}$$

Hence

$$\tilde{h}_t(\theta) - h_t(\theta) = \omega + \alpha y_{t-1}^2 + \beta\tilde{h}_{t-1}(\theta) = \beta^t(\tilde{h}_1 - h_1(\theta)) \tag{16}$$

and by (16),

$$\mathrm{E}\sup_{\theta\in\Theta^0}\left|h_t(\theta) - \tilde{h}_t(\theta)\right|^p \ \leq\ \beta^t(\tilde{h}_1^p + \mathrm{E}\sup_{\theta\in\Theta^0}|h_1(\theta)|^p) \leq K\bar{\beta}^t \tag{17}$$

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

$$\mathrm{E}(\bar{\omega} + \bar{\alpha}y_{t-1-i}^2)^p < \infty \tag{18}$$

and

$$\mathrm{E}\left(\sup_{\theta\in\Theta}\left|\tilde{h}_t(\theta)\right|^p\right) \leq \sum\nolimits_{i=0}^{t-1}\mathrm{E}(\bar{\omega} + \bar{\alpha}y_{t-1-i}^2)^p\bar{\beta}^{ip} + \bar{\beta}^{pt}\tilde{h}_1^p < \infty$$

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

$$\mathrm{E}\left(\sup_{\theta\in\Theta^0}\left|\dot{h}_{it}(\theta)\right|^p\right) < \infty \tag{19}$$

By (13) and the fact that $y_{t-1-i}^2 \leq \alpha_0^{-1}h_{0t}$,

$$\frac{\partial h_t(\theta)}{\partial\omega} \ \leq\ \frac{1}{1-\underline{\beta}} \tag{20}$$

$$\frac{\partial h_t(\theta)}{\partial\alpha} \ =\ \sum\nolimits_{i=0}^{\infty}\beta^i y_{t-1-i}^2 \ \leq\ \frac{1}{\alpha}\left[\sum\nolimits_{i=0}^{\infty}\alpha\beta^i y_{t-1-i}^2\right] \ \leq\ \frac{1}{\underline{\alpha}}h_t(\theta) \tag{21}$$

$$\frac{\partial h_t(\theta)}{\partial\beta} \ =\ \sum\nolimits_{i=1}^{\infty}i\beta^i(\omega + \alpha y_{t-1-i}^2) \tag{22}$$

$$\leq\ \sum\nolimits_{i=1}^{\infty}i\beta^i\left(\omega + \frac{\alpha}{\alpha_0}h_{0t}\right) \ \leq\ \bar{\omega}\sum\nolimits_{i=1}^{\infty}i\bar{\beta}^i + \frac{\bar{\alpha}}{\alpha_0}\sum\nolimits_{i=0}^{\infty}\bar{\beta}^i h_{0t}$$

The term in (20) is bounded and admits moments of any order. As for (21)-(22), the result follows directly from the c_r inequality and Lemma 1(i). In view of (16), almost surely,

$$\sup_{\theta\in\Theta^0}\left|\dot{h}_{it}(\theta)-\dot{\tilde{h}}_{it}(\theta)\right| \;\leq\; t\bar{\beta}^{(t-1)}(\tilde{h}_1+\sup_{\theta\in\Theta^0}h_1(\theta))+\bar{\beta}^t\sup_{\theta\in\Theta^0}|\dot{h}_{i1}(\theta)| \;\leq\; K\bar{\beta}^t$$

the desired result follows by (14), (19) and the c_r inequality.

iii) From (20)-(22) and direct calculations we get,

$$\frac{\partial^2 h_t}{\partial\omega^2}=\frac{\partial^2 h_t}{\partial\alpha^2}=\frac{\partial^2 h_t}{\partial\omega\partial\alpha}=0, \qquad \frac{\partial^2 h_t}{\partial\omega\partial\beta}\frac{1}{\beta}\leq\sum\nolimits_{i=1}^{\infty}i\bar{\beta}^i \tag{23}$$

