CREATES

Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2009-6

On IGARCH and convergence of the QMLE for misspecified GARCH models

Anders Tolver Jensen and Theis Lange

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Aarhus School of Business
University of Aarhus

Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF COPENHAGEN

On IGARCH and convergence of the QMLE for misspecified GARCH models

Anders Tolver Jensen
Department of Natural Sciences
University of Copenhagen

Theis Lange
Department of Economics
University of Copenhagen
&
CREATES

Abstract: We address the IGARCH puzzle by which we understand the fact that a GARCH(1,1) model fitted by quasi maximum likelihood estimation to virtually any financial dataset exhibit the property that $\hat{\alpha}+\hat{\beta}$ is close to one. We prove that if data is generated by certain types of continuous time stochastic volatility models, but fitted to a GARCH(1,1) model one gets that $\hat{\alpha}+\hat{\beta}$ tends to one in probability as the sampling frequency is increased. Hence, the paper suggests that the IGARCH effect could be caused by misspecification. The result establishes that the stochastic sequence of QMLEs do indeed behave as the deterministic parameters considered in the literature on filtering based on misspecified ARCH models, see e.g. Nelson (1992). An included study of simulations and empirical high frequency data is found to be in very good accordance with the mathematical results.

Keywords: GARCH; Integrated GARCH; Misspecification; High frequency exchange rates.
JEL codes: C13; C22; C51.

Theis Lange gratefully acknowledges the research support of grant no. 2114-04-0001 from the "Danish Social Sciences Research Council" and of CREATES (funded by the Danish National Research Foundation).

1 Introduction

A complete characterization of the volatility of financial assets has long been one of the main goals of financial econometrics. Since the seminal papers of Engle (1982) and Bollerslev (1986) the class of generalized autoregressive heteroskedastic (GARCH) models has been a key tool when modeling time dependent volatility. Indeed the GARCH(1,1) model has become so widely used that it is often referred to as "the workhorse of the industry" (Lee & Hansen 1994).

Recall that given a sequence of returns $(y_t)_{t=0,\ldots,T}$ the GARCH(1,1) model defines the conditional volatility as

$$\sigma_t^2(\theta) = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta), \tag{1}$$

for some non-negative parameters $\theta = (\omega, \alpha, \beta)'$. Quasi maximum likelihood estimation of GARCH(1,1) models on financial returns almost always indicates that $\hat{\alpha}$ is small, $\hat{\beta}$ is close to unity, and the sum of $\hat{\alpha}$ and $\hat{\beta}$ is very close to one and approaches one as the sampling frequency is increased, see e.g. Engle & Bollerslev (1986), Bollerslev & Engle (1993), Baillie, Bollerslev & Mikkelsen (1996), Ding & Granger (1996), Andersen & Bollerslev (1997), and Engle & Patton (2001). This feature seems to be present independently of the considered asset class or sample. Engle & Bollerslev (1986) proposed the integrated GARCH (IGARCH) model specifically to reflect this fact. Also in the recent literature on quasi maximum likelihood estimation in GARCH models it has been paramount to allow for $\alpha + \beta$ to be close to or even exceeding one, see e.g. Jensen & Rahbek (2004) and Francq & Zakoïan (2004). IGARCH implies that the return series is not covariance stationary and multiperiod forecasts of volatility will trend upwards. Recently it has been suggested that either long memory, see e.g. Mikosch & Stărică (2004), or parameter changes, see e.g. Hillebrand (2005), in the data generating process can give the impression of IGARCH.

In a series of seminal papers Nelson (1992), Nelson & Foster (1994), and Nelson & Foster (1995) explore the consequences of applying ARCH type filters on discrete samples from continuous time stochastic volatility models. One important result demonstrates the existence of a deterministic sequence of parameters for the GARCH(1,1) model such that the difference between the GARCH conditional volatility estimates based on (1) and the true volatility converges to zero in probability as the sampling frequency is increased. This may explain the success of ARCH type models at recovering and forecasting volatility even though they are, no doubt, misspecified. The result, however, depends on the fact that the chosen parameters have the IGARCH property in the limit and do not depend on the data. Indeed Nelson & Foster (1994) stress the need for extending their results to cover filters based on quasi maximum likelihood estimators (QMLEs) from ARCH type models. In addition a number of papers (see, e.g. Drost & Nijman (1993), Drost & Werker (1996), and Francq & Zakoïan (2000)) explore the connection between continuous time stochastic volatility models and discrete time GARCH models. They establish a link between parameters of the two classes of models and consequently suggest that one estimates parameters in the continuous time model from the GARCH estimates. However, Wang (2002) warns against applying statistical inference based on a GARCH model to its continuous time counterpart as this may lead to inefficient inference. The concerns of Nelson & Foster (1995) and Wang (2002) illustrate the need to understand better the behaviour of QMLEs for misspecified GARCH models.

In this paper we propose a simple stochastic volatility type model that enables us to study the statistical properties of the QMLE based on the GARCH(1,1) model (1). We prove in Theorem 1 that as the sampling frequency is increased certain data generating processes will spuriously lead to the conclusion of IGARCH. The employed infill asymptotics has recently been much used in the literature on realized volatility, see e.g. Andersen, Bollerslev, Diebold & Labys (2003) and Barndorff-Nielsen & Shephard (2001).

The paper also provides a more intuitive explanation of the IGARCH puzzle by exposing similarities between the GARCH model and non-parametric estimation of a volatility process, see Stărică (2003) for a related study. The GARCH model provides a filter for computing the present volatility as, roughly speaking, a weighted average of past squared observations and a constant. Examination of the weights and the shape of the quasi likelihood function makes it plausible to believe that the performance of the filter is optimized when α and β sum to one.

Finally, since the theoretical results not only establish that the sum of the GARCH parameters will tend to one, but also indicate that they will do so at a polynomial rate, an illustration using high frequency exchange rates as well as simulated data is provided. The results are found to be in remarkably good accordance with the theoretical results and furthermore indicates that Theorem 1 is valid for other models than the ones covered by the present proof.

The rest of the paper is organized as follows. Section 2 presents the main result and explores connections between the GARCH(1,1) model and non-parametric estimation of volatility. Section 3 illustrates our results by both simulations and empirical data, while Section 4 concludes and presents ideas for future research. All technical lemmas are deferred to the Appendix.

2 Main Results

Based on a large class of volatility models this section initially provides a more heuristic explanation of the IGARCH puzzle by exposing similarities between the GARCH model and non-parametric estimation of a volatility process. In the second part of the section we present a mathematical setup where these heuristic arguments can be formalized and we state our main theorem.

2.1 An Intuitive Explanation of the IGARCH Puzzle

Essentially all volatility models for a sequence $(y_t)_{t=0,\dots,T}$ can be captured by the formulation

$$y_t = \sqrt{f_t} \cdot z_t, \tag{2}$$

where z_t is a sequence of zero mean random variables with unit variance and $(f_t)_{t=0,\dots,T}$ a sequence of stochastic volatilities such that z_t is independent of $(f_t, y_{t-1}, \dots, y_0)$. Define $\sigma_t^2(\theta)$ to be the conditional variance process corresponding to the GARCH(1,1) model with parameters $\theta = (\omega, \alpha, \beta)'$

$$\begin{aligned} \sigma_t^2(\theta) &= \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta) \\ &= \omega \sum_{i=0}^{t-1} \beta^i + \alpha \sum_{i=0}^{t-1} \beta^i y_{t-1-i}^2 + \beta^t \sigma_0^2, \end{aligned} \tag{3}$$

with σ_0^2 a fixed constant. Consider the usual quasi log-likelihood function

$$l_T(\theta) = -\frac{1}{T}\sum_{t=1}^{T}(\log(\sigma_t^2(\theta)) + \frac{y_t^2}{\sigma_t^2(\theta)}) \tag{4}$$

and note that under the data generating process given by (2) the likelihood function may be rewritten as

$$l_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}(1 - z_t^2)\frac{f_t}{\sigma_t^2(\theta)} - \frac{1}{T}\sum_{t=1}^{T}(\log(\sigma_t^2(\theta)) + \frac{f_t}{\sigma_t^2(\theta)}). \tag{5}$$

Strictly speaking this is not a likelihood function, but just an objective function for the GARCH(1,1) model, but the terminology emphasizes the connection to the literature on estimation of GARCH models. Since the first term has zero mean (if finite) and the function $x \mapsto -\log(x) - a/x$ has a unique maximum at $x = a$, the decomposition (5) suggests that for a large class of data generating processes it is plausible that the likelihood function is optimized when the conditional variance process is close to the true unobserved volatility process f_t.

For large values of t the conditional variance process in (3) can be viewed as a kernel estimator of the unobserved volatility at time t with kernel weights $\alpha\beta^i, i = 0, \ldots, t-1$ on past observations $y_{t-1}^2, \ldots, y_0^2$ plus the constant $\frac{\omega}{1-\beta}$. In order for this to be an unbiased estimator of the non-constant volatility f on average over the entire sample one must in general have $\sum_{i=0}^{\infty} \alpha\beta^i = \frac{\alpha}{1-\beta} \approx 1$ and the constant $\frac{\omega}{1-\beta}$ small. Hence, when considering the conditional variance process, $\sigma_t^2(\theta)$, as a non-parametric estimator of the unobserved volatility one must have $\alpha + \beta \approx 1$ and ω small in order to avoid introducing a systematical bias. Clearly, the method above is not always the optimal way to match the conditional variance process, $\sigma_t^2(\theta)$, with the volatility process, f_t. For instance if the data generating process is in fact the GARCH(1,1) model one should choose θ to be the true parameter value and hence obtain $\sigma_t^2(\theta) = f_t$.

