An Equation-by-Equation Estimator of a Multivariate Log-GARCH-X Model of Financial Returns[1]

Christian Francq[2] and Genaro Sucarrat[3]

30th August 2016

Abstract

Estimation of large financial volatility models is plagued by the curse of dimensionality. As the dimension grows, joint estimation of the parameters becomes unfeasible in practice. This problem is compounded if covariates or conditioning variables ("X") are added to each volatility equation. The problem is especially acute for non-exponential volatility models, e.g., GARCH models, since the variables and parameters in these cases are restricted to be positive. Here, we propose an estimator for a multivariate log-GARCH-X model that avoids these problems. The model allows for feedback among the equations, admits several stationary regressors as conditioning variables in the X-part (including leverage terms), and allows for time-varying conditional covariances of unknown form. Strong consistency and asymptotic normality of an equation-by-equation least squares estimator are proved, and the results can be used to undertake inference both within and across equations. The flexibility and usefulness of the estimator is illustrated in two empirical applications.

JEL Classification: C13, C22, C32, C51, C58

Keywords: Exponential GARCH; Multivariate log-GARCH-X; VARMA-X; Equation-by-Equation Estimation (EBEE); Least Squares

[1] We are grateful to participants at the CFE 2015 (London) for useful comments, suggestions and questions.

[2]CREST and Université de Lille. Email christian.francq at univ-lille3.fr

[3]Corresponding author. Department of Economics, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway. Email genaro.sucarrat at bi.no, phone +47 46410779, fax +47 23264788.

1 Introduction

Estimating the volatility of financial returns is crucial, in particular for risk management (McNeil and Embrechts (2005)), asset pricing (Duan (1995)) and portfolio optimisation (Santos and Moura (2014)). The standard Autoregressive Conditional Heteroscedasticity (ARCH) models explain the volatility by using past returns only, and do not incorporate covariates that could convey relevant exogenous information.

Covariates often help in explaining and forecasting financial variability. Examples of such covariates include, among others, measures of information arrival, e.g., Clark (1973), Tauchen and Pitts (1983), Lamoureux and Lastrapes (1990), interest rates, e.g., Brenner et al. (1996), Hagiwara and Herce (1999), central bank interventions (Dominguez (1998)), bid-ask spreads (Bollerslev and Melvin (1994)), macroeconomic fundamentals (Apergis and Rezitis (2011)), cross-sectional volatility (Hwang and Satchell (2005)) and volatility proxies made up of high-frequency data (Engle and Gallo (2006), Hansen et al. (2012), Shephard and Sheppard (2010)). An example of a study that combines several of these covariates in a single analysis is Bauwens et al. (2006). Despite the widespread use and usefulness of such additional information in explaining and forecasting volatility, there are relatively few results on ARCH models with covariates, i.e., ARCH-X models, where the assumptions on the X-part are non-restrictive and of general practical interest.

In the univariate case, Han and Kristensen (2014) prove the Consistency and Asymptotic Normality (CAN) of the Gaussian QMLE for specifications contained in the GARCH(1,1)-X model, where the X-part consists of a single variable only. Francq and Thieu (2015) also prove the CAN for the QMLE, but for a much broader model-class: The

Asymmetric Power GARCH(p, q)-X model, where the X-part can contain more than one variable. However, as is common in GARCH-specifications that are not exponential, all terms (parameters, the variables in the X-part, etc.) are restricted to being non-negative in both works. Chen and Song (2015) prove the CAN of a QMLE for a log-GARCH(1,1) model with no X-part in the classical sense, but where the ARCH parameter varies over time and is driven by the past values of two covariates that are independent of the standardised error. This independence assumption is somewhat restrictive, however, since it excludes feedback effects between the covariates and the log-volatility process (this is usually not fulfilled in empirical practice). Also, it is not clear what the economic motivation is in making the ARCH parameter – the part of a log-GARCH(1,1) that usually accounts for the smallest portion of the variation in volatility – time-varying and dependent on past covariates whose properties are usually not fulfilled in empirical practice. Finally, to the best of our knowledge, there is no proof of the CAN for multivariate GARCH-X models.

