

# EXPONENTIAL-TYPE GARCH MODELS WITH LINEAR-IN-VARIANCE RISK PREMIUM

Hafner, C. and D. Kyriakopoulou

REPRINT | 2020 / 29

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

Christian M. Hafner\* and Dimitra Kyriakopoulou†

Université catholique de Louvain

September 16, 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 an empirical application to stock market indices. Section 5 concludes. Proofs of the main Theorems, 3.1 and 3.2, are collected in the Appendix. The remaining proofs are put in Supplementary Material (SM in what follows).

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  $\Theta$  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  $\Theta$ , with the identity  $\sup_{\Theta} |\cdot| = \vee_{k=1}^N \sup_{\mathcal{V}(\theta_{0k})} |\cdot|$ . 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

$$Y_t = \lambda_1 + \lambda_2 h_t + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} Z_t, \quad (2.1)$$

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  $\sigma$ -algebra generated by the past information up to time  $t - 1$ , and the error term  $\varepsilon_t$  is a zero mean process with conditional variance  $h_t$ . We would expect the risk premium parameter  $\lambda_2$  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

$$\log h_t(\theta) = g(Z_{t-1}) + \beta \log h_{t-1}(\theta), \quad t \in \mathbb{Z} \quad (2.2)$$

where  $g(\cdot)$  is a known parametric “news impact” function, and  $\theta$  is the parameter vector containing the parameters of  $g(\cdot)$ , the persistence parameter  $\beta$ , and the risk premium parameters  $\lambda_1$  and  $\lambda_2$ . 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 (2.2) 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

$$\log h_t(\theta) = \omega + \gamma Z_{t-1} + \delta |Z_{t-1}| + \beta \log h_{t-1}(\theta), \quad t \in \mathbb{Z} \quad (2.3)$$

where  $\omega, \gamma, \delta \in \mathbb{R}$  and  $|\beta| < 1$ . The parameter  $\gamma$  is governing the asymmetry and  $\delta$  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),

$$\log h_t(\theta) = \omega_t + (\alpha_+ \mathbf{1}_{\{\varepsilon_{t-1} > 0\}} + \alpha_- \mathbf{1}_{\{\varepsilon_{t-1} < 0\}}) \log \varepsilon_{t-1}^2 + \tilde{\beta} \log h_{t-1}(\theta), \quad t \in \mathbb{Z} \quad (2.4)$$

where the time-varying intercept  $\omega_t = \omega + \omega_- \mathbf{1}_{\{\varepsilon_{t-1} < 0\}}$  is introduced here to make the volatility stable by scaling<sup>1</sup>. The symmetric log-GARCH model corresponds to the case of  $\alpha_+ = \alpha_- = \alpha$  and  $\omega_- = 0$  and can be rewritten in the form (2.2) with  $g(z) = \omega + \alpha \log z^2$  and  $\beta = \alpha + \tilde{\beta}$ . The asymmetric log-GARCH model corresponds to (2.2) only by letting  $\beta$  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  $\varepsilon_{t-1}$ , 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 Escibano (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 (2.1) and (2.3), or (2.1) and (2.4), allows the conditional variance of each observation to affect the mean of that observation. The most popular specification for the risk-return relationship is (2.1), i.e. to

---

<sup>1</sup>As noted by Francq *et al.* (2018), stability by scaling is a desirable property of volatility models. If the error process,  $\varepsilon_t$ , 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  $\varepsilon_t$ , 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  $\theta_0$  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  $\theta$  belongs to a parameter

space  $\Theta$ .

At this point, the existence of a stationary solution of (2.2) should be considered. We will work under the following assumptions.

**Assumption 2.1** *The parameter space  $\Theta$  is a compact set.*

**Assumption 2.2** (i) *It holds that  $0 \leq \beta_0 < 1$ .*

(ii) *It holds that  $\delta_0 \geq |\gamma_0|$ .*

Assumption 2.2(i) ensures that the model (2.2) admits a unique stationary solution with finite moments. Assumption 2.2(ii) 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,  $\Theta$  is a compact subset of  $\mathbb{R}^3 \times D \times [0, 1)$ , where  $D = \{(\gamma, \delta)^\top \in \mathbb{R}^2 | \gamma \in \mathbb{R}, \delta \geq |\gamma|\}$ . In the Log-GARCH-M case,  $\Theta$  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  $\phi_t$  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  $(\phi_t)$  if it satisfies the following stochastic difference equation:*

