University of Heidelberg

Department of Economics

Discussion Paper Series | No. 579

Asymptotics for parametric GARCH-in-Mean Models

Christian Conrad
and Enno Mammen

January 2015

Asymptotics for parametric GARCH-in-Mean Models*

Christian Conrad† and Enno Mammen‡

January 13, 2015

Abstract

In this paper we develop an asymptotic theory for the parametric GARCH-in-Mean model. The asymptotics is based on a study of the volatility as a process of the model parameters. The proof makes use of stochastic recurrence equations for this random function and uses exponential inequalities to localize the problem. Our results show why the asymptotics for this specification is quite complex although it is a rather standard parametric model. Nevertheless, our theory does not yet treat all standard specifications of the mean function.

Keywords: GARCH-in-Mean, stochastic recurrence equations, risk-return relationship

JEL Classification: C13, C22, C51, G12

*We would like to thank Karin Loch and Matthias Hartmann for helpful comments and suggestions.

†Faculty of Economics and Social Studies, Heidelberg University, Bergheimer Strasse 58, 69115 Heidelberg, Germany. E-mail: christian.conrad@awi.uni-heidelberg.de

‡Institute for Applied Mathematics, Heidelberg University, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany; Laboratory of Stochastic Analysis and its Applications, Higher School of Economics, 26, Ulitsa Shabolovka, Moscow, Russia. E-mail: mammen@math.uni-heidelberg.de

1 Introduction

The aim of this paper is to develop an asymptotic theory for the Quasi-Maximum Likelihood Estimator (QMLE) in GARCH-in-Mean (GARCH-M) models for the special case of GARCH(1,1)-innovations. This model was suggested in Engle et al. (1987) and has been frequently used in empirical finance for investigating the risk-return trade-off implied by Merton's (1973) intertemporal CAPM (see, among many others, French et al., 1987 or Lundblad, 2007). Despite of its popularity in empirical applications, up to now there is no asymptotic theory for the QMLE of the GARCH-M. We will explain why the proof of the asymptotic normality of the QMLE is so difficult in this simple classical parametric model. There is also a mathematical motivation for investigating this model, because difficulties in the study of the model arise from nonstationarities of derivatives of the likelihood function that create some nonstandard mathematical problems.

The GARCH(1,1)-M model is given by

$$Y_t = m_{\gamma_0}(\bar{h}_t(\theta_0)) + \bar{h}_t(\theta_0)^{1/2} Z_t, \tag{1}$$

$$\bar{h}_t(\theta) = \omega + \alpha(Y_{t-1} - m_\gamma(\bar{h}_{t-1}(\theta)))^2 + \beta \bar{h}_{t-1}(\theta) \tag{2}$$

with i.i.d. mean zero variables Z_t with variance equal to one. Here, $\theta = (\psi, \gamma)$ is the unknown parameter, consisting of the regression parameter γ and the GARCH parameter $\psi = (\omega, \alpha, \beta)$. The true parameters are denoted by θ_0, ψ_0, γ_0, ω_0, α_0 and β_0. We also write h_t for $\bar{h}_t(\theta_0)$ and m_0 for m_{γ_0}. The function $\bar{h}_t(\theta)$ is defined as the strictly stationary and ergodic solution of (2). Below we will state conditions under which such a solution exists and is unique (see Lemma 1). The existence and uniqueness of such a solution at the value $\theta = \theta_0$ is guaranteed if $E[\ln(\alpha_0 Z_t^2 + \beta_0)] < 0$, see Nelson (1990). We also write $\hat{h}_t(\theta)$ for a solution of (2) with fixed starting value ζ_0, that is

$$\hat{h}_t(\theta) = \omega + \alpha(Y_{t-1} - m_\gamma(\hat{h}_{t-1}(\theta)))^2 + \beta \hat{h}_{t-1}(\theta) \text{ with } \hat{h}_0(\theta) \equiv \zeta_0. \tag{3}$$

In the following, the quasi-likelihood function will be based on $\hat{h}_t(\theta)$. Lee and Hansen (1994) and Lumsdaine (1996) were the first to derive the distribution theory for the QMLE of the GARCH(1,1) model. The theory has been extended to the general GARCH(p, q) case by Berkes et al. (2003) and Francq and Zakoïan (2004), among others. The result for the GARCH(1,1) can be easily extended to a GARCH(1,1) model with a constant function m_γ. Also, one can use results from the GARCH(1,1) literature to study the properties of the GARCH(1,1)-M model. For example, in Carrasco and Chen (2002) it has been shown that h_t in the GARCH(1,1) model is β-mixing with exponentially decaying mixing coefficients. A detailed discussion of the dependence structure of Y_t and h_t is provided in Conrad and Karanasos (2014). Some properties of the volatility process h_t follow directly from the ARCH(∞) representation of h_t. Christensen et al. (2012) give a complete asymptotic analysis for a GARCH(1,1)-M model with modified recurrence equation (2). In their model it is assumed that $h_t(\theta) = w + \alpha Y_{t-1}^2 + \beta h_{t-1}(\theta)$. Then, by construction, the ARCH(∞) representation of h_t does no longer depend on m. This allows them to develop a detailed theory, also for nonparametric m. Alternative estimators for a nonparametric m have been studied in Linton and Perron (2003) and Conrad and Mammen (2008).

For a parametric m function, it is standard to assume that the conditional mean can be written as $m_\gamma(x) = \gamma_1 + \gamma_2 g(x)$ with a fixed function g. The original specification of Engle et al. (1987) assumes either $g(h_t(\theta_0)) = h_t(\theta_0)$ or $g(h_t(\theta_0)) = \sqrt{h_t(\theta_0)}$, while some authors also use $g(h_t(\theta_0)) = \ln(h_t(\theta_0))$. The linear specification is directly motivated by Merton's (1973) intertemporal CAPM, which suggests that the expected excess return on the market should be proportional to the conditional variance of the market return. As noted by Pagan and Hong (1990), the log specification may be unsatisfactory, since as $h_t(\theta_0) \to 0$ the conditional variance in logs takes very large negative values and the relationship between the conditional variance and Y_t may be overstated.

In this paper we will develop an asymptotic theory for GARCH-M models. For doing so, we will assume that (2) behaves like a "stochastic contraction". Our approach will cover the specifications $g(x) = \sqrt{x}$ and $g(x) = \ln(x)$ but will not apply when $g(x) = x$.

2 Asymptotics for GARCH(1,1)-M models

Our main result is a theorem on the asymptotic normality of the QMLE $\hat{\theta}$. The proof of this result proceeds in several steps where in the first step consistency is shown. In the second step, we derive rates of convergence for $\hat{\theta}$. In the final step, this result is used to get local expansions of the quasi-likelihood function that allow to establish asymptotic normality.

In the first step, our treatment of the quasi-maximum likelihood estimator is based on a stochastic recurrence equations approach as developed in Straumann (2005) and Straumann and Mikosch (2006). In those papers, stochastic recurrence equations of the quasi-likelihood function and of its derivatives have been used to show that they have a stationary ergodic functional solution. In the GARCH(1,1)-M model, we can use these arguments to show that the quasi-likelihood function has a stationary ergodic functional solution. But this argument does not apply for the derivatives of the quasi-likelihood function, at least under reasonable assumptions. We argue that the derivatives of the quasi-likelihood functions show exploding behavior in a neighborhood of the true parameter and that they only have a stable behavior in a shrinking neighborhood. For this reason, in a second step we have to show convergence rates for $\hat{\theta}$. Having these rates, we only have to consider the derivatives of the quasi-likelihood functions in shrinking neighborhoods.

We make the following assumptions.

Assumption 1. *The parameter set Θ is compact and equal to the closure of its interior. The true parameter θ_0 lies in the interior of Θ. The function $(\gamma, u) \rightsquigarrow m_\gamma(u)$ is continuous with respect to γ and differentiable with respect to u. It holds that $\omega \geq \omega_* > 0$, $\alpha \geq \alpha_* > 0$, $\beta \geq \beta_* > 0$ for all $\theta \in \Theta$. The innovations Z_t are i.i.d. with $E[Z_t] = 0$ and $E[Z_t^2] = 1$.*

Assumption 2. *It holds that $\alpha_0 + \beta_0 < 1$.*

Assumptions 1 and 2 imply that $\sqrt{h_t} Z_t$ is a covariance-stationary process with unconditional variance equal to $\omega_0/(1 - \alpha_0 - \beta_0)$ (see Bollerslev, 1986). Further, they imply that $E[\ln(\alpha_0 Z_t^2 + \beta_0)] < 0$. As mentioned above, this guarantees that (2) has a strictly stationary and ergodic solution h_t for $\theta = \theta_0$. In the proof of consistency of the quasi-maximum likelihood estimator, we make use of the theory on stochastic recurrence equations. The essential assumption needed in this approach is stated below. In the following, we denote derivatives of functions $m_\gamma(u)$ w.r.t. the argument u by $m'_\gamma(u)$, $m''_\gamma(u)$, ... Derivatives w.r.t. the parameter γ are denoted by $\dot{m}_\gamma(u)$, $\ddot{m}_\gamma(u)$, ...

