



2019/13

DP

Christian M. Hafner and
Dimitra Kyriakopoulou

Exponential-type GARCH
models with linear-in-
variance risk premium



CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

[https://uclouvain.be/en/research-institutes/
lidam/core/discussion-papers.html](https://uclouvain.be/en/research-institutes/lidam/core/discussion-papers.html)

Exponential-type GARCH models with linear-in-variance risk premium

Christian M. Hafner* and Dimitra Kyriakopoulou†

Université catholique de Louvain

September 10, 2019

Abstract

One of the implications of the intertemporal capital asset pricing model (CAPM) is that the risk premium of the market portfolio is a linear function of its variance. Yet, estimation theory of classical GARCH-in-mean models with linear-in-variance risk premium requires strong assumptions and is incomplete. We show that exponential-type GARCH models such as EGARCH or Log-GARCH are more natural in dealing with linear-in-variance risk premia. For the popular and more difficult case of EGARCH-in-mean, we derive conditions for the existence of a unique stationary and ergodic solution and invertibility following a stochastic recurrence equation approach. We then show consistency and asymptotic normality of the quasi maximum likelihood estimator under weak moment assumptions. An empirical application estimates the dynamic risk premia of a variety of stock indices using both EGARCH-M and Log-GARCH-M models.

Keywords: GARCH-in-Mean, EGARCH, Log-GARCH, CAPM, risk premium, maximum likelihood, stochastic recurrence equation

JEL Classification: C13, C22, C51, G12.

*Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA), Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium, e-mail: christian.hafner@uclouvain.be.

†Corresponding author. Center for Operations Research and Econometrics (CORE), Université catholique de Louvain, Voie du Roman Pays 34, Bte L1.03.01, B-1348 Louvain-la-Neuve, Belgium, e-mail: dimitra.kyriakopoulou@uclouvain.be.

1 Introduction

The investigation of the trade-off between risk and expected return is one of the fundamental problems in financial econometrics. The functional form of this trade-off is widely debated in the empirical literature, but the leading theoretical case is that of a linear relation between the expected return and the variance of the market portfolio's return, as suggested by Merton's (1973) Intertemporal Capital Asset Pricing Model (ICAPM). Estimation and testing of this model remains one of the econometric challenges. Empirical evidence on the risk-return relation is mixed. For example, the early studies of French *et al.* (1987) and Chou (1988) find a statistically significant positive relation between volatility and excess returns on stock market indices. More recently, Lundblad (2007) concludes that a large data span is required to reliably detect the risk return tradeoff.

The econometric framework typically employed is that of a generalized autoregressive conditional heteroskedasticity (GARCH) type, the so-called GARCH-in-mean (GARCH-M) model introduced by Engle *et al.* (1987), who found that risk premia are not time invariant when studying the expected excess returns of bond holding yields. They considered the linear function as their specification for the risk-return relationship. Surprisingly, the GARCH-M model with linear-in-variance risk premium does not seem to have an available complete estimation theory. Iglesias and Phillips (2012) showed that to obtain asymptotic normality of quasi maximum likelihood estimators, finite fourth unconditional moments of returns are necessary, a strong assumption that is often not justified in practice. Using a stochastic recurrence equation approach, Conrad and Mammen (2016) cover the case of GARCH-M with risk premium that does not increase faster than the conditional standard deviation, and hence excludes the case of a linear-in-variance risk premium. The problem of the linear-in-variance case in the GARCH-M model is that the behavior of the volatility process in a fixed neighborhood of the true parameter is explosive and difficult to control. Christensen *et al.* (2012) estimate a modified version of a semiparametric GARCH-in-mean model, where the modification is in the volatility equation in which the news impact term is driven by observed returns rather than residuals. This makes the model easier to handle as volatility does not depend on the conditional mean. Nyberg (2012) augments the GARCH-M model by a regime-switching component, driven by a binary

business cycle indicator, and finds a positive risk-return relationship independent of the state of the economy.

In this paper we derive the estimation theory of exponential-type GARCH models, the leading case being the exponential GARCH (EGARCH) model of Nelson (1991), with linear-in-variance risk premia, which are compatible with the ICAPM theory. We will show that exponential-type GARCH models are easier to handle in the linear-in-variance risk premium case than classical GARCH models. They also have the advantage of ensuring a positive conditional variance irrespective of parameter constraints. They can be made invariant with respect to power transformations of the observations, and accommodate the so-called leverage effect, i.e. an asymmetric news impact function. Our work strongly builds on recent progress on the asymptotic theory of the recalcitrant classical EGARCH model, see Straumann (2005), Straumann and Mikosch (2006), and Wintenberger (2013). Hafner and Linton (2017) show asymptotic properties of a moment-based estimator in the EGARCH case. Linton and Perron (2003) study a semiparametric version of the EGARCH-M model without a-priori information on the functional form of the risk premium, and found that the relationship between the risk and the volatility is nonlinear and nonmonotonic. However, no asymptotic theory was provided.

Another example of an exponential-type model is the Log-GARCH model, originally introduced by Geweke (1986), Pantula (1986), and Milhøj (1987). An asymmetric version of the Log-GARCH model was studied by Francq *et al.* (2013) and an extension of that model, which is both asymmetric and also stable by scaling, was recently proposed by Francq *et al.* (2018).

The paper is organized as follows. In Section 2, we present the class of exponential-type GARCH-in-Mean models and discuss their properties and estimation by quasi maximum likelihood. We show consistency and asymptotic normality of the estimator in Section 3. Section 4 delivers simulation results and an empirical application to stock market indices. Section 5 concludes. All proofs are collected in the Appendix.

The following notation is applied throughout the paper. We work on the complete metric space of continuous functions from a local fixed neighborhood of the true parameter $V(\theta_0) := \{\theta \in \Theta : \|\theta - \theta_0\| \leq \delta, \delta > 0\}$, where Θ denotes the parameter space, to some subset $K_\Theta \subset \mathbb{R}$ equipped with the supremum norm denoted by $\sup_{V(\theta_0)} |\cdot|$ or equivalently $\|\cdot\|_{V(\theta_0)}$. We also denote by $\sup_\Theta |\cdot|$ or $\|\cdot\|_\Theta$ the uniform norm over the parameter space Θ , with the identity

$\sup_{\Theta} |\cdot| = \bigvee_{k=1}^N \sup_{\mathcal{V}(\theta_{0k})} |\cdot|$. The norm $\|\cdot\|$ denotes the Euclidean l_2 -norm when applied to vectors, and the Frobenius norm when applied to matrices. The notation x^+ is used for the positive part of x , i.e. $x^+ = \max(x, 0)$, for any $x \in \mathbb{R}$. The operator \circ denotes function composition. The indicator function is denoted by $\mathbf{1}_{\{A\}}$, i.e., it takes the value 1 if the logical expression A is true, and 0 otherwise.

2 Exponential-type GARCH-in-mean models

We consider a financial return series denoted by $\{Y_t, t \in \mathbb{Z}\}$, which is given by

$$Y_t = \lambda_1 + \lambda_2 h_t + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} Z_t, \quad (1)$$

where $\{Z_t, t \in \mathbb{Z}\}$ is a sequence of independent, identically distributed (i.i.d.) random variables with zero mean and unit variance, h_t is F_{t-1} -measurable, where F_{t-1} is the σ -algebra generated by the past information up to time $t - 1$, and the error term ε_t is a zero mean process with conditional variance h_t . We would expect the risk premium parameter λ_2 to be positive, a finding that we will confirm in our empirical section for a number of stock indices.

The volatility dynamics are defined by exponential-type GARCH models of the form

$$\log h_t(\theta) = g(Z_{t-1}) + \beta \log h_{t-1}(\theta), \quad t \in \mathbb{Z} \quad (2)$$

where $g(\cdot)$ is a known parametric “news impact” function, and θ is the parameter vector containing the parameters of $g(\cdot)$, the persistence parameter β , and the risk premium parameters λ_1 and λ_2 . Higher order models including more lags of Z_t or of the autoregressive term, are much less popular and therefore not considered in this paper for simplicity. The model class (2) specifies log volatility as a linear autoregressive process, where the innovation term is some nonlinear function of lagged returns, Z_{t-1} . Because log volatility has an additive structure, this implies that volatility itself has a multiplicative structure.

A leading example of an exponential-type GARCH model is the exponential GARCH model of Nelson (1991) given by

$$\log h_t(\theta) = \omega + \gamma Z_{t-1} + \delta |Z_{t-1}| + \beta \log h_{t-1}(\theta), \quad t \in \mathbb{Z} \quad (3)$$

where $\omega, \gamma, \delta \in \mathbb{R}$ and $|\beta| < 1$. The parameter γ is governing the asymmetry and δ measures the size effect. The EGARCH corresponds to the news impact function $g(z) = \omega + \gamma z + \delta |z|$. If $\gamma < 0$, negative shocks increase volatility more than positive shocks of the same size. For reviews of various GARCH models including EGARCH, see Bera and Higgins (1993), Hentschel (1995), Francq and Zakoian (2010), and Rodriguez and Ruiz (2012). The moment and autocorrelation structure of the EGARCH(1, 1) model are analyzed in He *et al.* (2002), while more generally, Karanasos and Kim (2003) discussed the autocorrelation function of the EGARCH(p, q) of general order.

Another exponential-type GARCH model is the following version of the Log-GARCH model proposed by Francq *et al.* (2018),

$$\log h_t(\theta) = \omega_t + (\alpha_+ \mathbf{1}_{\{\varepsilon_{t-1} > 0\}} + \alpha_- \mathbf{1}_{\{\varepsilon_{t-1} < 0\}}) \log \varepsilon_{t-1}^2 + \tilde{\beta} \log h_{t-1}(\theta), \quad t \in \mathbb{Z} \quad (4)$$

where the time-varying intercept $\omega_t = \omega + \omega_- \mathbf{1}_{\{\varepsilon_{t-1} < 0\}}$ is introduced here to make the volatility stable by scaling¹. The symmetric log-GARCH model corresponds to the case of $\alpha_+ = \alpha_- = \alpha$ and $\omega_- = 0$ and can be rewritten in the form (2) with $g(z) = \omega + \alpha \log z^2$ and $\beta = \alpha + \tilde{\beta}$. The asymmetric log-GARCH model corresponds to (2) only by letting β depend on the sign of Z_{t-1} , and hence on time. For a recent discussion on the Log-GARCH model we refer to Sucarrat *et al.* (2016).

A well known problem in the Log-GARCH model is the occurrence of exact zeros in ε_{t-1} , in which case the term $\log \varepsilon_{t-1}^2$ would make $\log h_t$ degenerate. Exact zero returns often occur in illiquid assets, tranquil periods, or when the price granularity is not sufficiently fine. See Sucarrat and Escribano (2018) for a detailed study of this problem and algorithms to solve it. In the case of the log-GARCH-M model, the problem is alleviated by the presence of the risk premium, so that $\varepsilon_t = Y_t - \lambda_1 - \lambda_2 h_t$ is different from zero a.s. even though Y_t might be zero, provided that the risk premium parameters are sufficiently far from zero.

The formulation given by the mean and log-variance equations, i.e. either of (1) and (3), or (1) and (4), allows the conditional variance of each observation to affect the mean of that observation. The most popular specification for the risk-return relationship is (1), i.e. to

¹As noted by Francq *et al.* (2018), stability by scaling is a desirable property of volatility models. If the error process, ε_t , is scaled by a constant $c > 0$, it is important that the dynamics of the volatility process be not affected by the choice of scale c .

consider a linear function of h_t , as the original specification of the GARCH-M model of Engle *et al.* (1987), while other typical choices are $\sqrt{h_t}$ or $\log h_t$. Pagan and Hong (1991) commented that the use of $\log(h_t)$ is problematic since for $h_t < 1$, $\log(h_t)$ will be negative and also when h_t approaches zero, the conditional variance in logs takes very large negative values and the relationship between the conditional variance and Y_t may be overstated.

A particularly important example where h_t is introduced in the mean equation is the intertemporal CAPM of Merton (1973), which suggests that the expected excess return on the market should be proportional to the conditional variance of the market return. Iglesias and Phillips (2012) showed that in the case of the GARCH-M model, the existence of $E(Z_t^4)$, which is usually assumed for deriving the asymptotic properties of the QML estimator, implies the existence of $E(\varepsilon_t^4)$, i.e. the unconditional fourth moment of ε_t , which is a strong assumption. This is an important difference with traditional GARCH models, and also with exponential-type GARCH-in-Mean models, in which the existence of $E(Z_t^4)$ does not require the existence of $E(\varepsilon_t^4)$.

Although we concentrate on the theoretically important linear-in-variance case, many of our results could be generalized to a risk premium that is a function of volatility. For the estimation theory of a general EGARCH-M model, the crucial condition to check is invertibility, which we will discuss in Section 3.1. for the linear case. Adapting this condition to other functional forms of the risk premium, for example quadratic, may lead to more restrictive parameter regions. In the linear case, as we will show, the condition is satisfied for a range of risk premium parameters sufficiently close to zero.

Let $\theta = (\lambda_1, \lambda_2, \omega, \gamma, \delta, \beta)^\top$ be the vector of unknown parameters in the EGARCH-M model and $\theta = (\lambda_1, \lambda_2, \omega, \omega_-, \alpha_+, \alpha_-, \beta)^\top$ in the Log-GARCH-M model, respectively. We denote by θ_0 the true parameter vector, and add analogous subscripts for the true individual parameters. Since we do not impose any positivity restrictions on the volatility coefficients, we can generally assume that each parameter can take values in \mathbb{R} . In standard GARCH estimation, the need to impose non-negativity constraints on the parameters often creates numerical as well as statistical difficulties in estimation, especially inferential when the estimated parameters fall on a boundary of the constraints, while the exponential versions of the GARCH model solve these problems by construction. We assume that the parameter vector θ belongs to a parameter

space Θ .

At this point, the existence of a stationary solution of (2) should be considered. We will work under the following assumptions.

Assumption 1 The parameter space Θ is a compact set.

Assumption 2

a. It holds that $0 \leq \beta_0 < 1$.

b. It holds that $\delta_0 \geq |\gamma_0|$.

Assumption 2a ensures that the model (2) admits a unique stationary solution with finite moments. Assumption 2b ensures that log-volatility has a universal deterministic lower bound, i.e. there exists a strictly positive constant $s^* = \omega_0/(1 - \beta_0)$ such that $\inf_{\Theta} \log(h_t) \geq s^*$ a.s., see Straumann (2005, pp. 60-61) for details. As a consequence, in the EGARCH-M case, Θ is a compact subset of $\mathbb{R}^3 \times D \times [0, 1)$, where $D = \{(\gamma, \delta)^\top \in \mathbb{R}^2 \mid \gamma \in \mathbb{R}, \delta \geq |\gamma|\}$. In the Log-GARCH-M case, Θ is a compact subset of $\mathbb{R}^6 \times [0, 1)$.

2.1 Properties

The stochastic properties of the processes can be studied using a particular class of Markov chains which are recursions of i.i.d. random Lipschitz functions. These iterations are known as stochastic recurrence equations (SREs) and their stability is usually deduced via appropriate contraction conditions involving Lipschitz continuity. This approach was developed in Bougerol (1993) and for volatility models in Straumann (2005) and Straumann and Mikosch (2006). We refer to Diaconis and Freedman (1999) for an overview of SREs.

Definition 1 (Stochastic recurrence equation system) *Let a stationary ergodic process $\{\phi_t, t \in \mathbb{Z}\}$ be defined on a probability space, where each ϕ_t is a random Lipschitz function that is jointly measurable and Lipschitz continuous. The process $(\mathcal{X}_t)_{t \geq 1}$ is called a stochastic recurrence equation (SRE) associated with (ϕ_t) if it satisfies the following stochastic difference equation:*

$$\mathcal{X}_t = \phi_t(\mathcal{X}_{t-1}). \tag{5}$$

The process (\mathcal{X}_t) is referred to as a solution to the SRE (5).

The parametric SRE representation of (3) and (4) is given by

$$\log h_{t+1}(\theta) = \phi_t(\log h_t(\theta)), \quad t \in \mathbb{Z}$$

where ϕ_t is a function of the innovations Z_t , with

$$\phi_t(s) = \omega + \gamma Z_t + \delta |Z_t| + \beta s \quad (6)$$

for the EGARCH model and

$$\phi_t(s) = \omega_t + (\alpha_+ \mathbf{1}_{\{Z_t > 0\}} + \alpha_- \mathbf{1}_{\{Z_t < 0\}}) \log Z_t^2 + \left(\alpha_+ \mathbf{1}_{\{Z_t > 0\}} + \alpha_- \mathbf{1}_{\{Z_t < 0\}} + \tilde{\beta} \right) s \quad (7)$$

for the Log-GARCH process, substituting $\log \varepsilon_t^2$ by $\log h_t + \log Z_t^2$ in (4).

Equation (6) can be written in the form

$$\log h_{t+1} = A_t \log h_t + B_t,$$

where $\{(A_t, B_t)\}$ constitutes an i.i.d. sequence. The conditions $E(\log |A_0|) < 0$ and $E(\log^+ |B_0|) < \infty$ guarantee the existence and uniqueness of a strictly stationary and ergodic solution of (3) at the point $\theta = \theta_0$ provided $\{(A_t, B_t)\}$ is a stationary and ergodic sequence, which is true since it is a measurable function of the i.i.d. innovations Z_t . These conditions and $E Z_t^2 = 1$ imply

$$|\beta_0| < 1 \quad \text{and} \quad E[\log^+(\omega_0 + \gamma_0 Z_t + \delta_0 |Z_t|)] < \infty,$$

see, for instance, Straumann and Mikosch (2006). Theorem 2.6.1 of Straumann (2005) implies that any other solution converges almost surely fast to the unique stationary and ergodic solution. The stationarity guarantees the existence of higher order moments and, in particular, of the kurtosis. If Y_t is stationary, the unconditional variance, kurtosis and the autocorrelation function of squares can be derived using the results in Nelson (1991), see also He *et al.* (2002). Nelson (1991) proved that if the distribution of Z_t is a member of the Generalized Error Distribution (GED) family and is not too fat tailed, and if $|\beta_0| < 1$, then $\{h_t\}$ and $\{\varepsilon_t\}$ have finite unconditional moments of arbitrary order.