$$\mathcal{X}_t = \phi_t(\mathcal{X}_{t-1}). \quad (2.5)$$

*The process  $(\mathcal{X}_t)$  is referred to as a solution to the SRE (2.5).*



The parametric SRE representation of (2.3) and (2.4) is given by

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

where  $\phi_t$  is a function of the innovations  $Z_t$ , with

$$\phi_t(s) = \omega + \gamma Z_t + \delta |Z_t| + \beta s \quad (2.6)$$

for the EGARCH model and

$$\phi_t(s) = \omega_t + (\alpha_+ \mathbf{1}_{\{Z_t > 0\}} + \alpha_- \mathbf{1}_{\{Z_t < 0\}}) \log Z_t^2 + \left( \alpha_+ \mathbf{1}_{\{Z_t > 0\}} + \alpha_- \mathbf{1}_{\{Z_t < 0\}} + \tilde{\beta} \right) s \quad (2.7)$$

for the Log-GARCH process, substituting  $\log \varepsilon_t^2$  by  $\log h_t + \log Z_t^2$  in (2.4).

Equation (2.6) 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 (2.3) 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  $EZ_t^2 = 1$  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 (2.4) is

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

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 (2.3) admits an MA( $\infty$ ) representation of the form

$$\log h_t(\theta) = \omega (1 - \beta)^{-1} + \sum_{k=0}^{\infty} \beta^k (\gamma Z_{t-1-k} + \delta |Z_{t-1-k}|), \quad (2.8)$$

which is an almost sure representation of  $\log h_t$ , i.e. the a.s. absolute convergence of the series. From (2.8) and Assumption 2.2(ii), 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 (2.3) is by construction latent, another solution of (2.3) 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

$$\log \hat{h}_t(\theta) = \omega + \gamma \frac{Y_{t-1} - \lambda_1 - \lambda_2 \hat{h}_{t-1}(\theta)}{\sqrt{\hat{h}_{t-1}(\theta)}} + \delta \left| \frac{Y_{t-1} - \lambda_1 - \lambda_2 \hat{h}_{t-1}(\theta)}{\sqrt{\hat{h}_{t-1}(\theta)}} \right| + \beta \log \hat{h}_{t-1}(\theta), \quad t \geq 1 \quad (2.9)$$

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  $\{\log \hat{h}_t(\theta), t \geq 0\}$ , approximates the true volatility for a given parameter  $\theta$ . It is well defined but generally nonstationary due to its dependence on initial conditions.

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

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

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

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

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 3.1 and 3.2 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  $\log \hat{h}_0$ , we can approximate the filtered log-volatility process,  $\log \hat{h}_t(\theta)$ , 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  $\theta$  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 (2.6) is a function of the innovations  $Z_t$  which are not observable. For the asymptotic normality part, we need to establish differentiability of  $\log \hat{h}_t(\theta)$ , analyzing the first and second derivative of the log-likelihood function with respect to the parameter  $\theta$ . 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*

$$\Lambda(f) = \sup_{x, y \geq 0, x \neq y} \frac{|f(x) - f(y)|}{|x - y|}, \quad (3.1)$$

*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 3.1 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 (3.1) 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 (2.5) converges, so that it admits a stationary ergodic solution  $(\mathcal{Y}_t)_{t \in \mathbb{Z}}$  and if there is any other solution to (2.5),  $(\tilde{\mathcal{Y}}_t)_{t \in \mathbb{Z}}$ , then almost surely*

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

*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 (2.3) that is driven by a function of the observations  $Y_t$

$$s_{t+1} = [\Phi_t(s_t)](\theta), \quad t \in \mathbb{Z}, \quad \forall \theta \in \Theta \quad (3.2)$$

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

$$g_\theta(y, s) = \omega + [\gamma(y - \lambda_1 - \lambda_2 \exp(s)) + \delta(|y - \lambda_1 - \lambda_2 \exp(s)|)] \exp(-s/2) + \beta s, \quad (3.3)$$

for each  $\theta \in \Theta$ . Observe that  $\log \hat{h}_t$  defined by (2.9) is a solution of the previous equation.

**Definition 4 (Invertibility)** *The volatility model (2.3) is called invertible (or its unique stationary ergodic solution is invertible) if it admits a unique stationary solution and*

$$\left| \log \hat{h}_t(\theta) - \log h_t(\theta) \right| \xrightarrow{P} 0, \quad t \rightarrow \infty \text{ and } \theta = \theta_0. \quad (3.4)$$

*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 (2.9) generated by past observed values, where  $\log h_t(\theta)_{\theta=\theta_0}$  is given by (2.6), 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.)<sup>2</sup> and also irrespective of the initial value of volatility. However, definition (3.4) is not sufficient to

