November 30, 2007

Addressing the IGARCH puzzle

Anders Tolver Jensen & Theis Lange

Department of Natural Sciences & Department of Mathematical Sciences

University of Copenhagen

Abstract: We address the IGARCH puzzle, by which we understand the fact that a GARCH(1,1) model fitted to virtually any financial dataset exhibit the property that $\hat{\alpha} + \hat{\beta}$ is close to one. We do this by proving that if data is generated by a stochastic volatility model but fitted to a GARCH(1,1) model one would get that $\hat{\alpha} + \hat{\beta}$ tends to one in probability as the sampling frequency is increased. We also demonstrate that the conditional variance based on the GARCH(1,1) model converges in probability to the true unobserved volatility process even when the model is misspecified. An included study of simulations and empirical high frequency data is found to be in very good accordance with the mathematical results. The paper establishes that the IGARCH effect is apparently merely a consequence of the mathematical structure of a GARCH model and not a property of the true data generating mechanism.

Keywords: GARCH; Integrated GARCH; Misspecification; High frequency exchange rates

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 modelling 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 $(y_t)_{t=0,\dots,T}$ the GARCH(1,1) model defines the conditional volatility as

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

for some non-negative parameters $\theta = (\omega, \alpha, \beta)'$. 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 sample is increased, see e.g. Engle & Bollerslev (1986), Bollerslev & Engle (1993), Baillie, Bollerslev & Mikkelsen (1996), Ding & Granger (1996), Andersen & Bollerslev (1997), Engle & Patton (2001). This phenomenon, where $\hat{\alpha} + \hat{\beta}$ is close to one, seems to be present independently of the considered asset class or sampling frequency. Engle & Bollerslev (1986) proposed the integrated GARCH (IGARCH) model specifically to reflect this fact. Also in the recent litterateur 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 this paper we prove that a very large class of data generating processes will spuriously lead to the conclusion of IGARCH. Specifically, in Theorem 2 we address the IGARCH puzzle by proving that if the returns were in fact generated by a stochastic volatility model, but fitted to a GARCH(1,1) model the estimated parameters would exhibit the IGARCH property. Furthermore, in Theorem 1 we establish that the conditional variance process based on the GARCH(1,1) model converges to the actual unobserved volatility process as the sampling frequency is increased. This result provides additional theoretical justification for the recent literature on realized volatility, see e.g. Andersen, Bollerslev, Diebold & Labys (2003) and Barndorff-Nielsen & Shephard (2001), who have found empirical evidence indicating that a GARCH(1,1) model is an impressively accurate predictor of the realized volatility and hence also of the actual unobserved volatility process.

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 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. Hence our paper contributes by addressing a very well known puzzle in financial econometrics in a mathematically rigorous way and by providing increased theoretical justification for well established practices in the industry.

The rest of the paper is organized as follows. Section 2 presents the main results in Theorem 1 and 2 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. Finally, all technical lemmas are deferred to the Appendix.

2 Main Results

Based on a very 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 two main theorems.

2.1 An Intuitive Explanation of the IGARCH Puzzle

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

$$y_t = \sqrt{f_t} \cdot z_t, \tag{1}$$

where z_t is an i.i.d. sequence of zero mean random variables with unit variance and $(f_t)_{t=0,\dots,T}$ a sequence of volatilities, which are measurable with respect to the past and may be stochastic. Define $\sigma_t^2(\theta)$ to be the conditional variance process corresponding to the GARCH(1,1) model with parameters $\theta = (\omega, \alpha, \beta)'$

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

with σ_0^2 a fixed constant. Corresponding to the GARCH(1,1) model define the usual quasi log-likelihood function

$$l_T(\theta) = -\frac{1}{T} \sum_{t=1}^{T} (\log(\sigma_t^2(\theta)) + \frac{y_t^2}{\sigma_t^2(\theta)}) \tag{3}$$

and note that under the data generating process given by (1) the likelihood function can be rewritten as

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

Strictly speaking this is not a likelihood function, but just an objective function for the GARCH(1,1) model. However, to ease comparison with the literature on estimation of GARCH models we will refer to it as the likelihood function. Since the first term is the average of a martingale difference sequence and the function $x \mapsto -\log(x) - a/x$ has a unique maximum at $x = a$, this decomposition 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 (2) can be viewed as a kernel estimator of the unobserved volatility at time t with kernel weights $\alpha\beta^i, i = 0, \dots, t-1$ on past observations $y_{t-1}^2, \dots, y_0^2$ 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 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 2 we discuss possible generalizations.