2.2 A Mathematical Explanation of the IGARCH Puzzle

In the following we introduce a mathematical framework allowing us to formalize the considerations above. Clearly, we cannot give unified mathematical proofs of our results covering all interesting stochastic volatility models. However, the framework below offers a compromise between flexibility of the model class and clarity of the formal mathematical arguments. Following Theorem 1 we discuss possible generalizations.

Let the continuous time process $(S_u)_{u\in[0,1]}$ be a solution to the stochastic differential equation

$$d\log S_u = \sqrt{f(u)}dW_u, \tag{6}$$

where W is a standard Brownian motion and f is a stochastic function on the unit interval, which is strictly positive, continuous, and independent of W (f could for instance be a solution to a stochastic differential equation). Usually one considers a discrete sample $(r_t)_{t=1,\ldots,T}$ of returns

$$r_t = \log S_{t/T} - \log S_{(t-1)/T},$$

and indeed many of the results by D.B. Nelson concerning misspecified ARCH models, see e.g. Nelson (1992), are based on discrete samples from models of the type captured by (6). By independence of W and f then conditionally on the volatility, f, we have that

$$r_t = \log S_{t/T} - \log S_{(t-1)/T} \sim N(0, \int_{(t-1)/T}^{t/T} f(u)du).$$

In particular, for large T the distribution of the returns (scaled by $\sqrt{T}$) will resemble the distribution of a sequence $(y_t)_{t=1,\ldots,T}$ generated by

$$y_t = \sqrt{f(t/T)} \cdot z_t, \tag{7}$$

where z_t is an i.i.d. sequence of zero mean unit variance Gaussian random variables. This justifies the use of (7) as the relevant data generating process when deriving asymptotic properties for QMLE based on high frequency sampling of the stochastic volatility model (6).

Consider the sequence of parameters $\theta_T = (0, T^{-d}, 1 - T^{-d})'$ and introduce the stochastic processes

$$h_T(u) = \sigma^2_{\lfloor Tu \rfloor}(\theta_T)$$

on $u \in [0,1]$, where $\sigma_t(\theta)$ is given by the GARCH(1,1) recursion (3). Here and throughout the paper $\lfloor x \rfloor$ denotes the integer part of x. Further, let $D([a,b])$ denote the space of càdlàg functions on the interval $[a,b]$.

Lemma 1. *Under the data generating process given by (7) and if $E[z_t^8] < \infty$ then for any $d \in]1/2, 1[$ and $\gamma \in]0,1]$ the process $h_T \xrightarrow{P} f$ in the uniform norm on $D([\gamma, 1])$ as T tends to infinity.*

From the proof of the lemma it is evident that the lower bound on the parameter d and the moment requirement on z_t are mostly technically motivated and can probably be relaxed considerably (to $d > 0$ and $E[z_t^4] < \infty$) at the price of a significantly more complex proof. However, the present version suffices for our intended application.

The lemma establishes that there exists a sequence of parameters such that the conditional variance process associated with the GARCH(1,1) model gets arbitrarily close to the unobserved volatility process when the sampling frequency is increased. The lemma is an analogue to Theorem 3.1 of Nelson (1992), however, our result is given for the uniform norm, but assuming a somewhat simpler data generating process. Note that the chosen parameter sequence is by no means the only sequence for which the result holds, see Nelson (1992), however, this choice simplifies the proof and is adequate for the intended application.

Proof of Lemma 1. All arguments are given conditionally on the realization of f. Introduce the notation $g_T(u) := \mathbb{E}[h_T(u)]$ for $u \in [0,1]$. For $\gamma, \eta > 0$

$$\begin{aligned} & \mathbb{P}(\sup_{u\in[\gamma,1]} |h_T(u) - f(u)| > \eta) \\ \leq\ & \mathbb{P}(\sup_{u\in[\gamma,1]} |h_T(u) - g_T(u)| > \eta/2) + \mathbb{P}(\sup_{u\in[\gamma,1]} |g_T(u) - f(u)| > \eta/2). \end{aligned}$$

By Lemma 3 in the Appendix the last term converges to zero as T tends to infinity. To handle the first term note that by Lemma 2 in the Appendix it holds that

$$
\begin{aligned}
& \mathbb{P}(\sup_{u\in[\gamma,1]} |h_T(u) - g_T(u)| > \eta/2) \\
= \; & \mathbb{P}(\max_{t=\lfloor T\gamma\rfloor-1,\ldots,T} |h_T(t/T) - g_T(t/T)| > \eta/2) \\
\leq \; & \sum_{t=\lfloor T\gamma\rfloor-1}^{T} \mathbb{P}(|h_T(t/T) - g_T(t/T)| > \eta/2) \\
\leq \; & A\eta^{-4}T\alpha_T^2
\end{aligned}
$$

which converges to zero as T tends to infinity since $\alpha_T = T^{-d}$ with $d > 1/2$. $\square$

Before stating our main theorem define the parameter set

$$\Theta = \{(\omega, \alpha, \beta)' \in \mathbb{R}^3 \mid 0 \leq \omega, 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1\} \tag{8}$$

and let $\hat{\theta}_T = (\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)' = \arg\max_{\theta\in\Theta} l_T(\theta)$ be the usual quasi maximum likelihood estimator based on (4).

Theorem 1. *Suppose that the data generating process is given by (7) where the stochastic function f is almost surely non-constant and $\mathbb{E}[z_t^8] < \infty$. Then the QMLE based on (4) satisfies that $(\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)' \xrightarrow{P} (0, 0, 1)'$ as T tends to infinity.*

Remark 1. The initial value σ_0^2 for the conditional volatility process $\sigma_t^2(\theta)$ does not need to be a constant. For instance Theorem 1 still holds if σ_0^2 is merely bounded in probability as T tends to infinity. This includes defining σ_0^2 as the unconditional variance of the full sample, which is implemented in many software packages.

Remark 2. When defining σ_0^2 as the unconditional variance of the full sample it is a simple consequence of the GARCH(1,1) recursion that any scaling of the observations only affects the estimate of the scale parameter ω. Hence, if the QMLE is based on the unscaled returns, r_t, from the continuous time model (6) Theorem 1 remains valid. Also in this case the result should be read as conditional on the sample path of the volatility process.

Remark 3. To facilitate the presentation we have assumed that the volatility process f is a continuous function. However, the proofs can be extended to cover a finite number of discontinuities at the price of a somewhat more cumbersome notation and Theorem 1 therefore remains valid.

Remark 4. The proof is given for the case of Gaussian innovations z_t, however, it can easily be adapted to most other distributions such as the t-distribution as long as the moment condition in Theorem 1 is met. Another generalization is to allow for some dependence in the sequence of innovations. For instance including an autoregressive structure on z_t would permit modeling leverage effects, but leads to considerably more complicated proofs.

Proof of Theorem 1. All arguments are given conditionally on the realization of f. For $\omega_U > 0$ divide the full parameter space Θ defined in (8) into the compact subset

$$\Theta_{\omega_U} := \{\theta = (\omega, \alpha, \beta)' \in \Theta \mid \omega \leq \omega_U\}$$

and its complement $\Theta^c_{\omega_U}$. Let

$$V_\epsilon(0,0,1) = \{(\omega, \alpha, \beta)' \in \Theta \mid ||(\omega, \alpha, \beta)' - (0,0,1)'|| < \epsilon\}$$

and use Lemma 6 in the Appendix to construct a finite covering

$$\cup_{i=1}^k V(\theta_i) \supset \Theta_{\omega_U} \backslash V_\epsilon(0,0,1)$$

of the compact set $\Theta_{\omega_U} \backslash V_\epsilon(0,0,1)$ with open subsets of Θ and let $\gamma_{\theta_1}, \ldots, \gamma_{\theta_k} > 0$ be constants such that according to Lemma 6

$$\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^* \in V(\theta_i)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - \gamma_{\theta_i}) = 1$$

for $i = 1, \dots, k$. With $\gamma = \min(\gamma_{\theta_1}, \dots, \gamma_{\theta_k})$ we conclude that

$$\begin{aligned} 1 &\geq \mathbb{P}(\sup_{\theta \in \Theta \setminus V_\epsilon(0,0,1)} l_T(\theta) < -\int_0^1 \log(f(u))du - 1 - \gamma) \\ &\geq \mathbb{P}(\sup_{\theta \in \cup_{i=1}^k V(\theta_i) \cup \Theta^c_{\omega_U}} l_T(\theta) < -\int_0^1 \log(f(u))du - 1 - \gamma) \\ &\geq 1 - \sum_{i=1}^k \mathbb{P}(\sup_{\theta \in V(\theta_i)} l_T(\theta) \geq -\int_0^1 \log(f(u))du - 1 - \gamma) \qquad (9) \\ &- \mathbb{P}(\sup_{\theta \in \Theta^c_{\omega_U}} l_T(\theta) \geq -\int_0^1 \log(f(u)))du - 1 - \gamma) \qquad (10) \end{aligned}$$