---

<sup>2</sup>The 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  $v_t \xrightarrow{e.a.s.} 0$ , if there exists  $\rho > 1$  such that  $\rho^t \|v_t\| \xrightarrow{a.s.} 0$ .

ensure the consistency of the estimator, because this condition at  $\theta = \theta_0$  does not necessarily ensure the stability of the approximation  $\log \hat{h}_t(\theta)$ ,  $\forall \theta \in \Theta$ . 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 (3.3) is continuously invertible on a compact set  $\Theta$  if*

$$\sup_{\Theta} \left| \log \hat{h}_t(\theta) - \log h_t(\theta) \right| \xrightarrow{e.a.s.} 0, \quad t \rightarrow \infty. \quad (3.5)$$

Since  $\log \hat{h}_t(\theta)$  approximates a continuous function  $\log h_t(\theta)$  that does not depend on the initialization  $\log \hat{h}_0$ , uniformly over the parameter space  $\Theta$ , the previous definition ensures the stability of the SRE (3.3) defined for the filtered log-volatility process (2.9). Continuous invertibility is stronger than invertibility as (3.5) implies (3.4). Sufficient conditions for continuous invertibility of the SRE (3.3) 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 3.1** *For each  $\theta \in \Theta$  it holds that*

$$E [\log \max \{\beta, \Psi_0\}] < 0, \quad (3.6)$$

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 3.1 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 (3.6) 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  $\log \hat{h}_0$ , the exponentially fast almost sure convergence of  $\log \hat{h}_t(\theta)$ , generated by (2.9), to a unique stationary and ergodic solution  $\log h_t(\theta)$ ,  $t \in \mathbb{Z}$ , which implies their asymptotic equivalence.

**Lemma 1** *Let Assumptions 2.1, 2.2 and 3.1 be satisfied. Then the SRE (3.3) 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 (2.3) only when  $\theta = \theta_0$ .

For illustration, Figure 1 shows the expectation in (3.6), evaluated at  $\theta_0$  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  $\delta_0$ , 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  $\lambda_2$  is around 0.03 (see Lundblad, 2007). For large values of  $\lambda_2$  and  $\delta$ , condition (3.6) is not satisfied, so that this is a restriction. By continuity of (3.6) in  $\theta$ , the condition is expected to hold at least in a neighborhood of the true parameter  $\theta_0$ .

**Assumption 3.2** *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 3.1.** *For the model given by (2.1) and (2.3), let Assumptions 2.1, 2.2, 3.1 and 3.2 be satisfied. Then it holds that*

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

## 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  $\theta_0$ .

If  $\hat{\theta}_T$  lies in the interior of  $\Theta$ , 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  $\theta_T^*$  on the line segment  $[\hat{\theta}_T, \theta_0]$  that converges to  $\theta_0$ . 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  $\theta_T^*$  and its limit, i.e.  $|\hat{L}_T(\theta_T^*) - \hat{L}_T(\theta_0)|$ , and its derivatives. For our next Theorem, we need the following additional assumptions.

**Assumption 3.3** *The true parameter  $\theta_0$  belongs to the interior  $\overset{\circ}{\Theta}$  of  $\Theta$ .*

**Assumption 3.4**  *$E(Z_t^4) < \infty$  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 3.2.** *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 2.1, 2.2, 3.1, 3.2, 3.3 and 3.4 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

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

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 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 1 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 2 plots the Hill (1975) tail index estimator together with 95% confidence bands for the returns of the SP500<sup>3</sup>. 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 estimate in Figure 2 is plotted as a function of the number of tail observations used for estimation. Our estimate of the tail index varies around three, which suggests that the fourth order moment is unlikely to exist. As discussed above, this would invalidate the estimation theory of classical GARCH-in-mean models.

Tables 2 and 3 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.

---

<sup>3</sup>To save space we show only this case as the other ones look very similar.

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  $\lambda_2$  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  $\omega$  is essentially a scaling parameter that determines the general level of (unconditional) volatility, and the link between  $\omega$  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 3 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 2 and 3. 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 4 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 4 shows the estimated news impact functions.

Table 4 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 3, 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 Vries, 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 (Heidel-

berg), 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.

## A Proofs

**Proof of Theorem 3.1.** 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)  $\sup_{\Theta} \left| \widehat{L}_T(\theta) - \overline{L}_T(\theta) \right| \rightarrow 0 \text{ a.s.},$
- (ii)  $\sup_{\Theta} \left| \overline{L}_T(\theta) - L(\theta) \right| \rightarrow 0 \text{ a.s.},$
- (iii)  $L(\theta) < L(\theta_0), \forall \theta \in \Theta : \theta \neq \theta_0.$

By the triangle inequality,

$$\sup_{\Theta} \left| \widehat{L}_T(\theta) - L(\theta) \right| \leq \sup_{\Theta} \left| \widehat{L}_T(\theta) - \overline{L}_T(\theta) \right| + \sup_{\Theta} \left| \overline{L}_T(\theta) - L(\theta) \right| \quad (\text{A.1})$$

the convergence of  $\widehat{L}_T(\theta)$  to  $L(\theta)$  follows by showing that the two terms on the right hand side of (A.1) 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 (2.8).