For the Log-GARCH model, assuming $E \log^+ |\log Z_t^2| < \infty$, a sufficient condition for the existence of a strictly stationary solution to the model of (4) is

$$\left| \tilde{\beta}_0 \right| + \frac{1}{2} (|\alpha_{0+}| + |\alpha_{0-}|) < 1,$$

see, for instance, Francq *et al.* (2013) and (2018). Due to asymmetric effects in the persistence coefficient, stationarity and existence of log-moments are more difficult to establish for the asymmetric Log-GARCH model. In the symmetric case, the condition reads $|\alpha_0 + \tilde{\beta}_0| < 1$, like in the classical GARCH model. As for GARCH-type models, where the strict stationarity condition entails the existence of moments for $|\varepsilon_t|$, it is also the case for $|\log \varepsilon_t^2|$ in the Log-GARCH model, see Proposition 2.1 in Francq *et al.* (2013).

The log-volatility specification in (3) admits an MA(∞) representation of the form

$$\log h_t(\theta) = \omega(1 - \beta)^{-1} + \sum_{k=0}^{\infty} \beta^k (\gamma Z_{t-1-k} + \delta |Z_{t-1-k}|), \quad (8)$$

which is an almost sure representation of $\log h_t$, i.e. the a.s. absolute convergence of the series. From (8) and Assumption 2b, which implies $\gamma Z_t + \delta |Z_t| \geq 0$, *a.s.* $\forall Z_t$, we can deduce that $\log h_t \geq \omega(1 - \beta)^{-1}$ *a.s.*, see Straumann (2005, pp. 60-61), which will be used in the proofs of the theory. However, this is not the case for the Log-GARCH and in order to avoid any inconsistency problems of the estimator we need to ensure that $E|\log \varepsilon_t^2| < \infty$. Studying the Log-GARCH model, we can observe that small values, and not only large ones, can have persistent effects on volatility. Moreover, volatility forecasts are more robust to jumps and outliers. These properties distinguish the Log-GARCH model from the EGARCH model, despite their quite similar form.

2.2 Quasi maximum likelihood estimation

We now discuss the quasi maximum likelihood (QML) estimator $\hat{\theta}_T$, where T denotes the sample size. To save space, the same applies to the estimator $\hat{\theta}_T$ for the Log-GARCH-M model. Given that the volatility process (3) is by construction latent, another solution of (3) with fixed starting value \hat{h}_0 is given by the filtered log-volatility by also reconstructing the unobserved volatility sequence from data as follows

$$\log \hat{h}_t(\theta) = \omega + \gamma \frac{Y_{t-1} - \lambda_1 - \lambda_2 \hat{h}_{t-1}(\theta)}{\sqrt{\hat{h}_{t-1}(\theta)}} + \delta \left| \frac{Y_{t-1} - \lambda_1 - \lambda_2 \hat{h}_{t-1}(\theta)}{\sqrt{\hat{h}_{t-1}(\theta)}} \right| + \beta \log \hat{h}_{t-1}(\theta), \quad t \geq 1 \quad (9)$$

and

$$\log \hat{h}_t(\theta) = \omega + \omega_- \mathbf{1}_{\{\varepsilon_t < 0\}} + (\alpha_+ \mathbf{1}_{\{\varepsilon_{t-1} > 0\}} + \alpha_- \mathbf{1}_{\{\varepsilon_{t-1} < 0\}}) \log \varepsilon_{t-1}^2 + \tilde{\beta} \log \hat{h}_{t-1}(\theta), \quad t \geq 1$$

which is fitted to the data and used for constructing the quasi log-likelihood function and its derivatives. The auxiliary log-volatility process $\{\log \widehat{h}_t(\theta), t \geq 0\}$, approximates the true volatility for a given parameter θ . It is well defined but generally nonstationary due to its dependence on initial conditions.

Given proper initial values for Z_0 and \widehat{h}_0 , the conditional Gaussian quasi log-likelihood function up to an additive constant is

$$\widehat{L}_T(\theta | Z_0, \widehat{h}_0) = -\frac{1}{2} \sum_{t=1}^T \left[\log(\widehat{h}_t(\theta)) + \frac{[Y_t - \lambda_1 - \lambda_2 \widehat{h}_t(\theta)]^2}{\widehat{h}_t(\theta)} \right],$$

and the quasi-maximum likelihood estimator $\widehat{\theta}_T$ of θ is defined as any measurable solution of

$$\widehat{\theta}_T = \arg \max_{\theta \in \Theta} \widehat{L}_T(\theta).$$

We also define the true likelihood function and its expectation as

$$\begin{aligned} \bar{L}_T(\theta) &= -\frac{1}{2} \frac{1}{T} \sum_{t=1}^T \left[\log(h_t(\theta)) + \frac{[Y_t - \lambda_1 - \lambda_2 h_t(\theta)]^2}{h_t(\theta)} \right] \\ L(\theta) &= -\frac{1}{2} E \left[\log(h_t(\theta)) + \frac{[Y_t - \lambda_1 - \lambda_2 h_t(\theta)]^2}{h_t(\theta)} \right]. \end{aligned}$$

3 Asymptotic theory

Our main results, Theorems 1 and 2 for strong consistency and asymptotic normality of the quasi-maximum likelihood estimator, build on the (SRE) approach developed for volatility models in Straumann (2005) and Straumann and Mikosch (2006). In the following subsections we focus on the EGARCH-M model, which is more difficult than the Log-GARCH-M model. The theory for the latter is more standard and could be derived by extending the results of Francq *et al.* (2013) to the in-mean case.

For the consistency part of the theory, it is essential to prove that, for any fixed initial condition $\log \widehat{h}_0$, we can approximate the filtered log-volatility process, $\log \widehat{h}_t(\theta)$, by a stationary ergodic sequence $\log h_t(\theta)$, such that the difference converges to zero sufficiently fast as $t \rightarrow \infty$ for any value of θ that belongs to the parameter space. This comprehends the invertibility of the EGARCH-in-mean volatility so that $\log h_t(\theta)$ can be represented in terms of past observations,

since (6) is a function of the innovations Z_t which are not observable. For the asymptotic normality part, we need to establish differentiability of $\log \widehat{h}_t(\theta)$, analyzing the first and second derivative of the log-likelihood function with respect to the parameter θ . The SREs defined for the derivatives of the quasi-likelihood function are then used to show that they have a stationary ergodic functional solution.

Definition 2 (Lipschitz coefficient) *Let f be any real valued function. Then the Lipschitz coefficient of that function, $\Lambda(f)$, is defined by the relation*

$$\Lambda(f) = \sup_{x,y \geq 0, x \neq y} \frac{|f(x) - f(y)|}{|x - y|}, \quad (10)$$

and is called a contraction if $\Lambda(f) < 1$. If $\Lambda(f)$ is finite, the map f is a Lipschitz map.

For the consistency of the QMLE, the essential assumption is that the Lipschitz coefficients have a finite log moment of order 1, so that they satisfy the contraction property of the SRE (see also Assumption 3 and Lemma 1) as a result of Bougerol (1993).

The following theorem due to Bougerol (1993) shows convergence properties of processes obtained by iteration of stationary Lipschitz functions.

Theorem 3.1 of Bougerol (1993). *Let $\{\phi_t, t \in \mathbb{Z}\}$ be a stationary ergodic sequence of Lipschitz maps taking values in a complete separable metric space, \mathcal{E} . $\Lambda(\phi_0)$ is given by the supremum in (10) and thus is measurable. We suppose that the following conditions hold:*

- (1) *For some $x \in \mathcal{E}$, $E[\log^+ |\phi_0(x) - x|] < \infty$.*
- (2) *$E[\log^+ \Lambda(\phi_0)] < \infty$ and for some integer $r \geq 1$,*

$$E\left[\log \Lambda\left(\phi_0^{(r)}\right)\right] = E\left[\log \Lambda\left(\phi_0 \circ \dots \circ \phi_{-r+1}\right)\right] < 0.$$

Then the SRE (5) converges, so that it admits a stationary ergodic solution $(\mathcal{Y}_t)_{t \in \mathbb{Z}}$ and if there is any other solution to (5), $(\widetilde{\mathcal{Y}}_t)_{t \in \mathbb{Z}}$, then almost surely

$$\overline{\lim}_{t \rightarrow \infty} \frac{1}{t} \log \left| \mathcal{Y}_t - \widetilde{\mathcal{Y}}_t \right| < 0,$$

where $\overline{\lim}_{t \rightarrow \infty}$ denotes the limit superior of the sequence.

Definition 3 *The SRE is called contractive if the associated Lipschitz map obeys the conditions given in Theorem 3.1. of Bougerol (1993).*

3.1 Strong consistency

Invertibility plays a crucial role for establishing consistency of the QML estimator in the EGARCH-M model. In the following we recall alternative concepts of invertibility and give sufficient conditions for continuous invertibility of the filtered log-volatility process, which converges to the unique stationary and ergodic solution.

We consider a new functional SRE of (3) that is driven by a function of the observations Y_t

$$s_{t+1} = [\Phi_t(s_t)](\theta), \quad t \in \mathbb{Z}, \quad \forall \theta \in \Theta \quad (11)$$

where (Φ_t) is a sequence of random Lipschitz functions $g_\theta(Y_t, s_t)$ given by

$$g_\theta(y, s) = \omega + [\gamma(y - \lambda_1 - \lambda_2 \exp(s)) + \delta(|y - \lambda_1 - \lambda_2 \exp(s)|)] \exp(-s/2) + \beta s, \quad (12)$$

for each $\theta \in \Theta$. Observe that $\log \widehat{h}_t$ defined by (9) is a solution of the previous equation.

Definition 4 (Invertibility) *The volatility model (3) is called invertible (or its unique stationary ergodic solution is invertible) if it admits a unique stationary solution and*

$$\left| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right| \xrightarrow{P} 0, \quad t \rightarrow \infty \text{ and } \theta = \theta_0. \quad (13)$$

In other words, if the model is invertible, the above algorithm converges and the sequence of the volatilities $(\log h_t(\theta))_{\theta=\theta_0}$ is adapted to the filtration (9) generated by past observed values, where $\log h_t(\theta)_{\theta=\theta_0}$ is given by (6), the unique stationary solution at $\theta = \theta_0$.

This notion has been introduced by Granger and Andersen (1978) in the context of a general nonlinear autoregressive moving average model. There is another notion of invertibility as an adaptation of the previous one, given by Straumann and Mikosch (2006). Following Theorem 3.1 of Bougerol (1993) who used Lyapunov conditions, their notion is based upon the contraction of the Lipschitz coefficients and the existence of logarithmic moments. Under their notion, the above convergence is given almost surely at an exponentially fast rate (e.a.s.)² and also irrespective of the initial value of volatility. However, definition (13) is not sufficient to

²The exponentially fast almost sure (e.a.s.) convergence is used in the definition of continuous invertibility and also later in the paper. It is essentially about the convergence of a sequence (v_t) of random elements, with values in a normed vector space, given by $v_t \xrightarrow{\text{e.a.s.}} 0$, if there exists $\rho > 1$ such that $\rho^t \|v_t\| \xrightarrow{\text{a.s.}} 0$.

ensure the consistency of the estimator, because this condition at $\theta = \theta_0$ does not necessarily ensure the stability of the approximation $\log \widehat{h}_t(\theta)$, $\forall \theta \in \Theta$. Another notion of continuous invertibility has been defined by Wintenberger (2013) by working on a space of continuous functions of the log-volatility, see also Blasques *et al.* (2016) for a discussion of this concept:

Definition 5 (Continuous invertibility) *The SRE (12) is continuously invertible on a compact set Θ if*

$$\sup_{\Theta} \left| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right| \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty. \quad (14)$$

Since $\log \widehat{h}_t(\theta)$ approximates a continuous function $\log h_t(\theta)$ that does not depend on the initialization $\log \widehat{h}_0$, uniformly over the parameter space Θ , the previous definition ensures the stability of the SRE (12) defined for the filtered log-volatility process (9). Continuous invertibility is stronger than invertibility as (14) implies (13). Sufficient conditions for continuous invertibility of the SRE (12) for each $\theta \in \Theta$ can be obtained assuming either some uniform or pointwise Lipschitz conditions on the SRE, see e.g. condition (CI) of Wintenberger (2013).

To prove consistency, we make the following additional assumptions:

Assumption 3 For each $\theta \in \Theta$ it holds that

$$E [\log \max \{\beta, \Psi_0\}] < 0, \quad (15)$$

where

$$\begin{aligned} \Psi_0 &= 2^{-1} [\gamma \varepsilon_0(\theta) + \delta |\varepsilon_0(\theta)|] \exp \left(-2^{-1} \inf_{\Theta} \omega (1 - \beta)^{-1} \right) \\ &\quad + \lambda_2 [\gamma + \delta \operatorname{sgn}(\varepsilon_0(\theta))] \exp \left(2^{-1} \inf_{\Theta} \omega (1 - \beta)^{-1} \right) - \beta, \end{aligned}$$

with $\varepsilon_0(\theta) := Y_0 - \lambda_1 - \lambda_2 h_0(\theta)$, and

$$\log h_0(\theta) = \omega + \gamma \frac{\varepsilon_{-1}(\theta)}{\sqrt{h_{-1}(\theta)}} + \delta \left| \frac{\varepsilon_{-1}(\theta)}{\sqrt{h_{-1}(\theta)}} \right| + \beta \log h_{-1}(\theta).$$

Assumption 3 is crucial for establishing invertibility of the EGARCH-M process, which ensures that the QML estimator converges to one point even if different initializations are considered. Condition (15) can be verified by numerical simulation methods. It can be viewed as an extension of condition (11) of Wintenberger (2013) to the EGARCH-M case.

The next lemma establishes for any fixed starting value $\log \widehat{h}_0$, the exponentially fast almost sure convergence of $\log \widehat{h}_t(\theta)$, generated by (9), to a unique stationary and ergodic solution $\log h_t(\theta)$, $t \in \mathbb{Z}$, which implies their asymptotic equivalence.

Lemma 1 *Let Assumptions 1-3 be satisfied. Then the SRE (12) is continuously invertible.*

The consequence of this lemma is that the effect of the initialization vanishes asymptotically at an exponential rate and this also implies that the true path, $\log h_t(\theta_0)$, can be asymptotically recovered, i.e. $\log h_t(\theta_0)$ is given by (3) only when $\theta = \theta_0$.

For illustration, Figure 1 shows the expectation in (15), evaluated at θ_0 using Monte-Carlo simulation and assuming that Z_t has a Generalized Error Distribution (GED) with parameter $\nu = 1.5$, as a function of $\lambda_{0,2}$, fixing $\beta_0 = 0.9$, $\omega_0 = -0.08$, and $\gamma_0 = \lambda_{0,1} = 0$. For δ_0 , two cases are considered: $\delta_0 = 0.15$ (the solid line), and $\delta_0 = 0.9$ (the dashed line), the typical range of empirical estimates being 0.1 to 0.5. Typical estimates of the risk premium parameter are close to zero, see our results in Section 4, where returns are defined as $100 \times \log(P_t/P_{t-1})$ - we have checked that the relative region for λ_2 is around 0.03 (see Lundblad, 2007). For large values of λ_2 and δ , condition (15) is not satisfied, so that this is a restriction. By continuity of (15) in θ , the condition is expected to hold at least in a neighborhood of the true parameter θ_0 .

Assumption 4 The distribution of Z_0 is not concentrated at two points and $\log h_t(\theta_0)$ has a non-degenerate distribution. Furthermore, $(\gamma_0, \delta_0) \neq (0, 0)$.

The next theorem states the consistency of the QMLE. This result holds for every initialization of the filter satisfying the conditions of Lemma 1.

Theorem 1. For the model given by (1) and (3), let Assumptions 1-4 be satisfied. Then it holds that

$$\widehat{\theta}_T \xrightarrow{a.s.} \theta_0 \text{ as } T \rightarrow \infty.$$

3.2 Asymptotic normality

Given consistency, we can establish asymptotic normality via the use of local approximations of the likelihood function. The appendix gives the expressions for the first and second order partial derivatives of the likelihood function, which are well defined. The local approximations

emerge from a second order Taylor expansion of the likelihood function in a neighborhood of the true parameter value θ_0 .

If $\hat{\theta}_T$ lies in the interior of Θ , then the QML estimator satisfies the first-order condition $\frac{\partial \hat{L}_T(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_T} = 0$. By a mean-value expansion of the score vector, $\frac{\partial \hat{L}_T(\theta)}{\partial \theta}$, around the point $\theta = \theta_0$, for which the estimator was proved consistent, it yields a valid Taylor expansion for sufficiently large T as follows

$$0 = \frac{\partial \hat{L}_T(\theta_0)}{\partial \theta} + \frac{\partial^2 \hat{L}_T(\theta_T^*)}{\partial \theta \partial \theta^\top} (\hat{\theta}_T - \theta_0),$$

for some sequence θ_T^* on the line segment $[\hat{\theta}_T, \theta_0]$ that converges to θ_0 . Multiplying by \sqrt{T} , setting $\hat{L}'_T(\theta_0) := \frac{\partial \hat{L}_T(\theta_0)}{\partial \theta}$ and $\hat{L}''_T(\theta_T^*) := \frac{\partial^2 \hat{L}_T(\theta_T^*)}{\partial \theta \partial \theta^\top}$ and rearranging the terms, we have

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

In order to prove the uniform convergence of the second derivative of the log-likelihood function by the strong law of large numbers (SLLN) theorem, a necessary condition is that its first order moment is uniformly bounded. We follow Wintenberger (2013) by using the SLLN uniformly only locally on some fixed neighborhood of the true parameter. This approach can provide sharp conditions for the EGARCH-M model. As soon as the consistency of the QMLE has been proved, a local uniform continuity argument with respect to the parameters applied to the Lipschitz coefficients of the underlying SRE can yield the asymptotic normality under the contraction of the SRE for the true parameter only. The fundamental SRE used to prove the next theorem is satisfied by absolute differences of the conditional log-likelihood function evaluated at any convergent sequence θ_T^* and its limit, i.e. $\left| \hat{L}_T(\theta_T^*) - \hat{L}_T(\theta_0) \right|$, and its derivatives. For our next Theorem, we need the following additional assumptions.