Sucarrat et al. (2015) propose a general framework for the estimation of, and inference in, univariate and multivariate log-GARCH-X models – with Dynamic Conditional Correlations (DCCs) of unknown form – via the (V)ARMA-X representation. However, they do not prove the CAN, neither in the univariate nor in the multivariate case. Here, we adopt their framework, but provide a proof of strong consistency and asymptotic normality under mild assumptions. Specifically, we do so for a least squares Equation-by-Equation (EBE) estimator of a multivariate log-GARCH(1,1)-X model that admits DCCs of unknown form. Moreover, the assumptions on the X-part are very general: It is not restricted to a single variable, the X-variables are allowed to be subject to feedback effects from volatility, i.e., the X-variables need not be exogenous, and the X-variables need not be independent of the standardised error. The latter means that asymmetry or leverage can be accommodated via the X-part. There are several advantages with the VARMA approach. First, it enables theoretical results of unprecedented economic generality and flexibility. In one of our applications, for example, we illustrate this in an empirical study of volatility spillover among stock markets. In ordinary multivariate GARCH models such tests require complicated restrictions on the parameters, and restrictive assumptions on the correlations, i.e., constant conditional correlations, see Conrad and Weber (2013), and Pedersen (2015). By contrast, in our model, complicated restrictions on the parameters are not needed, and tests are valid under time-varying DCCs of unknown form. Moreover, we can also test whether covariates, e.g., volatility proxies, provide additional – or alternative – channels of volatility spillover. Second, the EBE nature of our estimator, together with the VARMA-X representation, means that large systems can readily be estimated with software that is already widely available. We illustrate this in a second empirical application by estimating, in just over a minute, a 50-dimensional model that admits time-varying correlations, and where the X-part contains 5 conditioning variables in each equation. Next, a DCC model of the 1225 correlation paths is fitted. Third, the statistical theory we rely upon is much more tractable than for the EGARCH of Nelson (1991). Indeed, currently the only proof of CAN for a QMLE is for the univariate EGARCH(1,1) without covariates, see Straumann and Mikosch (2006), and Wintenberger (2013). Fourth, estimation via the VARMA-representation is likely to be more efficient when the standardised error is fat-tailed, since the application of the logarithm makes large (in absolute value) observations less influential. Finally, solutions to the log-of-zero (or inlier) problem is available when log-GARCH models are estimated via the (V)ARMA representation, see Sucarrat and Escribano (2014), and Sucarrat et al. (2015).

The rest of the paper is organised into four parts. First, in Section 2, we present the model and its associated notation. Section 3 contains our main theoretical results, while Section 4 gathers the proofs. Section 5 contains two empirical applications. Finally, Section 6 concludes. A Table and a Figure are located at the end.

2 Model and notation

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{Mt})^\top$ denote an $M \times 1$ vector of random variables that are nonzero with probability one, and let $\mathcal{F}_{t-1} = \sigma\{\epsilon_{ju}^2, X_{\ell u} : u < t, j = 1, \ldots, M, \ell = 1, \ldots, K\}$ be the σ-field generated by the past values of $\boldsymbol{\epsilon}_t^2 = (\epsilon_{1t}^2, \ldots, \epsilon_{Mt}^2)^\top$ and of some K-dimensional vector $\boldsymbol{X}_t = (X_{1t}, \ldots, X_{Kt})^\top$ of covariates. It should be noted that the covariates need *not* be exogenous (the exact assumptions on $\boldsymbol{X}_t$ are given below). Assume the existence of the $M \times M$ matrix $\boldsymbol{H}_t$, such that

$$\mathrm{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top | \mathcal{F}_{t-1}) \quad = \quad \boldsymbol{H}_t. \tag{2.1}$$

Assuming the nonsingularity of $\boldsymbol{D}_t^2 = \mathrm{diag}(\boldsymbol{H}_t)$, let the $M \times 1$ vector

$$\boldsymbol{\eta}_t \quad = \quad \boldsymbol{D}_t^{-1} \boldsymbol{\epsilon}_t. \tag{2.2}$$