Assumption 3. *It holds that*

$$E[\ln(U_t)] < 0, \quad D_1 < +\infty, \quad D_2 < +\infty,$$

where

$$\begin{aligned} U_t &= \sup_{\alpha,\beta} \quad 2\alpha D_1 |Z_t \sqrt{h_t} + m_0(h_t)| + D_2 + \beta, \\ D_1 &= \sup_{\gamma,u} \quad |m'_\gamma(u)|, \\ D_2 &= \sup_{\gamma,u} \quad |m_\gamma(u) m'_\gamma(u)|. \end{aligned}$$

Next, we explain why this assumption naturally arises here. For this purpose, we shortly come back to the classical assumption that $E[\ln(\alpha_0 Z_t^2 + \beta_0)] < 0$. We recall why it implies that there exists a stationary ergodic solution h_t of the GARCH equation. We will later explain why Assumption 3 will be useful for similar reasons. Afterwards, we will discuss how restrictive the assumption is. We start with a brief discussion of stochastic recurrence equations.

Consider first two sequences h_t^* and h_t^{**} with different starting values $\zeta_0^* > 0$ and $\zeta_0^{**} > 0$ that fulfill the recurrence equation of $h_t = \bar{h}_t(\theta_0)$:

$$\begin{aligned} h_t^* &= \omega_0 + h_{t-1}^*(\alpha_0 Z_{t-1}^2 + \beta_0), \\ h_t^{**} &= \omega_0 + h_{t-1}^{**}(\alpha_0 Z_{t-1}^2 + \beta_0). \end{aligned}$$

Then $h_t^{**} - h_t^* = (h_{t-1}^{**} - h_{t-1}^*)(\alpha_0 Z_{t-1}^2 + \beta_0)$ and the condition $E[\ln(\alpha_0 Z_t^2 + \beta_0)] < 0$ implies that $h_t^{**} - h_t^* \to 0$ a.s.. It can be shown that this result implies that there exists a unique stationary ergodic solution of the recurrence equation of $h_t = \bar{h}_t(\theta_0)$. The approach of stochastic recurrence equations has been generalized w.r.t. two aspects: First, one can consider nonlinear recurrence equations. Then one needs conditions of the type $E[\ln(\Lambda)] < 0$ where Λ is the (random) Lipschitz constant of the recurrence equation. Second, instead of real valued random variables one can consider random elements of function spaces.

We use the second approach with the random functions

$$\bar{h}_t(\cdot) = \omega + \alpha(Y_{t-1} - m_\gamma(\bar{h}_{t-1}(\cdot)))^2 + \beta\bar{h}_{t-1}(\cdot).$$

Consider two sequences $\bar{h}_t(\cdot)^{**}$ and $\bar{h}_t^*(\cdot)$ again with different starting values $\zeta_0^{**} > 0$ and $\zeta_0^* > 0$:

$$\begin{aligned} \bar{h}_t^{**}(\cdot) &= \omega + \alpha(Y_{t-1} - m_\gamma(\bar{h}_{t-1}^{**}(\cdot)))^2 + \beta\bar{h}_{t-1}^{**}(\cdot), \\ \bar{h}_t^{*}(\cdot) &= \omega + \alpha(Y_{t-1} - m_\gamma(\bar{h}_{t-1}^{*}(\cdot)))^2 + \beta\bar{h}_{t-1}^{*}(\cdot). \end{aligned}$$

One can show the following Lipschitz inequality:

$$\|\bar{h}_t^{**}(\cdot) - \bar{h}_t^*(\cdot)\|_\infty \leq U_t\|\bar{h}_{t-1}^{**} - \bar{h}_{t-1}^*\|_\infty$$

with U_t defined above and $\|...\|_\infty$ equal to the sup-norm. In our Assumption 3, we had assumed that $E[\ln(U_t)] < 0$. This assumption implies that the recurrence equation

$$\bar{h}_t(\theta) = \omega + \alpha(Y_{t-1} - m_\gamma(\bar{h}_{t-1}(\theta)))^2 + \beta\bar{h}_{t-1}(\theta)$$

has a stationary ergodic solution $\bar{h}_t(\theta)$. This is stated in the following lemma.

Lemma 1. *Let Assumptions 1–3 be satisfied. Then (2) has a solution $\bar{h}_t(\cdot)$ that is unique, stationary and ergodic. Furthermore, it holds that there exists a $\rho > 1$ such that*

$$\rho^t \sup_{\theta\in\Theta} |\hat{h}_t(\theta) - \bar{h}_t(\theta)| \to 0, \qquad a.s. \tag{4}$$

for the random function $\hat{h}_t$ that solves (3) for $t \geq 1$ with fixed starting value $\zeta_0 > 0$.

For the convergence statement in (4), one also says that $\hat{h}_t(\cdot)$ converges exponentially fast almost surely to $\bar{h}_t(\cdot)$.

Next, we discuss that Assumption 3 is rather restrictive. It is always fulfilled if $\beta < 1$ and D_1 and D_2 are small enough. The assumption $D_2 < +\infty$ states that our function m does not grow faster than $x \to a\sqrt{x}$. The treatment of functions with faster growth would require a different approach. Consider e.g. the recurrence equation for the linear function $m_\gamma(x) = \gamma_1 + \gamma_2 x$. Here, we get that

$$\bar{h}_t(\theta) - \bar{h}_t(\theta_0) = \omega - \omega_0 + ... + \alpha\gamma_2^2[\bar{h}_{t-1}(\theta) - \bar{h}_{t-1}(\theta_0)]^2 + ...$$

It needs a very careful check why the quadratic term in the recurrence equation does not lead to an explosive behavior during $0 \leq t \leq T$. The process is not stationary and it is to be expected that the process explodes for $t \to \infty$. In order to illustrate this behavior we simulate the GARCH(1,1)-M model with the following parameters: $\alpha_0 = 0.1$, $\beta_0 = 0.85$, $\gamma_{01} = 0$ and $\gamma_{02} = 0.5$. For the process $\bar{h}_t(\theta)$, all parameters but γ_2 are chosen as in $\bar{h}_t(\theta_0)$, while $\gamma_2 \in \{0.6, 0.7, 0.8, 0.9, 1\}$. We choose $T = 3000$ and consider $M = 1000$ replications. For different values of $\Delta\gamma_2 = \gamma_2 - \gamma_{02}$, the following tables show the fraction of cases in which $\bar{h}_t(\theta) - \bar{h}_t(\theta_0)$ is diverging (defined as $\bar{h}_t(\theta) - \bar{h}_t(\theta_0) > 100$) and the average point in time t when this is happening (explosion time). As Tables 1-3 clearly show, the fraction of cases in which $\bar{h}_t(\theta) - \bar{h}_t(\theta_0)$ is diverging (# divergence) is increasing in $\Delta\gamma_2$. Also, the larger $\Delta\gamma_2$ the earlier the difference $\bar{h}_t(\theta) - \bar{h}_t(\theta_0)$ explodes. Further, by considering different values of ω_0 it becomes evident that divergence occurs more often and earlier the larger is the expected value of $h_t = \bar{h}_t(\theta_0)$.

Table 1: Comparison of $\bar{h}_t(\theta) - \bar{h}_t(\theta_0)$ when $\omega_0 = 0.05$.

$\Delta\gamma_2$	0.1	0.2	0.3	0.4	0.5
# divergence	0.02	0.50	0.99	1	1
explosion time	1403.8	1316.5	645.96	187.67	82.73

Notes: The table reports the fraction of cases in which $\bar{h}_t(\theta) - \bar{h}_t(\theta_0)$ is diverging (# divergence). For those simulations for which we observe divergence the average explosion time t is reported.

Table 2: Comparison of $\bar{h}_t(\theta) - \bar{h}_t(\theta_0)$ when $\omega_0 = 0.1$.

$\Delta\gamma_2$	0.1	0.2	0.3	0.4	0.5
# divergence	0.41	1	1	1	1
explosion time	1327.5	363.15	90.85	40.91	25.09

Notes: see Table 1.

Table 3: Comparison of $\bar{h}_t(\theta) - \bar{h}_t(\theta_0)$ when $\omega_0 = 0.2$.

$\Delta\gamma_2$	0.1	0.2	0.3	0.4	0.5
# divergence	0.99	1	1	1	1
explosion time	513.39	69.49	27.84	16.78	11.67

Notes: see Table 1.

Next, the quasi-maximum likelihood estimator $\hat{\theta}$ is defined as

$$\hat{\theta} = arg \max_{\theta \in \Theta} \quad \hat{L}_T(\theta),$$

where $\hat{L}_T(\theta)$ is the quasi-likelihood function:

$$\hat{L}_T(\theta) = -\frac{1}{2} \sum_{t=1}^{T} [\ln(\hat{h}_t(\theta)) + \hat{h}_t(\theta)^{-1} \ (Y_t - m_\gamma(\hat{h}_t(\theta)))^2]. \tag{5}$$

For the consistency of $\hat{\theta}$, we need one further assumption.

Assumption 4. *The distribution of the random variable Z_t allows for a strictly positive density on an interval $[z^*, z^{**}]$ with $z^* < z^{**}$. The following identifiability condition holds:*

$$m_\gamma(h_0) = m_0(h_0) \quad a.s. \quad \textit{if and only if } \gamma = \gamma_0.$$

The next theorem states the asymptotic consistency of the QMLE. The proof makes essential use of the ergodicity of the process $\bar{h}(\theta)$. In particular, this also implies that the quasi-likelihood function converges to its expectation.