where by construction (9) converges to one as T tends to infinity. Further, as $\sigma_t^2(\theta) \geq \omega_U$ on $\Theta^c_{\omega_U}$ we get that

$$\sup_{\theta \in \Theta^c_{\omega_U}} l_T(\theta) = \sup_{\theta \in \Theta^c_{\omega_U}} -\frac{1}{T} \sum_{t=1}^T (\log(\sigma_t^2(\theta)) + \frac{y_t^2}{\sigma_t^2(\theta)}) \leq -\log(\omega_U)$$

hence the probability in (10) is zero if we choose ω_U large enough. By Lemma 4 in the Appendix it holds that $l_T(\theta_T) \xrightarrow{P} -\int_0^1 \log(f(u))du - 1$ and since $l_T(\hat{\theta}_T) \geq l_T(\theta_T)$ we conclude that for any $\epsilon > 0$

$$\lim_{T \to \infty} \mathbb{P}(\hat{\theta}_T \in V_\epsilon(0,0,1)) = 1.$$

$\square$

3 Illustrations

The main result (Theorem 1) establishes that for certain data generating processes the quasi maximum likelihood estimators for the GARCH(1,1) model will converge to $(0,0,1)'$ as the sampling frequency increases. In this section we illustrate the convergence results and go a step further by examining the rate of convergence as well. Based on Lemma 1 one could conjecture that $\hat{\alpha}_T$ and $1 - \hat{\beta}_T$ are proportional to T^{-d} for some $d \in (0,1)$. This assertion can be visualized by plotting $\log(\hat{\alpha}_T)$ and $\log(1 - \hat{\beta}_T)$ against $\log(T)$. If a linear relationship is found the parameter d can be estimated by ordinary least squares.

The first part of the study is based on high frequency recordings of the EUR-USD exchange rate. To increase the empirical relevance of the simulation part we use broadly applied continuous time models as data generating processes. However, formally these models do not satisfy the assumptions of Theorem 1. In this respect the simulation study actually demonstrates that the scope of the results might be extended to a wider class of models.

EUR-USD. Based on 30-minute recordings of the EUR-USD exchange rate spanning the period from the 2$^{\text{nd}}$ of February 1986 to the 30$^{\text{th}}$ of March 2007[1] log-returns are computed corresponding to 4 through 72 hour returns. This gives estimates $\hat{\theta}_T$ for T between 3,687 and 64,525.

Simulations. We consider three different simulation setups including the Heston model and the continuous GARCH model (obtained as the diffusion limit of a GARCH(1,1) model, see Nelson (1990)). The considered models can all be embedded in the formulation

$$dS_u = S_u V_u^{1/2} dW_{1u}, \quad dV_u = \kappa V_u^a(\mu - V_u)du + \sigma V_u^b dW_{2u},$$

where W_1 and W_2 are standard Brownian motions with a possibly non-zero correlation denoted by ρ. For ease of exposition we have omitted a drift term in the equation for dS_u. We will consider three configurations for the parameters a and b, corresponding to the Heston model, the continuous GARCH model, and the $3/2N$ model suggested in Christoffersen, Jacobs & Mimouni (2007). To make the simulations comparable to the empirical study we consider a fixed time span of 21 years. For the remaining parameters we choose the estimated values stated in Christoffersen et al. (2007), which are based on fitting the models to S&P-500 data. By this choice of time span and parameter values it is reasonable to compare the empirical study and the simulation study directly. GARCH(1,1) models are fitted to log-returns based on a discrete sample from the S process[2]. Table 1 summarizes the parameters (per annum) for the included models. We will con-

[1]Prior to January 1999 the series is generated from the DEM-USD exchange rate using a fixed exchange rate of 1.95583 DEM per EUR. Preceding the analysis the dataset has been cleaned as described in Andersen et al. (2003).

[2]The continuous time process is simulated by a standard Euler scheme using 10^8 data points.

Name	a	b	κ	μ	σ	ρ
Heston	0	1/2	6.5200	0.0352	0.4601	-0.7710
Continuous GARCH	0	1	3.9248	0.0408	2.7790	-0.7876
$3/2N$	1	3/2	60.1040	0.0837	12.4989	-0.7591

Table 1: Parameter values used in the simulation study.

sider log-returns corresponding to weekly through 5 minute returns, which gives estimates $\hat{\theta}_T$ for T between 1,000 and 300,000.

Figure 1 reports the correspondence between the estimates of α and T for the four setups. The conjectured linear relationship between $\log(\hat{\alpha}_T)$ and $\log(T)$ is clearly present. The corresponding plots for $1 - \hat{\beta}_T$ have been omitted since they are indistinguishable from Figure 1. In particular we have verified the IGARCH property, i.e. that $(\hat{\alpha}_T, \hat{\beta}_T) \to (0, 1)$. The estimated values for d are in all cases found to be between 0.25 and 0.5, but explaining this phenomenon is left for future research.

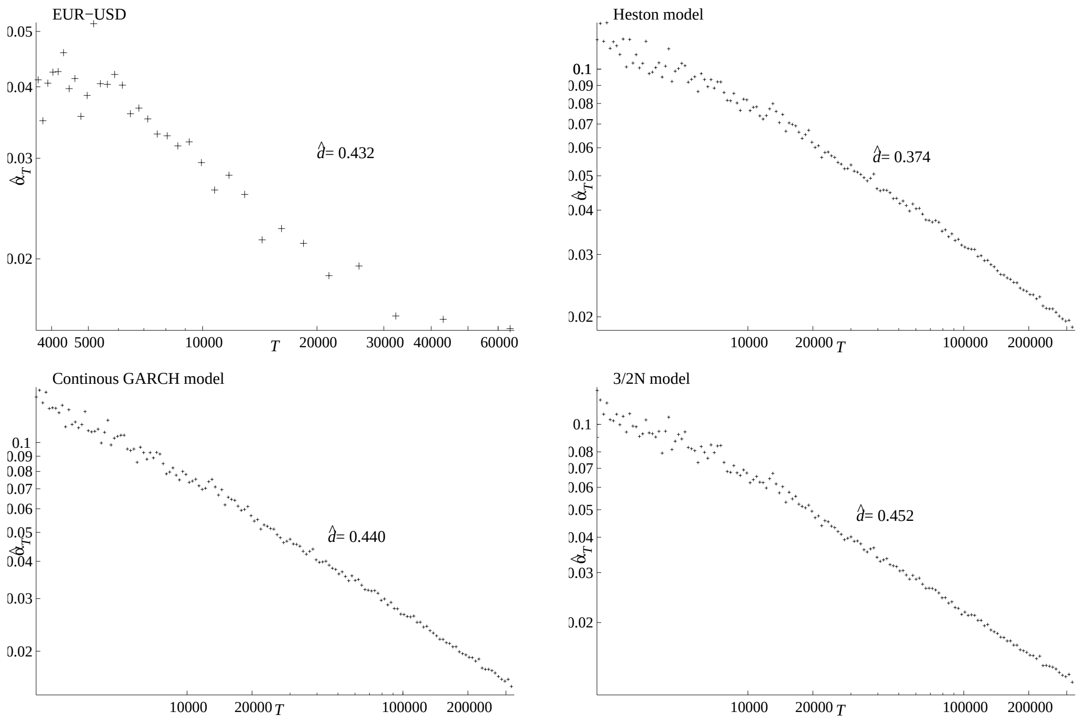

Figure 1: Correspondence between $\hat{\alpha}_T$ and T in log-scale for the four configurations. The estimate of d is obtained by regressing $\log(\hat{\alpha}_T)$ on $\log(T)$ and a constant.

The fact that none of the simulations satisfy the assumptions clearly indicates that Theorem 1 holds for a far larger class of models than those covered by the present version of our proof. This emphasizes that the IGARCH effect can be caused by the mathematical structure of a GARCH model alone and hence might not be a property of the true data generating mechanism. That the apparent polynomial convergence of the QMLEs is not only a property of the simulated series is illustrated by the striking similarities between plots based on simulated and real data.

4 Conclusion

In this paper we have established that when a GARCH(1,1) model is fitted to a discrete sample from a certain class of continuous time stochastic volatility models then the sum of the quasi maximum likelihood estimates of α and β will converge to one in probability as the sampling frequency is increased. Our results therefore indicate that the IGARCH property often found in empirical work could potentially be caused by misspecification.

The work of Nelson (1992) showed that it is possible to make the conditional variance process based on ARCH type models with deterministic parameters converge to the true unobserved volatility process. The parameters must here satisfy that $(\omega_T, \alpha_T, \beta_T) \to (0, 0, 1)$ as the number of sample points T tends to infinity. Our main result states that the same convergence holds for the *stochastic* sequence of quasi maximum likelihood estimators.