Assume that the sequence $(y_t)_{t=0,\ldots,T}$ is generated by

$$y_t = \sqrt{f(t/T)} \cdot z_t, \tag{4}$$

where f is a strictly positive continuous function on the unit interval. 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 $u \in [0, 1]$. 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 $[a, b]$.

Theorem 1. *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.*

The theorem 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.

Proof of Theorem 1. 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 2 in the Appendix the last term converges to zero as T tends to infinity. To handle the first term note that by Lemma 1 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

$$\Theta = \{(\omega, \alpha, \beta)' \in \mathbb{R}^3 \mid 0 \leq \omega, 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1\} \tag{5}$$

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 (3).

Theorem 2. *If f is non-constant and $\mathbb{E}[z_t^8] < \infty$ then $(\hat{\omega}_T, \hat{\alpha}_T, \hat{\beta}_T)' \xrightarrow{P} (0,0,1)'$ as T tends to infinity.*

Remark 1. To facilitate the presentation we have assumed that the volatility process f is a continuous function. However, the proofs of both Theorem 1 and Theorem 2 can be extended to cover a finite number of discontinuities at the price of a somewhat more cumbersome notation (Theorem 1 will in this case apply to every open interval where f is continuous).

Remark 2. The result also covers the case of a stochastic volatility f as long as the innovations z_t are independent conditionally on f. This includes Taylor's SV model as well as many diffusion based models, but excludes the GARCH(1,1) model itself for which the result obviously does not hold. When f is stochastic the result should be read as conditional on the given realization of f.

Remark 3. The initial value σ_0^2 for the conditional volatility process $\sigma_t^2(\theta)$ does not need to be a constant. For instance both theorems still hold 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 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. 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 2. For $\omega_U > 0$ divide the full parameter space Θ defined in (5) into the compact subset

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

and its complement $\Theta_{\omega_U}^c$. Let

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

and use Lemma 5 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 5

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

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

$$
\begin{aligned}
1 &\geq \mathbb{P}\Big(\sup_{\theta \in \Theta \setminus V_\epsilon(0,0,1)} l_T(\theta) < -\int_0^1 \log(f(u))du - 1 - \gamma\Big) \\
&\geq \mathbb{P}\Big(\sup_{\theta \in \cup_{i=1}^k V(\theta_i) \cup \Theta^c_{\omega_U}} l_T(\theta) < -\int_0^1 \log(f(u))du - 1 - \gamma\Big) \\
&\geq 1 - \sum_{i=1}^k \mathbb{P}\Big(\sup_{\theta \in V(\theta_i)} l_T(\theta) \geq -\int_0^1 \log(f(u))du - 1 - \gamma\Big) \qquad (6) \\
&- \mathbb{P}\Big(\sup_{\theta \in \Theta^c_{\omega_U}} l_T(\theta) \geq -\int_0^1 \log(f(u)))du - 1 - \gamma\Big) \qquad (7)
\end{aligned}
$$

where by construction (6) 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 (7) is zero if we choose ω_U large enough. By Lemma 3 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 $\epsilon > 0$

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

□

3 Illustrations

Below we illustrate the convergence results established in the previous section using both exchange rates and simulated data. The main result (Theorem 2) establishes that the quasi maximum likelihood estimators for the GARCH(1,1) will converge to $(0, 0, 1)'$ as the sampling frequency increases. However, in this section we go a step further and examine also the rate of convergence. Based on Theorem 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 examined 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 obtained by ordinary least squares.

EUR-USD Based on 30-minute recordings of the EUR-USD exchange rate spanning the period from the 2nd of February 1986 to the 30th 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.

Simulation A. Based on simulated data from (4) with $f(x) = \sin(x\pi) + 2$, $z_t \sim$ i.i.d.$N(0,1)$, and T between 500 and 10 million.

Simulation B. Based on simulated data from (4) with $f(x) = \sin(x\pi) + 2$, z_t i.i.d. standardized t-distributed with 5 degrees of freedom, and T between 500 and 10 million.