Assumption 5 The true parameter θ_0 belongs to the interior $\overset{\circ}{\Theta}$ of Θ .

Assumption 6 $E(Z_t^4) < \infty$ and

$$E \left[\left(\beta_0 - \left(2^{-1} (\gamma_0 Z_t + \delta_0 |Z_t|) + \lambda_{02} (\gamma_0 + \delta_0 \operatorname{sgn}(Z_t)) \sqrt{h_t} \right) \right)^2 \right] < 1,$$

such that the asymptotic variance of the Gaussian distribution of the QML estimator exists and is well defined.

Theorem 2 *It holds that $P\left(\widehat{\theta}_T \in \overset{\circ}{\Theta}\right) \rightarrow 1$ if the true value is in the interior of the parameter space. Under the Assumptions 1-6 and with $\delta > 0$ small enough, the QMLE $\widehat{\theta}_T$ is asymptotically normal as $T \rightarrow \infty$, i.e.*

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, V\left(\theta_0\right)\right),$$

where \xrightarrow{d} denotes convergence in distribution, $V\left(\theta_0\right) = F\left(\theta_0\right)^{-1} G\left(\theta_0\right) F\left(\theta_0\right)^{-1}$, where

$$\begin{aligned} F\left(\theta_0\right) &= E \frac{\partial \log \left(h_t\left(\theta_0\right)\right)}{\partial \theta} \frac{\partial \log \left(h_t\left(\theta_0\right)\right)}{\partial \theta^\top} + E \frac{2}{h_t\left(\theta_0\right)} \frac{\partial \varepsilon_t\left(\theta_0\right)}{\partial \theta} \frac{\partial \varepsilon_t\left(\theta_0\right)}{\partial \theta^\top} \\ \frac{\partial \varepsilon_t\left(\theta_0\right)}{\partial \theta} &= -e_{1n} - e_{2n} h_t\left(\theta_0\right) - \lambda_2 \frac{\partial h_t\left(\theta_0\right)}{\partial \theta} \\ G\left(\theta_0\right) &= E\left[\nabla l_t\left(\theta_0\right) \nabla l_t\left(\theta_0\right)^\top\right] \\ \nabla l_t\left(\theta_0\right) &= \frac{1}{2} \frac{\partial \log h_t\left(\theta_0\right)}{\partial \theta}\left(Z_t^2 - 1\right) + \frac{1}{\sqrt{h_t\left(\theta_0\right)}} Z_t\left(e_{1n} + h_t\left(\theta_0\right) e_{2n} + \lambda_{02} \frac{\partial h_t\left(\theta_0\right)}{\partial \theta}\right) \end{aligned}$$

and e_{jn} is the j -th unit vector, i.e. the j -th column of the identity matrix of dimension n .

Consistent estimators for $F\left(\theta_0\right)$ and $G\left(\theta_0\right)$ can be obtained by replacing the expectation operators by sample averages, and the true parameter by its estimate. When Z_t is Gaussian, we have that $F\left(\theta_0\right) = G\left(\theta_0\right)$ and the asymptotic covariance $V\left(\theta_0\right)$ is then equal to $F\left(\theta_0\right)^{-1}$. When there is no risk premium in the mean equation, then

$$F\left(\theta_0\right) = \frac{1}{2} E\left[\frac{\partial \log h_t\left(\theta_0\right)}{\partial \theta} \frac{\partial \log h_t\left(\theta_0\right)}{\partial \theta^\top}\right]$$

and

$$G\left(\theta_0\right) = \frac{1}{4}\left[E\left(Z_t^4\right) - 1\right] E\left[\frac{\partial \log h_t\left(\theta_0\right)}{\partial \theta} \frac{\partial \log h_t\left(\theta_0\right)}{\partial \theta^\top}\right],$$

so that the covariance is the one given for the EGARCH(1,1) model (see, for instance, Kyriakopoulou, 2015).

The first term in the expression for $F\left(\theta_0\right)$ is the same as that of the EGARCH model without risk premium, and it does not depend on the risk premium parameter. The second term is positive definite, which implies that, for a given set of parameters, the presence of a risk premium increases the asymptotic variance of the other parameters.

4 Numerical applications

4.1 Simulation study

In order to investigate the small, moderate and large sample behavior of the QMLE in the EGARCH(1, 1)-M and Log-GARCH(1, 1)-M models, we perform the following simulation study. We generate time series of length $T = 500, 1000, 2500, 5000$ and 10000 observations and we tie the data generating process parameterization for both models to empirically relevant values (see Section 4.2), i.e. the true parameter vector for the EGARCH(1, 1)-M model is given by $\theta_0 = (0, 0.04, -0.10, -0.12, 0.13, 0.98)^\top$, and for the Log-GARCH(1, 1)-M model by $\theta_0 = (0, 0.07, 0.024, 0.027, 0.016, 0.971)^\top$. For the innovation process, we assume that Z_t has a Generalized Error Distribution (GED) with tail-thickness parameter ν given by either 2, which yields the Gaussian case, or 1.4, which produces fatter tails. The latter choice corresponds to typical estimates in empirical applications, see e.g. Hafner and Linton (2017) for the S&P 500 index.

The QML estimation is applied in a second step to obtain parameter estimates. We simulate 1000 paths and hence get 1000 independent replicates of $\hat{\theta}_T$. In Figures 2 and 3 we depict the distributions of the vectors $\hat{\theta}_T$ for the EGARCH(1, 1)-M model, when the GED parameter ν is 2 and 1.4, respectively, by compiling boxplots for each estimated parameter. These numerical results illustrate the performance of the QMLE in finite samples. For both cases, the dispersion obviously shrinks as T increases, and the median parameter estimates are very close to the true parameter values even for small sample sizes. The corresponding boxplots for the Log-GARCH-M model are presented in Figures 4 and 5, respectively, and show a very similar behavior.

Tables 1 and 2 summarize the bias and RMSE performance of the QMLE of the EGARCH-M and Log-GARCH-M model parameters, respectively. As we can see from the results of Table 1, the bias is small even for small sample sizes and approaches zero as T increases. The RMSE drops quite fast for all estimated parameters as T gets larger, confirming consistency. While the bias is very similar for the Gaussian case and $\nu = 1.4$, the RMSE is slightly larger for most parameters in the case of fat tails. In terms of estimation precision, the parameter that is most difficult to estimate is the risk premium, as the RMSE is about 100% for $T = 500$, and for $T = 10000$ is still 19% and 23% for the Gaussian and GED(1.4) case, respectively.

As for the Log-GARCH-M results of Table 2, the results are similar to the EGARCH-M case, except that the volatility parameter estimates tend to have a higher RMSE, while the risk premium parameter has a lower RMSE across different sample sizes and distributions. This suggests that volatility estimates are more precise for EGARCH-M than log-GARCH-M, but that the reverse is true for the estimation of the risk premium.

Table 3 presents some parameter evaluations from each model, that is summary statistics of the estimated conditional standard deviations simulated by both the EGARCH-M and Log-GARCH-M models, under GED innovations and for the two cases, i.e. $\nu = 2$ and $\nu = 1.4$. As we can see, all statistics are very close for both models.

4.2 Empirical application

We consider daily return series of the stock market indices S&P 500, Nasdaq 100, FTSE 100, DAX, Nikkei 225, CAC 40, and Dow Jones Industrial Average for the period 2/1/1986 - 30/12/2016, except for CAC40 which is for the period 9/7/1987 - 30/12/2016. Table 4 reports sample moments for the data over the full sample and three subsamples, each containing approximately 10 years of observations and at least one major financial crisis. For example, the first subsample covers Black Monday of October 1987, the second the burst of the dot-com bubble, and the third the financial crisis of 2008-2009. The idea is to see whether or not our results are stable across different subsamples of roughly equal size.

Not surprisingly, there is strong evidence of leptokurtosis for each series, while the skewness is negative in most cases. Note that the higher order moments are strongly influenced by outliers, so that the first subsample, which includes the October 1987 crash, shows the highest kurtosis coefficient and also, except for the Nikkei, the highest negative skewness coefficients. The first order autocorrelation coefficient is mostly negative, but only for S&P 500, Nasdaq 100, Nikkei 225 and Dow Jones it is statistically significant.

Figure 6 plots the Hill (1975) tail index estimator together with 95% confidence bands for the returns of the stock market indices. The Hill estimator is known to be optimal only for i.i.d. data from an exact Pareto distribution, see e.g. Hsing (1991) and Wagner and Marsh (2004), but Hill (2010) has shown that it remains consistent and asymptotically normal for a wide range of dependent processes including GARCH-type processes. The Hill estimates in

Figure 6 are plotted as a function of the number of tail observations used for estimation. Our estimates of the tail index vary around three for all series, which suggests that fourth order moments are unlikely to exist. As discussed above, this would invalidate the estimation theory of classical GARCH-in-mean models.

Tables 5 and 6 report the estimated EGARCH-M and Log-GARCH-M models for each series. The persistence is very high throughout, given by the estimated parameter $\hat{\beta}$ for the EGARCH-M and $\hat{\beta} + 0.5(\hat{\alpha}_+ + \hat{\alpha}_-)$ for the log-GARCH-M model, and is generally highest for the second subsample, including the dot-com bubble. There is strong evidence of leverage in all series, justifying the use of asymmetric GARCH models. The leverage effect is strongest in the third subsample, which comprises the financial crisis of 2008/09. An ARCH LM test applied to standardized residuals showed that there are no remaining conditional heteroskedasticity effects after estimation.

Regarding the EGARCH-M model, the estimated risk premium is significantly positive for all full samples except for the DAX and the Nikkei, and for all subsamples including the financial crisis of 2008/09. It is not significant for the first subsample in all series, and only in the S&P 500 and the Dow Jones it is significant for the second subsample. The invertibility condition is generally satisfied. Moreover, the condition $\delta \geq |\gamma|$ holds throughout as well, i.e. the size effect is larger than the sign effect in the news impact function of volatility.

For the Log-GARCH-M model, the risk premium parameter estimate $\hat{\lambda}_2$ is positive and statistically significant for all series and subsamples, except for subsample III of the S&P 500. It tends to be higher than for the EGARCH-M model, with a higher estimate of λ_2 in 15 out of 21 subsamples across the different series. Using the full samples, the Log-GARCH-M estimate of the risk premium parameter is larger than that of EGARCH-M in all cases except for the CAC 40 and the Dow Jones. The estimated constant in the volatility process, $\hat{\omega}$, is positive for all series, as opposed to the EGARCH-M estimates. This can be explained by the fact that ω is essentially a scaling parameter that determines the general level of (unconditional) volatility, and the link between ω and the unconditional variance differs in both models.

In terms of goodness-of-fit, using the likelihood value as criterion, the EGARCH-M model clearly provides overall a better fit, as its likelihood is superior for all considered sub-samples. This may suggest that, for our data, the logarithmic increase of the news impact function of

log volatility in Log-GARCH-M is too slow compared with the linear increase of EGARCH-M.

In Figure 7 we plot the estimated log-volatilities of all stock market indices for both EGARCH-M and Log-GARCH-M using the estimation results of Tables 5 and 6. It is obvious that the two risk premia models estimate the log-volatility processes in a very similar manner. The only case where notable differences occur is Dow Jones, for which the estimated EGARCH-M volatility shows more variation than the log-GARCH-M volatility. The results are supplemented by Table 7 in which we present summary statistics of the estimated conditional standard deviation for all daily stock market returns. As we can see, the mean, median and minimum of the estimated conditional standard deviations are very close for the two models. However, we notice a substantial difference in terms of the maximum values of the estimated conditional standard deviations, mainly for Dow Jones and Nikkei225, with 7.258 and 8.329 for EGARCH-M versus 3.378 and 5.558 for Log-GARCH-M, respectively. The explanation is the faster increase of the news impact function for EGARCH-M for larger shocks. Figure 8 shows the estimated news impact functions.

Table 7 also reports the standard deviation of the estimated conditional standard deviations as a measure of fluctuation, as well as their correlation. As we expected from Figure 7, the EGARCH-M model has a higher standard deviation than Log-GARCH-M for the Dow Jones index. For this series, the correlation between the two estimated log-volatilities is only 0.727, while for the other stock indices it is closer to unity.

5 Conclusions

In this paper, we propose two models that belong to a class of exponential-type GARCH models compatible with the intertemporal CAPM of Merton (1973), i.e. including a linear-in-variance risk premium. We investigate the asymptotic properties for the popular EGARCH-in-Mean model using the stochastic recurrence equation approach, including estimation theory for the quasi maximum likelihood estimator. A simulation study shows that the QMLE of the risk premium parameters in the Log-GARCH-M model are more precise than in the EGARCH-M model, but that the reverse holds for the volatility parameters. Our empirical application to a number of stock indices shows that EGARCH-M is generally preferred in terms of goodness-

of-fit, which suggests that the slow increase of the news impact curve in the log-GARCH-M model does not sufficiently take into account the impact of large shocks on volatility.

Acknowledgement

We are grateful to Luc Bauwens, Christian Conrad, Christian Francq, Enno Mammen, Francisco Penaranda, Genaro Sucarrat, Frédéric Vrins, the participants of the 11th Annual Meeting of the Society for Financial Econometrics SoFiE 2018 (Lugano), the 11th International Conference on Computational and Financial Econometrics (London), the EC2 Conference on Time-Varying Parameter Models (Amsterdam), the seminar series on Econometrics at CORE (UCLouvain), the 25th Annual Meeting of Belgian Statistical Society (BSS) 2017 (Leuven), the Inaugural Conference of the HeiKaMEtrics network on Financial Econometrics (Heidelberg), 2017 Econometric Society European Meeting (ESEM) (Lisbon), 2017 Asian Meeting of the Econometric Society (Hong Kong), the conference on Stochastic Dynamical Models in Mathematical Finance, Econometrics, and Actuarial Sciences (Lausanne), New Methods for the Empirical Analysis of Financial Markets conference (Comillas), the seminar series on Quantitative Methods at the University of St. Gallen, the seminar series of the Laboratory of Economic Management, Université de Lille, and the SNDE 2017 (Paris) conference. All remaining errors are solely our responsibility. The second author would like to acknowledge financial support from the MOVE-IN Louvain Post-doctoral Fellowship, co-funded by the Marie Curie Actions of the European Commission.

Appendix A The likelihood derivatives

Recall that the quasi log-likelihood function is given by $\bar{L}_T(\theta) = -\frac{1}{2} \frac{1}{T} \sum_{t=1}^T l_t(\theta)$, where

$$l_t(\theta) = \log(h_t(\theta)) + \frac{[Y_t - \lambda_1 - \lambda_2 h_t(\theta)]^2}{h_t(\theta)}. \quad (16)$$

Let $\varepsilon_t(\theta) := Y_t - \lambda_1 - \lambda_2 h_t(\theta)$. The first and second derivatives of $l_t(\theta)$ with respect to θ are given by

$$\frac{\partial l_t(\theta)}{\partial \theta} = \left(1 - \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right) \frac{\partial \log(h_t(\theta))}{\partial \theta} + \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta},$$

and

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^\top} &= \frac{\varepsilon_t^2(\theta)}{h_t(\theta)} \frac{\partial \log(h_t(\theta))}{\partial \theta} \frac{\partial \log(h_t(\theta))}{\partial \theta^\top} - \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial \log(h_t(\theta))}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^\top} \\ &+ \frac{2}{h_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^\top} + \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta^\top} - \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial \log(h_t(\theta))}{\partial \theta^\top} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \\ &+ \left(1 - \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right) \frac{\partial^2 \log(h_t(\theta))}{\partial \theta \partial \theta^\top}, \end{aligned}$$

where, denoting e_{in} the i -th unit vector in \mathbb{R}^n with $n = \dim(\Theta)$,

$$\begin{aligned} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} &= -e_{1n} - e_{2n} h_t(\theta) - \lambda_2 \frac{\partial h_t(\theta)}{\partial \theta} \\ \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta^\top} &= -\frac{\partial h_t(\theta)}{\partial \theta} e_{2n}^\top - e_{2n} \frac{\partial h_t(\theta)}{\partial \theta^\top} - \lambda_2 \frac{\partial^2 h_t(\theta)}{\partial \theta \partial \theta^\top} \end{aligned}$$

Appendix B Proofs

Appendix B.1 Consistency

Proof of Lemma 1

Following Bougerol (1993) and Straumann and Mikosch (2006), there is a unique stationary and ergodic solution if the function ϕ_t in (6) is contracting on average, i.e. if the Lyapunov exponent of the mapping is negative. The contraction condition of Bougerol (1993) can be used to ensure model invertibility and bounded moments for the filtered sequence, $\log \hat{h}_t(\theta)$. We consider a new functional SRE that is driven by a function of the observations Y_t

$$s_{t+1} = [\Phi_t(s_t)](\theta), \quad (17)$$

where the sequence of random functions Φ_t are given by

$$[\Phi_t(s)](\theta) = \omega + [\gamma(Y_t - \lambda_1 - \lambda_2 \exp(s(\theta))) + \delta(|Y_t - \lambda_1 - \lambda_2 \exp(s(\theta))|)] \exp(-s(\theta)/2) + \beta s(\theta),$$

for each $\theta \in \Theta$. The functions Φ_t map continuous functions $s : \Theta \rightarrow [\inf_{\Theta} \omega(1 - \beta)^{-1}, +\infty)$ onto the class of such functions. Because Φ_0 is Lipschitz continuous, and in general a Lipschitz continuous function is differentiable pointwise almost everywhere due to Rademacher's theorem (Evans and Gariepy, 2015, p. 103), continuous differentiability is not necessary to prove that its first derivative is bounded. The Lipschitz constant Λ , in this case, is given by the essential

supremum of its derivative, ignoring any set of elements of Lebesgue measure zero where the derivative of the random functions Φ_t is not defined, which is the following

$$\left| \frac{\partial \Phi_0(s)}{\partial s} \right| = \left| \begin{array}{c} \beta - 2^{-1} [\gamma (Y_0 - \lambda_1 - \lambda_2 \exp(s(\theta))) + \delta (|Y_0 - \lambda_1 - \lambda_2 \exp(s(\theta))|)] \exp(-s(\theta)/2) \\ -\lambda_2 \exp(s(\theta)) [\gamma + \delta \operatorname{sgn}(Y_0 - \lambda_1 - \lambda_2 \exp(s(\theta)))] \exp(-s(\theta)/2) \end{array} \right|.$$