The simulations and the empirical study confirm the theoretical results and further suggest that: *i*) the assumptions of the main results may be weakened considerably and *ii*) that it may be possible to derive the exact rate of convergence of the estimators in specific mathematical frameworks. These questions are left for future research.

Appendix: Auxiliary lemmas

Since f is independent of the sequence $(z_t)_{t\in\mathbb{N}}$ all lemmas will be proved conditional on the realization of f, which can therefore be treated as a fixed strictly positive continuous function throughout.

Lemma 2. *If $\mathbb{E}[z_t^8] < \infty$ there exists some $A > 0$ such that for any $\eta > 0$*

$$\sup_{u\in[0,1]} \mathbb{P}[|h_T(u) - g_T(u)| > \eta] \leq A\eta^{-4}\alpha_T^2.$$

Proof. It follows from Chebychev's inequality that

$$\begin{aligned}
& \mathbb{P}(|h_T(u) - g_T(u)| > \eta) \\
\leq\ & \eta^{-4}\mathbb{E}[|h_T(u) - g_T(u)|^4] \\
\leq\ & \eta^{-4}\mathbb{E}[(\alpha_T \sum_{t=0}^{\lfloor Tu\rfloor-1} \beta_T^t f(\tfrac{\lfloor Tu\rfloor-1-t}{T})(z^2_{\lfloor Tu\rfloor-1-t} - 1))^4] \\
\leq\ & \eta^{-4}||f||_\infty^4\alpha_T^4\{\sum_{t=0}^{\lfloor Tu\rfloor-1} \beta_T^{4t}\kappa_4 + 6\sum_{t=1}^{\lfloor Tu\rfloor-1}\sum_{j=0}^{t-1}\beta_T^{2t+2j}\kappa_2^2\} \\
\leq\ & A_1\eta^{-4}\alpha_T^4(\sum_{t=0}^{\infty}\beta_T^{4t} + \sum_{t=1}^{\infty}\beta_T^{2t}\tfrac{1-\beta_T^{2t}}{1-\beta_T^2}),
\end{aligned}$$

where we make use of the fact that f is bounded and that $\kappa_1 = 0$ with $\kappa_r = \mathbb{E}[(z_t^2 - 1)^r]$. Evaluating the geometric series above, using that $\alpha_T = 1 - \beta_T$, and that the last expression does not depend on u one arrives at an inequality of the form stated in the lemma. $\square$

Lemma 3. *For any $\gamma > 0$ then $\sup_{u\in[\gamma,1]} |g_T(u)-f(u)| \to 0$ as T tends to infinity.*

Proof. For any sequence c_T and any $u \in [\gamma, 1]$ we get

$$\begin{aligned}
& |g_T(u) - f(u)| \\
= \; & |\beta_T^{\lfloor Tu \rfloor}\sigma_0^2 + \alpha_T \sum_{t=0}^{\lfloor Tu \rfloor - 1} \beta_T^t (f(\tfrac{\lfloor Tu \rfloor - t - 1}{T}) - f(u)) - \alpha_T \sum_{t=\lfloor Tu \rfloor}^{\infty} \beta_T^t f(u)| \\
\leq \; & \beta_T^{\lfloor Tu \rfloor}\sigma_0^2 + \alpha_T \sum_{t=0}^{c_T - 1} \beta_T^t |f(\tfrac{\lfloor Tu \rfloor - t - 1}{T}) - f(u)| + \alpha_T \sum_{t=c_T}^{\infty} \beta_T^t ||f||_\infty \\
\leq \; & \beta_T^{\lfloor T\gamma \rfloor}\sigma_0^2 + \alpha_T \frac{1 - \beta_T^{c_T}}{1 - \beta_T} \sup_{v \in [u - \frac{c_T}{T}, u]} |f(v) - f(u)| + \alpha_T \frac{\beta_T^{c_T}}{1 - \beta_T} ||f||_\infty .
\end{aligned}$$

If $c_T/T = o(1)$ the uniform continuity of f implies that the middle term can be made arbitrary small by choosing T adequately large and that the convergence is uniform over $u \in [\gamma, 1]$. To complete the proof note that

$$\log(\beta_T^{c_T}) = c_T \log(1 - T^{-d}) = -c_T T^{-d} \frac{\log(1 - T^{-d}) - \log(1)}{T^{-d}} \to -\infty$$

as T tends to infinity provided that we choose c_T so that c_T/T^d tends to infinity as T tends to infinity. $\square$

Lemma 4. *For $d > 1/2$ then*

$$l_T(\theta_T) \xrightarrow{P} -\int_0^1 \log(f(u))du - 1, \quad \textit{as} \quad T \to \infty.$$

Proof of Lemma 4. Rewriting the expression for $l_T(\theta_T)$ yields

$$l_T(\theta_T) = -\frac{1}{T} \sum_{t=1}^{T} (\log(\sigma_t^2(\theta_T)) + \frac{f(t/T)}{\sigma_t^2(\theta_T)}) \tag{11}$$

$$- \frac{1}{T} \sum_{t=1}^{T} \frac{f(t/T)}{\sigma_t^2(\theta_T)} (z_t^2 - 1) \tag{12}$$

By the law of large numbers for martingale difference sequences (12) $\xrightarrow{P} 0$. Formally, since $\mathbb{E}[z_t^2 - 1] = 0$ and $\sigma_t^2(\theta_T)$ is $\mathcal{F}_{t-1}-$measurable we get by applying

Chebechev's inequality that

$$
\begin{aligned}
& \mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta_T)}(z_t^2-1)| > \eta) \\
\leq \quad & \frac{B_1}{T^2}\sum_{i=1}^{T}\sum_{j=i}^{T}\mathbb{E}[\mathbb{E}[\frac{(z_i^2-1)(z_j^2-1)}{\sigma_i^2(\theta_T)\sigma_j^2(\theta_T)} \mid \mathcal{F}_{j-1}]] \\
= \quad & \frac{B_2}{T^2}\sum_{t=1}^{T}\mathbb{E}[\frac{1}{\sigma_t^4(\theta_T)}] \leq \frac{B_3}{T\alpha_T^2\beta_T^{2c_T}}\mathbb{E}[\frac{1}{(z_1^2+\ldots+z_{c_T}^2)^2}],
\end{aligned}
$$

where c_T is a sequence of positive integers. For T sufficiently large (Mathai & Provost (1992), p. 59)

$$
\mathbb{E}[\frac{1}{(z_1^2+\ldots+z_{c_T}^2})^2] \leq \frac{B_4}{c_T^2}
$$

hence

$$
0 \leq \limsup_{T\to\infty}\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta_T)}(z_t^2-1)| > \eta) \leq \limsup_{T\to\infty}\frac{B_5}{T\alpha_T^2\beta_T^{2c_T}c_T^2}
$$

and by choosing $c_T = \lfloor \alpha_T^{-1} \rfloor = \lfloor T^d \rfloor$ the right hand side is zero.

For any $\gamma > 0$ (11) may be written as

$$
\begin{aligned}
- \quad & \frac{1}{T}\sum_{t=1}^{\lfloor T\gamma \rfloor - 1}\log(\sigma_t^2(\theta_T)) - \frac{1}{T}\sum_{t=1}^{\lfloor T\gamma \rfloor - 1}\frac{f(t/T)}{\sigma_t^2(\theta_T)} \\
- \quad & \int_\gamma^1 \log(h_T(u))du - \int_\gamma^1 \frac{f(u)}{h_T(u)}du + \sum_{t=\lfloor T\gamma \rfloor}^{T}\int_{(t-1)/T}^{t/T}\frac{f(u)-f(t/T)}{h_T(u)}du,
\end{aligned}
$$

using that $h_T(u)$ is piecewise constant on intervals of the form $[(t-1)/T, t/T[$.

We deduce from Lemma 1 and the continuous mapping theorem that

$$\begin{aligned}
\int_\gamma^1 \log(h_T(u))du &\xrightarrow{P} \int_\gamma^1 \log(f(u))du \\
\int_\gamma^1 \frac{f(u)}{h_T(u)}du &\xrightarrow{P} 1-\gamma \\
\int_\gamma^1 \frac{1}{h_T(u)}du &\xrightarrow{P} \int_\gamma^1 \frac{1}{f(u)}du.
\end{aligned}$$

By the uniform continuity of f we conclude that

$$\sum_{t=\lfloor T\gamma\rfloor}^{T} \int_{(t-1)/T}^{t/T} \frac{f(u)-f(t/T)}{h_T(u)}du \xrightarrow{P} 0.$$

For $\eta > 0$ then

$$\begin{aligned}
\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor-1}\frac{1}{\sigma_t^2(\theta_T)}| > \eta) &\leq \mathbb{P}(\max_{t=1,\ldots,\lfloor T\gamma\rfloor-1}\frac{1}{T}\frac{1}{\sigma_t^2(\theta_T)} > \frac{\eta}{\lfloor T\gamma\rfloor - 1}) \\
&\leq \mathbb{P}(\min_{t=1,\ldots,\lfloor T\gamma\rfloor-1}\sigma_t^2(\theta_T) \leq \frac{\gamma}{B_6}) \\
&\leq \sum_{t=1}^{\lfloor T\gamma\rfloor}\mathbb{P}(\sigma_t^2(\theta_T) \leq \frac{\gamma}{B_6}).
\end{aligned}$$