Simulation C. Based on simulated data from (4) with $f(x) = 1 + 1_{\{x \leq 1/2\}}$, $z_t \sim$ i.i.d.$N(0,1)$, and T between 500 and 10 million.

Note that the only simulation setup formally covered by our mathematical results is Simulation A. For Simulation B the moment condition $\mathbb{E}[z_t^8] < \infty$ is not met and in Simulation C the volatility process is not continuous, but as mentioned we expect Theorem 1 and Theorem 2 to hold under much weaker conditions than those stated. Unreported simulation results show that the functional form of f does not affect the conclusions. Furthermore simulation studies based on high frequency sampling from the Heston (1993) square root model with parameters matched to the S&P-500 index yield qualitatively similar results.

Figure 1 reports the correspondence between the estimates of α and T for the four setups. The conjectured linear relationship between the logarithm of $\hat{\alpha}_T$ and T is clearly present. The estimated values for d are in all configurations found to be quite close to a half, but explaining this phenomenon is left for future research. The corresponding plots for $1-\hat{\beta}_T$ have been omitted since they are indistinguishable from Figure 1 and leads to the same estimates for d.

[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).

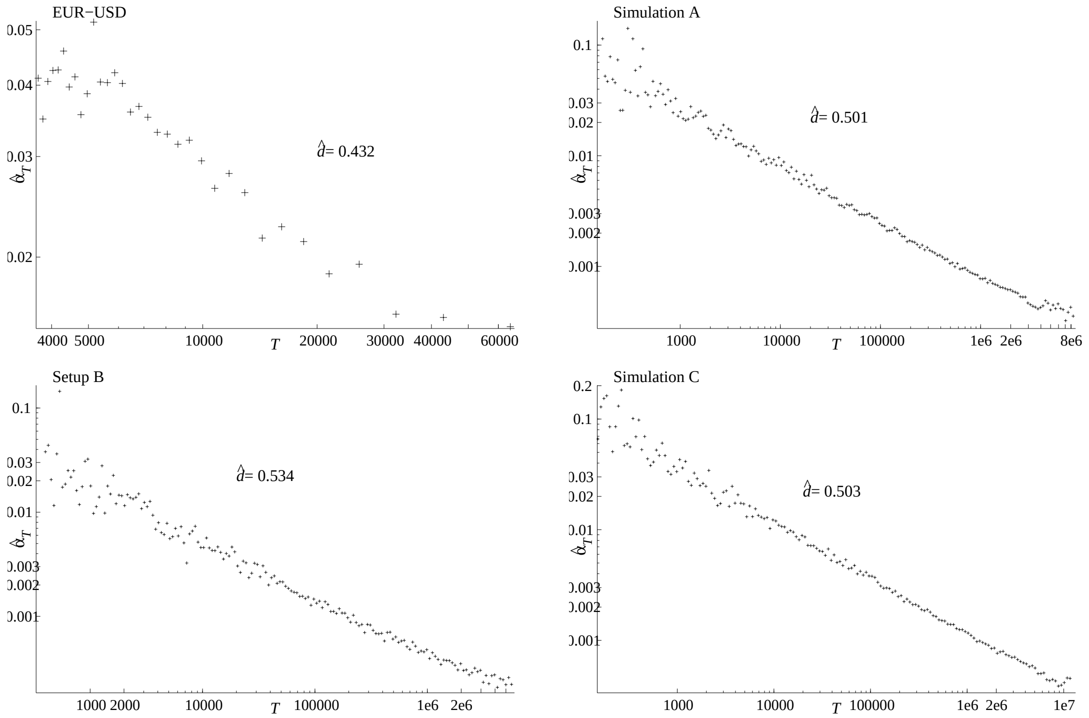

Figure 1: Correspondence between $\hat{\alpha}_T$ and T in log-scale for the four configurations. The estimate of d is obtained by regressing $\log(\hat{\alpha}_T)$ on $\log(T)$ and a constant.

The fact that only Simulation A satisfies the assumptions of the theorems clearly indicates that the results in Theorem 1 and Theorem 2 hold for a far larger class of models than for those covered by the present version of our proof. Even more important is the similarity between plots based on simulated and real data. This emphasizes that the IGARCH effect is apparently merely a consequence of the mathematical structure of a GARCH model and not a property of the true data generating mechanism.