Note that this implies $\mathrm{E}(\boldsymbol{\eta}_t^2) = (1, \ldots, 1)^\top$, where $\boldsymbol{\eta}_t^2 = (\eta_{1t}^2, \ldots, \eta_{Mt}^2)^\top$. Let $\boldsymbol{\sigma}_t^2$ be the M-dimensional vector equal to the diagonal of $\boldsymbol{H}_t$. If, for some vector $\boldsymbol{\sigma}$ with positive elements, $\ln \boldsymbol{\sigma}$ denotes the vector resulting from applying the natural log on $\boldsymbol{\sigma}$ element-wise, then the M-dimensional log-GARCH(1,1)-X specification with diagonal GARCH matrix and covariate-vector $\boldsymbol{X}_t$ is given by

$$\ln \boldsymbol{\sigma}_t^2 \quad = \quad \boldsymbol{\omega}_0 + \boldsymbol{\alpha}_0 \ln \boldsymbol{\epsilon}_{t-1}^2 + \boldsymbol{\beta}_0 \ln \boldsymbol{\sigma}_{t-1}^2 + \boldsymbol{\lambda}_0 \boldsymbol{X}_{t-1}, \tag{2.3}$$

where $\ln \boldsymbol{\sigma}_t^2 = (\ln \sigma_{1,t}^2, \ldots, \ln \sigma_{M,t}^2)^\top$, $\boldsymbol{\omega}_0 = (\omega_{01}, \ldots, \omega_{0M})^\top$,

$$\boldsymbol{\alpha}_0 = \begin{pmatrix} \alpha_{011} & \cdots & \alpha_{01M} \\ \vdots & \ddots & \vdots \\ \alpha_{0M1} & \cdots & \alpha_{0MM} \end{pmatrix}, \quad \boldsymbol{\lambda}_0 = \begin{pmatrix} \lambda_{011} & \cdots & \lambda_{01K} \\ \vdots & \ddots & \vdots \\ \lambda_{0M1} & \cdots & \lambda_{0MK} \end{pmatrix}$$

and the $M \times M$ matrix $\boldsymbol{\beta}_0 = \mathrm{diag}(\beta_{011}, \ldots, \beta_{0MM})$. In principle, it should be possible to extend the study to log-GARCH-X of higher orders, but at the price of more complicated notations and less explicit assumptions.

Under stationarity and moments conditions that will be discussed below, the VARMA-X representation of this model is given by

$$\ln \boldsymbol{\epsilon}_t^2 = \boldsymbol{c}_0 + \boldsymbol{\phi}_0 \ln \boldsymbol{\epsilon}_{t-1}^2 - \boldsymbol{\beta}_0 \boldsymbol{u}_{t-1} + \boldsymbol{\lambda}_0 \boldsymbol{X}_{t-1} + \boldsymbol{u}_t, \tag{2.4}$$

where $\boldsymbol{\phi}_0 = \boldsymbol{\alpha}_0 + \boldsymbol{\beta}_0$,

$$\boldsymbol{c}_0 = \boldsymbol{\omega}_0 + (\boldsymbol{I}_M - \boldsymbol{\beta}_0)\mathrm{E}(\ln \boldsymbol{\eta}_t^2), \qquad \boldsymbol{u}_t = \ln \boldsymbol{\eta}_t^2 - \mathrm{E}(\ln \boldsymbol{\eta}_t^2),$$

and where $\boldsymbol{I}_M$ is the identity matrix of dimension M. Accordingly, equation j in the VARMA(1,1)-X system can be written as

$$\begin{aligned} \ln \epsilon_{jt}^2 &= c_{0j} + \sum_{\ell=1}^{M} \phi_{0j\ell} \ln \epsilon_{\ell,t-1}^2 + \sum_{\ell=1}^{K} \lambda_{0j\ell} X_{\ell,t-1} - \beta_{0jj} u_{j,t-1} + u_{jt}, \\ c_{0j} &= \omega_{0j} + (1 - \beta_{0jj})\mathrm{E}(\ln \eta_{jt}^2). \end{aligned} \tag{2.5}$$

The norm of a matrix $M_1 \times M_2$ of generic element $M(i,j)$ is defined by $\|\boldsymbol{M}\| = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} |M(i,j)|$. The L_2-norm of a random variable X is defined by $\|X\|_2 = \sqrt{\mathrm{E}X^2}$.