Noting that $\mathbb{E}[\sigma_t^2(\theta_T)] \geq \min(\underline{f}, \sigma_0^2) \equiv \underline{\sigma}^2 > 0$ uniformly in t and T we find that for $\gamma > 0$ sufficiently small then

$$\begin{aligned}
\mathbb{P}(\sigma_t^2(\theta) \leq \frac{\gamma}{B_6}) &\leq \mathbb{P}(|\sigma_t^2(\theta_T) - \mathbb{E}[\sigma_t^2(\theta_T)]| \geq \mathbb{E}[\sigma_t^2(\theta_T)] - \frac{\gamma}{B_6}) \\
&\leq \mathbb{P}(|\sigma_t^2(\theta_T) - \mathbb{E}[\sigma_t^2(\theta_T)]| \geq \underline{\sigma}^2 - \frac{\gamma}{B_6})
\end{aligned}$$

we get by applying Lemma 2 that

$$\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor-1}\frac{1}{\sigma_t^2(\theta_T)}|>\eta)\le B_7\lfloor T\gamma\rfloor\alpha_T^2$$

which tends to zero as T tends to infinity. For $\eta>0$ given we get

$$\begin{aligned}
&\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor-1}\log(\sigma_t^2(\theta_T))|>\eta)\\
\le\ &\mathbb{P}(\max_{t=1,\dots,\lfloor T\gamma\rfloor-1}|\frac{1}{T}\log(\sigma_t^2(\theta_T))|>\frac{\eta}{\lfloor T\gamma\rfloor-1})\\
\le\ &\mathbb{P}(\max_{t=1,\dots,\lfloor T\gamma\rfloor-1}\sigma_t^2(\theta_T)\ge\exp(B_8/\gamma))+\mathbb{P}(\min_{t=1,\dots,\lfloor T\gamma\rfloor-1}\sigma_t^2(\theta_T)\le\exp(-B_8/\gamma))\\
\le\ &\sum_{t=1}^{\lfloor T\gamma\rfloor-1}\mathbb{P}(\sigma_t^2(\theta_T)\ge\exp(B_8/\gamma))+\sum_{t=1}^{\lfloor T\gamma\rfloor-1}\mathbb{P}(\sigma_t^2(\theta_T)\le\exp(-B_8/\gamma)).
\end{aligned}$$

From the previous argument we find that for $\gamma>0$ sufficiently small

$$\begin{aligned}
\mathbb{P}(\sigma_t^2(\theta_T)\ge\exp(B_8/\gamma)) &\le \mathbb{P}(|\sigma_t^2(\theta_T)-\mathbb{E}[\sigma_t^2(\theta_T)]|\ge\exp(B_8/\gamma)-\overline{\sigma}^2)\\
\mathbb{P}(\sigma_t^2(\theta_T)\le\exp(-B_8/\gamma)) &\le \mathbb{P}(|\sigma_t^2(\theta_T)-\mathbb{E}[\sigma_t^2(\theta_T)]|\ge\underline{\sigma}^2-\exp(-B_8/\gamma)),
\end{aligned}$$

where $\overline{\sigma}^2=\sigma_0^2+\|f\|_\infty$. From Lemma 2 we get that

$$\mathbb{P}(|\frac{1}{T}\sum_{t=1}^{\lfloor T\gamma\rfloor-1}\log(\sigma_t^2(\theta_T))|>\eta)\le B_9\lfloor T\gamma\rfloor\alpha_T^2$$

as T tends to infinity. $\square$

Lemma 5. *For any $\theta\in\Theta$ it holds that if f is non-constant there exists a constant $c_\theta>0$ such that*

$$\lim_{T\to\infty}\mathbb{P}(l_T(\theta)-\{-\int_0^1\log(f(u))du-1\}<-c_\theta)=1.$$

Proof of Lemma 5. Assume initially that θ is such that $\alpha\ne0$ and $\beta\ne0,1$ and rewrite the log-likelihood function as follows

$$
\begin{aligned}
& l_T(\theta) - \{-\int_0^1 \log(f(u))du - 1\} \\
= \quad & \int_0^1 \log(f(u))du - \frac{1}{T}\sum_{t=1}^{T}\log(f(t/T)) - \frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta)}(z_t^2 - 1) \qquad (13) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}\{\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}\}. \qquad (14)
\end{aligned}
$$

By the LLN for martingale differences (13) tends to zero in probability as T tends to infinity. Formally, since $\mathbb{E}[z_t^2 - 1] = 0$ and $\sigma_t^2(\theta)$ is measurable with respect to $\mathcal{F}_{t-1} = \mathcal{F}(z_0, ..., z_{t-1})$ we get by applying Chebechev's inequality that

$$
\begin{aligned}
& \mathbb{P}(|\frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta)}(z_t^2 - 1)| > \eta) \\
\leq \quad & \frac{C_1}{T^2}\sum_{i=1}^{T}\sum_{j=i}^{T}\mathbb{E}[\mathbb{E}[\frac{(z_i^2 - 1)(z_j^2 - 1)}{\sigma_i^2(\theta)\sigma_j^2(\theta)} \mid \mathcal{F}_{j-1}]] \\
= \quad & \frac{C_2}{T^2}\sum_{t=1}^{T}\mathbb{E}[\frac{1}{\sigma_t^4(\theta)}] \leq \frac{C_3}{T}\mathbb{E}[\frac{1}{(\alpha(z_5^2 + \beta z_4^2 + \ldots + \beta^4 z_1^2))^2}]
\end{aligned}
$$

and the expectation on the right hand side is finite if $\alpha, \beta > 0$ c.f. Mathai & Provost (1992).

Next turn to the expression in (14) which we decompose into

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}(\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) - \mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)})]) \qquad (15) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}(\mathbb{E}[\frac{f(t/T)}{\sigma_t^2(\theta)}] - \frac{f(t/T)}{\sigma_t^2(\theta)}) \qquad (16) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}] \qquad (17)
\end{aligned}
$$

Initially we will establish that (15) converges in probability to zero. For any

$\eta > 0$ direct calculations yield

$$\begin{aligned}
& \mathbb{P}(|T^{-1}\sum_{t=1}^{T}\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) - \mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)})]| > \eta) \\
= \;& \mathbb{P}(|T^{-1}\sum_{t=1}^{T}\log(\sigma_t^2(\theta)) - \mathbb{E}[\log(\sigma_t^2(\theta))]| > \eta) \\
\leq \;& \frac{2}{T^2\eta^2}\sum_{i=1}^{T}\sum_{j=i}^{T}|\mathrm{cov}(\log(\sigma_i^2(\theta)), \log(\sigma_j^2(\theta)))|. \qquad (18)
\end{aligned}$$

Utilizing the following inequalities

$$-\frac{1}{\sqrt{x}} \leq \log(x) \leq \sqrt{x}, \quad 0 \leq \log(1+x) \leq x,$$

which hold for all strictly positive x, it can be concluded that

$$\begin{aligned}
& |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(\sigma_j^2(\theta)))| \\
= \;& |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(\beta^{j-i}\sigma_i^2(\theta) + \underbrace{\omega\frac{1-\beta^{j-i}}{1-\beta} + \alpha\sum_{k=0}^{j-i-1}\beta^k y_{j-1-k}^2}_{:=Z(i,j)}))| \\
= \;& |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(Z(i,j)(1 + \frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)})))| \\
= \;& |\mathrm{Cov}(\log(\sigma_i^2(\theta)), \log(1 + \frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)}))| \\
\leq \;& \sqrt{\mathbb{E}[(\log(\sigma_i^2(\theta)))^2]}\sqrt{\mathbb{E}[(\log(1 + \frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)}))^2]} \\
\leq \;& \sqrt{\mathbb{E}[(\frac{1}{\sqrt{\sigma_i^2(\theta)}} + \sqrt{\sigma_i^2(\theta)})^2]}\sqrt{\mathbb{E}[(\frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)})^2]} \\
\leq \;& \beta^{j-i}\sqrt{\mathbb{E}[(\sigma_i^2(\theta) + \frac{1}{\sigma_i^2(\theta)} + 2)]}\sqrt{\mathbb{E}[\sigma_i^4(\theta)]}\sqrt{\mathbb{E}[\frac{1}{Z(i,j)^2}]}.
\end{aligned}$$