4 Conclusion

In this paper we have established that if a GARCH(1,1) model is fitted to data generated by a wide class of stochastic volatility model then the sum of the quasi maximum likelihood estimates of α and β will converge to one in probability. Our results therefore indicate that the IGARCH property often found in empirical work may indeed be an artifact caused by misspecification. We also establish that the conditional variance process, $\sigma_t^2(\theta)$, converges to the unobserved volatility process even if the data is not generated by a GARCH(1,1) model.

The simulations and the empirical studies confirmed the theoretical results and further suggested 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.

Even though our results indicate that the IGARCH property is in fact only an artifact caused by misspecification this is by no means bad news for the wide application of the GARCH(1,1) model. Indeed our Theorem 2 establishes that the GARCH(1,1) model is a well suited filter to extract the volatility process, for a wide class of data generating processes.

Acknowledgements

This work was supported by grant no. 2114-04-0001 from the "Danish Social Sciences Research Council". Furthermore, we would like to thank Søren Johansen and Anders Rahbek for helpful comments and discussions. Part of this work was conducted while the second author visited Tim Bollerslev at Duke University and his hospitality is gratefully acknowledged. Finally, we would like to thank the participants at the CREATES Christmas meeting at University of Aarhus for their helpful suggestions.

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.

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

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.

Heston, S. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *The Review of Financial Studies* **6**, 327–343.

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.

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.

Appendix: Auxilliary lemmas

Lemma 1. *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

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

where we make use of the fact that f is bounded and that $\kappa_1 = 0$ with $k_r := \mathbb{E}[(z_t^2-1)^r]$. 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 2. *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 3. *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 3. Rewriting the expression for $l_T(\theta_T)$ yields

$$\begin{aligned}
l_T(\theta_T) \;&=\; -\frac{1}{T} \sum_{t=1}^{T} (\log(\sigma_t^2(\theta_T)) + \frac{f(t/T)}{\sigma_t^2(\theta_T)}) && (8) \\
&-\; \frac{1}{T} \sum_{t=1}^{T} \frac{f(t/T)}{\sigma_t^2(\theta_T)} (z_t^2 - 1) && (9)
\end{aligned}$$

By the law of large numbers for martingale difference sequences (9) $\xrightarrow{P} 0$. 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 $\gamma > 0$ (8) 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 Theorem 1 and the continuous mapping theorem that

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

By the uniform continuity of f we conclude that

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

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 1 that

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

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

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

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

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

where $\overline{\sigma}^2=\sigma_0^2+\|f\|_\infty$. From Lemma 1 we get that

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

as T tends to infinity. □

Lemma 4. *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 4. Assume initially that θ is such that $\alpha\neq0$ and $\beta\neq0,1$ and rewrite the log-likelihood function as follows

$$
\begin{aligned}
& l_T(\theta) - \{-\int_0^1 \log(f(u))du - 1\} \\
= \quad & \int_0^1 \log(f(u))du - \frac{1}{T}\sum_{t=1}^{T}\log(f(t/T)) - \frac{1}{T}\sum_{t=1}^{T}\frac{f(t/T)}{\sigma_t^2(\theta)}(z_t^2 - 1) \qquad (10) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}\{\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}\}. \qquad (11)
\end{aligned}
$$

By the LLN for martingale differences (10) 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 (11) which we decompose into

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}(\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) - \mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)})]) \qquad (12) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}(\mathbb{E}[\frac{f(t/T)}{\sigma_t^2(\theta)}] - \frac{f(t/T)}{\sigma_t^2(\theta)}) \qquad (13) \\
+ \quad & \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}] \qquad (14)
\end{aligned}
$$

Initially we will establish that (12) converges in probability to zero. For any

$\eta > 0$ direct calculations yield

$$\begin{aligned}
& \mathbb{P}(|T^{-1}\sum_{t=1}^{T}\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) - \mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)})]| > \eta) \\
= \; & \mathbb{P}(|T^{-1}\sum_{t=1}^{T}\log(\sigma_t^2(\theta)) - \mathbb{E}[\log(\sigma_t^2(\theta))]| > \eta) \\
\leq \; & \frac{2}{T^2\eta^2}\sum_{i=1}^{T}\sum_{j=i}^{T}|\mathrm{cov}(\log(\sigma_i^2(\theta)), \log(\sigma_j^2(\theta)))|. \qquad (15)
\end{aligned}$$