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

$$\frac{\partial^2 h_t}{\partial\alpha\partial\beta} \;\leq\; \alpha\sum\nolimits_{i=1}^{\infty}i\beta^i y_{t-1-i}^2 \;\leq\; \frac{\bar{\alpha}}{\alpha_0}\sum\nolimits_{i=1}^{\infty}i\beta^i h_{0t} \tag{24}$$

$$\frac{\partial^2 h_t}{\partial\beta^2} \;=\; \frac{1}{\beta}\sum\nolimits_{i=2}^{\infty}i(i-1)(\omega+\alpha y_{t-1-i}^2)\beta^i \tag{25}$$

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

$$\mathrm{E}\left(\sup_{\theta\in\Theta^0}\left|\ddot{h}_{ijt}(\theta)\right|^p\right)<\infty \tag{26}$$

for all i,j. In view of (16), almost surely,

$$\begin{aligned}\sup_{\theta\in\Theta^0}\left|\ddot{h}_{ijt}(\theta)-\ddot{\tilde{h}}_{ijt}(\theta)\right| \;&\leq\; t(t-1)\bar{\beta}^{(t-2)}[\tilde{h}_1+\sup_{\theta\in\Theta^0}h_1(\theta)]\\ &+\; t\bar{\beta}^{(t-1)}\sup_{\theta\in\Theta^0}|\dot{h}_{j1}(\theta)|+\;t\bar{\beta}^{t-1}\sup_{\theta\in\Theta^0}|\dot{h}_{i1}(\theta)|\\ &+\;\bar{\beta}^t\sup_{\theta\in\Theta^0}|\ddot{h}_{ij1}(\theta)|\end{aligned}$$

and by (14), (19), (26) and the c_r inequality the desired result follows.

Lemma 3[7]: Under Assumptions A1-A4, for all $r\geq 1$

[7]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.

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.

Proof: i) Eq. (20) and (21) imply that the derivative of h_t with respect to ω and α (divided by h_t) are bounded and hence admits moments of any order. However, this is not true for the derivative with respect to β. From (13) 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

$$\begin{aligned}
\frac{\partial h_t}{\partial \beta}\frac{1}{h_t} &\leq \frac{1}{\beta}\sum\nolimits_{i=1}^{\infty} i\frac{(\omega + \alpha y_{t-1-i}^2)\beta^i}{\omega + (\omega + \alpha y_{t-1-i}^2)\beta^i} \qquad (27)\\
&\leq \frac{1}{\beta}\sum\nolimits_{i=1}^{\infty} i\left[\frac{(\omega + \alpha y_{t-1-i}^2)\beta^i}{\omega}\right]^{p/r}\\
&\leq \frac{1}{\underline{\beta}\underline{\omega}^{p/r}}\sum\nolimits_{i=1}^{\infty} i\bar{\beta}^{ip/r}(\bar{\omega} + \bar{\alpha} y_{t-1-i}^2)^{p/r}
\end{aligned}$$

Therefore, by (18) and Minkowski's inequality we get

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

ii) From (23)-(25), we observe that the relevant second derivatives satisfy

$$\frac{\partial^2 h_t}{\partial \beta^2}\frac{1}{h_t} \leq \frac{1}{\beta}\sum\nolimits_{i=2}^{\infty} i(i-1)\frac{(\omega + \alpha y_{t-1-i}^2)\beta^i}{\omega + (\omega + \alpha y_{t-1-i}^2)\beta^i} \qquad (28)$$

and

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

(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.

Lemma 4: Under Assumptions A1-A5,

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

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

$$\liminf_{n\to\infty} \inf_{\theta\in\mathrm{N}} \tfrac{1}{n}\sum_{t=1}^{n} \ell_t(\theta) \ \geq \ \mathrm{E}\inf_{\theta\in\mathrm{N}} \ell_t(\theta).$$

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

iv) $\mathrm{E}\left(|\sup_{\theta\in\Theta} \ln h_t(\theta)|^2\right) < \infty$ and $\mathrm{E}(z_t^2) < \infty$

Proof:

i) Note that

$$\begin{aligned}
\mathrm{E}(\ell_t(\theta)) - \mathrm{E}(\ell_t(\theta_0)) &= \tfrac{1}{2}\mathrm{E}\left[(z_t - \ln h_t(\theta))^2 - \eta_t^2\right] \qquad (29)\\
&= \tfrac{1}{2}\mathrm{E}\left[\ln(h_{0t} - \ln(h_t(\theta))\right]^2 + \mathrm{E}\left[\ln(h_{0t}/h_t)\right]\mathrm{E}(\eta_t)\\
&= \tfrac{1}{2}\mathrm{E}\left[\ln\left(h_t(\theta_0)/h_t(\theta)\right)\right]^2 \ \geq \ 0
\end{aligned}$$

with equality if and only if $h_t(\theta_0) = h_t(\theta)$a.s.