For $j > i+1$ the right hand side can be bounded by $\beta^{j-1}C_4$, where the constant C_4 does not depend on either i nor j. In the derivations it is used repeatedly that $\sigma_i^2(\theta)$ is independent of $Z(i,j)$. Since $T^{-2}\sum_{i=1}^{T}\sum_{j=i}^{T}\beta^{j-i}$ tends to zero as T

tends to infinity it can be concluded that (18) and hence also (15) tends to zero. To show that (16) tends to zero in probability note that

$$\begin{aligned}
& |\mathrm{Cov}(\frac{f(i/T)}{\sigma_i^2(\theta)}, \frac{f(j/T)}{\sigma_j^2(\theta)})| \\
= & |f(\frac{i}{T})f(\frac{j}{T})(\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}\frac{1}{\beta^{j-i}\sigma_i^2(\theta)+Z(i,j)}] - \mathbb{E}[\frac{1}{\sigma_i^2(\theta)}]\mathbb{E}[\frac{1}{\beta^{j-i}\sigma_i^2(\theta)+Z(i,j)}])| \\
\leq & f(\frac{i}{T})f(\frac{j}{T})\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}] \mid \mathbb{E}[\frac{1}{Z(i,j)}] - \mathbb{E}[\frac{1}{\beta^{j-i}\sigma_i^2(\theta)+Z(i,j)}]| \\
\leq & f(\frac{i}{T})f(\frac{j}{T})\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}]\mathbb{E}[\frac{\beta^{j-i}\sigma_i^2(\theta)}{Z(i,j)(\beta^{j-i}\sigma_i^2(\theta)+Z(i,j))}] \\
\leq & \beta^{j-i}f(\frac{i}{T})f(\frac{j}{T})\mathbb{E}[\frac{1}{\sigma_i^2(\theta)}]\mathbb{E}[\sigma_i^2(\theta)]\mathbb{E}[\frac{1}{Z(i,j)^2}].
\end{aligned}$$

As before if $j > i+4$ the expression can be bounded by $\beta^{j-1}C_5$, where the constant C_5 does not depend on either i nor j. Hence it can be concluded that (16) tends to zero. Before turning towards (17) note that for any $\eta > 0$ it holds that

$$\begin{aligned}
\mathbb{P}(\sigma_t^2(\theta) \notin [\underline{f}-\eta, \|f\|_\infty+\eta]) & \geq \mathbb{P}(\sigma_t^2(\theta) > \|f\|_\infty + \eta) \\
& \geq \mathbb{P}(\alpha\underline{f}z_t^2 > \|f\|_\infty + \eta) = C_6 > 0.
\end{aligned}$$

Furthermore since the function $x \mapsto \log(a/x) + (x-a)/x$ has a unique maximum at a with the value 0 and the function f is strictly positive and bounded there exists a constant $C_7 > 0$ such that

$$\sup_{a\in[\underline{f},\|f\|_\infty]} \sup_{x\in[0,a-\eta]\cup[a+\eta,\infty]} \log(a/x) + (x-a)/x < -C_7.$$

Finally it can be concluded that (17) can be bounded by

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}] \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T} -C_7\mathbb{P}(\sigma_t^2(\theta) \notin [\underline{f} - \eta, \|f\|_\infty + \eta]) \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T} -C_7 C_6 = -C_7 C_6 = c_\theta < 0,
\end{aligned}
$$

which verifies the claim of the lemma. For the special cases $\alpha = 0$ or $\beta = 0$ the lemma is trivially satisfied. If $\beta = 1$ the lemma follows from observing $\sigma_t^2(\theta)$ tends to infinity almost surely as t grows. $\square$

Lemma 6. *For $\theta \in \Theta \backslash (0, 0, 1)$ there exists an open subset of Θ around θ denoted $V(\theta)$ and a constant $\gamma_\theta > 0$ such that*

$$
\mathbb{P}(\sup_{\theta^* \in V(\theta)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - \gamma_\theta)
$$

tends to one as T tends to infinity.

Proof of Lemma 6. We divide the proof into seven cases mainly because we have to be very careful when θ lies on the boundary of Θ.

1. $\theta = (\omega, \alpha, \beta)' \in (0, \infty) \times [0, 1] \times [0, 1)$
2. $\theta = (\omega, \alpha, \beta)' \in (0, \infty) \times (0, 1] \times \{1\}$
3. $\theta = (\omega, \alpha, \beta)' \in (0, \infty) \times \{0\} \times \{1\}$
4. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times (0, 1] \times \{0\}$
5. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times (0, 1] \times (0, 1)$
6. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times (0, 1] \times \{1\}$
7. $\theta = (\omega, \alpha, \beta)' \in \{0\} \times \{0\} \times [0, 1)$

Case 1. Choose according to Lemma 5 a $c_\theta > 0$ such that

$$\lim_{T\to\infty} \mathbb{P}(l_T(\theta) - \{-\int_0^1 \log(f(u))du - 1\} \geq -c_\theta) = 0.$$

For $\epsilon > 0$ denote by

$$V_\epsilon(\theta) = \{\theta^* \in \Theta \mid ||\theta^* - \theta|| \leq \epsilon\}$$

and note that for T sufficiently large

$$\begin{aligned}
&\mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - c_\theta/2) \\
= \ & 1 - \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta) \geq -\int_0^1 \log(f(u))du - 1 - c_\theta/2) \\
\geq \ & 1 - \mathbb{P}(l_T(\theta) \geq -\int_0^1 \log(f(u))du - 1 - c_\theta) - \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)| \geq c_\theta/2).
\end{aligned}$$

To complete the proof we only need to show that for some sufficiently small $\epsilon > 0$ then

$$\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)| \geq c_\theta/2) = 0. \tag{19}$$

Note that this is much weaker than proving that

$$\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)|$$

converges to zero in probability since the probability in (19) should not necessarily converge to zero for this particular ϵ if c_θ is replaced by an arbitrarily small positive number. We proceed by showing that there exists a constant, $D_1 > 0$, such that for any small $\epsilon > 0$ then

$$\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)|$$

can be bounded above by something that converges in probability to $D_1\epsilon$ as T tends to infinity. In particular, the conclusion given by (19) holds for $\epsilon > 0$ such that $D_1\epsilon < c_\theta/2$.

Trivially, for ϵ sufficiently small we get the inequalities

$$\begin{aligned}
\sup_{\theta^*\in V_\epsilon(\theta)} |\beta^t - \beta^{*t}| &\leq \epsilon t(\beta+\epsilon)^{t-1} \\
\sup_{\theta^*\in V_\epsilon(\theta)} |\alpha\beta^t - \alpha^*\beta^{*t}| &\leq \epsilon\alpha t(\beta+\epsilon)^{t-1} + \epsilon(\beta+\epsilon)^t \\
\sup_{\theta^*\in V_\epsilon(\theta)} |\omega\sum_{i=0}^{t-1}\beta^i - \omega^*\sum_{i=0}^{t-1}\beta^{*i}| &\leq \epsilon\frac{1}{1-\beta} + \epsilon(\omega+\epsilon)\sum_{i=0}^{\infty} i(\beta+\epsilon)^{i-1}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq\ & D_1\epsilon + ||f||_\infty \epsilon \sum_{i=0}^{t-1} z_{t-1-i}^2 \underbrace{[\alpha i(\beta+\epsilon)^{i-1} + (\beta+\epsilon)^i]}_{:=c_i} + \epsilon t(\beta+\epsilon)^{t-1}\sigma_0^2 \qquad (20)
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{\theta^*\in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T} \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq\ & D_1\epsilon + ||f||_\infty \epsilon \sum_{t=1}^{T}\sum_{i=0}^{t-1} z_{t-1-i}^2 c_i + \frac{1}{T}\sum_{t=1}^{T} t(\beta+\epsilon)^{t-1}\sigma_0^2\epsilon \\
\leq\ & D_2\epsilon + ||f||_\infty \epsilon \{\sum_{i=0}^{\infty} c_i\} \frac{1}{T}\sum_{t=0}^{T-1} z_t^2 \xrightarrow{P} D_3\epsilon
\end{aligned}$$

as T tends to infinity. As $\sigma_t^2(\theta^*)$ is bounded below by $\omega-\epsilon$ on $V_\epsilon(\theta)$ the derivations just above demonstrate that

$$\sup_{\theta^*\in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} |\log(\sigma_t^2(\theta)) - \log(\sigma_t^2(\theta^*))| \leq \sup_{\theta^*\in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\omega-\epsilon} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)|$$

is bounded above by something that converges in probability to $D_4\epsilon$ as T tends to infinity. Consider now the decomposition

$$
\begin{aligned}
& \sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta) - l_T(\theta^*)| \\
\leq & \sup_{\theta^*\in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} |\log(\sigma_t^2(\theta)) - \log(\sigma_t^2(\theta^*))| \\
+ & ||f||_\infty \frac{1}{T}\sum_{t=1}^{T} z_t^2 \sup_{\theta^*\in V_\epsilon(\theta)} |\frac{1}{\sigma_t^2(\theta)} - \frac{1}{\sigma_t^2(\theta^*)}| \\
\leq & \sup_{\theta^*\in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\omega-\epsilon}|\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \qquad (21) \\
+ & \frac{||f||_\infty}{(\omega-\epsilon)^2}\frac{1}{T}\sum_{t=1}^{T}(z_t^2 - 1) \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \qquad (22) \\
+ & \frac{||f||_\infty}{(\omega-\epsilon)^2}\frac{1}{T}\sum_{t=1}^{T} \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)|. \qquad (23)
\end{aligned}
$$

It follows by previous computations that (21) and (23) can be bounded above by variables converging in probability to constants of the form $D\epsilon$. The remaining term (22) is a martingale difference and by (20) we find that for $\epsilon > 0$ sufficiently small

$$
\begin{aligned}
0 \leq & \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq & D_5\epsilon + D_6\epsilon \sum_{i=0}^{t-1}(z_{t-1-i}^2 - 1)c_i.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
& \mathbb{E}[(\frac{1}{T}\sum_{t=1}^{T}(z_t^2-1)\sup_{\theta^*\in V_\epsilon(\theta)}|\sigma_t^2(\theta)-\sigma_t^2(\theta^*)|)^2] \\
\leq\ & \kappa_2^2\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}[(\sup_{\theta^*\in V_\epsilon(\theta)}|\sigma_t^2(\theta)-\sigma_t^2(\theta^*)|)^2] \\
\leq\ & \frac{1}{T}D_5^2\epsilon^2+D_6^2\epsilon^2\frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}[(z_1^2-1)^2]\sum_{i=0}^{t-1}c_i^2 \\
\leq\ & \frac{1}{T}D_5^2\epsilon^2+\frac{1}{T}D_6^2\epsilon^2\kappa_2\sum_{i=0}^{\infty}c_i^2
\end{aligned}
$$

verifying that (22) tends to zero in probability which is much stronger that what we need.