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  $\log(\widehat{h}_t(\theta))$  are Lipschitz continuous functions, since they have a bounded first derivative by Assumption 2.2. 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  $\theta$  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 (8.2) in SM,  $\left\| \widehat{h}_t(\theta) - h_t(\theta) \right\|_{\Theta}$  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 (A.1), 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 2.2(i). This implies that  $\left[ \widehat{L}_T(\theta) - \overline{L}_T(\theta) \right]$  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

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

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 3.2. 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.  $\square$

To show asymptotic normality of the estimator, it is essential to analyze the asymptotic behavior of the sequences of continuous functions  $\partial \log \hat{h}_t(\theta) / \partial \theta$  and  $\partial^2 \log \hat{h}_t(\theta) / \partial \theta \partial \theta^\top$ . 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  $\partial \log \hat{h}_t(\theta) / \partial \theta$  and  $\partial^2 \log \hat{h}_t(\theta) / \partial \theta \partial \theta^\top$ , 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  $\theta$ , ignoring any set of elements of Lebesgue measure zero where the derivative of the random functions  $\Phi_t$  is not defined. The partial derivatives are given by

$$\begin{aligned} \frac{\partial g_\theta}{\partial s} &= \beta - \left\{ \begin{array}{l} 2^{-1} [\gamma(y - \lambda_1 - \lambda_2 \exp(s)) + \delta (|y - \lambda_1 - \lambda_2 \exp(s)|)] \\ + \lambda_2 [\gamma + \delta \operatorname{sgn}(y - \lambda_1 - \lambda_2 \exp(s))] \exp(s) \end{array} \right\} \exp(-s/2) \quad (\text{A.2}) \\ \frac{\partial g_\theta}{\partial \theta} &= \begin{pmatrix} -[\gamma + \delta \operatorname{sgn}(y - \lambda_1 - \lambda_2 \exp(s))] \exp(-s/2) \\ -[\gamma + \delta \operatorname{sgn}(y - \lambda_1 - \lambda_2 \exp(s))] \exp(s/2) \\ 1 \\ (y - \lambda_1 - \lambda_2 \exp(s)) \exp(-s/2) \\ (|y - \lambda_1 - \lambda_2 \exp(s)|) \exp(-s/2) \\ s \end{pmatrix}, \quad (\text{A.3}) \end{aligned}$$

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  $\partial \log \hat{h}_t(\theta) / \partial \theta$ , we can obtain a new linear SRE of the first order derivative as

$$\frac{\partial \log h_{t+1}(\theta)}{\partial \theta} = A_t \frac{\partial \log h_t(\theta)}{\partial \theta} + B_t, \quad (\text{A.4})$$

where

$$A_t = \beta - \{2^{-1} [\gamma \varepsilon_t(\theta) + \delta |\varepsilon_t(\theta)|] + \lambda_2 [\gamma + \delta \operatorname{sgn}(\varepsilon_t(\theta))] h_t(\theta)\} h_t(\theta)^{-1/2}$$

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

with  $\varepsilon_t(\theta) := Y_t - \lambda_1 - \lambda_2 h_t(\theta)$ . For the proof of Theorem 3.2, we will use the following lemmas.

**Lemma 2** *Let the Assumptions 2.1, 2.2, 3.1, 3.2, 3.3 and 3.4 hold. Eq. (A.4) has a unique stationary solution  $\partial \log h_t(\theta) / \partial \theta$  that is ergodic. There exists a  $\delta > 0$  such that*

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

**Lemma 3** *Under Assumptions 2.1, 2.2, 3.1, 3.2, 3.3 and 3.4, eq. (8.3) in SM 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*

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

**Lemma 4** *Let the Assumptions 2.1, 2.2, 3.1, 3.2, 3.3 and 3.4 hold. For any sequence  $\hat{\theta}_T$  such that  $\hat{\theta}_T \rightarrow \theta_0$  almost surely, it holds that*

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

**Lemma 5** *Let the Assumptions 2.1, 2.2, 3.1, 3.2, 3.3 and 3.4 hold. For any sequence  $\hat{\theta}_T$  such that  $\hat{\theta}_T \rightarrow \theta_0$  almost surely, it holds that*