3 Equation-by-equation estimation

Assume that

$$(\boldsymbol{\epsilon}_t^\top, \boldsymbol{X}_t^\top)^\top \text{ is stationary and ergodic,} \tag{3.1}$$

where $\boldsymbol{\epsilon}_t$ is centred and satisfies (2.1) and (2.3) with

$$\text{the spectral radius of } \boldsymbol{\phi}_0 \text{ being strictly less than 1, and } |\beta_{0jj}| < 1 \text{ for all } j. \tag{3.2}$$

Given a stationary and ergodic sequence $(\boldsymbol{X}_t)$ of covariates, one way to generate a process $(\boldsymbol{\epsilon}_t)$ satisfying the requirement (3.1) consists of drawing an iid sequence $(\boldsymbol{\eta}_t^*)$ with zero mean and variance $\boldsymbol{I}_M$, independently of $(\boldsymbol{X}_t)$. Under (3.2), it then suffices, for instance, to set $\boldsymbol{\epsilon}_t = \boldsymbol{\sigma}_t \boldsymbol{\eta}_t^*$, where the functions of vectors are defined element-wise, and

$$\ln \boldsymbol{\sigma}_t^2 = \sum_{k=0}^{\infty} \boldsymbol{\phi}_0^k \left(\boldsymbol{\omega}_0 + \boldsymbol{\alpha}_0 \ln \boldsymbol{\eta}_{t-k-1}^{*2} + \boldsymbol{\lambda}_0 \boldsymbol{X}_{t-k-1} \right).$$

Note that, in this case, we have $\boldsymbol{\eta}_t = \boldsymbol{\eta}_t^*$. More generally, from (2.2), it is equivalent to assume (3.1) or to assume that

$$(\boldsymbol{\eta}_t^\top, \boldsymbol{X}_t^\top)^\top \text{ is stationary and ergodic,}$$

Note also that the stationary and ergodic solution of (2.4) is then given by

$$\ln \boldsymbol{\epsilon}_t^2 = \sum_{k=0}^{\infty} \boldsymbol{\phi}_0^k \left(\boldsymbol{c}_0 + \boldsymbol{\lambda}_0 \boldsymbol{X}_{t-k-1} - \boldsymbol{\beta}_0 \boldsymbol{u}_{t-k-1} + \boldsymbol{u}_{t-k} \right).$$

Moreover, the model is invertible, in the sense that

$$\boldsymbol{u}_t = \sum_{k=0}^{\infty} \boldsymbol{\beta}_0^k \left(\ln \boldsymbol{\epsilon}_{t-k}^2 - \boldsymbol{\phi}_0 \ln \boldsymbol{\epsilon}_{t-k-1}^2 - \boldsymbol{c}_0 - \boldsymbol{\lambda}_0 \boldsymbol{X}_{t-k-1} \right).$$

Remark 3.1 *Note that, contrary to the log-GARCH model, the stationarity and invertibility conditions of the standard multivariate GARCH models are quite complicated (see Boussama et al. (2011) for the BEKK model), or remain unknown (as for the DCC-MGARCH model proposed by Engle, 2002). An advantage of the log-GARCH is that its ARMA representation has a noise which depends on $\boldsymbol{\eta}_t$ only. In the ARMA representations of the other GARCH formulations, the innovations generally depend on the past observations themselves (they are typically of the form $\boldsymbol{u}_t = \boldsymbol{\epsilon}_t^2 - \mathrm{E}(\boldsymbol{\epsilon}_t^2 \mid \mathcal{F}_{t-1})$), and thus these ARMA representations are of no use for finding stationarity conditions and are hardly usable for estimation purposes.*

Under the moment conditions

$$\mathrm{E} \left\| \ln \boldsymbol{\eta}_t^2 \right\|^2 < \infty \quad \text{and} \quad \mathrm{E} \|\boldsymbol{X}_t\|^2 < \infty \tag{3.3}$$

we have $\mathrm{E} \left\| \ln \boldsymbol{\epsilon}_t^2 \right\|^2 < \infty$ and $\mathrm{E}\|\boldsymbol{u}_t\|^2 < \infty$, and we can define the Hilbert space $\mathcal{H}_{t-1}$ that is generated by the linear combinations of the $\ln \epsilon_{js}^2$'s and the $X_{\ell,s}$'s for $s < t$, $j = 1, \ldots, M$, and $\ell = 1, \ldots, K$, and by their limits in L^2. Note that $\mathcal{H}_{t-1}$ is also equal to the Hilbert space generated by $\{X_{\ell s}, u_{js} : s < t\}$, and equivalently to the Hilbert space generated by $\{X_{\ell s}, \ln \eta_{js}^2 : s < t\}$. One can thus interpret $(\boldsymbol{u}_t)$ as a linear innovation process:

$$\boldsymbol{u}_t = \ln \boldsymbol{\epsilon}_t^2 - \mathrm{E}(\ln \boldsymbol{\epsilon}_t^2 \mid \mathcal{H}_{t-1}),$$

where $\mathrm{E}(\ln \boldsymbol{\epsilon}_t^2 \mid \mathcal{H}_{t-1})$ denotes the orthogonal projection of $\ln \boldsymbol{\epsilon}_t^2$ onto $\mathcal{H}_{t-1}$. Similarly, we define

$$\boldsymbol{v}_t = \boldsymbol{X}_t - \mathrm{E}(\boldsymbol{X}_t \mid \mathcal{H}_{t-1}).$$