Theorem 1. *For the model given by (1) - (2), let Assumptions 1 - 4 by satisfied. Then it holds that $\hat{\theta} \xrightarrow{P} \theta_0$.*

In a next step, we will show the asymptotic normality of the QMLE $\hat{\theta}$. Unfortunately, as mentioned above, the derivatives of the quasi-likelihood do not behave well in fixed neighborhoods of the true parameter θ_0. The basic reason is that under reasonable conditions the derivatives of $\hat{h}_t$ do not behave well at points $\theta \neq \theta_0$. Only for θ in a shrinking neighborhood of θ_0, one can control the asymptotic behavior of the derivatives. For this reason, we need a stronger result than Theorem 1. In our next theorem, we will show that $\hat{\theta}$ converges to θ_0 with nearly parametric rate $O_P(\ln(T)T^{-1/2})$. In a next step, we will show that the first two derivatives $\hat{h}_t^{(l)}(\theta)$ $(l \in \{1,2\})$ of $\hat{h}_t(\theta)$ converge to a stationary ergodic processes, uniformly over θ with $\|\theta - \theta_0\| \leq \ln(T)T^{-1/2}$. The limiting processes do not depend on θ in this shrinking neighborhood. This can be used to show asymptotic normality of the QMLE. For our next theorem, we need the following additional assumptions:

Assumption 5. *For some $D > 0$ it holds for $\|\theta - \theta_0\| \leq \delta$ that for the functions $g_1(s) = m_\gamma(s)\dot{m}_\gamma(s), g_1(s) = m_\gamma(s)m'_\gamma(s), g_2(s) = m'_\gamma(s), g_2(s) = m_\gamma(s), g_2(s) = \dot{m}_\gamma(s), g_2(s) = \dot{m}'_\gamma(s), g_2(s) = m''_\gamma(s), g_2(s) = \ddot{m}_\gamma(s), g_3(s) = \dot{m}_\gamma(s), g_4(s) = m'_\gamma(s), g_4(s) = \dot{m}'_\gamma(s)$ with some constant $D > 0$*

$$\begin{aligned}
\|g_1(s) - g_1(s')\| &\leq D|s - s'|, \\
\|g_2(s) - g_2(s')\| &\leq D|s - s'|s^{-1/2}, \\
\|g_3(s)\| &\leq D\sqrt{s}, \\
\|g_4(s)\| &\leq Ds^{-1/2}
\end{aligned}$$

for $s > s' \geq \omega_*$. Here, $\dot{f}_\gamma(s)$ and $\ddot{f}_\gamma(s)$ denote the first or second order partial derivative of a function $f_\gamma(s)$ with respect to γ.

Assumption 6. *There exists $\delta > 0, D_3 > 0$ such that for*

$$V_t = \sup_{\|\theta-\theta_0\|\leq\delta} \frac{4D_3\alpha|Z_t| + \alpha D_3^2 + 4DD_3\delta + 4\beta}{4\left(\alpha_0 Z_t^2 + \beta_0\right)}$$

it holds that

$$E\ln(V_t) < 0.$$

Here, D_3 is chosen such that

$$|m_\gamma(s) - m_\gamma(s^*)| \leq \frac{D_3}{2}\frac{|s-s^*|}{\sqrt{s}}$$

for $s, s^* \geq \omega_*$ and $\|\theta - \theta_0\| \leq \delta$. If V_t fulfills that $P[V_t > 1] > 0$, then we define $\kappa_1 > 0$ as the solution of the equation $EV_t^{\kappa_1} = 1$. If $P[V_t > 1] = 0$ we set $\kappa_1 = \infty$ and $\kappa_1^{-1} = 0$.

If V_t fulfils that $P[V_t > 1] > 0$, then there exists a unique solution $0 < \kappa_1 < \infty$ of the equation $EV_t^{\kappa_1} = 1$, see Theorem 2.1 in Mikosch and Stărică (2000).

We make the following assumption on the moments of Z_t.

Assumption 7. *The random variables Z_t fulfill the following moment condition:*

$$EZ_t^{\kappa_2} < \infty$$

for some $\kappa_2 > 4$.

Assumption 8. *The random variables Z_t fulfill the following moment condition:*

$$E(\alpha_0 Z_t^2 + \beta_0)^{\kappa_3/2} = 1$$

for some $\kappa_3 > 0$ with $2\kappa_3^{-1} + 2\kappa_1^{-1} < 1$.

Assumption 9. *The matrix $S = S_1 + S_2$ with*

$$\begin{aligned} S_1 &= E\left[\frac{\bar{h}_t'(\theta_0)\bar{h}_t'(\theta_0)^\top}{h_t^2}\right], \\ S_2 &= E\left[\frac{(\dot{m}_{\gamma_0}(h_t) + m_{\gamma_0}'(h_t)\bar{h}_t'(\theta_0))(\dot{m}_{\gamma_0}(h_t) + m_{\gamma_0}'(h_t)\bar{h}_t'(\theta_0))^\top}{h_t}\right] \end{aligned}$$

is non-singular.

We now state the following result on the convergence rate of the quasi-maximum likelihood estimator.

Proposition 1. *For the model given by (1) - (2), let Assumptions 1 - 9 with $\delta > 0$ small enough be satisfied. Then it holds that*

$$\hat{\theta} - \theta_0 = O_P\left(\ln(T)T^{-1/2}\right).$$

Proposition 1 allows to restrict attention to local expansions of the quasi-likelihood and this is the essential step to derive asymptotic normality of the maximum quasi-likelihood estimator as stated in our main theorem.

Theorem 2. *For the model (1) - (2), make the Assumptions 1 - 9 with $\delta > 0$ small enough. Then it holds that*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(\Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}),$$

where

$$\begin{aligned}
\Sigma_1 &= E\Bigg[\left\{\frac{1}{2}\frac{\bar{h}_t'}{\bar{h}_t}(Z_t^2-1) + \bar{h}_t^{-1/2}(\dot{m}_{\gamma_0}(\bar{h}_t) + m_{\gamma_0}'(\bar{h}_t)\bar{h}_t')Z_t\right\}\left\{\frac{1}{2}\frac{\bar{h}_t'}{\bar{h}_t}(Z_t^2-1) + \bar{h}_t^{-1/2}(\dot{m}_{\gamma_0}(\bar{h}_t)\right. \\
&\qquad \left. + m_{\gamma_0}'(\bar{h}_t)\bar{h}_t')Z_t\right\}^{\top}\Bigg], \\
\Sigma_2 &= E\left[\frac{1}{2}\frac{\bar{h}_t'(\bar{h}_t')^{\top}}{\bar{h}_t^2} + \frac{1}{\bar{h}_t}(\dot{m}_{\gamma_0}(\bar{h}_t) + m_{\gamma_0}'(\bar{h}_t)\bar{h}_t')(\dot{m}_{\gamma_0}(\bar{h}_t) + m_{\gamma_0}'(\bar{h}_t)\bar{h}_t')^{\top}\right].
\end{aligned}$$

Note that for Gaussian Z_t, we have $\Sigma_1 = \Sigma_2$ and the asymptotic covariance is equal to Σ_1^{-1}. On the other hand, if there is no mean function, i.e. $m_{\gamma_0}(h_t) = 0$, then Σ_1 and Σ_2 reduce to

$$\Sigma_1 = \frac{1}{2}E\left[\frac{\bar{h}_t'(\bar{h}_t')^{\top}}{\bar{h}_t^2}\right] \quad \text{and} \quad \Sigma_2 = \frac{1}{4}(E[Z_t^4]-1)E\left[\frac{\bar{h}_t'(h_t')^{\top}}{\bar{h}_t^2}\right]$$

and the covariance reduces to the one of the standard GARCH(1,1) (see Theorem 2.2 in Francq and Zakoïan, 2004).

3 Conclusions

Finding sufficient regularity conditions that ensure consistency and asymptotic normality of the QMLE for the GARCH-M model has been a long-standing problem in financial econometrics. We consider the special case of a parametric GARCH-M model with innovations that follow a GARCH(1,1) process, which is the specification most often used in empirical applications. Following Straumann (2005) and Straumann and Mikosch (2006), we make use of stochastic recurrence equations und employ exponential inequalities to show the consistency and asymptotic normality of the QMLE for certain specifications of the mean function that do not grow too fast.

Appendix

Proof of Lemma 1. Put

$$g_\theta(y,s) = \omega + \alpha(y - m_\gamma(s))^2 + \beta s \tag{6}$$

and consider the sequence of random functions ϕ_t with:

$$[\phi_t(s)](\theta) = g_\theta(Y_t, s(\theta)).$$

The functions ϕ_t map continuous functions $s : \Theta \to [0,\infty)$ onto the class of such functions. Note that

$$h_{t+1}(\theta) = [\phi_t(h_t)](\theta).$$

Because Y_t is a stationary and ergodic sequence, the same holds for ϕ_t.