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



### Proof of Theorem 3.2.

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  $\nabla$  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

$$\|H_t(\theta)\|_{\mathcal{V}(\theta_0)} \leq \left\| \frac{\partial \log(h_t(\theta))}{\partial \theta} \frac{\partial \log(h_t(\theta))}{\partial \theta^\top} \right\|_{\mathcal{V}(\theta_0)} + \left\| \frac{2}{h_t(\theta)} \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta^\top} \right\|_{\mathcal{V}(\theta_0)}. \quad (\text{A.5})$$

In the sequel we can use the triangle inequality

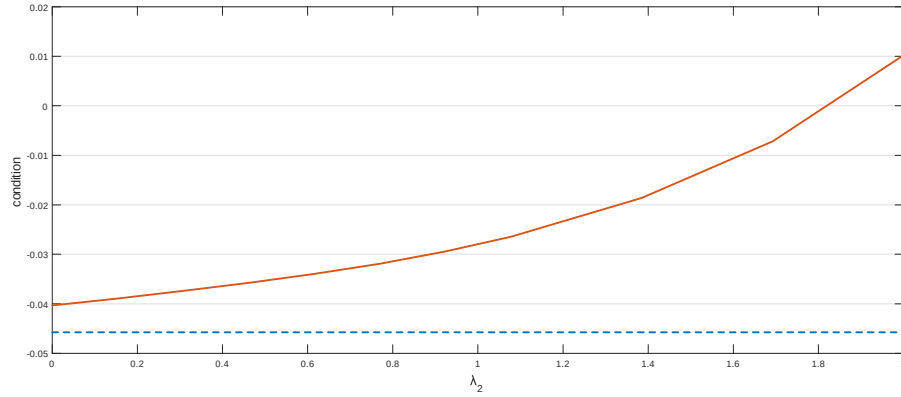
$$\left\| \widehat{L}_T''(\widehat{\theta}_T) - F(\theta_0) \right\| \leq \left\| \widehat{L}_T''(\widehat{\theta}_T) - \overline{L}_T''(\theta_0) \right\| + \left\| \overline{L}_T''(\theta_0) - F(\theta_0) \right\|,$$

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 3.1 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 (A.5) 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

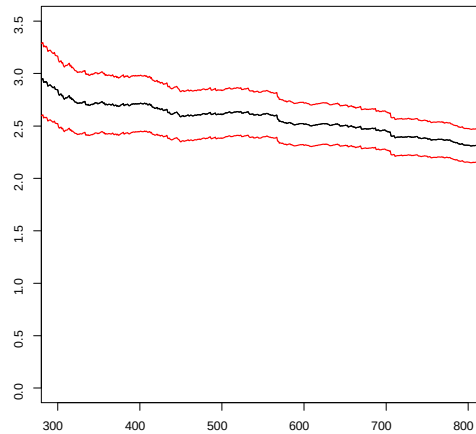
concentrated on two points, the proof is analogous to that for the EGARCH(1,1) model (see Wintenberger, 2013 and Lemma 8.2 of Straumann and Mikosch, 2006). Finally, it suffices to apply Slutsky's lemma to obtain the statement of Theorem 3.2.  $\square$

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



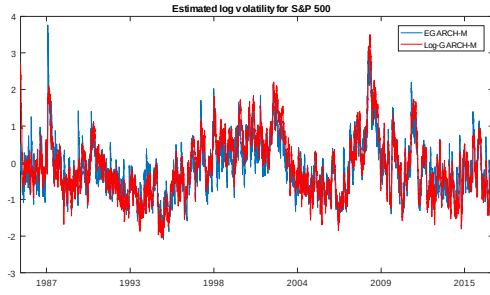
Notes: The verification of the condition is done by numerical approximation. The solid red line corresponds to the case of  $\delta_0 = 0.15$  and the dashed blue line to the case of  $\delta_0 = 0.9$ .

Figure 2: Hill plots of the returns of the SP500.

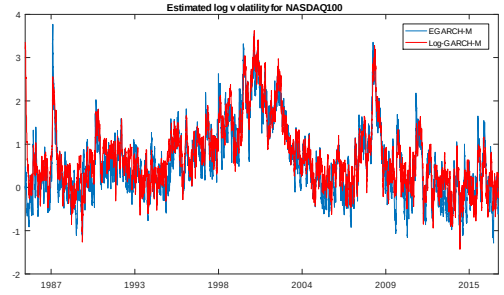


Notes: The horizontal axis shows the number of tail observations. The red lines correspond to the 95% confidence bands.