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

$$\liminf_{n\to\infty} \inf_{\theta\in\mathrm{N}} \frac{1}{n}\sum_{t=1}^{n} \ell_t(\theta) \ \geq \ \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \inf_{\theta\in\mathrm{N}} \ell_t(\theta) \qquad (30)$$

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

$$\begin{aligned}
\liminf_{n\to\infty} \inf_{\theta\in\mathrm{N}} \tfrac{1}{n}\sum_{t=1}^{n} \ell_t(\theta) &\geq \liminf_{n\to\infty} \tfrac{1}{n}\sum_{t=1}^{n} \inf_{\theta\in\mathrm{N}} \ell_t(\theta) \qquad (31)\\
&\geq \mathrm{E}\left(\inf_{\theta\in\mathrm{N}} \ell_t(\theta)\right)
\end{aligned}$$

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.

iv) Notice, that since $0 < \underline{\omega} \leq h_t(\theta)$ for any $p > 0$,

$$\ln(\underline{\omega}) \leq \left|\sup_{\theta\in\Theta} \ln h_t(\theta)\right| \leq K + \left|\sup_{\theta\in\Theta} h_t(\theta)\right|^{p/2}$$

By (14) we obtain that $\mathrm{E}\left(\left|\sup_{\theta\in\Theta} \ln h_t(\theta)\right|^2\right) < \infty$. This result and Lemma 1(iii) also imply that $\mathrm{E}(z_t^2)$ is finite.

Lemma 5: Under Assumptions A1-A4,

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

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 $\lambda > 0$,

$$\sum_{t=1}^{\infty} \mathrm{P}(\sup_{\theta\in\Theta} |A_t(\theta)| > \lambda) \leq \sum_{t=1}^{\infty} \mathrm{E}\sup_{\theta\in\Theta} |A_t(\theta)|^q/\lambda^q < \infty \qquad (32)$$

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

$$|\ln \tilde{h}_t(\theta) - \ln h_t(\theta)| \leq K|\tilde{h}_t(\theta) - h_t(\theta)| \qquad (33)$$

So, from (4), (8) and the c_r inequality, for some $p \in (0, 1)$

$$\begin{aligned}
\mathrm{E}\sup_{\theta\in\Theta} |\tilde{\ell}_t(\theta) - \ell_t(\theta)|^{p/4} &\leq \mathrm{E}\sup_{\theta\in\Theta} \left[\left|\ln \tilde{h}_t(\theta) - \ln h_t(\theta)\right|^{p/4} \right. \\
&\quad \times \left. \left|\ln \tilde{h}_t(\theta) - \ln h_t(\theta) + 2\left(z_t + \ln h_t(\theta)\right)\right|^{p/4} \right] \\
&\leq \mathrm{E}\left[\sup_{\theta\in\Theta} \left|\tilde{h}_t(\theta) - h_t(\theta)\right|^{p/2} \left|1 + z_t + \sup_{\theta\in\Theta} \ln h_t(\theta)\right|^{p/4} \right] \\
&\leq K\mathrm{E}\left(\sup_{\theta\in\Theta} \left|\tilde{h}_t(\theta) - h_t(\theta)\right|^{p} \right) = O\left(\bar{\beta}^t\right)
\end{aligned}$$

The second inequality holds by (33). The third inequality holds by the c_r and Cauchy-Schwarz inequalities and Lemma 4(iv). The last equality holds by Lemma 2(i).

Lemma 6: Under Assumptions A1-A5,

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

ii) $n^{-1/2} \sum_{t=1}^{n} \nabla \ell_t(\theta_0) \underset{D}{\rightarrow} N(0, \kappa J)$ where J is positive definite and $\kappa = \mathrm{E}(\eta_t^2)$
.