Consider functions s, s^* with $\|s - s^*\|_\infty \le \delta$. It holds that

$$\begin{aligned} |\phi_t(s) - \phi_t(s^*)|(\theta) &\le 2\alpha|\sqrt{h_t}Z_t + m_{\gamma_0}(h_t)||m_\gamma(s) - m_\gamma(s^*)| + \alpha|m_\gamma(s)^2 - m_\gamma(s^*)^2| \\ &\quad +\beta|s - s^*|(\theta) \\ &\le U_t|s - s^*|(\theta). \end{aligned}$$

The lemma follows from $\mathrm{E}[\ln(U_t)] < 0$ by application of Theorem 3.1 in Bougerol (1993), see also Proposition 5.2.12 in Straumann (2005). See also the discussion before the statement of Lemma 1. □

Proof of Theorem 1. The theorem can be shown by similar arguments as in Theorem 5.3.1. in Straumann (2005). There, the proof is based on the comparison of $\hat{L}_T(\theta)$, $\bar{L}_T(\theta)$ and $L(\theta)$, where $\hat{L}_T(\theta)$ is defined in (5) and

$$\begin{aligned} \bar{L}_T(\theta) &= -\frac{1}{2}\sum_{t=1}^{T} \ln(\bar{h}_t(\theta)) + \bar{h}_t(\theta)^{-1}(Y_t - m_\gamma(\bar{h}_t(\theta)))^2, \\ L(\theta) &= -\frac{1}{2}E\left[\ln(\bar{h}_t(\theta)) + \bar{h}_t(\theta)^{-1}(Y_t - m_\gamma(\bar{h}_t(\theta)))^2\right]. \end{aligned}$$

The proof in Straumann (2005) is based on showing:

$$\frac{1}{T}||\hat{L}_T - \bar{L}_T||_\infty \to 0 \text{ (in probability)}, \quad ||\frac{1}{T}\bar{L}_T - L||_\infty \to 0 \text{ (in probability)}, \quad L(\theta) < L(\theta_0) \quad \text{for} \quad \theta \ne \theta_0.$$

The second claim follows from the fact that $\bar{h}_t$ is a stationary ergodic process (see Straumann, 2005). For the first claim, one uses the bound

$$\begin{aligned} \frac{1}{T}|\hat{L}_T(\theta) - \bar{L}_T(\theta)| &\le \frac{c}{T}\sum_{T=1}^{T}\Delta_t(\theta)\left\{\frac{(Y_t - m_\gamma(\hat{h}_t(\theta))^2}{s_t(\theta)} + \frac{(Y_t - m_\gamma(\bar{h}_t(\theta))^2}{s_t(\theta)} + 1\right\} \\ &\quad + \frac{c}{T}\sum_{t=1}^{T}\frac{|m_\gamma(\hat{h}_t(\theta)) - m_\gamma(\bar{h}_t(\theta))|\ |Y_t|}{s_t(\theta)} \\ &\quad + \frac{c}{T}\sum_{t=1}^{T}\frac{|m_\gamma(\hat{h}_t(\theta)) - m_\gamma(\bar{h}_t(\theta))|^2}{s_t(\theta)}, \end{aligned}$$

where $s_t(\theta) = \hat{h}_t(\theta) + \bar{h}_t(\theta)$, $\Delta_t(\theta) = |\hat{h}_t(\theta) - \bar{h}_t(\theta)|$ and $c > 0$ is a constant, not depending on θ. Using Assumption 3, we have $m_\gamma^2(x) \le c'\ x$, $|m_\gamma(x) - m_\gamma(y)| \le c'\ |x-y|$, $|m_\gamma(x)^2 - m_\gamma(y)^2| \le c'\ |x-y|$ with a constant $c' > 0$. Using $s_t(\theta) \ge 2\omega_*$ (see Assumption 1) we get with a constant c'':

$$\frac{1}{T}[\hat{L}_T(\theta) - \bar{L}_T(\theta)] \le \frac{c''}{T}\sum_{t=1}^{T}\Delta_t(\theta)\left\{1 + h_t + h_t\ Z_t^2 + Z_t^2\right\}. \tag{7}$$

Because of Assumptions 1 and 2, we have that $E[h_t Z_t^2] = E[h_t] < \infty$. Further, because of the ergodicity of h_t, $h_t Z_t^2$ and Z_t^2, this implies that $\frac{1}{s}\sum_{t=1}^{s}[1 + h_t + h_t Z_t^2 + Z_t^2] = O_P(1)$ for $s \to \infty$. We apply this result with $s = T$ and $s = C \ln T$ with $C > 0$ large enough. With this bound, (7) and (4), we get if C is chosen large enough that

$$\begin{aligned}\frac{1}{T}[\hat{L}_T(\theta) - \bar{L}_T(\theta)] &\le O_P\left(\frac{C\ln T}{T}\right)\frac{1}{C\ln T}\sum_{t=1}^{C\ln T}\left\{1 + h_t + h_t\ Z_t^2 + Z_t^2\right\} \\ &\quad + O_P\left(\rho^{-C\ln T}\right)\frac{1}{T}\sum_{t=1}^{T}\left\{1 + h_t + h_t\ Z_t^2 + Z_t^2\right\} \\ &= O_P\left(\frac{\ln T}{T}\right),\end{aligned} \tag{8}$$

uniformly over $\theta \in \Theta$. In particular, we have that $\frac{1}{T}[\hat{L}_t(\theta) - \bar{L}_t(\theta)]$ converges almost surely to 0, uniformly over $\theta \in \Theta$.

It remains to check the last claim: $L(\theta) < L(\theta_0)$ for $\theta \ne \theta_0$. For the proof of this claim, it suffices to check that

$$\bar{h}_0(\theta) = h_0 \text{ a.s.} \quad \text{implies } \theta = \theta_0.$$

From $\bar{h}_0(\theta) = h_0$ a.s. and stationarity of $(\bar{h}_t(\theta), h_t)$, we get that $\bar{h}_1(\theta) = h_1$ a.s.. Thus, we have that

$$0 = \omega - \omega_0 + (\alpha - \alpha_0)Y_0^2 - 2Y_0(\alpha m_\gamma(h_0) - \alpha_0 m_0(h_0)) + \alpha m_\gamma^2(h_0) - \alpha_0 m_0^2(h_0) + (\beta - \beta_0)h_0 \text{ a.s.}$$

Using Assumption 4 and considering the first two derivatives of the right hand side with respect to the y value, we get that $\alpha = \alpha_0$, $m_\gamma = m_0$ on the support of h_0 (a.s.), and $\omega = \omega_0$. □

For the proof of Proposition 1, we will make use of the following lemmas.

Lemma 2. *Make the assumptions of Proposition 1. Then with $\hat{h}_t = \hat{h}_t(\hat{\theta})$ and $\hat{m} = m_{\hat{\gamma}}$ it holds that*

$$\begin{aligned}&\frac{1}{2}\sum_{t=1}^{T}\frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2 \vee h_t^2} + \sum_{t=1}^{T}\frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t} \\ &\quad \le -\sum_{t=1}^{T}\frac{(\hat{h}_t - h_t)}{\hat{h}_t}(Z_t^2 - 1) + 2\sum_{t=1}^{T}(\hat{m}(\hat{h}_t) - m(h_t))\frac{\sqrt{h_t}}{\hat{h}_t}Z_t + O_P(\ln T),\end{aligned} \tag{9}$$

where $a \vee b$ denotes the maximum of the real numbers a and b.

Proof of Lemma 2. Note that by definition of the quasi-likelihood estimator $\hat{\theta}$, we have that $\hat{L}_T(\hat{\theta}) \ge \hat{L}_T(\theta_0)$. Because of (8), this implies that $\bar{L}_T(\hat{\theta}) \ge \bar{L}_T(\theta_0) + O_P(\ln T)$. We make use of the inequality

$\ln(1+x) \le x - x^2(1+(x)_+)^{-1}$, where $(x)_+$ is the positive part of x. This inequality follows easily from a Taylor expansion around $x = 0$. From these two inequalities, we get that

$$\begin{aligned} O_P(\ln T) &\le \frac{1}{2}\sum_{t=1}^{T} \ln\left(\frac{h_t}{\hat{h}_t}\right) - \frac{1}{2}\sum_{t=1}^{T}\left[\hat{h}_t^{-1}(h_t^{1/2}Z_t + m_0(h_t) - \hat{m}(\hat{h}_t))^2 - Z_t^2\right] \\ &\le \frac{1}{2}\sum_{t=1}^{T}\frac{h_t - \hat{h}_t}{\hat{h}_t} - \frac{1}{4}\sum_{t=1}^{T}\frac{(h_t-\hat{h}_t)^2}{\hat{h}_t^2 \vee h_t^2} - \frac{1}{2}\sum_{t=1}^{T}\frac{h_t-\hat{h}_t}{\hat{h}_t}Z_t^2 \\ &\quad - \sum_{t=1}^{T}\frac{h_t^{1/2}}{\hat{h}_t}(m_0(h_t) - \hat{m}(\hat{h}_t))Z_t - \frac{1}{2}\sum_{t=1}^{T}\hat{h}_t^{-1}(m_0(h_t) - \hat{m}(\hat{h}_t))^2. \end{aligned}$$

The claim of the lemma follows by rearrangement of the terms. □

Lemma 3. *Make the assumptions of Proposition 1. There exist random variables W_t with $\sup_{1\le t\le T}|W_t| = O_P(T^{1/\kappa_1})$ such that for $\|\theta-\theta_0\| \le \delta$*

$$\left|\frac{\bar{h}_t(\theta) - h_t}{h_t}\right| \le \|\theta-\theta_0\|W_t, \qquad (10)$$

$$\left|\frac{\hat{h}_t(\theta) - \hat{h}_t}{h_t}\right| \le \|\theta-\theta_0\|W_t. \qquad (11)$$