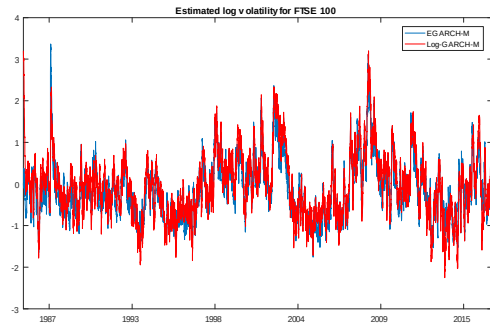
Figure 3: Estimated log volatilities for the stock market indices.



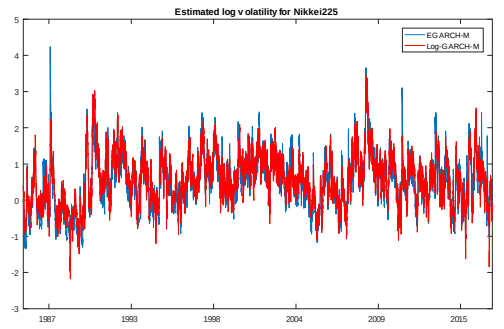
(a) SP500



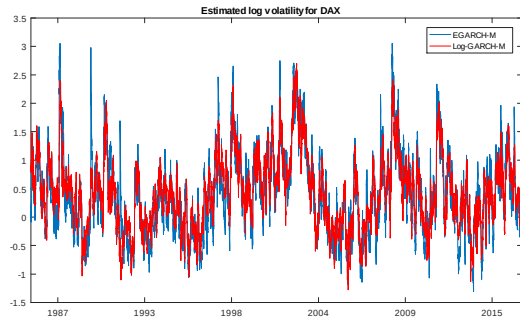
(b) Nasdaq100



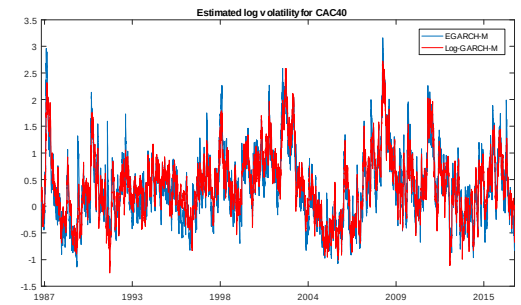
(c) FTSE100



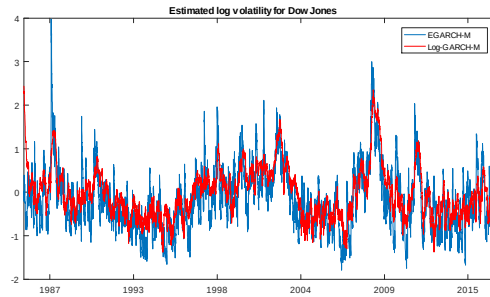
(d) Nikkei225



(e) DAX

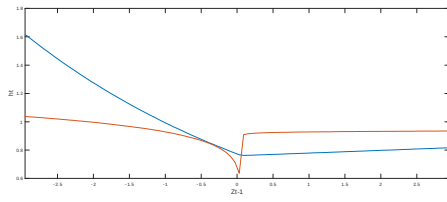


(f) CAC40

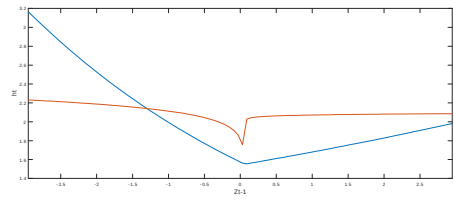


(g) Dow Jones

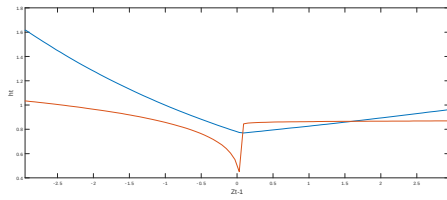
Figure 4: News Impact Curves (NIC) for the stock market indices.