Hence,

$$\Lambda = \sup_{\Theta} \left| \begin{array}{c} \beta - \frac{1}{2} [\gamma (Y_0 - \lambda_1 - \lambda_2 \exp(s(\theta))) + \delta (|Y_0 - \lambda_1 - \lambda_2 \exp(s(\theta))|)] \exp(-\frac{1}{2} \inf_{\Theta} \omega (1 - \beta)^{-1}) \\ -\lambda_2 [\gamma + \delta \operatorname{sgn}(Y_0 - \lambda_1 - \lambda_2 \exp(s(\theta)))] \exp(2^{-1} \inf_{\Theta} \omega (1 - \beta)^{-1}) \end{array} \right|$$

The condition $E[\log \Lambda] < 0$, which implies continuous invertibility, is pointwise and reads

$$E[\log \max\{\beta, \Psi_0\}] < 0,$$

where

$$\begin{aligned} \Psi_0 &= 2^{-1} [\gamma \varepsilon_0(\theta) + \delta |\varepsilon_0(\theta)|] \exp\left(-2^{-1} \inf_{\Theta} \omega (1 - \beta)^{-1}\right) \\ &\quad + \lambda_2 [\gamma + \delta \operatorname{sgn}(\varepsilon_0(\theta))] \exp\left(2^{-1} \inf_{\Theta} \omega (1 - \beta)^{-1}\right) - \beta, \end{aligned}$$

with $\varepsilon_0(\theta) := Y_0 - \lambda_1 - \lambda_2 h_0(\theta)$ and

$$\log h_0(\theta) = \omega + \gamma \frac{\varepsilon_{-1}(\theta)}{\sqrt{h_{-1}(\theta)}} + \delta \left| \frac{\varepsilon_{-1}(\theta)}{\sqrt{h_{-1}(\theta)}} \right| + \beta \log h_{-1}(\theta).$$

Since the uniform log moments exist by continuity of the Lipschitz coefficient in θ , having excluded the zero point of discontinuity, and because $E[\log^+(Y_t - \lambda_1 - \lambda_2 h_t(\theta))] < \infty$, together with the last result for each $\theta \in \Theta$ the condition (CI) for continuous invertibility of Wintenberger (2013) is satisfied.

The lemma follows from an application of Theorem 3.1 in Bougerol (1993), as its conditions are met since $E[\log \Lambda] < 0$ which implies $E[\log^+ \Lambda] < \infty$ and also $E[\log + \|\Phi_t(s) - s\|] < \infty$ for a constant function s by the fact that Y_t has a bounded first moment. The invertibility condition is sufficient for

$$\sup_{\Theta} \left| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right| \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty.$$

Also, since $\log h_t(\theta)$ is stationary and because

$$E(\log h_0) \leq (\omega + \delta)(1 - \beta)^{-1} < \infty,$$

this implies that

$$\sup_{\Theta} \left| \widehat{h}_t(\theta) - h_t(\theta) \right| \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty, \quad (18)$$

by the mean value theorem and Lemma 2.5.4. of Straumann (2005). \square

Proof of Theorem 1. The proof of consistency is obtained by the uniform convergence of the criterion function and the identifiable uniqueness of $\theta_0 \in \Theta$. The proof is given by showing that the following statements hold:

- (i) $\sup_{\Theta} \left| \widehat{L}_T(\theta) - \overline{L}_T(\theta) \right| \rightarrow 0 \quad a.s.$,
- (ii) $\sup_{\Theta} \left| \overline{L}_T(\theta) - L(\theta) \right| \rightarrow 0 \quad a.s.$,
- (iii) $L(\theta) < L(\theta_0), \forall \theta \in \Theta : \theta \neq \theta_0$.

By the triangle inequality,

$$\sup_{\Theta} \left| \widehat{L}_T(\theta) - L(\theta) \right| \leq \sup_{\Theta} \left| \widehat{L}_T(\theta) - \overline{L}_T(\theta) \right| + \sup_{\Theta} \left| \overline{L}_T(\theta) - L(\theta) \right| \quad (19)$$

the convergence of $\widehat{L}_T(\theta)$ to $L(\theta)$ follows by showing that the two terms on the right hand side of (19) converge, and that the limit $L(\theta) := E\overline{L}_T(\theta)$ has a unique maximum at $\theta = \theta_0$. The convergence of the second term, $\sup_{\Theta} \left| \overline{L}_T(\theta) - L(\theta) \right|$ follows by using the ergodic theorem for separable Banach spaces (Theorem 2.7 in Straumann and Mikosch, 2006) and from the fact that $\log h_t$ is a stationary ergodic process with $E |\log(h_t)| < \infty$ by its almost sure representation (8).

To show (i), note that the components of the conditional log-likelihood function $\widehat{L}_T = \frac{1}{T} \sum_{t=1}^T \left[\log \left(\widehat{h}_t(\theta) \right) + \frac{(Y_t - \lambda_1 - \lambda_2 \widehat{h}_t(\theta))^2}{\widehat{h}_t(\theta)} \right]$ are given by $\log \left(\widehat{h}_t(\theta) \right)$, $\frac{1}{\widehat{h}_t(\theta)} (Y_t + Y_t^2)$, $\widehat{h}_t(\theta)$, and Y_t . The functions $\frac{1}{\widehat{h}_t(\theta)}$ and $\log \left(\widehat{h}_t(\theta) \right)$ are Lipschitz continuous functions, since they have a bounded first derivative by Assumption 2. An application of the mean value theorem to $\left| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right|$ and $\left| \frac{1}{\widehat{h}_t(\theta)} - \frac{1}{h_t(\theta)} \right|$ implies that, for any $\theta \in \Theta$, there exists a constant $c > 0$ which does not depend on θ such that

$$\sup_{\Theta} \left| \widehat{L}_T(\theta) - \overline{L}_T(\theta) \right| \leq \frac{c}{T} \left[\sum_{t=1}^T (1 + Y_t^2) \left\| \widehat{h}_t(\theta) - h_t(\theta) \right\|_{\Theta} \right],$$

by using that $1 + |x| + x^2 \leq 2(1 + x^2)$. But, by (18), $\left\| \widehat{h}_t(\theta) - h_t(\theta) \right\|_{\Theta}$ uniformly converges to zero e.a.s. We also need the integrability condition $E \log^+(1 + Y_t^2) < \infty$ to ensure the a.s. uniform convergence to zero of the first term of (19), by applying Lemma 2.1

of Straumann and Mikosch (2006). There exists a constant c such that $E \log^+(1 + Y_t^2) \leq cE \log^+(Y_t^2) \leq cE(Y_t^2) < \infty$, where the last inequality follows by Assumption 2a. This implies that $[\widehat{L}_T(\theta) - \bar{L}_T(\theta)]$ converges almost surely to 0, uniformly over $\theta \in \Theta$.

For the last claim, i.e. $L(\theta) < L(\theta_0)$ for $\theta \neq \theta_0$, we check that $\log h_t(\theta) = \log h_t(\theta_0)$ *a.s.* implies $\theta = \theta_0$. Recall that $\log h_t(\theta)$ is a stationary process, and let $h_t := h_t(\theta_0)$. Then,

$$\begin{aligned} \log h_1(\theta) - \log h_1 &= \omega - \omega_0 + \gamma \frac{Y_0 - \lambda_1 - \lambda_2 h_0}{\sqrt{h_0}} - \gamma_0 \frac{Y_0 - \lambda_{01} - \lambda_{02} h_0}{\sqrt{h_0}} \\ &+ \delta \left| \frac{Y_0 - \lambda_1 - \lambda_2 h_0}{\sqrt{h_0}} \right| - \delta_0 \left| \frac{Y_0 - \lambda_{01} - \lambda_{02} h_0}{\sqrt{h_0}} \right| + (\beta - \beta_0) \log h_0 \end{aligned}$$

After substituting for Y_0 and setting $\log h_1(\theta) = \log h_1$ *a.s.* one obtains

$$\begin{aligned} 0 &= \omega - \omega_0 + \gamma \left(Z_0 - \frac{\lambda_1 - \lambda_{01}}{\sqrt{h_0}} - (\lambda_2 - \lambda_{02})\sqrt{h_0} \right) - \gamma_0 Z_0 \\ &+ \delta \left| Z_0 - \frac{\lambda_1 - \lambda_{01}}{\sqrt{h_0}} - (\lambda_2 - \lambda_{02})\sqrt{h_0} \right| - \delta_0 |Z_0| + (\beta - \beta_0) \log h_0 \quad \textit{a.s.} \end{aligned}$$

Following the argument of Straumann and Mikosch (2006), suppose that $\lambda_1 \neq \lambda_{01}$, $\lambda_2 \neq \lambda_{02}$, or $\beta \neq \beta_0$. In that case, h_0 would be both independent of Z_0 and an implicit function of it, which implies that it is deterministic. However, $Var(\log h_0) = Var(\gamma_0 Z_0 + \delta_0 |Z_0|)/(1 - \beta^2)$, which is strictly positive because of Assumption 4. By contradiction, necessarily $\lambda_1 = \lambda_{01}$, $\lambda_2 = \lambda_{02}$, and $\beta = \beta_0$. The remaining equalities $\omega = \omega_0$, $\delta = \delta_0$ and $\gamma = \gamma_0$ then follow exactly as in Straumann and Mikosch (2006). Thus, $\log h_t(\theta) = \log h_t(\theta_0)$ *a.s.* implies that $\theta = \theta_0$, which completes the proof. \square

Appendix B.2 Asymptotic normality

To show asymptotic normality of the estimator, it is essential to analyze the asymptotic behavior of the sequences of continuous functions $\partial \log \widehat{h}_t(\theta) / \partial \theta$ and $\partial^2 \log \widehat{h}_t(\theta) / \partial \theta \partial \theta^\top$. Applying Proposition 5.5.1 of Straumann (2005), we show the existence, stationarity and ergodicity of $\partial \log h_t(\theta) / \partial \theta$ and $\partial^2 \log h_t(\theta) / \partial \theta \partial \theta^\top$, and that they are approximations of $\partial \log \widehat{h}_t(\theta) / \partial \theta$ and $\partial^2 \log \widehat{h}_t(\theta) / \partial \theta \partial \theta^\top$, respectively. As in Lemma 1, we use the functional SRE $[\Phi_t(s)](\theta) = g_\theta(Y_t, s(\theta))$ for the filtered log-volatility process, where

$$g_\theta(y, s) = \omega + [\gamma(y - \lambda_1 - \lambda_2 \exp(s)) + \delta |y - \lambda_1 - \lambda_2 \exp(s)|] \exp(-s/2) + \beta s,$$

which is twice differentiable pointwise almost everywhere with respect to s and θ , ignoring any set of elements of Lebesgue measure zero where the derivative of the random functions Φ_t is not defined. The partial derivatives are given by

$$\begin{aligned} \frac{\partial g_\theta}{\partial s} &= \beta - \left\{ \begin{array}{l} 2^{-1} [\gamma (y - \lambda_1 - \lambda_2 \exp(s)) + \delta (|y - \lambda_1 - \lambda_2 \exp(s)|)] \\ + \lambda_2 [\gamma + \delta \operatorname{sgn}(y - \lambda_1 - \lambda_2 \exp(s))] \exp(s) \end{array} \right\} \exp(-s/2), \quad (20) \\ \frac{\partial g_\theta}{\partial \theta} &= \begin{pmatrix} -[\gamma + \delta \operatorname{sgn}(y - \lambda_1 - \lambda_2 \exp(s))] \exp(-s/2) \\ -[\gamma + \delta \operatorname{sgn}(y - \lambda_1 - \lambda_2 \exp(s))] \exp(s/2) \\ 1 \\ (y - \lambda_1 - \lambda_2 \exp(s)) \exp(-s/2) \\ (|y - \lambda_1 - \lambda_2 \exp(s)|) \exp(-s/2) \\ s \end{pmatrix}, \quad (21) \end{aligned}$$

where the sign function is known to have bounded derivatives defined pointwise almost everywhere except at a point of Lebesgue measure zero.

In order to find a stationary approximation of $\partial \log \widehat{h}_t(\theta) / \partial \theta$, we can obtain a new linear SRE of the first order derivative as

$$\frac{\partial \log h_{t+1}(\theta)}{\partial \theta} = A_t \frac{\partial \log h_t(\theta)}{\partial \theta} + B_t, \quad (22)$$

where

$$\begin{aligned} A_t &= \beta - \left\{ 2^{-1} [\gamma \varepsilon_t(\theta) + \delta |\varepsilon_t(\theta)|] + \lambda_2 [\gamma + \delta \operatorname{sgn}(\varepsilon_t(\theta))] h_t(\theta) \right\} h_t(\theta)^{-1/2} \\ B_t &= \begin{pmatrix} -[\gamma + \delta \operatorname{sgn}(\varepsilon_t(\theta))] h_t(\theta)^{-1/2} \\ -[\gamma + \delta \operatorname{sgn}(\varepsilon_t(\theta))] h_t(\theta)^{1/2} \\ 1 \\ \varepsilon_t(\theta) h_t(\theta)^{-1/2} \\ |\varepsilon_t(\theta)| h_t(\theta)^{-1/2} \\ \log h_t(\theta) \end{pmatrix}^\top. \end{aligned}$$

with $\varepsilon_t(\theta) := Y_t - \lambda_1 - \lambda_2 h_t(\theta)$. For the proof of Theorem 2, we will use the following lemmas.

Lemma 2 *Let the Assumptions 1-6 hold. Eq. (22) has a unique stationary solution $\partial \log h_t(\theta) / \partial \theta$ that is ergodic. There exists a $\delta > 0$ such that*

$$\sup_{\|\theta - \theta_0\| \leq \delta} \left\| \frac{\partial \log \widehat{h}_t(\theta)}{\partial \theta} - \frac{\partial \log h_t(\theta)}{\partial \theta} \right\| \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty.$$

Proof of Lemma 2

The proof follows the lines of Proposition 5.5.1 of Straumann (2005) and Theorem 8 of Wintenberger (2013), which essentially show the existence of a unique stationary solution to the stochastic recurrence equation (SRE), which is ergodic. We need similar contraction techniques as in the proof of Lemma 1. Equation (22) is a linear SRE, $\partial \log h_{t+1}(\theta) / \partial \theta = \Phi'_t [\partial \log h_t(\theta) / \partial \theta]$, where $\Phi'_t(s) := \partial \Phi_t(s) / \partial s$ is the stationary approximation of $\partial \widehat{\Phi}_t(s) / \partial s$. By Assumption 2, log-volatility is bounded from below, i.e. there exists a constant s^* such that $\inf_{\Theta} \log(h_t) \geq s^*$ a.s. Note also that except for a point with Lebesgue measure zero, both A_t and B_t are continuous functions in (s, θ) and that g_θ in (12) is differentiable pointwise almost everywhere. There exists a $\delta > 0$ such that

$$\left\| \widehat{\Phi}'_t(0) - \Phi'_t(0) \right\| = \left\| \widehat{B}_t - B_t \right\|_{\|\theta - \theta_0\| \leq \delta} \leq \left\| \sup_{s \geq s^*} \frac{\partial^2 g_\theta}{\partial s \partial \theta} \right\|_{\|\theta - \theta_0\| \leq \delta} \left\| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right\|_{\|\theta - \theta_0\| \leq \delta},$$

where

$$\frac{\partial^2 g_\theta}{\partial \theta \partial s} = \begin{pmatrix} 2^{-1} [\gamma + \delta \operatorname{sgn}(g(s))] \exp(s/2) \\ -2^{-1} [\gamma + \delta \operatorname{sgn}(g(s))] \exp(s/2) \\ 0 \\ [-2^{-1}g(s) - \lambda_2 \exp(s)] \exp(-s/2) \\ \{[-2^{-1}g(s) - \lambda_2 \exp(s)] \exp(-s/2)\} \operatorname{sgn}(g(s)) \\ 1 \end{pmatrix},$$

where $g(s) = y - \lambda_1 - \lambda_2 \exp(s)$. Because

$$E \left[\log^+ \left\| \sup_{s \geq s^*} \frac{\partial^2 g_\theta}{\partial \theta \partial s} \right\| \right] < \infty,$$

as $E \left[\log^+ \left| Y_t - \lambda_1 - \lambda_2 \widehat{h}_t \right| \right] < \infty$ and $E \left[\log^+ \widehat{h}_t \right] < \infty$, then for any $|\rho| < 1$,

$$\sum_{t=0}^{\infty} \left(\rho^t \left\| \sup_{s \geq s^*} \frac{\partial^2 g_\theta}{\partial \theta \partial s} \right\| \right) < \infty,$$

by an application of the Borel-Cantelli lemma (see Lemma 2.2 of Berkes *et al.* 2003). This means that $\sum_{t=0}^{\infty} \left\| \sup_{s \geq s^*} \frac{\partial^2 g_\theta}{\partial \theta \partial s} \right\| \left\| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right\|_{\|\theta - \theta_0\| \leq \delta}$ converges a.s. and hence

$$\left\| \widehat{\Phi}'_t(0) - \Phi'_t(0) \right\| \rightarrow 0 \quad e.a.s.$$

because $\left\| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right\|_{\|\theta - \theta_0\| \leq \delta} \rightarrow 0$ e.a.s. by Lemma 1.