Proof of Lemma 3. We only show claim (10). Claim (11) follows by similar arguments. For the proof of claim (10), we show that for some constant $C > 0$ for $\|\theta-\theta_0\| \le \delta$

$$\left|\frac{\bar{h}_{t+1}(\theta) - h_{t+1}}{h_{t+1}}\right| \le C\|\theta-\theta_0\| + V_t\left|\frac{\bar{h}_t(\theta) - h_t}{h_t}\right|. \qquad (12)$$

For a proof of this claim, write $\bar{h}_t = \bar{h}_t(\theta)$ and $\Delta\omega = |\omega-\omega_0|$, $\Delta\alpha = |\alpha-\alpha_0|$ and $\Delta\beta = |\beta-\beta_0|$. We get that for some constants $C_1 > 0$ for $\|\theta-\theta_0\| \le \delta$

$$\begin{aligned} |\bar{h}_{t+1} - h_{t+1}| &\le \Delta\omega + \Delta\alpha h_t Z_t^2 + \alpha\left|(h_t^{1/2}Z_t + m_0(h_t) - m_\gamma(\bar{h}_t))^2 - h_t Z_t^2\right| + \Delta\beta h_t + \beta|\bar{h}_t - h_t| \\ &\le C_1\|\theta-\theta_0\|(1 + h_t + h_t Z_t^2) + 2\alpha|m_\gamma(h_t) - m_\gamma(\bar{h}_t)|h_t^{1/2}Z_t \\ &\qquad +2\alpha|m_\gamma(h_t) - m_\gamma(\bar{h}_t)|\;|m_\gamma(h_t) - m_0(h_t)| + \alpha|m_\gamma(h_t) - m_\gamma(\bar{h}_t)|^2 + \beta|\bar{h}_t - h_t| \\ &\le C_1\|\theta-\theta_0\|(1 + h_t + h_t Z_t^2) + |\bar{h}_t - h_t|\left[\alpha D_3 Z_t + \alpha D D_3\delta + \frac{\alpha D_3^2}{4} + \beta\right]. \end{aligned}$$

If we divide both sides of this inequality by h_{t+1}, we get equation (12), because of $h_{t+1} \ge (\alpha_0 Z_t^2 + \beta_0)h_t$. For the proof of the lemma, it remains to show that (12) implies (10). Put $\Delta_t = |\bar{h}_t - h_t|/h_t$ and define Δ_t^* as the stationary solution of the recurrence equation $\Delta_{t+1}^* = 1 + V_t\Delta_t^*$. We have that

$$\Delta_{t+1} - C\|\theta-\theta_0\|\Delta_{t+1}^* \le V_t(\Delta_t - C\|\theta-\theta_0\|\Delta_t^*).$$

This implies that $(\Delta_{t+1} - C\|\theta-\theta_0\|\Delta_{t+1}^*)_+ \le V_t(\Delta_t - C\|\theta-\theta_0\|\Delta_t^*)_+$, where $(x)_+$ denotes the positive part of x. Because $(\Delta_t - C\|\theta-\theta_0\|\Delta_t^*)_+$ is stationary and $E\ln(V_t) < 0$, we get that $(\Delta_t - C\|\theta-\theta_0\|\Delta_t^*)_+ = 0$ a.s.. Thus for $W_t = C\Delta_t^*$ we have that $\Delta_t \le \|\theta-\theta_0\|W_t$ a.s.. For the proof of the lemma, it remains to be shown that $\sup_{1\le t\le T}|W_t| = O_P(T^{1/\kappa_1})$. If $P[V_t > 1] = 0$, we can bound V_t by a random variable V_t^* with $P[V_t^* > 1] = 0$ and $E(V_t^*)^{\kappa_1^*} = 1$ with κ_1^* as large as we like. Thus, w.l.o.g. we can assume that $P[V_t > 1] > 0$. For this case we get from Theorem 4.1 in Goldie (1991) that $P(W_t \ge x) \sim cx^{-\kappa_1}$ for $x \to \infty$ for some constant $c > 0$. This implies $\sup_{1\le t\le T}|W_t| = O_P(T^{1/\kappa_1})$. Note also that V_t is bounded by definition. □

Denote by $\hat{h}_t'(\theta)$ the solution of

$$\hat{h}_{t+1}'(\theta) = \partial_\theta g_\theta(Y_t, \hat{h}_t(\theta)) + \partial_s g_\theta(Y_t, \hat{h}_t(\theta))\hat{h}_t'(\theta) \tag{13}$$

with deterministic starting value $\hat{h}_0'(\theta) = \zeta_1$. The function g_θ was defined in (6). Furthermore, $\partial_\theta g_\theta$ and $\partial_s g_\theta$ are the partial derivatives of g_θ with respect to θ or s, respectively. We also define $\hat{h}_t''(\theta)$ as the solution of

$$\hat{h}_{t+1}''(\theta) = \partial_{\theta\theta} g_\theta(Y_t, \hat{h}_t(\theta)) + 2\partial_{\theta s} g_\theta(Y_t, \hat{h}_t(\theta))\hat{h}_t'(\theta) + \partial_{ss} g_\theta(Y_t, \hat{h}_t(\theta))\hat{h}_t''(\theta) \tag{14}$$

with deterministic starting value $\hat{h}_0''(\theta) = \zeta_2$. Here $\partial_{\theta\theta} g_\theta$, $\partial_{\theta s} g_\theta$ and $\partial_{ss} g_\theta$ denote second order partial derivatives of g_θ.

The next lemma states that

$$d_{t+1}^*(\theta) = \partial_\theta g_\theta(Y_t, \bar{h}_t(\theta)) + \partial_s g_\theta(Y_t, \bar{h}_t(\theta))d_t^*(\theta) \tag{15}$$

has a unique stationary solution $d_t^*(\cdot)$. We denote this solution by $\bar{h}_t'(\cdot) = d_t^*(\cdot)$.

Lemma 4. *Make the assumptions of Proposition 1. Equation (15) has a unique stationary solution $\bar{h}_t'(\cdot) = d_t^*(\cdot)$ that is ergodic. For $\delta > 0, \rho > 1$ small enough it holds that*

$$\rho^t \sup_{\|\theta-\theta_0\|\le\delta} \|\bar{h}_t'(\theta) - \hat{h}_t'(\theta)\| \to 0, \qquad a.s.$$

Furthermore, it holds that $\bar{h}_t'$ is identical to the derivative of $\bar{h}_t$, a.s., and that it is continuous.

Proof of Lemma 4. According to Proposition 5.5.1 of Straumann (2005), it suffices for the statement of the lemma to verify that: (i) $g_\theta(y, s)$ is continuously differentiable with respect to θ and s for y fixed. (ii) For some $\kappa > 0$ and a stationary process C_t with $E[\ln^+(C_t)] < \infty$ it holds that $\|\Delta g_\theta(Y_t, s) - \Delta g_\theta(Y_t, s^*)\| \le C_t|s - s^*|^\kappa$ for $s, s^* \ge \omega_*$, $\|\theta - \theta_0\| \le \delta$. (iii) $E[\ln^+(\sup_{\|\theta-\theta_0\|\le\delta} \Delta g_\theta(Y_0, \bar{h}_0(\theta))] < \infty$.

Here, $\Delta g_\theta(y, s)$ denotes the vector of the first order derivatives of $g_\theta(y, s)$ with respect to θ and s for y fixed. We now check (i)–(iii). Condition (i) directly follows from our Assumption 1. For the check of (ii) we note that from Assumption 1 we get by direct calculations that (ii) holds with $C_t = C(1+\sqrt{h_t}+\sqrt{h_t}|Z_t|)$ if the deterministic constant C is chosen large enough. The condition $E[\ln^+(C_t)] < \infty$ follows from $EZ_t^2 < \infty$ and $Eh_t < \infty$.

For the proof of (iii), one shows the following bound for $\|\theta - \theta_0\|$ small enough and $C > 0$ large enough

$$\|\Delta g_\theta(Y_0, \bar{h}_0(\theta)) - \Delta g_{\theta_0}(Y_0, \bar{h}_0(\theta_0))\| \leq C[1 + h_0 + h_0|Z_0| + (1 + |Z_0|)|\bar{h}_0(\theta) - h_0|]$$
$$\leq C[1 + h_0 + h_0|Z_0| + (1 + |Z_0|)h_0 W_0],$$

where in the last inequality Lemma 3 has been used. Now, by direct calculations with $C^* > 0$ large enough

$$\|\Delta g_{\theta_0}(Y_0, \bar{h}_0(\theta_0))\| \leq C^*[1 + h_0 + h_0|Z_0|].$$

This gives for $\|\theta - \theta_0\|$ small enough and $C^{**} > 0$ large enough

$$\|\Delta g_\theta(Y_0, \bar{h}_0(\theta))\| \leq C^{**} h_0(1 + Z_0^2)W_0.$$

Claim (iii) follows from $E[\ln^+(h_0)] < \infty$, $E[\ln^+(1 + Z_0^2)] < \infty$ and $E[\ln^+(W_0)] < \infty$. This concludes the proof of Lemma 4. □