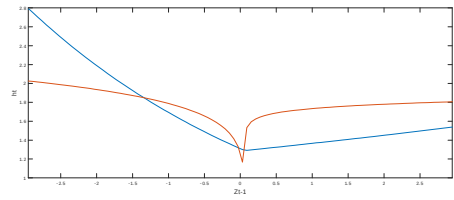
(a) SP500



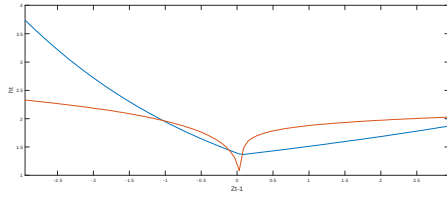
(b) Nasdaq100



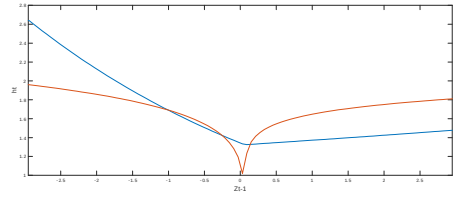
(c) FTSE100



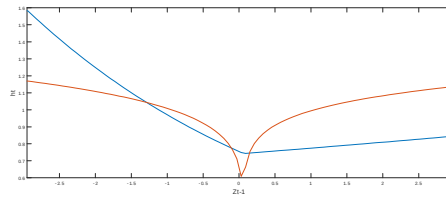
(d) DAX



(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

Table 1: 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   | <b>-0.042***</b> |
|            | 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   | <b>-0.103***</b> |
| Nasdaq 100 | Full sample | 7815         | 0.060 | 2.838    | 0.142    | 10.907   | <b>-0.019*</b>   |
|            | I           | 2527         | 0.067 | 1.643    | -0.850   | 17.484   | <b>0.143***</b>  |
|            | II          | 2520         | 0.067 | 5.068    | 0.334    | 7.0290   | <b>-0.052***</b> |
|            | III         | 2769         | 0.048 | 1.881    | 0.031    | 11.266   | <b>-0.074***</b> |
| 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   | <b>0.062***</b>  |
|            | II          | 2527         | 0.023 | 1.284    | -0.096   | 5.666    | 0.002            |
|            | III         | 2780         | 0.016 | 1.524    | 0.039    | 10.467   | <b>-0.039**</b>  |
| 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   | <b>-0.025**</b>  |
|            | I           | 2462         | 0.027 | 1.931    | 0.047    | 14.762   | <b>0.036*</b>    |
|            | II          | 2459         | 0.002 | 2.090    | 0.057    | 4.973    | <b>-0.039*</b>   |
|            | III         | 2745         | 0.020 | 2.580    | -0.258   | 10.300   | <b>-0.053***</b> |
| 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    | <b>-0.039*</b>   |
| Dow Jones  | Full sample | 7815         | 0.039 | 1.252    | -1.076   | 32.158   | <b>-0.037***</b> |
|            | 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   | <b>-0.102***</b> |

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 2: 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) | <b>0.037*** (0.013)</b> | -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) | <b>0.058** (0.027)</b>  | -3618    |
|               | III | -0.102 (0.010) | -0.178 (0.010) | 0.125 (0.012)  | 0.964 (0.003) | <b>0.054*** (0.016)</b> | -3748    |
| Nasdaq 100    | ALL | -0.113 (0.005) | -0.076 (0.004) | 0.161 (0.007)  | 0.981 (0.001) | <b>0.024*** (0.007)</b> | -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) | <b>0.050*** (0.016)</b> | -4213    |
| FTSE 100      | ALL | -0.131 (0.006) | -0.085 (0.003) | 0.163 (0.008)  | 0.975 (0.002) | <b>0.035** (0.014)</b>  | -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) | <b>0.069*** (0.019)</b> | -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) | <b>0.032* (0.017)</b>   | -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) | <b>0.033* (0.017)</b>   | -4738    |
| CAC 40        | ALL | -0.095 (0.004) | -0.096 (0.004) | 0.134 (0.006)  | 0.974 (0.001) | <b>0.033*** (0.012)</b> | -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) | <b>0.046** (0.018)</b>  | -4104    |
| Dow Jones     | ALL | -0.116 (0.005) | -0.104 (0.003) | 0.148 (0.006)  | 0.972 (0.001) | <b>0.036*** (0.014)</b> | -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) | <b>0.065** (0.029)</b>  | -3558    |
|               | III | -0.115 (0.012) | -0.171 (0.010) | 0.136 (0.015)  | 0.964 (0.003) | <b>0.048*** (0.018)</b> | -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  $\lambda_2$  we indicate statistical significance using bold face and asterisks.