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 (15) and hence also (12) tends to zero. To show that (13) 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 (13) tends to zero. Before turning towards (14) 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 (14) can be bounded by

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\log(\frac{f(t/T)}{\sigma_t^2(\theta)}) + \frac{\sigma_t^2(\theta) - f(t/T)}{\sigma_t^2(\theta)}] \\
\leq & \frac{1}{T}\sum_{t=1}^{T} -C_7\mathbb{P}(\sigma_t^2(\theta) \notin [\underline{f} - \eta, \|f\|_\infty + \eta]) \\
\leq & \frac{1}{T}\sum_{t=1}^{T} -C_7C_6 = -C_7C_6 = c_\theta < 0,
\end{aligned}
$$

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 5. *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 5. 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 4 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

$$\lim_{T\to\infty} \mathbb{P}(\sup_{\theta^*\in V_\epsilon(\theta)} |l_T(\theta^*) - l_T(\theta)| \geq c_\theta/2) = 0. \tag{16}$$

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 (16) 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 (16) 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

$$\begin{aligned}
&\sup_{\theta^*\in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \\
\leq\ & D_1\epsilon + ||f||_\infty \epsilon \sum_{i=0}^{t-1} z_{t-1-i}^2 \underbrace{[\alpha i(\beta+\epsilon)^{i-1} + (\beta+\epsilon)^i]}_{:=c_i} + \epsilon t(\beta+\epsilon)^{t-1}\sigma_0^2 \quad (17)
\end{aligned}$$

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

$$
\begin{aligned}
& \sup_{\theta^* \in V_\epsilon(\theta)} |l_T(\theta) - l_T(\theta^*)| \\
\leq & \sup_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T} \sum_{t=1}^{T} |\log(\sigma_t^2(\theta)) - \log(\sigma_t^2(\theta^*))| \\
+ & ||f||_\infty \frac{1}{T} \sum_{t=1}^{T} z_t^2 \sup_{\theta^* \in V_\epsilon(\theta)} |\frac{1}{\sigma_t^2(\theta)} - \frac{1}{\sigma_t^2(\theta^*)}| \\
\leq & \sup_{\theta^* \in V_\epsilon(\theta)} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\omega - \epsilon} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \qquad (18) \\
+ & \frac{||f||_\infty}{(\omega - \epsilon)^2} \frac{1}{T} \sum_{t=1}^{T} (z_t^2 - 1) \sup_{\theta^* \in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)| \qquad (19) \\
+ & \frac{||f||_\infty}{(\omega - \epsilon)^2} \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta^* \in V_\epsilon(\theta)} |\sigma_t^2(\theta) - \sigma_t^2(\theta^*)|. \qquad (20)
\end{aligned}
$$

It follows by previous computations that (18) and (20) can be bounded above by variables converging in probability to constants of the form $D\epsilon$. The remaining term (19) is a martingale difference and by (17) 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 (19) 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

$$
\begin{aligned}
& -\log(\alpha-\epsilon)-\log(\underline{f})-k\log(1-\epsilon)-\frac{1}{T}\sum_{t=1}^{T}\log(\sum_{i=0}^{t\wedge k-1}z_{t-i-1}^2) \\
\xrightarrow{P}\ & -\log(\alpha-\epsilon)-\log(\underline{f})-k\log(1-\epsilon)-\mathbb{E}[\log(U_k)] \qquad (21)
\end{aligned}
$$

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 (21) 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

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

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

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\epsilon\sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2}{(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2} \tag{22}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\alpha\epsilon\sum_{i=1}^{t-1}i(\beta+\epsilon)^{i-1} z_{t-1-i}^2}{(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2} \tag{23}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\epsilon z_t^2\sum_{i=0}^{t-1}(\beta+\epsilon)^i z_{t-1-i}^2}{[(\alpha-\epsilon)\sum_{i=1}^{k}(\beta-\epsilon)^i z_{t-1-i}^2]^2} \tag{24}$$

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

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

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

where

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

and the right hand side is finite for $k\geq 5$, c.f. Mathai & Provost (1992). This shows that asymptotically for T large then (22) and (23) may be bounded above in probability by ϵD. To show that (24) and (25) 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

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

□