It will be assumed that the matrix

$$\boldsymbol{\Sigma} := \mathrm{E} \boldsymbol{w}_1 \boldsymbol{w}_1^\top, \text{ where } \boldsymbol{w}_1 = (\boldsymbol{u}_1^\top, \boldsymbol{v}_1^\top)^\top, \text{ is positive definite.} \tag{3.4}$$

Note that this assumption rules out the possibility of exact linear relations between the explanatory variables involved in the log-volatility $\ln \boldsymbol{\sigma}_t^2$, which is obviously a necessary identifiability condition.

In order to accommodate certain types of dynamic conditional correlation (DCC) models, we assume (instead of the usual assumption that $(\boldsymbol{\eta}_t)$ is iid) that

$$\text{for any } j \in \{1, \ldots, M\}, \ \eta_{jt} \text{ is independent of } \mathcal{F}_{t-1}. \tag{3.5}$$

This is a mild assumption. If, for example, $\boldsymbol{\eta}_t = \boldsymbol{R}^{1/2}(\Delta_t)\boldsymbol{\xi}_t$, where (Δ_t) and $(\boldsymbol{\xi}_t)$ are two independent processes, $\boldsymbol{\xi}_t$ is independent of $\mathcal{F}_{t-1}$, $\boldsymbol{R}(\Delta_t)$ is a correlation matrix for any value of Δ_t, and $(\boldsymbol{\xi}_t)$ is an independent sequence of, say, $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M)$-distributed vectors, then (3.5) is satisfied. Indeed, conditionally on Δ_t and $\mathcal{F}_{t-1}$, the variable η_{jt} is $\mathcal{N}(0,1)$-distributed. Since this distribution does not depend on $\mathcal{F}_{t-1}$ nor on Δ_t, (3.5) holds true. The $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_M)$-distribution can be replaced by any other spherical distribution, as shown in Proposition 3.1 of Francq and Zakoïan (2015). A special case of (3.5) is

$$(\boldsymbol{\eta}_t) \text{ is an iid sequence and } \boldsymbol{\eta}_t \text{ independent of } \{\boldsymbol{X}_u, u < t\} . \tag{3.6}$$

In this case, we have a constant conditional correlation (CCC) model, such that

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t, \quad \boldsymbol{R} := \mathrm{E}\boldsymbol{\eta}_t \boldsymbol{\eta}_t^\top.$$

Note that (3.6) is not satisfied in the previous example where $\boldsymbol{\eta}_t = \boldsymbol{R}^{1/2}(\Delta_t)\boldsymbol{\xi}_t$, since Δ_t is not independent of $\mathcal{F}_{t-1}$. Moreover, empirical evidence of non-constant conditional correlations are often found, meaning that (3.6) is generally not satisfied.

Remark 3.2 *The VARMA-X representation (2.4) can be used to make linear h-step ahead predictions $\widehat{\boldsymbol{Z}}_{t+h|t} = E(\boldsymbol{Z}_{t+h} \mid \mathcal{H}_t)$ of $\boldsymbol{Z}_{t+h} := \ln \boldsymbol{\epsilon}_{t+h}^2$. Since the covariates $\boldsymbol{X}_t$ are exogenous, their predictions cannot be obtained from the log-GARCH-X model. We thus assume that, at time t, the value of $\boldsymbol{X}_{t+h-1}$ (or its prediction) is available. For some applications, this implies that $\boldsymbol{X}_t$ actually corresponds to a variable measured at time $t-h+1$, or before that date. Under (3.5), the predictions are then recursively obtained, for $k = 1, \dots, h$, by*

$$\widehat{\boldsymbol{Z}}_{t+k|t} = \boldsymbol{c}_0 + \boldsymbol{\phi}_0 \widehat{\boldsymbol{Z}}_{t+k-1|t} - \boldsymbol{\beta}_0 \widehat{\boldsymbol{u}}_{t+k-1|t} + \boldsymbol{\lambda}_0 \boldsymbol{X}_{t+k-1},$$