Table 3: 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) | <b>0.087***</b> (0.001) | -10966   |
|                                                             | I   | 0.091 (0.000)  | 0.058 (0.000)    | 0.057 (0.000)    | 0.929 (0.000) | <b>0.110***</b> (0.000) | -3150    |
|                                                             | II  | 0.042 (0.000)  | 0.025 (0.000)    | 0.035 (0.000)    | 0.962 (0.000) | <b>0.054***</b> (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) | <b>0.062***</b> (0.000) | -13687   |
|                                                             | I   | 0.070 (0.000)  | 0.022 (0.000)    | 0.063 (0.000)    | 0.930 (0.000) | <b>0.117***</b> (0.000) | -3983    |
|                                                             | II  | 0.030 (0.000)  | 0.007 (0.000)    | 0.037 (0.000)    | 0.976 (0.000) | <b>0.008***</b> (0.000) | -5186    |
|                                                             | III | 0.103 (0.000)  | 0.052 (0.000)    | 0.073 (0.000)    | 0.908 (0.000) | <b>0.148***</b> (0.000) | -4332    |
| FTSE 100                                                    | ALL | 0.052 (0.003)  | 0.003 (0.002)    | 0.090 (0.005)    | 0.928 (0.003) | <b>0.126***</b> (0.008) | -11091   |
|                                                             | I   | 0.035 (0.000)  | 0.003 (0.000)    | 0.055 (0.000)    | 0.952 (0.000) | <b>0.106***</b> (0.000) | -3333    |
|                                                             | II  | 0.029 (0.000)  | 0.005 (0.000)    | 0.044 (0.000)    | 0.973 (0.000) | <b>0.076***</b> (0.000) | -3537    |
|                                                             | III | 0.096 (0.006)  | 0.064 (0.005)    | 0.066 (0.004)    | 0.915 (0.004) | <b>0.052***</b> (0.002) | -4029    |
| DAX                                                         | ALL | 0.072 (0.000)  | 0.018 (0.000)    | 0.060 (0.000)    | 0.934 (0.000) | <b>0.093***</b> (0.000) | -12971   |
|                                                             | I   | 0.047 (0.000)  | 0.006 (0.001)    | 0.035 (0.001)    | 0.948 (0.000) | <b>0.125***</b> (0.000) | -3994    |
|                                                             | II  | 0.053 (0.000)  | 0.025 (0.000)    | 0.045 (0.000)    | 0.953 (0.000) | <b>0.076***</b> (0.000) | -4387    |
|                                                             | III | 0.044 (0.000)  | 0.014 (0.000)    | 0.036 (0.000)    | 0.967 (0.000) | <b>0.089***</b> (0.000) | -4647    |
| Nikkei 225                                                  | ALL | 0.106 (0.015)  | 0.034 (0.009)    | 0.083 (0.007)    | 0.908 (0.012) | <b>0.011***</b> (0.001) | -13009   |
|                                                             | I   | 0.138 (0.000)  | 0.073 (0.000)    | 0.086 (0.000)    | 0.883 (0.000) | <b>0.001***</b> (0.000) | -3969    |
|                                                             | II  | 0.034 (0.000)  | 0.014 (0.000)    | 0.021 (0.000)    | 0.969 (0.000) | <b>0.074***</b> (0.000) | -4294    |
|                                                             | III | 0.054 (0.000)  | 0.003 (0.001)    | 0.045 (0.002)    | 0.950 (0.000) | <b>0.093***</b> (0.000) | -4905    |
| CAC 40                                                      | ALL | 0.093 (0.001)  | 0.042 (0.002)    | 0.071 (0.002)    | 0.918 (0.001) | <b>0.020***</b> (0.000) | -12197   |
|                                                             | I   | 0.166 (0.000)  | 0.075 (0.000)    | 0.100 (0.000)    | 0.807 (0.000) | <b>0.074***</b> (0.000) | -3822    |
|                                                             | II  | 0.033 (0.000)  | 0.003 (0.000)    | 0.036 (0.000)    | 0.971 (0.000) | <b>0.083***</b> (0.000) | -4166    |
|                                                             | III | 0.064 (0.000)  | 0.019 (0.000)    | 0.058 (0.000)    | 0.939 (0.000) | <b>0.025***</b> (0.000) | -4227    |
| Dow Jones                                                   | ALL | 0.101 (0.000)  | 0.059 (0.001)    | 0.071 (0.000)    | 0.910 (0.001) | <b>0.034***</b> (0.000) | -10692   |
|                                                             | I   | 0.048 (0.005)  | 0.003 (0.008)    | 0.060 (0.010)    | 0.962 (0.003) | <b>0.066***</b> (0.010) | -3307    |
|                                                             | II  | 0.064 (0.000)  | 0.039 (0.002)    | 0.048 (0.003)    | 0.939 (0.002) | <b>0.015***</b> (0.000) | -3639    |
|                                                             | III | 0.161 (0.000)  | 0.124 (0.002)    | 0.087 (0.001)    | 0.857 (0.001) | <b>0.100***</b> (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  $\lambda_2$  we indicate statistical significance using bold face and asterisks.



Table 4: 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., Gariepy, 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.