We also have

$$\Lambda \left(\widehat{\Phi}'_t - \Phi'_t \right) \leq \left\| \widehat{A}_t - A_t \right\|_{\|\theta - \theta_0\| \leq \delta} \leq \left\| \sup_{s \geq s^*} \frac{\partial^2 g_\theta}{\partial s^2} \right\|_{\|\theta - \theta_0\| \leq \delta} \left\| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right\|_{\|\theta - \theta_0\| \leq \delta},$$

where

$$\begin{aligned} \frac{\partial^2 g_\theta}{\partial s^2} &= \left\{ 4^{-1} [\gamma g(s) + 2^{-1} \lambda_2 \exp(s) + (\delta g(s) + 2^{-1} \lambda_2 \exp(s)) \operatorname{sgn}(g(s))] \right\} \exp(-s/2) \\ &\quad + \left\{ 2^{-1} \lambda_2 [\gamma + \delta \operatorname{sgn}(g(s))] \right\} \exp(s/2). \end{aligned}$$

Again, because

$$E \left[\log^+ \left\| \sup_{s \geq s^*} \frac{\partial^2 g_\theta}{\partial s^2} \right\| \right] < \infty,$$

as $E \left[\log^+ \left| Y_t - \lambda_1 - \lambda_2 \widehat{h}_t \right| \right] < \infty$, then $\Lambda \left(\widehat{\Phi}'_t - \Phi'_t \right) \rightarrow 0$ e.a.s. by an application of the Borel-Cantelli lemma as before, since we also have that

$$\left\| \log \widehat{h}_t(\theta) - \log h_t(\theta) \right\|_{\|\theta - \theta_0\| \leq \delta} \rightarrow 0 \quad e.a.s..$$

Now, because (i) the SRE (22) has a unique stationary ergodic solution as it is contractive by virtue of $E(\log \Lambda(\Phi'_t)) = E(\log \|A_t\|) < 0$ from Lemma 1, (ii) the filtered log-volatility process is invertible, and (iii) $\log h_t(\theta)$ is differentiable as a consequence of Proposition 5.5.1 of Straumann (2005) and Theorem 8 of Wintenberger (2013), it follows that $\partial \log \widehat{h}_t / \partial \theta$ converges to the unique stationary ergodic solution uniformly locally in a neighborhood of θ_0 , that is

$$\left\| \frac{\partial \log \widehat{h}_t}{\partial \theta} - \frac{\partial \log h_t}{\partial \theta} \right\|_{\|\theta - \theta_0\| \leq \delta} \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty.$$

□

We now consider the second order derivatives and obtain a new linear SRE, that is

$$\frac{\partial^2 \log h_{t+1}(\theta)}{\partial \theta \partial \theta^\top} = A_t \frac{\partial^2 \log h_t(\theta)}{\partial \theta \partial \theta^\top} + C_t, \quad (23)$$

where $A_t := \frac{\partial^2 g_\theta}{\partial s \partial s^\top}$ and $C_t := \frac{\partial^2 g_\theta}{\partial \theta \partial \theta^\top} + 2 \frac{\partial^2 g_\theta}{\partial s \partial \theta} \frac{\partial \log h_t(\theta)}{\partial \theta}$. The next lemma implies the existence and uniqueness of a stationary and ergodic solution of the SRE (23).

Lemma 3 Under Assumptions 1-6, Eq. (23) has a unique stationary solution $\partial^2 \log h_t(\theta) / \partial \theta \partial \theta^\top$ that is ergodic. There exists a $\delta > 0$ such that

$$\sup_{\|\theta - \theta_0\| \leq \delta} \left\| \frac{\partial^2 \log \widehat{h}_t(\theta)}{\partial \theta \partial \theta^\top} - \frac{\partial^2 \log h_t(\theta)}{\partial \theta \partial \theta^\top} \right\| \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty.$$

Proof of Lemma 3

The proof is analogous to the proof of Lemma 2 for the first derivative, applying Proposition 5.5.1 of Straumann (2005) and Theorem 8 of Wintenberger (2013). Notice that the stationary solution $\partial^2 \log h_t(\theta) / \partial \theta \partial \theta^\top$ is continuous, as it is the locally uniform limit law of $\partial^2 \log \widehat{h}_t(\theta) / \partial \theta \partial \theta^\top$ that is continuous by definition. \square

The following two lemmas state that for the limiting processes of the derivatives of $\log \widehat{h}_t(\theta)$, the contraction condition is satisfied for their functional SRE restricted to $\mathcal{V}(\theta_0) := \{\theta : \|\theta - \theta_0\| \leq \delta\}$, having $\theta_T^* \in \mathcal{V}(\theta_0)$ for T sufficiently large.

Lemma 4 Let the Assumptions 1-6 hold. For any sequence $\widehat{\theta}_T$ such that $\widehat{\theta}_T \rightarrow \theta_0$ almost surely, it holds that

$$\left\| \frac{\log h_t(\widehat{\theta}_T)}{\partial \theta} - \frac{\log h_t(\theta_0)}{\partial \theta} \right\| = o_p(1).$$

Proof of Lemma 4

From (22) with $h_t := h_t(\theta_0)$ and letting $\theta = \theta_0$, we get

$$\begin{aligned} A_t|_{\theta_0} &= \beta_0 - \left[2^{-1} (\gamma_0 Z_t + \delta_0 |Z_t|) + \lambda_{02} [\gamma_0 + \delta_0 \operatorname{sgn}(Z_t)] \sqrt{h_t} \right], \quad \text{and} \\ B_t|_{\theta_0} &= \left(-[\gamma_0 + \delta_0 \operatorname{sgn}(Z_t)] h_t^{-1/2}, -[\gamma_0 + \delta_0 \operatorname{sgn}(Z_t)] h_t^{1/2}, 1, Z_t, |Z_t|, \log h_t \right)^\top. \end{aligned}$$

Let $\nabla g_t(\theta) := \partial \log h_t(\theta) / \partial \theta$. The SRE (22) can be written as

$$\nabla g_{t+1}(\theta) = \Phi'_t(g_t(\theta), \theta) \nabla g_t(\theta) + \nabla_\theta \Phi_t(g_t(\theta), \theta), \quad t \in \mathbb{Z}$$

where $\Phi'_t(g_t(\theta), \theta) =: A_t$ in (20) and $\nabla_\theta \Phi_t(g_t(\theta), \theta) =: B_t$ in (21).

Notice that we can rewrite (22), letting

$$\zeta_t(\theta) := \frac{\varepsilon_t(\theta)}{\sqrt{h_t(\theta)}} = \frac{\sqrt{h_t} Z_t + \lambda_{01} + \lambda_{02} h_t - \lambda_1 - \lambda_2 h_t(\theta)}{\sqrt{h_t(\theta)}}$$

with $\zeta_t(\theta_0) = Z_t$, as

$$\begin{aligned} A_t &= \beta - 2^{-1} [\gamma \zeta_t(\theta) + \delta |\zeta_t(\theta)|] + \lambda_2 [\gamma + \delta \operatorname{sgn}(\zeta_t(\theta))] \sqrt{h_t(\theta)} \\ B_t &= (-[\gamma + \delta \operatorname{sgn}(\zeta_t(\theta))] h_t(\theta)^{-1/2}, -[\gamma + \delta \operatorname{sgn}(\zeta_t(\theta))] h_t(\theta)^{1/2}, 1, \zeta_t(\theta), |\zeta_t(\theta)|, \log h_t(\theta))^\top \end{aligned}$$

By similar arguments as in the proof of Lemma 3 in Wintenberger (2013) we can argue that there exists a positive random variable a such that

$$|A_t - A_t|_{\theta_0}| + \|B_t - B_t|_{\theta_0}\| \leq a (\|\theta - \theta_0\| + |\log h_t(\theta) - \log h_t(\theta_0)|),$$

that is, for any θ that belongs to a compact neighborhood of θ_0 , $\mathcal{V}(\theta_0)$, we can apply a local continuity argument with respect to the parameters to the Lipschitz coefficients of the SRE (22).

Thus, for any sequence $(\hat{\theta}_T)$ such that $\|\hat{\theta}_T - \theta_0\| \xrightarrow{a.s.} 0$ as $T \rightarrow \infty$ and $\hat{\theta}_T \in \mathcal{V}(\theta_0)$ for T sufficiently large, using the consistency result of Theorem 1, we have

$$\begin{aligned} \left\| \frac{\partial \log h_{t+1}(\hat{\theta}_T)}{\partial \theta} - \frac{\partial \log h_{t+1}(\theta_0)}{\partial \theta} \right\| &= \left\| A_t|_{\hat{\theta}_T} \frac{\partial \log h_t(\hat{\theta}_T)}{\partial \theta} + B_t|_{\hat{\theta}_T} - A_t|_{\theta_0} \frac{\partial \log h_t}{\partial \theta} - B_t|_{\theta_0} \right\| \\ &\leq |A_t|_{\theta_0}| \left\| \frac{\partial \log h_t(\hat{\theta}_T)}{\partial \theta} - \frac{\partial \log h_t}{\partial \theta} \right\| \\ &\quad + \left\| \frac{\partial \log h_t(\hat{\theta}_T)}{\partial \theta} \right\| |A_t|_{\hat{\theta}_T} - A_t|_{\theta_0}| + \|B_t|_{\hat{\theta}_T} - B_t|_{\theta_0}\| \\ &\leq |A_t|_{\theta_0}| \left\| \frac{\partial \log h_t(\hat{\theta}_T)}{\partial \theta} - \frac{\partial \log h_t}{\partial \theta} \right\| + a \times \\ &\quad \left(\left\| \frac{\partial \log h_t}{\partial \theta} \right\|_{\mathcal{V}(\theta_0)} + 1 \right) (\|\hat{\theta}_T - \theta_0\| + |\log h_t(\hat{\theta}_T) - \log h_t|) \end{aligned}$$

Now $\|\partial \log h_t(\hat{\theta}_T)/\partial \theta - \partial \log h_t(\theta_0)/\partial \theta\|$ is Césaro summable, i.e. its arithmetic mean of the first T partial sums tends to a limit as $T \rightarrow \infty$, and there exist a positive random variable a^* and a random continuous function b satisfying $b(\theta_0) = 0$ a.s.,

$$T^{-1} \sum_{t=1}^T \left\| \frac{\partial \log h_t(\hat{\theta}_T)}{\partial \theta} - \frac{\partial \log h_t(\theta_0)}{\partial \theta} \right\| \leq a^* \|\hat{\theta}_T - \theta_0\| + b(\hat{\theta}_T).$$

Hence, for any sequence $(\widehat{\theta}_T)$ converging a.s. to θ_0 ,

$$P \left[\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \left\| \frac{\partial \log h_t(\widehat{\theta}_T)}{\partial \theta} - \frac{\partial \log h_t(\theta_0)}{\partial \theta} \right\| = 0 \right] = 1$$

which implies that, as stated, $\left\| \frac{\partial \log h_t(\widehat{\theta}_T)}{\partial \theta} - \frac{\partial \log h_t(\theta_0)}{\partial \theta} \right\| = o_p(1)$. \square

Lemma 5 *Let the Assumptions 1-6 hold. For any sequence $\widehat{\theta}_T$ such that $\widehat{\theta}_T \rightarrow \theta_0$ almost surely, it holds that*

$$\left\| \frac{\partial^2 \log h_t(\widehat{\theta}_T)}{\partial \theta \partial \theta^\top} - \frac{\partial^2 \log h_t(\theta_0)}{\partial \theta \partial \theta^\top} \right\| = o_p(1).$$

Proof of Lemma 5 By Lemma 4, taking the derivative in $\left\| \frac{\partial \log h_{t+1}(\widehat{\theta}_T)}{\partial \theta} - \frac{\partial \log h_{t+1}(\theta_0)}{\partial \theta} \right\|$, the second derivative of the log volatility is also Césaro summable, such that

$$T^{-1} \sum_{t=1}^T \left\| \frac{\partial^2 \log h_t(\widehat{\theta}_T)}{\partial \theta \partial \theta^\top} - \frac{\partial^2 \log h_t(\theta_0)}{\partial \theta \partial \theta^\top} \right\| \leq a' \|\widehat{\theta}_T - \theta_0\| + b'(\widehat{\theta}_T),$$

with a' a positive random variable and b' a random continuous function which satisfies $b'(\theta_0) = 0$ a.s. Then

$$\lim_{n \rightarrow \infty} \sum_{t=I}^n \left\| \frac{\partial^2 \log h_t(\widehat{\theta}_T)}{\partial \theta \partial \theta^\top} - \frac{\partial^2 \log h_t(\theta_0)}{\partial \theta \partial \theta^\top} \right\| = 0$$

a.s. for a random integer $I > 1$. Due to the fact that $\partial^2 \log h_t / \partial \theta \partial \theta^\top$ is continuous in its arguments, for any sequence $(\widehat{\theta}_T)$ converging a.s. to θ_0 , $\frac{\partial^2 \log h_t(\widehat{\theta}_T)}{\partial \theta \partial \theta^\top} - \frac{\partial^2 \log h_t(\theta_0)}{\partial \theta \partial \theta^\top} \rightarrow 0$ a.s. \square

Proof of Theorem 2.

Sufficient conditions for the asymptotic normality of the estimator are the following:

$$(a) \sqrt{T} \widehat{L}'_T(\theta_0) - \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\frac{1}{2} \frac{\partial \log h_t(\theta_0)}{\partial \theta} (Z_t^2 - 1) + \frac{1}{\sqrt{h_t(\theta_0)}} Z_t \left(e_{1n} + h_t(\theta_0) e_{2n} + \lambda_{02} \frac{\partial h_t(\theta_0)}{\partial \theta} \right) \right] = o_p(1),$$

and $\sqrt{T} \widehat{L}'_T(\theta_0) \rightarrow \mathcal{N}(0, G(\theta_0))$, where $G(\theta_0) := E \left[\nabla l_t(\theta_0) \nabla l_t(\theta_0)^\top \right]$ with ∇ the gradient vector of l_t ,

$$(b) \left\| \widehat{L}''_T(\widehat{\theta}_T) - F(\theta_0) \right\| \xrightarrow{a.s.} 0, \text{ for some random sequence } \widehat{\theta}_T \text{ converging a.s. to } \theta_0 \text{ as } T \rightarrow \infty,$$

and $F(\theta_0)$ is invertible, where $F(\theta_0) := -E[H_t(\theta_0)]$ with H_t the Hessian matrix of l_t ,

$$(c) \sqrt{T} \left\| \widehat{L}'_T(\widehat{\theta}_T) - \overline{L}'_T(\widehat{\theta}_T) \right\| \xrightarrow{a.s.} 0.$$

We have that

$$\widehat{L}'_T(\theta_0) = \frac{1}{2} \frac{1}{T} \sum_{t=1}^T (Z_t^2 - 1) \frac{\partial \log \widehat{h}_t(\theta_0)}{\partial \theta} + \frac{1}{T} \sum_{t=1}^T \frac{1}{\sqrt{\widehat{h}_t(\theta_0)}} Z_t \left(e_{1n} + \widehat{h}_t(\theta_0) e_{2n} + \lambda_{02} \frac{\partial \widehat{h}_t(\theta_0)}{\partial \theta} \right).$$

The weak convergence of the first part of (a) follows by an application of Lemmas 1 and 2, since the exponential almost sure rate in Lemma 2 implies that

$$\left\| \widehat{L}'_T(\theta_0) - \overline{L}'_T(\theta_0) \right\| \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty,$$

by using the mean-value theorem on $\left\| \widehat{L}'_T(\theta_0) - \overline{L}'_T(\theta_0) \right\|$. The last result in turn implies that $\sqrt{T} \left\| \widehat{L}'_T(\theta_0) - \overline{L}'_T(\theta_0) \right\| \xrightarrow{a.s.} 0$ as $T \rightarrow \infty$. Since $\widehat{\theta}_T \rightarrow \theta_0$ by consistency, this also proves statement (c). By an application of a martingale central limit theorem on $\widehat{L}'_T(\theta_0)$, see Hall and Heyde (1980), the second part of statement (a) follows because

$$G(\theta_0) := E \left[\nabla l_t(\theta_0) \nabla l_t(\theta_0)^\top \right]$$

exists by Assumption 6, where

$$\nabla l_t(\theta_0) = \frac{1}{2} \frac{\partial \log h_t(\theta_0)}{\partial \theta} (Z_t^2 - 1) + \frac{1}{\sqrt{h_t(\theta_0)}} Z_t \left(e_{1n} + h_t(\theta_0) e_{2n} + \lambda_{02} \frac{\partial h_t(\theta_0)}{\partial \theta} \right).$$

For the statement (b), let us first recall that $\widehat{L}''_T(\theta) = -\frac{1}{2} \frac{1}{T} \sum_{t=1}^T H_t(\theta)$, where $H_t(\theta) := \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^\top}$ and

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^\top} &= \frac{\varepsilon_t^2(\theta)}{h_t(\theta)} \frac{\partial \log(h_t(\theta))}{\partial \theta} \frac{\partial \log(h_t(\theta))}{\partial \theta^\top} - \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial \log(h_t(\theta))}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^\top} \\ &\quad + \frac{2}{h_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^\top} + \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta^\top} - \frac{2\varepsilon_t(\theta)}{h_t(\theta)} \frac{\partial \log(h_t(\theta))}{\partial \theta^\top} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \\ &\quad + \left(1 - \frac{\varepsilon_t^2(\theta)}{h_t(\theta)} \right) \frac{\partial^2 \log(h_t(\theta))}{\partial \theta \partial \theta^\top}, \end{aligned}$$

where

$$\begin{aligned} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} &= -e_{1n} - e_{2n} h_t(\theta) - \lambda_2 \frac{\partial h_t(\theta)}{\partial \theta} \\ \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta^\top} &= -\frac{\partial h_t(\theta)}{\partial \theta} e_{2n}^\top - e_{2n} \frac{\partial h_t(\theta)}{\partial \theta^\top} - \lambda_2 \frac{\partial^2 h_t(\theta)}{\partial \theta \partial \theta^\top} \end{aligned}$$

with expectation given by

$$F(\theta_0) := E[H_t(\theta_0)] = E \frac{\partial \log(h_t(\theta))}{\partial \theta} \frac{\partial \log(h_t(\theta))}{\partial \theta^\top} + E \frac{2}{h_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^\top}$$