Lemma 5. *Make the assumptions of Proposition 1. It holds for $\rho \to 0$ that*

$$\sup_{\|\theta-\theta_0\|\leq\rho} \|\theta - \theta_0\|^{-2} \left| E\left[\frac{(\bar{h}_t(\theta) - h_t)^2}{\bar{h}_t(\theta)^2 \vee h_t^2}\right] - (\theta - \theta_0)^\top S_1(\theta - \theta_0)\right| \to 0, \tag{16}$$

$$\sup_{\|\theta-\theta_0\|\leq\rho} \|\theta - \theta_0\|^{-2} \left| E\left[\frac{(\bar{h}_t(\theta) - h_t)^2}{h_t^2}\right] - (\theta - \theta_0)^\top S_1(\theta - \theta_0)\right| \to 0, \tag{17}$$

$$\sup_{\|\theta-\theta_0\|\leq\rho} \|\theta - \theta_0\|^{-2} \left| E\left[\frac{(m_\gamma(\bar{h}_t(\theta)) - m_0(h_t))^2}{\bar{h}_t(\theta)}\right] - (\theta - \theta_0)^\top S_2(\theta - \theta_0)\right| \to 0, \tag{18}$$

$$\sup_{\|\theta-\theta_0\|\leq\rho} \|\theta - \theta_0\|^{-2} \left| E\left[\frac{(m_\gamma(\bar{h}_t(\theta)) - m_0(h_t))^2}{h_t}\right] - (\theta - \theta_0)^\top S_2(\theta - \theta_0)\right| \to 0. \tag{19}$$

Proof of Lemma 5. Choose θ_n with $\theta_n \to \theta_0$ and $\|\theta_n - \theta_0\|^{-1}(\theta_n - \theta_0) \to e$ for a unit vector e. For claim (16), we have to show that

$$\left| \|\theta_n - \theta_0\|^{-2} E\left[\frac{(\bar{h}_t(\theta_n) - h_t)^2}{\bar{h}_t^2(\theta_n) \vee h_t^2}\right] - e^\top E\left[\frac{\bar{h}'_t(\theta_0)\bar{h}'_t(\theta_0)^\top}{h_t^2}\right] e\right| \to 0.$$

For a proof of this claim first note that because of Lemma 4 we have that

$$\|\theta_n - \theta_0\|^{-2} \left[(\bar{h}_t(\theta_n) - h_t)^2(\bar{h}_t^2(\theta_n) \vee h_t^2)^{-1}\right] \to e^\top \left[\bar{h}'_t(\theta_0)\bar{h}'_t(\theta_0)^\top h_t^{-2}\right] e \text{ a.s.}$$

Thus, the claim follows by dominated convergence since

$$\|\theta_n - \theta_0\|^{-2}[(\bar{h}_t(\theta_n) - h_t)^2(\bar{h}_t(\theta_n)^2 \vee h_t^2)^{-1}] \leq W_t^2 h_t \bar{h}_t(\theta_n)^{-1} \leq W_t^2 h_t \omega_*^{-1} \leq W_t^4 \omega_*^{-1} + h_t^2 \omega_*^{-1}, \tag{20}$$

$EW_t^4 < \infty$ and $Eh_t^2 < \infty$, see also Lemma 3.

Claims (17)–(19) can be shown by similar arguments. □

Lemma 6. *Make the assumptions of Proposition 1. With some constants $c^+ > c^- > 0$, it holds that*

$$c^-\|\theta-\theta_0\|^2 \le \frac{1}{T}\sum_{t=1}^T \frac{(\bar h_t(\theta)-h_t)^2}{\bar h_t^2(\theta)\vee h_t^2} + \frac{1}{T}\sum_{t=1}^T \frac{(m_\gamma(\bar h_t(\theta))-m(h_t))^2}{\bar h_t(\theta)} \le c^+\|\theta-\theta_0\|^2$$

for all $\|\theta-\theta_0\|\le\delta$ with probability tending to one.

Proof of Lemma 6. Put $\varphi_t(\theta_0)=0$ and define for $\theta\neq\theta_0$

$$\begin{aligned}
\varphi_t(\theta) &= \|\theta-\theta_0\|^{-2}\left[\frac{(\bar h_t(\theta)-h_t)^2}{\bar h_t^2(\theta)\vee h_t^2} + \frac{(m_\gamma(\bar h_t(\theta))-m(h_t))^2}{\bar h_t(\theta)} - (\theta-\theta_0)^\top R_t(\theta-\theta_0)\right],\\
R_t &= \frac{\bar h_t'(\theta_0)\bar h_t'(\theta_0)^\top}{h_t^2} + \frac{(\dot m_{\gamma_0}(h_t)+m'_{\gamma_0}(h_t)\bar h_t'(\theta_0))(\dot m_{\gamma_0}(h_t)+m'_{\gamma_0}(h_t)\bar h_t'(\theta_0))^\top}{h_t}.
\end{aligned}$$

Because of Lemmas 1 and 4, we have that φ_t is ergodic and stationary. Using the bound (20) for the first term of φ_t and a similar bound for the second term we get that $E[\sup_{\|\theta-\theta_0\|\le\delta}|\varphi_t(\theta)|]<\infty$. Thus, we have that

$$\sup_{\|\theta-\theta_0\|\le\delta}\left|\frac{1}{T}\sum_{t=1}^T\varphi_t(\theta)-E[\varphi_t(\theta)]\right| = o_P(1).$$

From Lemma 5, we know that $\sup_{\|\theta-\theta_0\|\le\rho}\|E[\varphi_t(\theta)]\|\to 0$ for $\rho\to 0$. Here, $\|\cdot\|$ denotes the spectral norm of a matrix, i.e. $\theta\to E[\varphi_t(\theta)]$ is continuous in $\theta=\theta_0$. The statement of the lemma now follows from $\frac{1}{T}\sum_{t=1}^T R_t = S+o_P(1)$, see Lemma 4, and our assumption that S is non-singular. Here, we make the assumption that δ is chosen small enough. □

Our next lemma contains an exponential inequality for martingales. This inequality is a modification of e.g. Lemma 8.9 in van de Geer (2000).

Lemma 7. *For random variables $...,e_{-1},e_0,e_1,...,e_T$ suppose that e_t is $\mathcal{F}_t$-measurable for an increasing σ-field $\mathcal{F}_t$, that $\mathbf{E}[e_t|\mathcal{F}_{t-1}]=0$ and that $\sup_t\mathbf{E}[\exp(c|e_t|)|\mathcal{F}_{t-1}]<\infty$ (a.s.) for a constant $c>0$ small enough. Consider a sequence of random variables $r_1,r_2,...$ where r_t is measurable with respect to the σ-field generated by $\mathcal{F}_{t-1}$. Assume that $\max_{1\le t\le T}|r_t|\le c/2$ (a.s.). Then it holds that*

$$\mathbf{E}\left[\exp\left(\sum_{t=1}^T r_te_t\right)\right] \le \left\{\mathbf{E}\left[\exp(C\sum_{t=1}^T r_t^2)\right]\right\}^{1/2},$$

where C is a deterministic a.s. bound of $\mathbf{E}\left[2e_t^2\exp(c|e_t|)\,|\mathcal{F}_{t-1}\right]$.

We will make use of this lemma in the proof of the following lemma. For completeness we will give a proof of Lemma 7, although proofs of related versions of the result must be available elsewhere.

Proof of Lemma 7. We will show that for $0\le s\le T$

$$\mathbf{E}\left[\exp\left(\sum_{t=1}^T r_te_t\right)\right] \le \left\{\mathbf{E}\left[\exp\left(\sum_{t=1}^s r_te_t\right)\sqrt{\mathbf{E}_{s+1}\left[\exp(C\sum_{t=s+1}^T r_t^2)\right]}\right]\right\}, \qquad (21)$$

where $\mathbf{E}_{s+1}[...]$ denotes the conditional expectation $\mathbf{E}[...|\mathcal{F}_s]$. Note that claim (21) with $s = 0$ implies the statement of the lemma because of

$$\mathbf{E}\left[\sqrt{\mathbf{E}_1\left[\exp(C\sum_{t=1}^{T} r_t^2)\right]}\right] \leq \left\{\mathbf{E}\left[\mathbf{E}_1\left[\exp(C\sum_{t=1}^{T} r_t^2)\right]\right]\right\}^{1/2} = \left\{\mathbf{E}\left[\exp(C\sum_{t=1}^{T} r_t^2)\right]\right\}^{1/2}.$$

Furthermore, (21) with $s = T$ holds trivially. We will show that (21) for $s = u+1$ implies that (21) holds for $s = u$, where $u = 1, ..., T-1$. Thus by an induction argument we get (21) with $s = 0$ and this implies the statement of the lemma.