Case 2 and 6. Note initially that for ϵ adequately small

$$\inf_{\theta^*\in V_\epsilon(\theta)}\sigma_t^2(\theta^*)\geq(\alpha-\epsilon)\sum_{i=0}^{t-1}(1-\epsilon)^i\underline{f}z_{t-1-i}^2\equiv\underline{\sigma}_t^2(\epsilon).$$

Hence

$$\sup_{\theta^*\in V_\epsilon(\theta)}l_T(\theta^*)=\sup_{\theta^*\in V_\epsilon(\theta)}-\frac{1}{T}\sum_{t=1}^{T}(\log(\sigma_t^2(\theta^*))+\frac{y_t^2}{\sigma_t^2(\theta^*)})\leq\frac{1}{T}\sum_{t=1}^{T}-\log(\underline{\sigma}_t^2(\epsilon)),$$

which can be bounded by

$$
\begin{aligned}
& -\log(\alpha-\epsilon)-\log(\underline{f})-k\log(1-\epsilon)-\frac{1}{T}\sum_{t=1}^{T}\log(\sum_{i=0}^{t\wedge k-1}z_{t-i-1}^2) \\
\xrightarrow{P}\ & -\log(\alpha-\epsilon)-\log(\underline{f})-k\log(1-\epsilon)-\mathbb{E}[\log(U_k)] \qquad (24)
\end{aligned}
$$

where the convergence is due to the the law of large numbers and $U_k=z_1^2+\cdots+z_k^2$. Now choose $k\in\mathbb{N}$ and ϵ so small that (24) is strictly less then $\int_0^1\log(f(u))-1du$ as desired.

Case 3. Note initially that for ϵ adequately small

$$\inf_{\theta^* \in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \geq (\omega - \epsilon) \sum_{i=0}^{t-1} (1-\epsilon)^i \equiv \underline{\underline{\sigma}}_t^2(\epsilon).$$

Hence for suitably large T

$$\sup_{\theta^* \in V_\epsilon(\theta)} l_T(\theta^*) \leq \frac{1}{T} \sum_{t=1}^{T} -\log(\underline{\underline{\sigma}}_t^2(\epsilon)) \leq -\log(\omega - \epsilon) + \log(2) + \log(\epsilon),$$

and since the right hand side converges to minus infinity as ϵ tends to zero the desired result has been established.

Case 4. Note that for ϵ sufficiently small then $\inf_{\theta^* \in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \geq (\alpha - \epsilon) y_{t-1}^2$. In particular

$$\begin{aligned}
l_T(\theta^*) &\leq -\frac{1}{T} \sum_{t=1}^{T} (\log((\alpha - \epsilon) y_{t-1}^2) + \frac{y_t^2}{\sigma_t^2(\theta^*)}) \\
&= -\log(\alpha - \epsilon) - \frac{1}{T} \sum_{t=1}^{T} \log(f(\frac{t-1}{T})) - \frac{1}{T} \sum_{t=1}^{T} \log(z_{t-1}^2) - \frac{1}{T} \sum_{t=1}^{T} \frac{y_t^2}{\sigma_t^2(\theta^*)}.
\end{aligned}$$

Now, working on a probability space where we have a doubly infinite sequence, $(z_t)_{t \in \mathbb{Z}}$, of innovations we get that

$$\begin{aligned}
\inf_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T} \sum_{t=1}^{T} \frac{y_t^2}{\sigma_t^2(\theta^*)} &\geq \frac{1}{T} \sum_{t=1}^{T} \frac{y_t^2}{\frac{\epsilon}{1-\epsilon} + (\alpha + \epsilon) \sum_{i=0}^{t-1} \epsilon^i y_{t-1-i}^2 + \epsilon^t \sigma_0^2} \\
&\geq D_7 \frac{1}{T} \sum_{t=1}^{T} \frac{z_t^2}{\epsilon + D_8 \sum_{i=0}^{t-1} \epsilon^i z_{t-1-i}^2} \\
&\geq D_7 \frac{1}{T} \sum_{t=1}^{T} \frac{z_t^2}{\epsilon + D_8 \sum_{i=0}^{\infty} \epsilon^i z_{t-1-i}^2}.
\end{aligned}$$

By the ergodic theorem the right hand side converges in probability towards its mean, and since by Fatou's lemma

$$
\begin{aligned}
& \liminf_{\epsilon\to 0} \mathbb{E}[\frac{1}{T}\sum_{t=1}^{T} \frac{z_t^2}{\epsilon + D_8 \sum_{i=0}^{\infty} \epsilon^i z_{t-1-i}^2}] \\
= \; & \liminf_{\epsilon\to 0} \mathbb{E}[\frac{z_t^2}{\epsilon + D_8 \sum_{i=0}^{\infty} \epsilon^i z_{t-1-i}^2}] \\
\geq \; & \mathbb{E}[\liminf_{\epsilon\to 0} \frac{z_t^2}{\epsilon + D_8 \sum_{i=0}^{\infty} \epsilon^i z_{t-1-i}^2}] = \mathbb{E}[\frac{z_t^2}{D z_{t-1}^2}] = +\infty
\end{aligned}
$$

we conclude that for $\epsilon > 0$ sufficiently small

$$
\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) - \{-\int_0^1 \log(f(u))du - 1\} < -1) = 1.
$$

Case 5. Since for $\epsilon > 0$ sufficiently small

$$
\begin{aligned}
& \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta^*) - \sigma_t^2(\theta)| \\
\leq \; & \frac{\epsilon}{1-(\beta+\epsilon)} + \epsilon \sum_{i=0}^{t-1} (\beta+\epsilon)^i y_{t-1-i}^2 + \alpha \sum_{i=1}^{t-1} i\epsilon(\beta+\epsilon)^{i-1} y_{t-1-i}^2 + (\beta+\epsilon)^t \sigma_0^2
\end{aligned}
$$

and for any $k \in \mathbb{N}$

$$
\inf_{\theta^*\in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \geq (\alpha-\epsilon) \sum_{i=1}^{k} (\beta-\epsilon)^i y_{t-1-i}^2
$$

we deduce from previous arguments that

$$
\begin{aligned}
& \sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)| \\
\leq \; & \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\inf_{\theta^*\in V_\epsilon(\theta)} \sigma_t^2(\theta^*)} \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta^*) - \sigma_t^2(\theta)| \\
+ \; & \frac{1}{T}\sum_{t=1}^{T} \frac{y_t^2}{(\inf_{\theta^*\in V_\epsilon(\theta)} \sigma_t^2(\theta^*))^2} \sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta^*) - \sigma_t^2(\theta)|
\end{aligned}
$$

In particular, to demonstrate that $\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)|$ is bounded in probability by ϵD we only need to work with terms of the form

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\epsilon\sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2}{(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2} \tag{25}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\alpha\epsilon\sum_{i=1}^{t-1}i(\beta+\epsilon)^{i-1} z_{t-1-i}^2}{(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2} \tag{26}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\epsilon z_t^2\sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2}{[(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2]^2} \tag{27}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\alpha\epsilon z_t^2\sum_{i=1}^{t-1}i(\beta+\epsilon)^{i-1} z_{t-1-i}^2}{[(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2]^2}. \tag{28}$$

As in the proof of Case 4 introduce a doubly infinite sequence, $(z_t)_{t\in\mathbb{Z}}$, of innovations and note that for $\rho_1,\rho_2\in(0,1)$ then by the ergodic theorem