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 (4) and (8) are given by

$$\nabla \tilde{\ell}_{it}(\theta_0) = (z_t - \ln \tilde{h}_{0t}) \frac{\dot{\tilde{h}}_{0it}}{\tilde{h}_{0t}}, \;\; \nabla \ell_{it}(\theta_0) = (z_t - \ln h_{0t}) \frac{\dot{h}_{0it}}{h_{0t}} = \eta_t \frac{\dot{h}_{0it}}{h_{0t}} \tag{34}$$

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

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

and

$$n^{-1/2} \sum_{t=1}^{n} B_t \leq n^{-1/2} K \sum_{t=1}^{n} \eta_t \left(\dot{h}_{0it} - \dot{\tilde{h}}_{0it} \right) + \frac{\dot{\tilde{h}}_{0it}}{\tilde{h}_{0t}} \left(h_{0it} - \tilde{h}_{0it} \right) \tag{35}$$

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 (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 (35) tends to zero a.s. as $n \rightarrow \infty$ and the desired result follows

ii) From (34)

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

where $\mathrm{F}_t = \sigma(y_t, y_{t-1}, \ldots)$ 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$\{\nabla\ell_t(\theta_0), \mathrm{F}_t\}$ 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. Thus, Theorem 23.1 of Billingsley (1968) and the Cramér-Wold device imply that $n^{-1/2}\sum_{t=1}^{n}\nabla^2\ell_t(\theta_0) \underset{D}{\rightarrow} N(0,\kappa J)$.

Lemma 7: Under Assumptions A1-A5,

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

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

Proof:

i) First, let $C_t(\theta) = \nabla^2\ell_{ijt}(\theta) - \nabla^2\tilde{\ell}_{ijt}(\theta)$. 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 (4) and (8) the second derivatives are

$$\nabla^2\tilde{\ell}_{ijt}(\theta) = (z_t - \ln\tilde{h}_t)\frac{\ddot{\tilde{h}}_{ijt}}{\tilde{h}_t} - (z_t - \ln\tilde{h}_t + 1)\frac{\dot{\tilde{h}}_{it}\dot{\tilde{h}}_{jt}}{\tilde{h}_t^2} \tag{36}$$

and

$$\nabla^2\ell_{ijt}(\theta) = (z_t - \ln h_t)\frac{\ddot{h}_{ijt}}{h_t} - (z_t - \ln h_t + 1)\frac{\dot{h}_{it}\dot{h}_{jt}}{h_t^2} \tag{37}$$

Third, note

$$\frac{\dot{h}_{it}\dot{\tilde{h}}_{jt}}{h_t^2} - \frac{\dot{\tilde{h}}_{it}\dot{\tilde{h}}_{jt}}{\tilde{h}_t^2} \leq K\left\{\frac{\dot{h}_{it}}{h_t}\left[\dot{\tilde{h}}_{it} - \dot{h}_{it}\right] + \frac{\dot{\tilde{h}}_{jt}}{\tilde{h}_t}\left[\dot{\tilde{h}}_{jt} - \dot{h}_{jt}\right]\right\} \tag{38}$$

Finally, using (36)-(38) 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 and Minkowski inequalities with Lemmas 2, 3 and 4(iv), 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 (37), $\mathrm{E}\left(\nabla^2\ell_t(\theta_0)\right) = -J$ 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 Lemma 1(iii), Lemmas 3 and 4(iv). From the ergodic theorem (see e.g., Billingsley (1995)),

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

Hence, given $\varepsilon>0$

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

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.

References

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

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

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

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

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

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

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

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] Francq C. and J-M. Zakoian (2009). A Tour in the asymptotic theory of GARCH estimation, *Handbook of Financial Time Series*, 85-111, Berlin, Springer-Verlag.

[14] Francq C. and J-M. Zakoian (2013). Estimating the marginal law of a time series with Applications to heavy-tailed distributions, *Journal of Business and Economic Statistics*, 31(4), 412-425.

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

[16] 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.

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

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

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

[20] 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(1), 1-28.

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

[22] 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.

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

[24] Nelson D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6(3), 318-334.

[25] Mukherjee K. (2008). M-estimation in GARCH models. *Econometric Theory*, 24(6), 1530-1553.

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

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

[28] Rekkasa M. and A. Wong (2008). Implementing likelihood-based inference for fat-tailed distributions. *Finance Research Letters*, 5(1), 32-46.

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

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

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

[32] 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.

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

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