Suppose that (21) with $s = u + 1$ for some $u = 1, ..., T - 1$. then we get by application of the Cauchy-Schwartz inequality that

$$\begin{aligned}
\mathbf{E}\left[\exp\left(\sum_{t=1}^{T} r_t e_t\right)\right] &\leq \mathbf{E}\left[\exp\left(\sum_{t=1}^{u+1} r_t e_t\right)\sqrt{\mathbf{E}_{u+2}\left[\exp(C\sum_{t=u+2}^{T} r_t^2)\right]}\right] \\
&= \mathbf{E}\left[\mathbf{E}_{u+1}\left[\exp\left(\sum_{t=1}^{u+1} r_t e_t\right)\sqrt{\mathbf{E}_{u+2}\left[\exp(C\sum_{t=u+2}^{T} r_t^2)\right]}\right]\right] \\
&\leq \mathbf{E}\left[\left\{\mathbf{E}_{u+1}\left[\exp\left(\sum_{t=1}^{u+1} 2r_t e_t\right)\right]\right\}^{1/2}\left\{\mathbf{E}_{u+1}\left[\mathbf{E}_{u+2}\left[\exp(C\sum_{t=u+2}^{T} r_t^2)\right]\right]\right\}^{1/2}\right] \\
&= \mathbf{E}\left[\exp\left(\sum_{t=1}^{u} r_t e_t\right)\left\{\mathbf{E}_{u+1}\left[\exp\left(2r_{u+1}e_{u+1}\right)\right]\right\}^{1/2}\left\{\mathbf{E}_{u+1}\left[\exp(C\sum_{t=u+2}^{T} r_t^2)\right]\right\}^{1/2}\right].
\end{aligned}$$

We now argue that

$$\mathbf{E}_{u+1}\left[\exp\left(2r_{u+1}e_{u+1}\right)\right] \leq \exp(Cr_{u+1}^2). \tag{22}$$

If one plugs this into the last inequality one gets (21) with $s = u$. This shows the statement of the lemma. Thus it remains to show (22). This claim follows by a simple Taylor expansion. One gets with $|\eta_{u+1}| \leq |r_{u+1}|\ |e_{u+1}| \leq c/2\ |e_{u+1}|$ that

$$\begin{aligned}
\mathbf{E}_{u+1}\left[\exp\left(2r_{u+1}e_{u+1}\right)\right] &= \mathbf{E}_{u+1}\left[1 + 2r_{u+1}e_{u+1} + 2r_{u+1}^2 e_{u+1}^2 \exp\left(2\eta_{u+1}\right)\right] \\
&= \mathbf{E}_{u+1}\left[1 + 2r_{u+1}^2 e_{u+1}^2 \exp\left(2\eta_{u+1}\right)\right] \\
&\leq \mathbf{E}_{u+1}\left[1 + Cr_{u+1}^2\right] \\
&= 1 + Cr_{u+1}^2 \\
&\leq \exp(Cr_{u+1}^2).
\end{aligned}$$

□

Lemma 8. *Make the assumptions of Proposition 1. It holds that*

$$\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)}{\hat{h}_t}(Z_t^2 - 1) + \sum_{t=1}^{T} (\hat{m}(\hat{h}_t) - m(h_t)) \frac{\sqrt{h_t}}{\hat{h}_t} Z_t \tag{23}$$

$$= O_P(\ln(T)) \left[\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2} + \sum_{t=1}^{T} \frac{h_t(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t^2} \right]^{1/2},$$

$$\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)}{\hat{h}_t}(Z_t^2 - 1) + \sum_{t=1}^{T} (\hat{m}(\hat{h}_t) - m(h_t)) \frac{\sqrt{h_t}}{\hat{h}_t} Z_t \tag{24}$$

$$= O_P(\ln(T) T^{1/\kappa_3}) \left[\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2 \vee h_t^2} + \sum_{t=1}^{T} \frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t} \right]^{1/2}.$$

Proof of Lemma 8. We will show that for $\delta > 0$ small enough

$$\sup_{\|\theta - \theta_0\| \le \delta} \left[\sum_{t=1}^{T} \frac{(\bar{h}_t(\theta) - h_t)^2}{\bar{h}_t(\theta)^2} + \sum_{t=1}^{T} \frac{h_t(m_\gamma(\bar{h}_t(\theta)) - m(h_t))^2}{\bar{h}_t(\theta)^2} \right]^{-1/2} \tag{25}$$

$$\times \left[\sum_{t=1}^{T} \frac{(\bar{h}_t(\theta) - h_t)}{\bar{h}_t(\theta)}(Z_t^2 - 1) + \sum_{t=1}^{T} (m_\gamma(\bar{h}_t(\theta)) - m(h_t)) \frac{\sqrt{h_t}}{\bar{h}_t(\theta)} Z_t \right] = O_P(\ln(T)).$$

Because of Lemma 1 and consistency of $\hat{\theta}$ this implies (23).

For the proof of (25) we will apply Lemma 7 with $e_t = e_t^* - E[e_t^*]$, $e_t^* = (Z_t^2 - 1) I[|Z_t| \le T^{1/\kappa_2}]$ and $\bar{r}_t(\theta) = r_t^*(\theta) I[|r_t^*(\theta)| \le c T^{2/\kappa_1}]$, $r_t^*(\theta) = \|\theta - \theta_0\|^{-1}(\bar{h}_t(\theta) - h_t)/h_t$ for θ in a δ-neighborhood of θ_0. Note that with probability tending to one $e_t = e_t^*$ for $t = 1, ..., T$. Furthermore, we have that for all $\varepsilon > 0$ the constant c can be chosen such that with probability $\ge 1 - \varepsilon$ it holds that $r_t(\theta) = r_t^*(\theta)$ for $t = 1, ..., T$ and for all θ in a δ-neighborhood of θ_0. We now show that for $\rho > 0$, one can choose a constant $c_\rho > 0$ such that

$$P\left(\sum_{t=1}^{T} \bar{r}_t(\theta) e_t > c_\rho \ln(T) \left[T + \sum_{t=1}^{T} \bar{r}_t(\theta)^2 \right]^{1/2} \right) \le C T^{-\rho} \tag{26}$$

for $T \ge T_0$ with constant C not depending on θ and ε and T_0 not depending on θ. For a proof of (26), we use the inequality

$$P\left(\sum_{t=1}^{T} \bar{r}_t(\theta) e_t > c_\rho \ln(T) \left[T + \sum_{t=1}^{T} \bar{r}_t(\theta)^2 \right]^{1/2} \right)$$
$$\le \mathbf{E}\left[\exp\left(\sum_{t=1}^{T} r_t(\theta) e_t \right) \right] \exp\left(-c_\rho \ln(T) \right)$$

with $r_t(\theta) = \left[T + \sum_{t=1}^{T} \bar{r}_t(\theta)^2 \right]^{-1/2} \bar{r}_t(\theta)$. Application of the last lemma with $r_t = r_t(\theta)$ gives (26).

In a second step, we apply Lemma 7 with $e_t = e_t^* - E[e_t^*]$, $e_t^* = Z_t I[|Z_t| \le T^{1/\kappa_2}]$ and $\bar{r}_t(\theta) = r_t^*(\theta) I[|r_t^*(\theta)| \le c T^{2/\kappa_1}]$, $r_t^*(\theta) = \|\theta - \theta_0\|^{-1}(m_\gamma(\bar{h}_t(\theta)) - m(h_t)) \frac{\sqrt{h_t}}{\bar{h}_t(\theta)} Z_t$ for θ in a δ-neighborhood of θ_0. Also, with these choices (26) holds.

We now note that it suffices to show (25) with the supremum running only over a grid of polynomially many θ-values. This follows by using rough estimates for neighbored values of θ. Thus, (25) follows from equation (26) with the two choices of e_t and $\bar{r}_t(\theta)$. At this stage, also Lemma 6 is used. This concludes the proof of (23). Claim (24) follows from (23) by using the bound $\sup_{1\le t\le T} h_t/\hat{h}_t \le \omega_*^{-2} \sup_{1\le t\le T} h_t = O_P(T^{1/\kappa_3})$. Here, the last equality follows from Assumption 8, see Theorem 2.1 in Mikosch and Stărică (2000) and the arguments at the end of the proof of Lemma 3. □

Proof of Proposition 1. From Lemmas 2 and 8, we get that

$$\begin{aligned}
&\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2 \vee h_t^2} + \sum_{t=1}^{T} \frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t} \\
&\quad \le - \sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)}{\hat{h}_t}(Z_t^2 - 1) + 2\sum_{t=1}^{T} (\hat{m}(\hat{h}_t) - m(h_t)) \frac{\sqrt{h_t}}{\hat{h}_t} Z_t + O_P(\ln T) \\
&\quad \le O_P(\ln(T) T^{1/\kappa_3}) \left\{ \left[\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2 \vee h_t^2} \right]^{1/2} + \left[\sum_{t=1}^{T} \frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t} \right]^{1/2} \right\}.
\end{aligned}$$

This implies that

$$\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2 \vee h_t^2} + \sum_{t=1}^{T} \frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t} = O_P(\ln(T)^2 T^{2/\kappa_3}).$$

Because of Lemma 6, this shows that

$$\|\hat{\theta} - \theta_0\|^2 = O_P(\ln(T)^2 T^{-1+(2/\kappa_3)}).$$

With Lemma 3, we get from this bound that

$$\left| \frac{\hat{h}_t - h_t}{h_t} \right| = O_P(\ln(T)^2 T^{-(1/2)+(1/\kappa_1)+(1/\kappa_3)}) = o_P(1).$$

Thus,

$$\left[\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2 \vee h_t^2} + \sum_{t=1}^{T} \frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t} \right] \left[\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{h_t^2} + \sum_{t=1}^{T} h_t \frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t^2} \right]^{-1} \to 1,$$

in probability and we get by using the above arguments with (23) instead of (24) that

$$\sum_{t=1}^{T} \frac{(\hat{h}_t - h_t)^2}{\hat{h}_t^2 \vee h_t^2} + \sum_{t=1}^{T} \frac{(\hat{m}(\hat{h}_t) - m(h_t))^2}{\hat{h}_t} = O_P(\ln(T)^2).$$