We have for a fixed local neighborhood $\mathcal{V}(\theta_0) := \|\theta - \theta_0\| \leq \delta$ that

$$\|H_t(\theta)\|_{\mathcal{V}(\theta_0)} \leq \left\| \frac{\partial \log(h_t(\theta))}{\partial \theta} \frac{\partial \log(h_t(\theta))}{\partial \theta^\top} \right\|_{\mathcal{V}(\theta_0)} + \left\| \frac{2}{h_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^\top} \right\|_{\mathcal{V}(\theta_0)}. \quad (24)$$

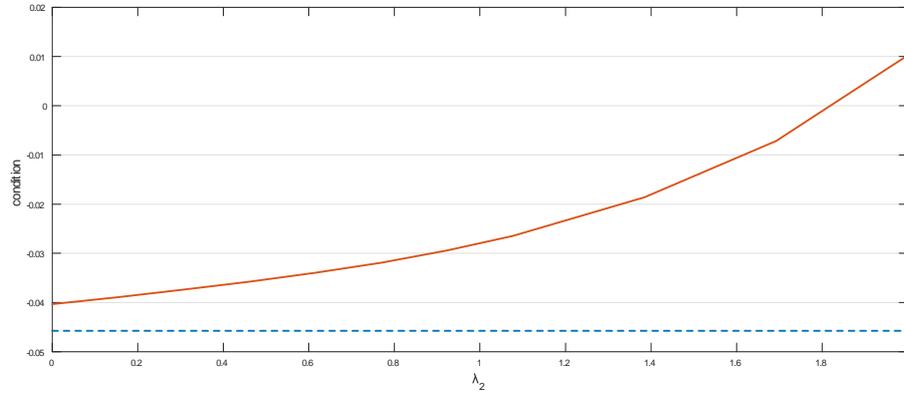
In the sequel we can use the triangle inequality

$$\left\| \widehat{L}_T''(\widehat{\theta}_T) - F(\theta_0) \right\| \leq \left\| \widehat{L}_T''(\widehat{\theta}_T) - \overline{L}_T''(\theta_0) \right\| + \left\| \overline{L}_T''(\theta_0) - F(\theta_0) \right\|,$$

and arguing that for any random sequence $\widehat{\theta}_T \rightarrow \theta_0$ a.s., the first term on the r.h.s. vanishes a.s. because $\left\| \widehat{L}_T''(\widehat{\theta}_T) - \widehat{L}_T''(\theta_0) \right\| \xrightarrow{a.s.} 0$ as $T \rightarrow \infty$ and $\left\| \widehat{\theta}_T - \theta_0 \right\| \rightarrow 0$, by an application of Theorem 1 and applying the continuous mapping theorem for the second derivative of the likelihood function. The second term on the r.h.s. converges under a locally uniform law of large numbers, by applying Lemmas 1, 2, 3, 4 and 5, the Hölder inequality in (24) and the ergodic theorem to the process $(H_t(\theta_0))$, that is integrable under Assumption 6. Now \widehat{L}_T'' is proved to be a stationary and ergodic sequence, as a measurable function of $\partial \log h_t / \partial \theta$ and $\partial^2 \log h_t / \partial \theta \partial \theta^\top$ (see Krengel, 1985 Proposition 4.3). Thus, we obtain that $\left\| \widehat{L}_T''(\widehat{\theta}_T) - F(\theta_0) \right\| \xrightarrow{a.s.} 0$.

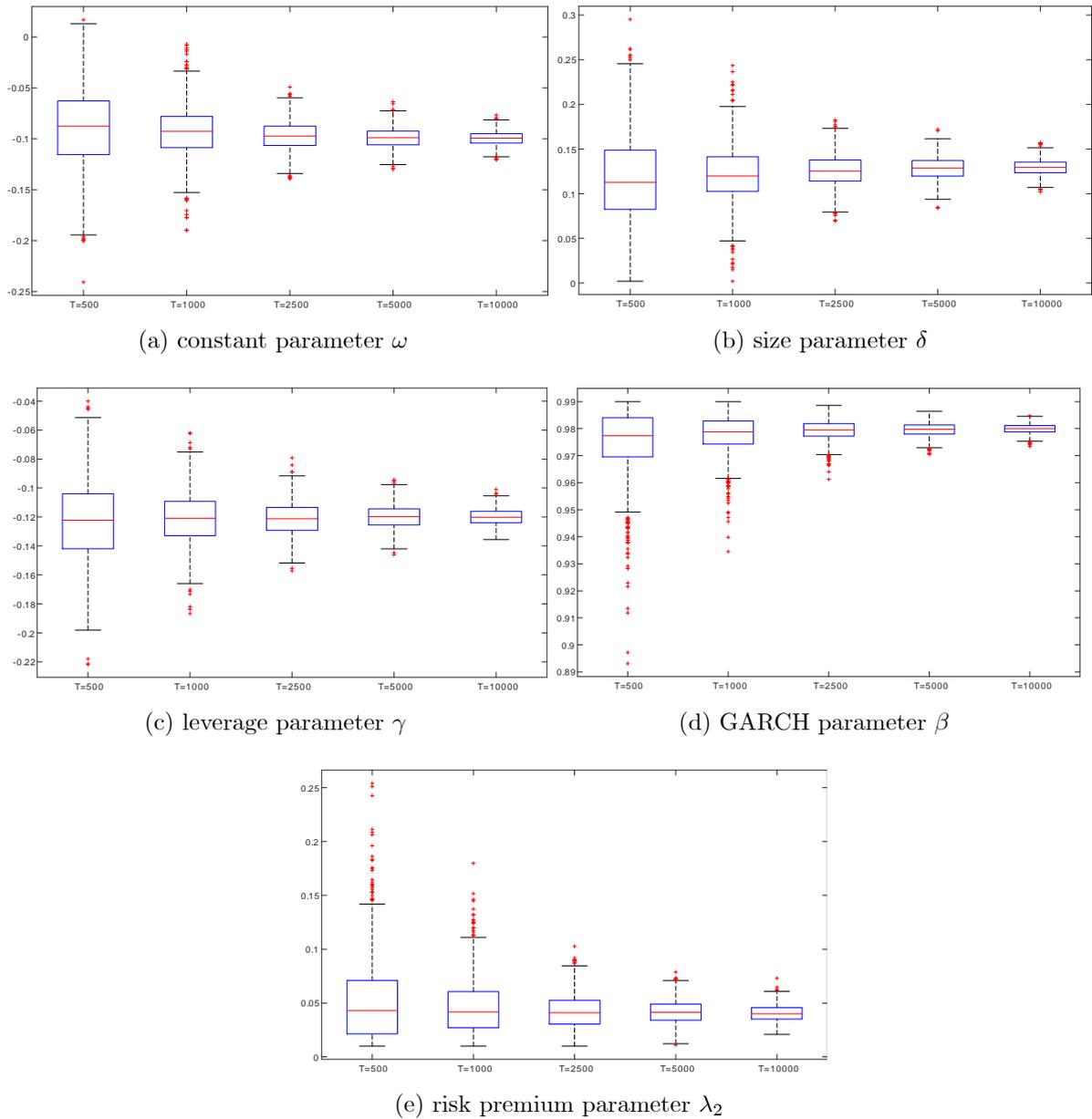
It remains to show that the covariance matrix V exists and is invertible. This requires that the matrix $F(\theta_0)$ is non-singular. By standard arguments, it is sufficient to prove that the elements of the score vector are linearly independent. Because the density of Z_t is not concentrated on two points, the proof is analogous to that for the EGARCH(1, 1) model (see Wintenberger, 2013 and Lemma 8.2 of Straumann and Mikosch, 2006). Finally, it suffices to apply Slutsky's lemma to obtain the statement of Theorem 2. \square

Figure 1: Verification of the invertibility condition in Assumption 3.



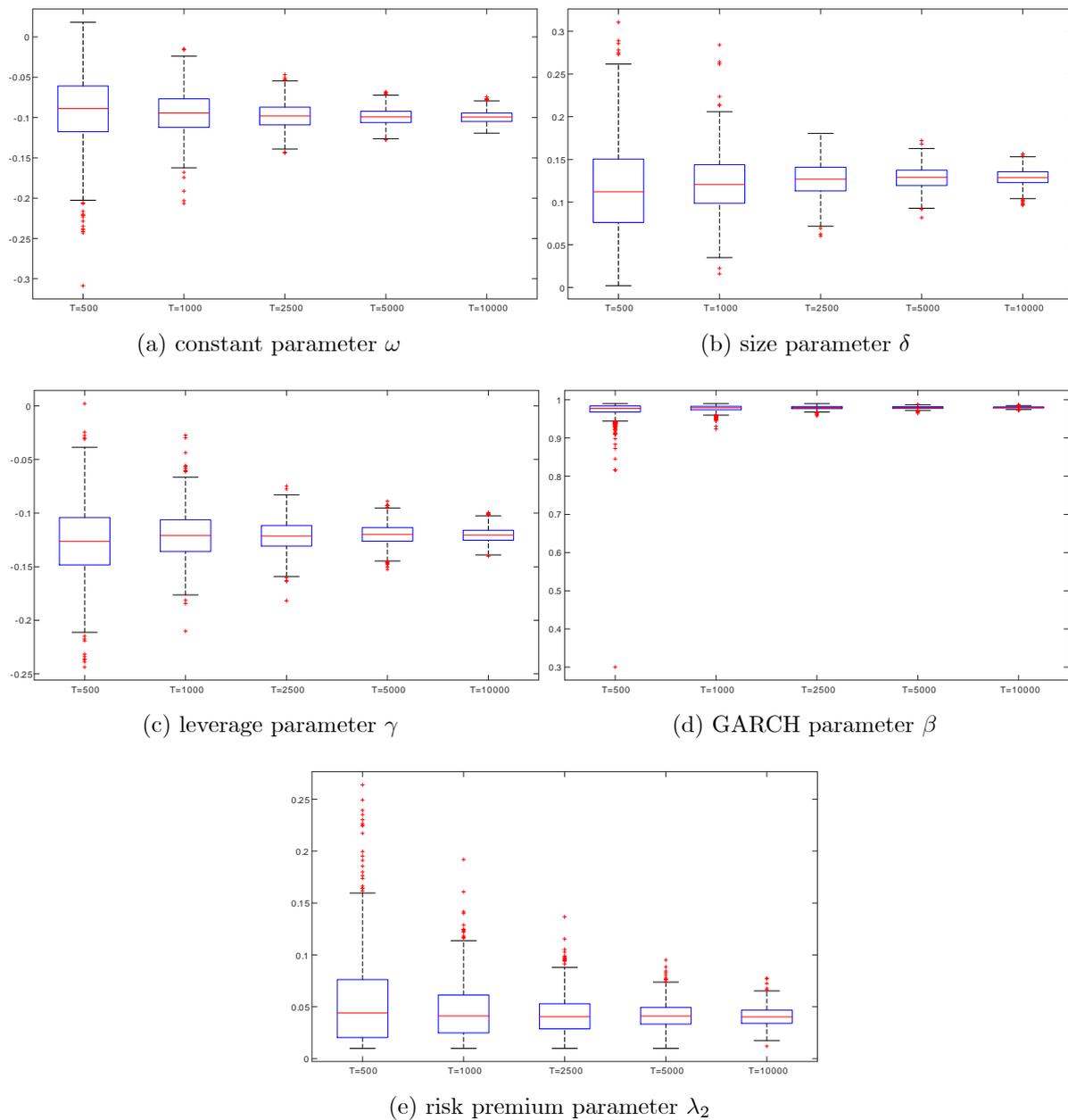
Notes: The verification of the condition is done by numerical approximation. The solid red line corresponds to the case of $\delta_0 = 0.15$ and the dashed blue line to the case of $\delta_0 = 0.9$.

Figure 2: Boxplots of independent realizations of the QML estimators of the EGARCH(1,1)-M parameters with GED innovations.



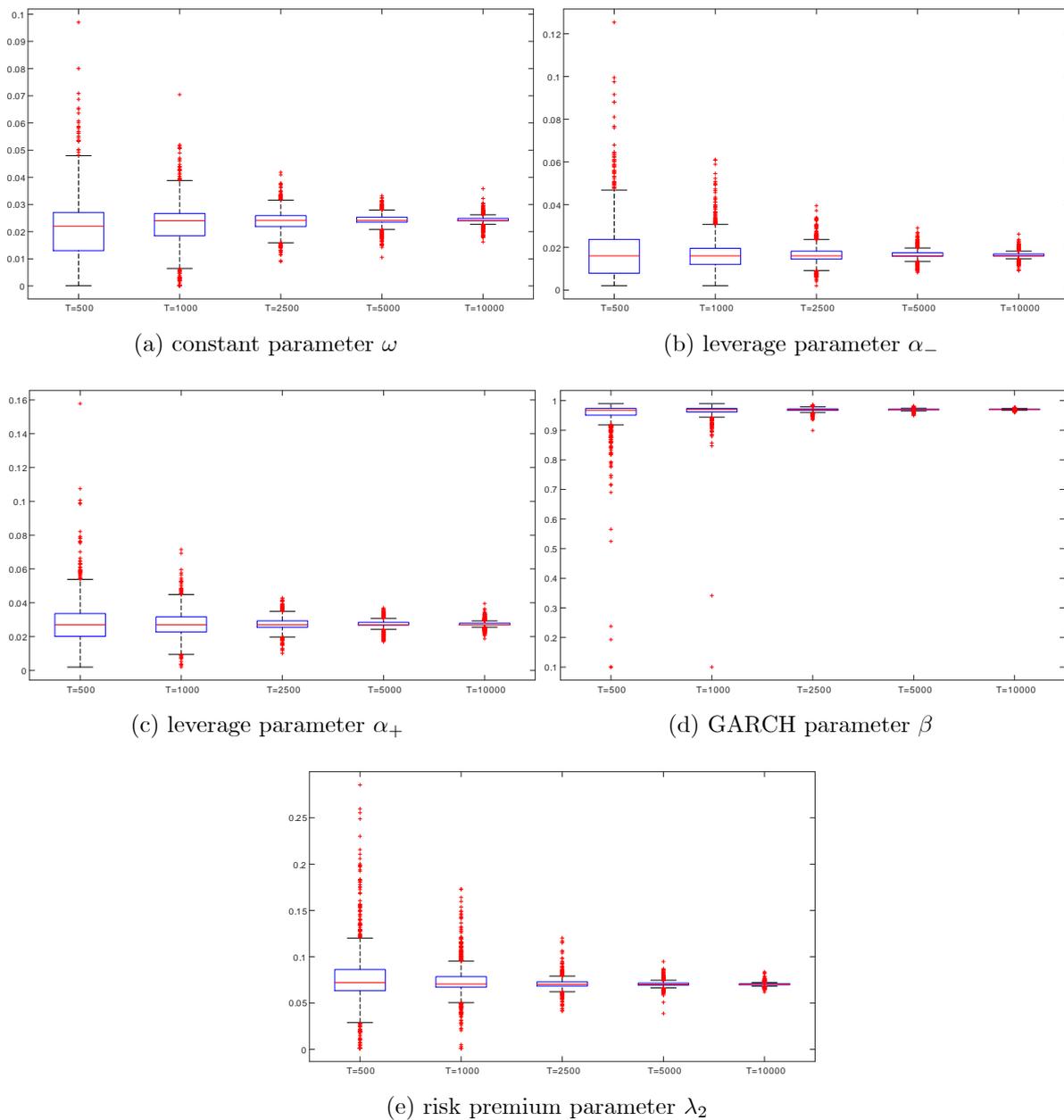
Notes: GED errors with $\nu = 2$. Various sample sizes are compared. The red line represents the median.

Figure 3: Boxplots of independent realizations of the QML estimators of the EGARCH(1,1)-M parameters with GED innovations.