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=0}^{\infty}i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2} \quad\xrightarrow{P}\quad \mathbb{E}[\frac{\sum_{i=0}^{\infty}i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]$$

where

$$\begin{aligned}
&\mathbb{E}[\frac{\sum_{i=0}^{\infty}i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]\\
= \quad & \sum_{i=0}^{k}\mathbb{E}[\frac{i\rho_1^i z_{t-1-i}^2}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]+\mathbb{E}[\frac{1}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]\mathbb{E}[\sum_{i=k+1}^{\infty}i\rho_1^i z_{t-1-i}^2]\\
\leq \quad & \sum_{i=1}^{k}i(\rho_1/\rho_2)^i+\mathbb{E}[\frac{1}{\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2}]\sum_{i=k+1}^{\infty}i\rho_1^i
\end{aligned}$$

and the right hand side is finite for $k\geq 5$, c.f. Mathai & Provost (1992). This shows that asymptotically for T large then (25) and (26) may be bounded above in probability by ϵD. To show that (27) and (28) may be bounded in probability by ϵD note that

$$\frac{1}{T}\sum_{t=1}^{T}\frac{z_t^2\sum_{i=0}^{\infty} i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2} \xrightarrow{P} \mathbb{E}[\frac{z_t^2\sum_{i=0}^{\infty} i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}]$$

where

$$\begin{aligned}
& \mathbb{E}[\frac{z_t^2\sum_{i=0}^{\infty} i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}] \\
\leq\ & \sum_{i=1}^{k}\mathbb{E}[\frac{i\rho_1^i z_{t-1-i}^2}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}] + \mathbb{E}[\frac{1}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}]\mathbb{E}[\sum_{i=k+1}^{\infty} i\rho_1^i z_{t-1-i}^2] \\
\leq\ & \sum_{i=1}^{k}\{\frac{1}{2}\mathbb{E}[(i\rho_1^i z_{t-1-i}^2)^2] + \frac{1}{2}\mathbb{E}[\frac{1}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^4}]\} \\
+\ & \mathbb{E}[\frac{1}{(\sum_{i=1}^{k}\rho_2^i z_{t-1-i}^2)^2}]\sum_{i=k+1}^{\infty} i\rho_1^i
\end{aligned}$$

with the right hand side finite for k large enough.

Case 7. For $\theta = (0,0,\beta)', 0 \leq \beta < 1$ and $\epsilon > 0$ small enough we get that

$$\sup_{\theta^*\in V_\epsilon(\theta)} \sigma_t^2(\theta^*) \leq \frac{1}{1-(\beta+\epsilon)}\epsilon + \epsilon||f||_\infty \sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2 + (\beta+\epsilon)^t\sigma_0^2 := \overline{\sigma}_t^2(\epsilon).$$

Using the inequality $-1/x \leq 2\log(x)$ we get that

$$\begin{aligned}
\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) &= \sup_{\theta^*\in V_\epsilon(\theta)} \frac{1}{T}\sum_{t=1}^{T}(-\log(\sigma_t^2(\theta^*)) - \frac{y_t^2}{\sigma_t^2(\theta^*)}) \\
&\leq \sup_{\theta\in V_\epsilon(\theta)} \sum_{t=1}^{T}(\log(\sigma_t^2(\theta^*)) - 2\log(y_t^2)) \\
&\leq \frac{1}{T}\sum_{t=1}^{T}(\log(\overline{\sigma}_t^2(\epsilon)) - 2\log(z_t^2) - 2\log(f(t/T))) \\
&\leq \log(\frac{1}{T}\sum_{t=1}^{T}\overline{\sigma}_t^2(\epsilon)) - \frac{2}{T}\sum_{t=1}^{T}\log(z_t^2) - \frac{2}{T}\sum_{t=1}^{T}\log(f(t/T)).
\end{aligned}$$

Clearly, the last two terms tend to a constant and since

$$\frac{1}{T}\sum_{t=1}^{T}\overline{\sigma}_t^2(\epsilon) \quad \leq \quad \frac{\epsilon}{1-(\beta+\epsilon)} + \frac{\epsilon}{1-(\beta+\epsilon)}\frac{||f||_\infty}{T}\sum_{t=1}^{T} z_t^2 + \frac{1}{T}\frac{1}{1-(\beta+\epsilon)}\sigma_0^2$$

we conclude that for $\epsilon > 0$ small and a suitable $\gamma_\theta > 0$ then

$$\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} l_T(\theta^*) < -\int_0^1 \log(f(u))du - 1 - \gamma_\theta) = 1.$$

□

References

Andersen, T.G. & T. Bollerslev (1997), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**, 115–158.

Andersen, T.G., T. Bollerslev, F.X. Diebold & P. Labys (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Baillie, R.T., T. Bollerslev & H.O. Mikkelsen (1996), 'Fractional integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **74**, 3–30.

Barndorff-Nielsen, O.E. & N. Shephard (2001), 'Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion)', *Journal of the Royal Statistical Society, Series B* **63**, 167–241.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. & R.F. Engle (1993), 'Common persistence in conditional variance', *Econometrica* **61**, 167–186.

Christoffersen, P., K. Jacobs & K. Mimouni (2007), Models for S&P-500 dynamics: Evidence from realized volatility, daily returns, and option prices. Working paper McGill Univeristy.

Ding, Z. & C.W.J. Granger (1996), 'Modeling volatility persistence of speculative returns: A new approach', *Journal of Econometrics* **73**, 185–215.

Drost, F.C. & B.J.M. Werker (1996), 'Closing the GARCH gap: Continuous time GARCH modeling', *Journal of Econometrics* **74**(1), 31–57.

Drost, F.C. & T.E. Nijman (1993), 'Temporal aggregation of GARCH processes', *Econometrica* **61**(4), 909–927.

Engle, R.F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**, 987–1007.

Engle, R.F. & A.J. Patton (2001), 'What good is a volatility model?', *Quantitative Finance* **1**, 237–245.

Engle, R.F. & T. Bollerslev (1986), 'Modelling the persistence of conditional variance', *Econometric Reviews* **5**, 1–50.

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

Francq, C. & J. Zakoïan (2000), 'Estimating weak GARCH representations', *Econometric Theory* **16**(5), 692–728.

Hillebrand, E. (2005), 'Neglecting parameter changes in GARCH models', *Journal of Econometrics* **129**, 121–138.

Jensen, S.T. & A. Rahbek (2004), 'Asymptotic inference for nonstationary GARCH', *Econometric Theory* **20**, 1203–1226.

Lee, S.W. & B. Hansen (1994), 'Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator', *Econometric Theory* **10**, 29–53.

Mathai, A.M. & S.B. Provost (1992), *Quadratic Forms in Random Variables*, Marcel Dekker, New York.

Mikosch, T. & C. Stărică (2004), 'Non-stationarities in financial time series, the long run dependence and the IGARCH effect', *Review of Economics and Statistics* **86**, 378–390.

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

Nelson, D.B. (1992), 'Filtering and forecasting with misspecified ARCH models. I. Getting the right variance with the wrong model', *Journal of Econometrics* **52**, 61–90.

Nelson, D.B. & P. Foster (1994), 'Asymptotic filtering theory for univariate ARCH models', *Econometrica* **62**(1), 1–41.

Nelson, D.B. & P. Foster (1995), 'Filtering and forecasting with misspecified ARCH models. II. Making the right forecast with the wrong model', *Journal of Econometrics* **67**(2), 303–335.

Stărică, C. (2003), Is GARCH(1,1) as good a model as the accolades of the Nobel prize would imply? Department of Mathematical Statistics, Chalmers University of Technology.

Wang, Y. (2002), 'Asymptotic nonequivalence of GARCH models and diffusions', *Annals of Statistics* **30**(3), 754–783.

Research Papers 2008

2008-54: Andrew J. Patton and Allan Timmermann: The Resolution of Macroeconomic Uncertainty: Evidence from Survey Forecast

2008-55: Carlos Capistrán and Allan Timmermann: Forecast Combination With Entry and Exit of Experts

2008-56: Carlos Capistrán and Allan Timmermann: Disagreement and Biases in Inflation Expectations

2008-57: Almut E. D. Veraart: Impact of time–inhomogeneous jumps and leverage type effects on returns and realised variances

2008-58: Dennis Kristensen and Yongseok Shin: Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood

2008-59: Per Frederiksen and Frank S. Nielsen: Testing for long memory in potentially nonstationary perturbed fractional processes

2008-60: Thomas Q. Pedersen: Intertemporal Asset Allocation with Habit Formation in Preferences: An Approximate Analytical Solution

2008-61: Jean Jacod, Mark Podolskij and Mathias Vetter: Limit theorems for moving averages of discretized processes plus noise

2008-62: Giuseppe Cavaliere, David I. Harvey, Stephen J. Leybourne and A.M. Robert Taylor: Testing for Unit Roots in the Presence of a Possible Break in Trend and Non-Stationary Volatility

2008-63: Ole E. Barndorff-Nielsen, Peter Reinhard Hansen, Asger Lunde and Neil Shephard: Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading

2009-01: Roman Frydman, Michael D. Goldberg, Søren JohanseN and Katarina Juselius: A Resolution of the Purchasing Power Parity Puzzle: Imperfect Knowledge and Long Swings

2009-02: Morten Ørregaard Nielsen: Nonparametric Cointegration Analysis of Fractional Systems With Unknown Integration Orders

2009-03: Andrés González, Kirstin Hubrich and Timo Teräsvirta: Forecasting inflation with gradual regime shifts and exogenous information

2009-4: Theis Lange: First and second order non-linear cointegration models

2009-5: Tim Bollerslev, Natalia Sizova and George Tauchen: Volatility in Equilibrium: Asymmetries and Dynamic Dependencies

2009-6: Anders Tolver Jensen and Theis Lange: On IGARCH and convergence of the QMLE for misspecified GARCH models