Because of Lemma 6, this proves the statement of the proposition. □

For the proof of Theorem 2 we will make use of the following expansions for functions $\theta \to s(\theta)$:

$$\partial_\theta g_\theta(y, s(\theta)) = \begin{pmatrix} 1 \\ v^2 \\ s \\ -2\alpha v\dot{m} \end{pmatrix} + (\beta - 2\alpha v m')s', \tag{27}$$

$$\partial_{\theta\theta} g_\theta(y, s(\theta)) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2v\dot{m} \\ 0 & 0 & 0 & 0 \\ 0 & -2v\dot{m} & 0 & 2\alpha\dot{m}\dot{m}^\top - 2\alpha v\ddot{m} \end{pmatrix} + \begin{pmatrix} 0 \\ -4vm' \\ 2 \\ 4\alpha m'\dot{m} - 4\alpha v\dot{m}' \end{pmatrix} s' + (\beta - 2\alpha v m')s''. \tag{28}$$

Here, we denote by ∂_θ and $\partial_{\theta\theta}$ the first and second order partial derivatives with respect to θ. Furthermore, we define $v = y - m_\gamma(s)$ and we write s instead of $s(\theta)$. These equations can be used to show for $h_t(\theta) = \hat{h}_t(\theta)$ or $h_t(\theta) = \bar{h}_t(\theta)$ that

$$\begin{aligned} \left\| h'_{t+1}(\theta) - h'_{t+1}(\theta_0) \right\| \leq\; & C\, [\|\theta - \theta_0\|(1 + h_t(\theta_0)|Z_t| + h_t + h'_t(\theta_0)|Z_t|) \\ & + |h_t(\theta) - h_t(\theta_0)|(1 + |Z_t| + h'_t(\theta_0)|Z_t|) \\ & + \|h'_t(\theta) - h'_t(\theta_0)\|\, V_t(\alpha_0 Z_t^2 + \beta_0)] \end{aligned} \tag{29}$$

for some positive constant C.

Using the last inequality and Lemma 3, we get the statement of the following lemma.

Lemma 9. *Make the assumptions of Proposition 1. For a constant $C > 0$ it holds that*

$$\begin{aligned} \sup_{\|\theta - \theta_0\| \leq C\ln(T)T^{-1/2}} \sup_{1 \leq t \leq T} h_t^{-1}(\theta_0)\, \|h'_t(\theta) - h'_t(\theta_0)\| &= o_P(1), \\ \sup_{1 \leq t \leq T} h_t^{-1}(\theta_0)\, \|h'_t(\theta_0)\| &= O_P(T^{1/\kappa_1}). \end{aligned}$$

The next lemma states that

$$d^{**}_{t+1} = \partial_{\theta\theta} g_{\theta_0}(Y_t, \hat{h}_t(\theta_0)) + 2\partial_{\theta s} g_{\theta_0}(Y_t, \hat{h}_t(\theta_0))\hat{h}'_t(\theta_0) + \partial_{ss} g_{\theta_0}(Y_t, \hat{h}_t(\theta_0)) d^{**}_{t+1}(\theta) \tag{30}$$

has a unique stationary solution d^{**}_t. We denote this solution by $\bar{h}''_t = d^{**}_t$. Note that this is a random value and not a random function.

Lemma 10. *Make the assumptions of Proposition 1. Equation (30) has a unique stationary solution $\bar{h}''_t = d^{**}_t$ that is ergodic. For $\rho > 1$ small enough it holds that*

$$\rho^t \|\bar{h}''_t - \hat{h}''_t(\theta_0)\| \to 0, \qquad a.s.$$

Equation (28) can be used to show that

$$\begin{aligned}\left\|\hat{h}''_{t+1}(\theta) - \hat{h}''_{t+1}(\theta_0)\right\| \quad \leq \quad & C\left[\|\theta - \theta_0\|(1 + |Z_t|(1 + h_t + \|h'_t\| + + \|h'_t\|^2 + \|h''_t\|)\right. \\ & + |\hat{h}_t(\theta) - \hat{h}_t(\theta_0)|(1 + |Z_t|)(1 + \|h'_t\| + \|h'_t\|^2 + \|h''_t\|) \\ & + \|\hat{h}'_t(\theta) - \hat{h}'_t(\theta_0)\|(1 + |Z_t|)(1 + \|h'_t\|) \\ & \left. + \left\|\hat{h}''_t(\theta) - \hat{h}''_t(\theta_0)\right\| V_t(\alpha_0 Z_t^2 + \beta_0)\right]\end{aligned}$$

for some positive constant C. Using the last inequality, we get the statement of the following lemma.

Lemma 11. *Make the assumptions of Proposition 1. For constants $C > 0$ it holds that*

$$\sup_{\|\theta-\theta_0\| \leq C \ln(T) T^{-1/2}} \sup_{1 \leq t \leq T} h_t^{-1}(\theta_0) \left\|\hat{h}''_t(\theta) - \hat{h}''_t(\theta_0)\right\| \quad = \quad o_P(1).$$

By making use of the derived results, we now get the statement of Theorem 2.

Proof of Theorem 2. We make use of $0 = \hat{L}'_T(\hat{\theta}) = \hat{L}'_T(\theta_0) + \hat{L}''_T(\hat{\theta}^*)(\hat{\theta} - \theta_0)$ for some random $\hat{\theta}^*$ with $\hat{\theta}^* = O_P(\ln(T)T^{-1/2})$. This gives:

$$\sqrt{T}(\hat{\theta} - \theta_0) = (T^{-1}\hat{L}''_T(\hat{\theta}^*))^{-1} T^{-1/2} \hat{L}'_T(\theta_0).$$

Using the above discussions, we get that

$$\begin{aligned} T^{-1}\hat{L}''_T(\hat{\theta}^*) \quad & \to \quad \Sigma_2, \text{ in probability}, \\ T^{-1/2}\hat{L}'_T(\theta_0) \quad & = \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{2} \frac{\bar{h}'_t}{\bar{h}_t}(Z_t^2 - 1) + \bar{h}_t^{-1/2}(\dot{m}_{\gamma_0}(\bar{h}_t) + m'_{\gamma_0}(\bar{h}_t)\bar{h}'_t) Z_t + o_P(1). \end{aligned}$$

The theorem follows by application of a martingale central limit theorem, see Hall and Heyde (1980). □

References

[1] Berkes, I., L. Horváth, and P. Kokoszka (2003). "GARCH processes: structure and estimation." *Bernoulli* 9, 201–227.

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

[3] Bougerol, P. (1993). "Kalman filtering with random coefficients and contractions." *SIAM J. Control Optim.* 31, 942–959.

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

[5] Christensen, B. J., C. M. Dahl, and E. M. Iglesias (2012). "Semiparametric inference in a GARCH-in-Mean model." *Journal of Econometrics* 167, 458–472.

[6] Conrad, C., and M. Karanasos (2014). "On the transmission of memory in GARCH-in-Mean models." *Journal of Time Series Analysis*, forthcoming.

[7] Conrad, C., and E. Mammen (2008). "Nonparametric regression on a generated covariate with an application to semiparametric GARCH-in-Mean models." Department of Economics, Discussion Paper No. 473, University of Heidelberg.

[8] Engle, R. F., D. M. Lilien, and R. P. Robins (1987). "Estimating time varying risk premia in the term structure." *Journal of Business and Economic Statistics* 9, 345–359.

[9] Francq, C., and J-M. Zakoïan (2004). "Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes." *Bernoulli* 10, 605-637.

[10] French, K. R., G. W. Schwert, and R. F. Stambaugh (1987). "Expected stock returns and volatility." *Journal of Financial Economics* 19, 3–29.

[11] Goldie, C.M. (1991). "Implicit renewal theory and tails of solutions of random equations." *The Annals of Applied Probability* 1, 126–166.

[12] Hall, P., and C. C. Heyde (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

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

[14] Linton, O., and B. Perron (2003). "The shape of the risk premium: evidence from a semiparametric generalized autoregressive conditional heteroscedasticity model." *Journal of Business & Economic Statistics* 21, 354–367.

[15] Lundblad, C. (2007). "The risk return tradeoff in the long run: 1836–2003." *Journal of Financial Economics* 85, 123-150.

[16] Lumsdaine, R. L. (1996). "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models." *Econometrica* 64, 575–596.

[17] Merton, R. C. (1973). "An intertemporal capital asset pricing model." *Econometrica* 41, 867–886.

[18] Mikosch, T. and Stărică, R. (2000). "Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process." *Annals of Statistics* 28, 1427–1451.

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

[20] Pagan, A. R., and Y. S. Hong (1990). "Nonparametric estimation and the risk premium." In Barnett, W. A., J. Powell, and G. E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, 51–75.

[21] Straumann, D. (2005). "Estimation in conditionally heteroskedastic time series models." *Lecture Notes in Statistics 181.* Springer, Berlin.

[22] Straumann, D. and Mikosch, T. (2006). "Quasi-maximum-likelihood estimation in conditionally heteroskedastic time series: a stochastic recurrence equations approach." *The Annals of Statistics* 34, 2449–2495.

[23] Van de Geer, S. (2000). *Empirical processes in M-estimation.* Cambridge University Press, Cambridge.