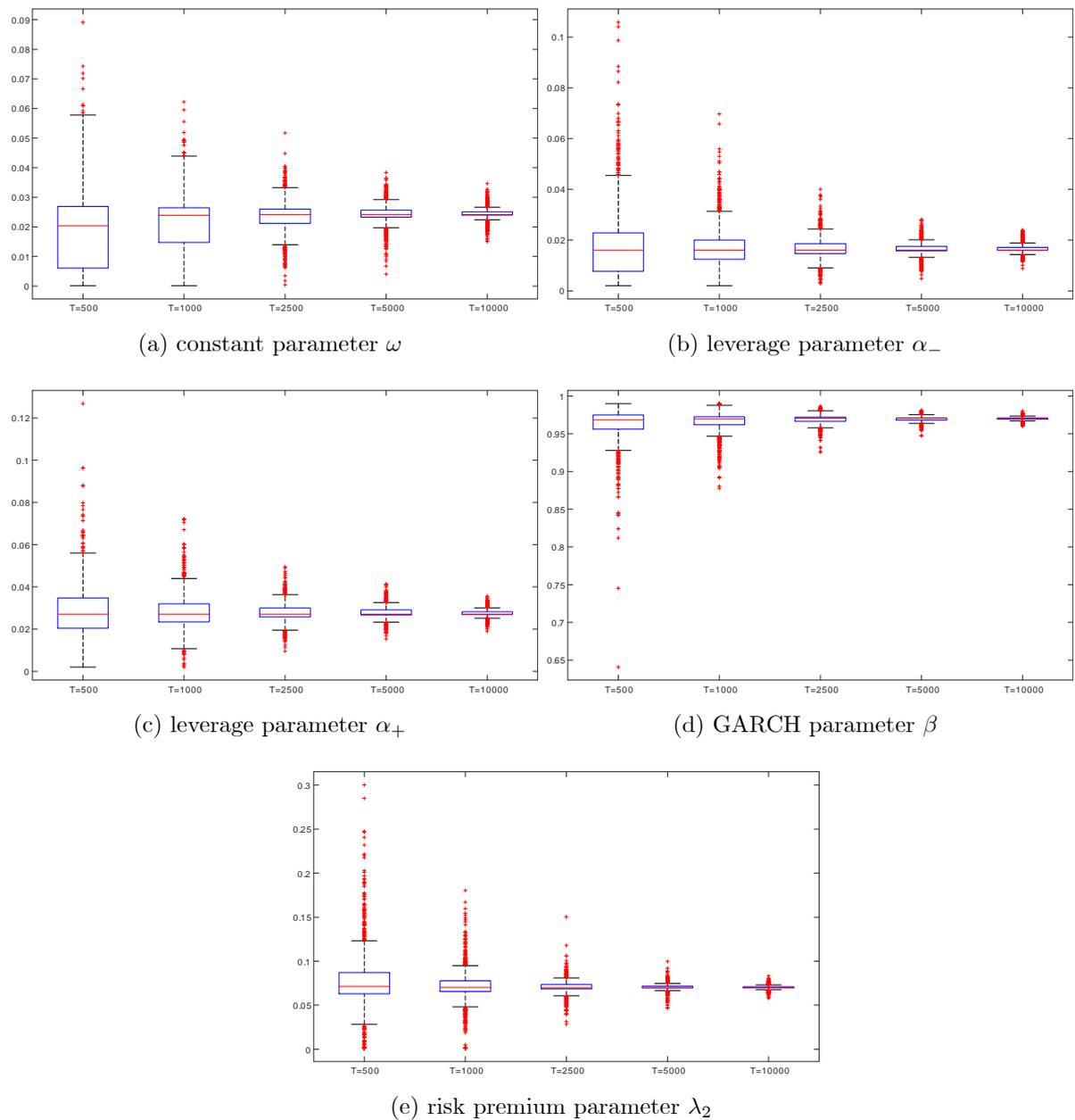
Notes: GED errors with $\nu = 1.4$. Various sample sizes are compared. The red line represents the median.

Figure 4: Boxplots of independent realizations of the QML estimators of the Log-GARCH(1,1)-M parameters with GED innovations.



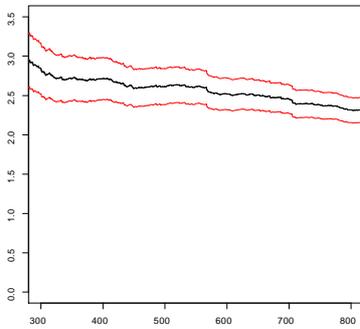
Notes: GED errors with $\nu = 2$. Various sample sizes are compared. The red line represents the median.

Figure 5: Boxplots of independent realizations of the QML estimators of the Log-GARCH(1,1)-M parameters with GED innovations.

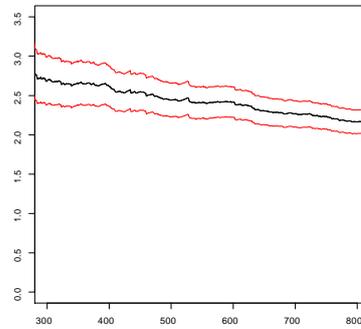


Notes: GED errors with $\nu = 1.4$. Various sample sizes are compared. The red line represents the median.

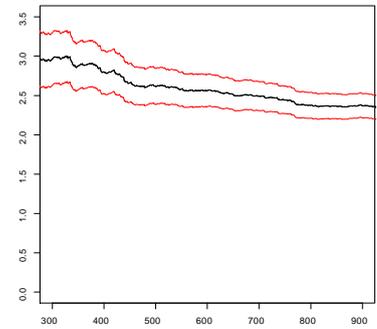
Figure 6: Hill plots of the returns of the stock market indices.



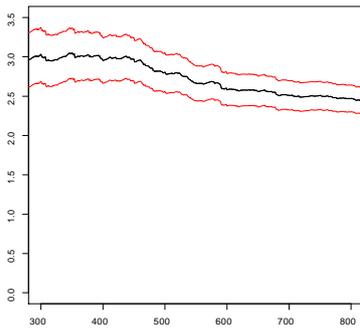
(a) SP500



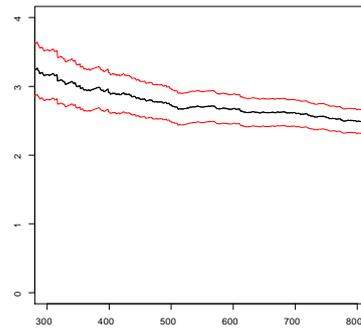
(b) Nasdaq100



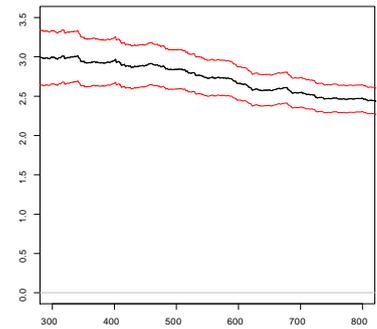
(c) FTSE100



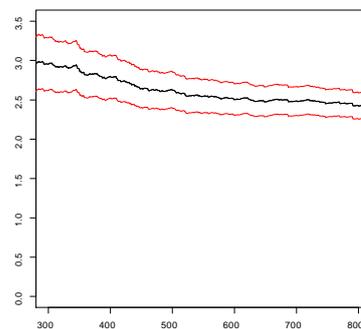
(d) DAX



(e) Nikkei225



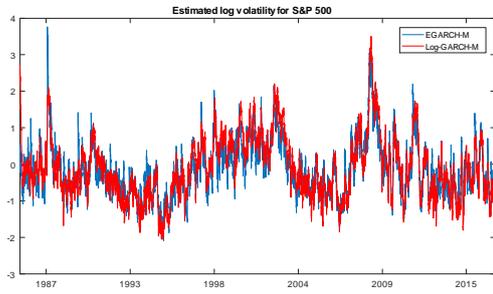
(f) CAC40



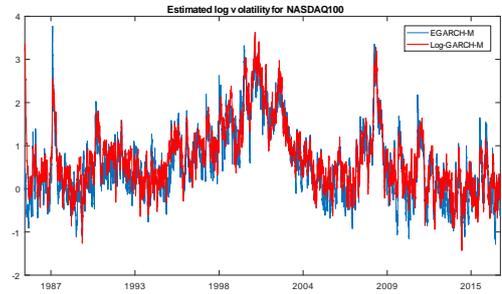
(g) Dow Jones

Notes: The horizontal axis shows the number of tail observations and the vertical one shows the Hill estimator of the tail index. The red lines correspond to the 95% confidence bands.

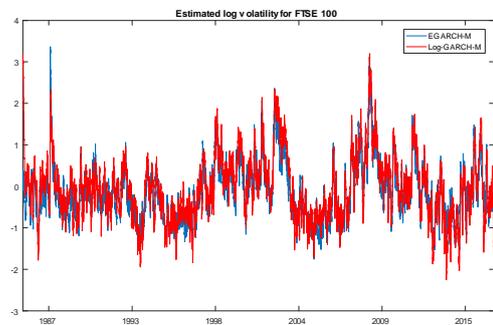
Figure 7: Estimated log volatilities for the stock market indices.



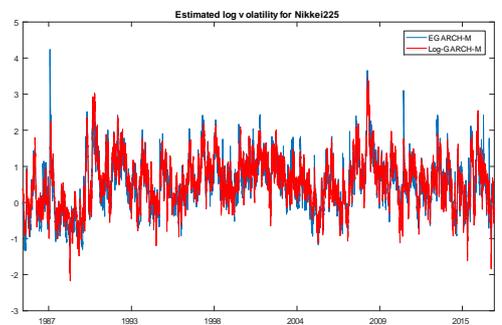
(a) SP500



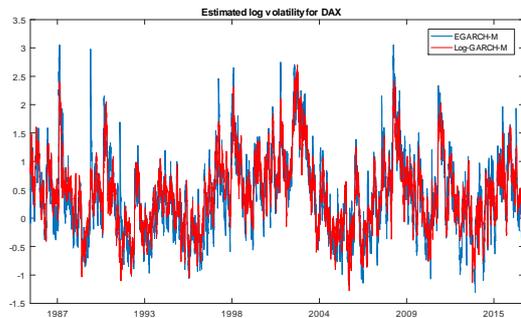
(b) Nasdaq100



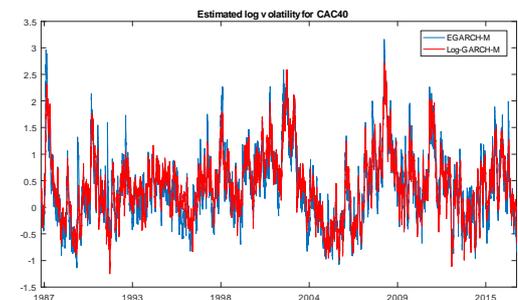
(c) FTSE100



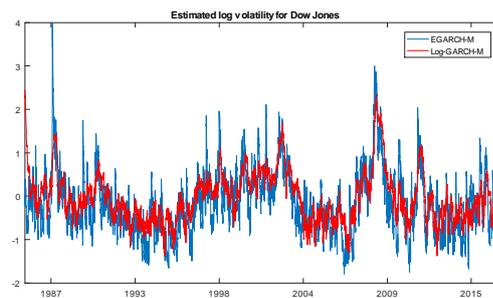
(d) Nikkei225



(e) DAX

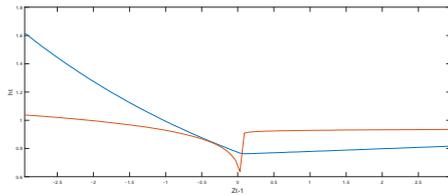


(f) CAC40

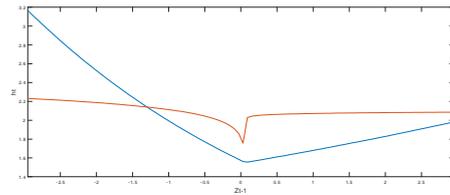


(g) Dow Jones

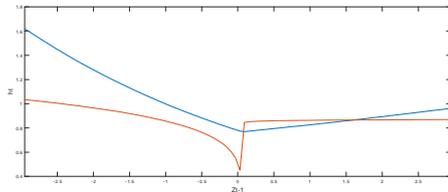
Figure 8: Estimated News Impact Curves (NIC) for the stock market indices.



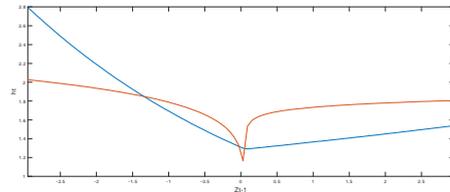
(a) SP500



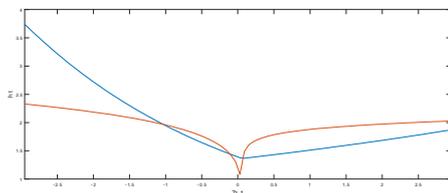
(b) Nasdaq100



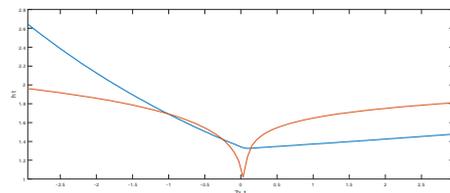
(c) FTSE100



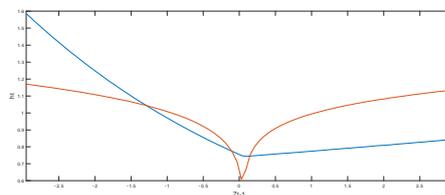
(d) DAX



(e) Nikkei225



(f) CAC40



(g) Dow Jones

Notes: The bold blue line corresponds to the EGARCH-M case and the red line to the Log-GARCH-M case.

The graphs show the estimated news impact functions g in eq. (2).

Table 1: Bias and RMSE in % of the true parameter values: key characteristics of the finite sample distribution of the QMLE for the EGARCH(1,1)-M model, for various sample sizes and two types of GED errors.

EGARCH(1,1)-M						
$\nu = 2$ (Gaussian)						
	T	$\hat{\omega}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\beta}$	$\hat{\lambda}_2$
Bias	500	-0.129	0.0267	-0.1308	-0.0050	0.2875
	1000	-0.062	0.0100	-0.0646	-0.0021	0.1375
	2500	-0.028	0.0108	-0.0292	-0.0007	0.0550
	5000	-0.011	-0.0025	-0.0123	-0.0004	0.0425
	10000	-0.005	0.0008	-0.0054	-0.0001	0.0125
RMSE	500	0.437	0.2317	0.4254	0.0139	1.0025
	1000	0.253	0.1542	0.2485	0.0076	0.6550
	2500	0.147	0.0983	0.1423	0.0039	0.4100
	5000	0.01	0.0675	0.0977	0.0027	0.2825
	10000	0.071	0.0483	0.0692	0.0018	0.1900
$\nu = 1.4$						
	T	$\hat{\omega}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\beta}$	$\hat{\lambda}_2$
Bias	500	-0.099	0.0542	-0.1338	-0.0074	0.3450
	1000	-0.053	0.0083	-0.0677	-0.0024	0.1350
	2500	-0.02	0.0117	-0.0238	-0.0008	0.0450
	5000	-0.008	-0.0008	-0.0115	-0.0005	0.0450
	10000	-0.006	0.0050	-0.0069	-0.0002	0.0175
RMSE	500	0.473	0.2875	0.4700	0.0287	1.1050
	1000	0.273	0.1867	0.2738	0.0083	0.6900
	2500	0.159	0.1175	0.1600	0.0046	0.4625
	5000	0.102	0.0800	0.1023	0.0031	0.3150
	10000	0.076	0.0583	0.0762	0.0021	0.2325

Notes: The number of simulations is 1000. The true values of the parameters are: $\omega = -0.10, \gamma = -0.12, \delta = 0.13, \beta = 0.98, \lambda_2 = 0.04$. We also assume that $\lambda_1 = 0$.

Table 2: Bias and RMSE in % of the true parameter values: key characteristics of the finite sample distribution of the QMLE for the Log-GARCH-M model, for various sample sizes and two types of GED errors.

Log-GARCH(1, 1)-M						
$\nu = 2$ (Gaussian)						
	T	$\hat{\omega}$	$\hat{\alpha}_-$	$\hat{\alpha}_+$	$\hat{\beta}$	$\hat{\lambda}_2$
Bias	500	-0.1250	0.1438	0.0333	-0.0194	0.0986
	1000	-0.0542	0.0250	0.0148	-0.0075	0.0500
	2500	-0.0042	0.0313	0.0074	-0.0020	0.0114
	5000	0.0083	0.0313	0.0111	-0.0012	0.0071
	10000	0.0125	0.0250	0.0111	-0.0008	0.0029
RMSE	500	0.5333	0.9313	0.5296	0.0683	0.4843
	1000	0.3417	0.5250	0.3037	0.0388	0.2614
	2500	0.1708	0.2813	0.1556	0.0067	0.0986
	5000	0.1042	0.1563	0.1037	0.0038	0.0500
	10000	0.0667	0.1000	0.0667	0.0023	0.0243
$\nu = 1.4$						
	T	$\hat{\omega}$	$\hat{\alpha}_-$	$\hat{\alpha}_+$	$\hat{\beta}$	$\hat{\lambda}_2$
Bias	500	-0.2167	0.1438	0.0519	-0.0101	0.0986
	1000	-0.1125	0.0625	0.0407	-0.0053	0.0229
	2500	-0.0125	0.0438	0.0259	-0.0021	0.0114
	5000	0.0083	0.0313	0.0259	-0.0013	0.0057
	10000	0.0125	0.0313	0.0148	-0.0008	0.0029
RMSE	500	0.6167	0.9375	0.5074	0.0234	0.5200
	1000	0.4333	0.5500	0.3296	0.0153	0.2729
	2500	0.2333	0.2938	0.1815	0.0072	0.1157
	5000	0.1500	0.1813	0.1222	0.0041	0.0586
	10000	0.0875	0.1188	0.0741	0.0026	0.0329

Notes: The number of simulations is 1000. The true values of the parameters are: $\omega = 0.024, \alpha_- = 0.016, \alpha_+ = 0.027, \beta = 0.971, \lambda_2 = 0.07$. We also assume that $\lambda_1 = 0$.

Table 3: Summary statistics of the estimated conditional standard deviation for simulated EGARCH-M and Log-GARCH-M models.

		Mean	Min.	Max.	Median	St.Dev.
EGARCH-M	$\nu = 2$	0.148	0.122	1.007	0.126	0.091
	$\nu = 1.4$	0.152	0.109	0.841	0.144	0.053
Log-GARCH-M	$\nu = 2$	0.149	0.124	1.007	0.128	0.090
	$\nu = 1.4$	0.069	0.048	0.715	0.062	0.052

Table 4: Descriptive statistics for the daily returns of the stock market indices, for the full sample and three subsamples as follows: I: 3/1/1986-29/12/1995, II: 2/1/1996-30/12/2005, III: 3/1/2006-31/12/2016, while for the CAC40 Index it is I: 10/7/1987-30/6/1997, II: 1/7/1997-29/6/2007, III: 2/7/2007-31/12/2016.

Statistics		Observations	Mean	Variance	Skewness	Kurtosis	Autocorr
S&P 500	Full sample	7815	0.037	1.321	-0.826	23.981	-0.042***
	I	2527	0.048	0.990	-3.700	81.416	0.032
	II	2519	0.035	1.332	-0.006	5.876	-0.016
	III	2769	0.029	1.615	-0.093	13.849	-0.103***
Nasdaq 100	Full sample	7815	0.060	2.838	0.142	10.907	-0.019*
	I	2527	0.067	1.643	-0.850	17.484	0.143***
	II	2520	0.067	5.068	0.334	7.0290	-0.052***
	III	2769	0.048	1.881	0.031	11.266	-0.074***
FTSE 100	Full sample	7847	0.027	1.246	-0.297	11.615	-0.001
	I	2540	0.042	0.904	-1.386	24.331	0.062***
	II	2527	0.023	1.284	-0.096	5.666	0.002
	III	2780	0.016	1.524	0.039	10.467	-0.039**
DAX	Full sample	7830	0.037	2.073	-0.130	8.487	-0.003
	I	2505	0.028	1.581	-0.596	12.942	0.006
	II	2530	0.047	2.573	-0.124	5.725	-0.017
	III	2795	0.037	2.062	0.139	9.144	0.004
Nikkei 225	Full sample	7666	0.016	2.213	-0.087	10.031	-0.025**
	I	2462	0.027	1.931	0.047	14.762	0.036*
	II	2459	0.002	2.090	0.057	4.973	-0.039*
	III	2745	0.020	2.580	-0.258	10.300	-0.053***
CAC 40	Full sample	7464	0.026	1.958	-0.018	8.248	-0.009
	I	2487	0.034	1.444	-0.433	10.138	0.013
	II	2542	0.040	2.020	0.001	5.703	0.011
	III	2435	0.003	2.419	0.173	8.745	-0.039*
Dow Jones	Full sample	7815	0.039	1.252	-1.076	32.158	-0.037***
	I	2527	0.053	1.114	-3.986	93.890	0.019
	II	2519	0.036	1.262	-0.125	6.719	-0.010
	III	2769	0.029	1.369	0.121	14.094	-0.102***

Notes: Returns have been multiplied by 100. Autocorr is the first order autocorrelation. Asterisks denote statistical significance; three asterisks at 1 %, two asterisks at 5 % and one asterisk only at 10 %.

Table 5: EGARCH(1,1)-in-Mean model fitted by QMLE on daily returns of stock market indices.

EGARCH(1,1)-M		$\hat{\omega}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\beta}$	$\hat{\lambda}_2$	Log-lik.
S&P 500	ALL	-0.108 (0.005)	-0.113 (0.004)	0.137 (0.006)	0.973 (0.001)	0.037*** (0.013)	-10459
	I	-0.116 (0.009)	-0.067 (0.008)	0.150 (0.011)	0.979 (0.002)	0.007 (0.035)	-3065
	II	-0.091 (0.011)	-0.113 (0.008)	0.116 (0.014)	0.972 (0.004)	0.058** (0.027)	-3618
	III	-0.102 (0.010)	-0.178 (0.010)	0.125 (0.012)	0.964 (0.003)	0.054*** (0.016)	-3748
Nasdaq 100	ALL	-0.113 (0.005)	-0.076 (0.004)	0.161 (0.007)	0.981 (0.001)	0.024*** (0.007)	-13328
	I	-0.122 (0.010)	-0.074 (0.008)	0.181 (0.010)	0.944 (0.007)	0.019 (0.030)	-3921
	II	-0.093 (0.011)	-0.062 (0.006)	0.130 (0.014)	0.991 (0.002)	0.014 (0.011)	-5131
	III	-0.091 (0.011)	-0.158 (0.011)	0.128 (0.014)	0.959 (0.004)	0.050*** (0.016)	-4213
FTSE 100	ALL	-0.131 (0.006)	-0.085 (0.003)	0.163 (0.008)	0.975 (0.002)	0.035** (0.014)	-10606
	I	-0.166 (0.014)	-0.060 (0.006)	0.193 (0.016)	0.943 (0.009)	-0.028 (0.045)	-3212
	II	-0.089 (0.013)	-0.083 (0.008)	0.111 (0.016)	0.987 (0.002)	0.026 (0.024)	-3446
	III	-0.115 (0.012)	-0.146 (0.008)	0.144 (0.015)	0.966 (0.004)	0.069*** (0.019)	-3901
DAX	ALL	-0.111 (0.004)	-0.098 (0.003)	0.159 (0.005)	0.971 (0.002)	0.017 (0.010)	-12644
	I	-0.107 (0.007)	-0.106 (0.007)	0.155 (0.009)	0.955 (0.005)	-0.000 (0.027)	-3817
	II	-0.132 (0.013)	-0.069 (0.008)	0.183 (0.017)	0.979 (0.003)	0.014 (0.016)	-4313
	III	-0.089 (0.011)	-0.131 (0.009)	0.132 (0.015)	0.967 (0.003)	0.032* (0.017)	-4493
Nikkei 225	ALL	-0.151 (0.006)	-0.113 (0.003)	0.222 (0.006)	0.962 (0.002)	-0.000 (0.010)	-12798
	I	-0.213 (0.009)	-0.152 (0.007)	0.300 (0.010)	0.955 (0.005)	-0.014 (0.017)	-3821
	II	-0.120 (0.015)	-0.073 (0.010)	0.175 (0.020)	0.973 (0.005)	-0.002 (0.025)	-4216
	III	-0.132 (0.012)	-0.119 (0.007)	0.210 (0.015)	0.951 (0.005)	0.033* (0.017)	-4738
CAC 40	ALL	-0.095 (0.004)	-0.096 (0.004)	0.134 (0.006)	0.974 (0.001)	0.033*** (0.012)	-11969
	I	-0.105 (0.014)	-0.075 (0.010)	0.148 (0.018)	0.952 (0.007)	0.030 (0.032)	-3744
	II	-0.098 (0.010)	-0.073 (0.007)	0.131 (0.013)	0.985 (0.003)	0.020 (0.019)	-4083
	III	-0.067 (0.011)	-0.177 (0.012)	0.114 (0.014)	0.959 (0.005)	0.046** (0.018)	-4104
Dow Jones	ALL	-0.116 (0.005)	-0.104 (0.003)	0.148 (0.006)	0.972 (0.001)	0.036*** (0.014)	-10376
	I	-0.116 (0.007)	-0.065 (0.006)	0.153 (0.010)	0.976 (0.002)	0.015 (0.035)	-3226
	II	-0.093 (0.011)	-0.097 (0.007)	0.118 (0.014)	0.978 (0.004)	0.065** (0.029)	-3558
	III	-0.115 (0.012)	-0.171 (0.010)	0.136 (0.015)	0.964 (0.003)	0.048*** (0.018)	-3558

Notes: The estimated standard errors are reported in brackets. Three asterisks indicate statistical significance at 1 %, two asterisks at 5 % and one asterisk at 10 %. For λ_2 we indicate statistical significance using bold face and asterisks.

Table 6: Log-GARCH(1,1)-in-Mean model fitted by QMLE on daily returns of stock market indices.

		Log-GARCH(1,1)-M with $\hat{\alpha}_+$ and $\hat{\alpha}_-$					
		$\hat{\omega}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\beta}$	$\hat{\lambda}_2$	Log-lik.
S&P 500	ALL	0.042 (0.001)	0.003 (0.002)	0.053 (0.001)	0.954 (0.000)	0.087*** (0.001)	-10966
	I	0.091 (0.000)	0.058 (0.000)	0.057 (0.000)	0.929 (0.000)	0.110*** (0.000)	-3150
	II	0.042 (0.000)	0.025 (0.000)	0.035 (0.000)	0.962 (0.000)	0.054*** (0.000)	-3714
	III	0.101 (0.012)	0.066 (0.010)	0.057 (0.008)	0.919 (0.009)	0.001 (0.004)	-3893
Nasdaq 100	ALL	0.026 (0.000)	0.003 (0.000)	0.026 (0.000)	0.979 (0.000)	0.062*** (0.000)	-13687
	I	0.070 (0.000)	0.022 (0.000)	0.063 (0.000)	0.930 (0.000)	0.117*** (0.000)	-3983
	II	0.030 (0.000)	0.007 (0.000)	0.037 (0.000)	0.976 (0.000)	0.008*** (0.000)	-5186
	III	0.103 (0.000)	0.052 (0.000)	0.073 (0.000)	0.908 (0.000)	0.148*** (0.000)	-4332
FTSE 100	ALL	0.052 (0.003)	0.003 (0.002)	0.090 (0.005)	0.928 (0.003)	0.126*** (0.008)	-11091
	I	0.035 (0.000)	0.003 (0.000)	0.055 (0.000)	0.952 (0.000)	0.106*** (0.000)	-3333
	II	0.029 (0.000)	0.005 (0.000)	0.044 (0.000)	0.973 (0.000)	0.076*** (0.000)	-3537
	III	0.096 (0.006)	0.064 (0.005)	0.066 (0.004)	0.915 (0.004)	0.052*** (0.002)	-4029
DAX	ALL	0.072 (0.000)	0.018 (0.000)	0.060 (0.000)	0.934 (0.000)	0.093*** (0.000)	-12971
	I	0.047 (0.000)	0.006 (0.001)	0.035 (0.001)	0.948 (0.000)	0.125*** (0.000)	-3994
	II	0.053 (0.000)	0.025 (0.000)	0.045 (0.000)	0.953 (0.000)	0.076*** (0.000)	-4387
	III	0.044 (0.000)	0.014 (0.000)	0.036 (0.000)	0.967 (0.000)	0.089*** (0.000)	-4647
Nikkei 225	ALL	0.106 (0.015)	0.034 (0.009)	0.083 (0.007)	0.908 (0.012)	0.011*** (0.001)	-13009
	I	0.138 (0.000)	0.073 (0.000)	0.086 (0.000)	0.883 (0.000)	0.001*** (0.000)	-3969
	II	0.034 (0.000)	0.014 (0.000)	0.021 (0.000)	0.969 (0.000)	0.074*** (0.000)	-4294
	III	0.054 (0.000)	0.003 (0.001)	0.045 (0.002)	0.950 (0.000)	0.093*** (0.000)	-4905
CAC 40	ALL	0.093 (0.001)	0.042 (0.002)	0.071 (0.002)	0.918 (0.001)	0.020*** (0.000)	-12197
	I	0.166 (0.000)	0.075 (0.000)	0.100 (0.000)	0.807 (0.000)	0.074*** (0.000)	-3822
	II	0.033 (0.000)	0.003 (0.000)	0.036 (0.000)	0.971 (0.000)	0.083*** (0.000)	-4166
	III	0.064 (0.000)	0.019 (0.000)	0.058 (0.000)	0.939 (0.000)	0.025*** (0.000)	-4227
Dow Jones	ALL	0.101 (0.000)	0.059 (0.001)	0.071 (0.000)	0.910 (0.001)	0.034*** (0.000)	-10692
	I	0.048 (0.005)	0.003 (0.008)	0.060 (0.010)	0.962 (0.003)	0.066*** (0.010)	-3307
	II	0.064 (0.000)	0.039 (0.002)	0.048 (0.003)	0.939 (0.002)	0.015*** (0.000)	-3639
	III	0.161 (0.000)	0.124 (0.002)	0.087 (0.001)	0.857 (0.001)	0.100*** (0.001)	-3728

Notes: The estimated standard errors are reported in brackets. Three asterisks indicate statistical significance at 1 %, two asterisks at 5 % and one asterisk at 10 %. For λ_2 we indicate statistical significance using bold face and asterisks.

Table 7: Summary statistics of the estimated conditional standard deviation for daily stock market returns.

		Nasdaq 100	S&P 500	Dow Jones	Nikkei 225	FTSE 100	DAX	CAC 40
EGARCH-M	Mean	1.470	1.000	0.986	1.375	0.998	1.303	1.258
Log-GARCH-M		1.558	1.044	1.029	1.398	1.064	1.305	1.268
EGARCH-M	Min.	0.522	0.354	0.414	0.512	0.415	0.519	0.441
Log-GARCH-M		0.490	0.359	0.500	0.338	0.324	0.528	0.499
EGARCH-M	Max.	6.554	6.512	7.258	8.329	5.376	4.607	4.854
Log-GARCH-M		6.126	5.758	3.378	5.558	4.953	3.852	3.897
EGARCH-M	Median	1.240	0.874	0.875	1.277	0.892	1.169	1.153
Log-GARCH-M		1.345	0.901	0.947	1.306	0.953	1.218	1.185
EGARCH-M	St.Dev.	0.739	0.477	0.454	0.561	0.432	0.542	0.489
Log-GARCH-M		0.739	0.535	0.356	0.522	0.479	0.447	0.436
	Corr.	0.910	0.880	0.727	0.885	0.886	0.915	0.929

Notes: Corr is the sample correlation of the estimated conditional standard deviations.

References

- [1] Bera, A. K., Higgins, M. L. (1993). ARCH models: properties, estimation and testing. *Journal of Economic Surveys* 7: 305-366.
- [2] Berkes, I., Horváth, L., Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* 9: 201–227.
- [3] Blasques, F., Gorgi, P., Koopman, S. J., Wintenberger, O. (2016). Feasible invertibility conditions and maximum likelihood estimation for observation-driven models. *Tinbergen Institute Discussion Paper 16-082/III*.
- [4] Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31: 942-959.
- [5] Chou, R. Y. (1988). Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics* 3: 279-294.
- [6] Christensen, B. J., Dahl, C. M., Iglesias E. (2012). Semiparametric inference in a GARCH-in-mean model. *Journal of Econometrics* 167: 458-472.
- [7] Conrad, C., Mammen, E. (2016). Asymptotics for parametric GARCH-in-mean models. *Journal of Econometrics* 194: 319-329.
- [8] Diaconis, P., Freedman, D. (1999). Iterated random functions. *SIAM review* 41: 45-76.
- [9] Engle, R. F., Lilien, D. M., Robins, R. P. (1987). Estimating time-varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55: 391-407.
- [10] Evans, L. C., Gariépy, R.F. (2015). *Measure theory and fine properties of functions*. Studies in Advanced Mathematics, CRC Press, Revised Edition.
- [11] Francq, C, Wintenberger, O., Zakoian, J.-M. (2013). GARCH models without positivity constraints: exponential or log GARCH?. *Journal of Econometrics* 177: 34-46.
- [12] _____ (2018). Goodness-of-fit tests for log and exponential GARCH models. *TEST* 27: 27-51.

- [13] Francq, C, Zakoian, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.
- [14] French, K. R., Schwert, G. W., Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics* 19: 3-29.
- [15] Geweke, J. (1986). Modeling the persistence of conditional variances: a comment. *Econometric Reviews* 5: 57-61.
- [16] Granger, C. W. J., Andersen, A. (1978). On the invertibility of time series models. *Stochastic Processes and their Applications* 8: 87-92.
- [17] Hafner, C. M., Linton, O. (2017). An almost closed form estimator for the EGARCH model. *Econometric Theory* 33: 1013-1038.
- [18] Hall, P., Heyde, C.C. (1980). *Martingale Limit Theory and its Application*. Academic press, New York.
- [19] He, C., Teräsvirta, T., Malmsten, H. (2002). Moment structure of a family of first-order exponential GARCH models. *Econometric Theory* 18: 868-885.
- [20] Hentschel, L. (1995). All in the family nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39: 71-104.
- [21] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 3: 1163-1174.
- [22] Hill, J.B. (2010). On tail index estimation for dependent, heterogenous data. *Econometric Theory*, 26: 1398-1436.
- [23] Hsing, T. (1991). On tail index estimation using dependent data. *The Annals of Statistics*, 19: 1547-1569.
- [24] Iglesias, E. M., Phillips, G. D. A. (2012). Estimation, testing, and finite sample properties of quasi-maximum likelihood estimators in GARCH-M models. *Econometric Reviews* 31: 532-557.

- [25] Karanasos, M., Kim, J. (2003). Moments of the ARMA-EGARCH model. *Econometrics Journal* 6: 146-166.
- [26] Krengel, U. (1985). *Ergodic Theorems*. de Gruyter, Berlin.
- [27] Kyriakopoulou, D. (2015). Asymptotic normality of the QML estimator of the EGARCH(1,1) model. *Discussion Paper*.
- [28] Linton, O., Perron, B. (2003). The shape of the risk premium: evidence from a semiparametric generalized autoregressive conditional heteroskedasticity model. *Journal of Business & Economic Statistics* 21: 354-367.
- [29] Lundblad, C. (2007). The risk return tradeoff in the long run: 1836-2003. *Journal of Financial Economics* 85: 123-150.
- [30] Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica* 41: 867-886.
- [31] Milhøj, A. (1987). A multiplicative parameterization of ARCH models. *Working Paper*, Department of Statistics, University of Copenhagen.
- [32] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59: 347-370.
- [33] Nyberg, H. (2012). Risk-return tradeoff in U.S. stock returns over the business cycle. *Journal of Financial and Quantitative Analysis*, 47: 137-158.
- [34] Pagan, A. R., Hong, Y. S. (1991). Nonparametric estimation and the risk premium. In Barnett, W. A., Powell, J., and Tauchen, G. E. (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, 51-75.
- [35] Pantula, S. G. (1986). Modeling the persistence of conditional variances: a comment. *Econometric Reviews* 5: 71-74.
- [36] Pierre, E. F. St. (1998). Estimating EGARCH-M models: science or art?. *The Quarterly Review of Economics and Finance* 38: 167-180.

- [37] Rodríguez, M. J., Ruiz, E. (2012). Revisiting several popular GARCH models with leverage effect: differences and similarities. *Journal of Financial Econometrics* 10: 637-668.
- [38] Straumann, D. (2005). *Estimation in Conditionally Heteroskedastic Time Series Models*. Springer.
- [39] Straumann, D., Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroskedastic time series: a stochastic recurrence equations approach. *Annals of Statistics* 34: 2449-2495.
- [40] Sucarrat, G. and Escribano, A. (2018). Estimation of log-GARCH models in the presence of zero returns. *The European Journal of Finance*, 24: 809-827.
- [41] Sucarrat, G., Grønneberg, S., Escribano, A. (2016). Estimation and inference in univariate and multivariate Log-GARCH-X models when the conditional density is unknown. *Computational Statistics and Data Analysis* 100: 582-594.
- [42] Wagner, N. and Marsh, T.A. (2004). Tail index estimation in small samples: Simulation results for independent and ARCH-type financial return models. *Statistical Papers*, 45: 545-561.
- [43] Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40: 846-867.