where $\widehat{\boldsymbol{Z}}_{t|t} = \boldsymbol{Z}_t$, $\widehat{\boldsymbol{u}}_{t|t} = \boldsymbol{u}_t$ and, for $k \geq 2$,

$$\widehat{\boldsymbol{u}}_{t+k-1|t} = \widehat{\boldsymbol{Z}}_{t+k-1|t} - \boldsymbol{c}_0 - \boldsymbol{\phi}_0 \widehat{\boldsymbol{Z}}_{t+k-2|t} - \boldsymbol{\beta}_0 \widehat{\boldsymbol{u}}_{t+k-2|t} - \boldsymbol{\lambda}_0 \boldsymbol{X}_{t+k-2}.$$

3.1 Estimator and strong consistency

Denote by

$$\boldsymbol{\vartheta}_0^{(j)} = (c_{0j}, \phi_{0j1}, \dots, \phi_{0jM}, \beta_{0jj}, \lambda_{0j1}, \dots, \lambda_{0jK})^\top$$

the vector of the unknown parameters involved in the j-th equation (2.5) of the VARMA-X model. This parameter of dimension $d = M + K + 2$ is assumed to belong to some compact parameter space $\Theta \subset \mathbb{R}^{M+1} \times (-1, 1) \times \mathbb{R}^K$ that does not depend on j. Let B be the backshift operator. For any $\underline{\boldsymbol{\vartheta}} = (c, \phi_1, \dots, \phi_M, \beta, \lambda_1, \dots, \lambda_K)^\top \in \Theta$, let

$$u_{jt}(\underline{\boldsymbol{\vartheta}}) = \frac{1 - \phi_j B}{1 - \beta B} \ln \epsilon_{jt}^2 - \frac{c}{1-\beta} - \sum_{\ell \neq j} \frac{\phi_\ell B}{1 - \beta B} \ln \epsilon_{\ell t}^2 - \frac{B}{1 - \beta B} \sum_{\ell=1}^{K} \lambda_\ell X_{\ell,t}.$$

Given observations $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$ and $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$, the $u_{jt}(\underline{\boldsymbol{\vartheta}})$'s are approximated by the recursions

$$\widetilde{u}_{jt}(\underline{\boldsymbol{\vartheta}}) = \ln \epsilon_{jt}^2 - c - \sum_{\ell=1}^{M} \phi_\ell \ln \epsilon_{\ell,t-1}^2 - \sum_{\ell=1}^{K} \lambda_\ell X_{\ell,t-1} + \beta \widetilde{u}_{j,t-1}(\underline{\boldsymbol{\vartheta}}),$$

for $t = 2, \ldots, n$, with the initial value $\widetilde{u}_{j1}(\underline{\boldsymbol{\vartheta}}) = 0$. An equation-by-equation least squares estimator of the VARMA-(1,1)-X model is then defined by

$$\widehat{\boldsymbol{\vartheta}}_n^{(j)} = \arg\min_{\underline{\boldsymbol{\vartheta}} \in \Theta} \widetilde{Q}_n^{(j)}(\underline{\boldsymbol{\vartheta}}), \quad \widetilde{Q}_n^{(j)}(\underline{\boldsymbol{\vartheta}}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{u}_{jt}^2(\underline{\boldsymbol{\vartheta}}). \tag{3.7}$$

Denote by $\boldsymbol{\vartheta}_0 = \left(\boldsymbol{\vartheta}_0^{(1)^\top}, \ldots, \boldsymbol{\vartheta}_0^{(M)^\top}\right)^\top$ the vector of all the parameters of the VARMA-X equation (2.4). This parameter vector belongs to the parameter space Θ^M, whose generic element is denoted by $\boldsymbol{\vartheta} = \left(\boldsymbol{\vartheta}^{(1)^\top}, \ldots, \boldsymbol{\vartheta}^{(M)^\top}\right)^\top$. The least squares estimator of the whole parameter $\boldsymbol{\vartheta}_0$ is defined by

$$\widehat{\boldsymbol{\vartheta}}_n = \arg\min_{\boldsymbol{\vartheta} \in \Theta^M} \sum_{t=1}^{n} \widetilde{\boldsymbol{u}}_t^\top(\boldsymbol{\vartheta})\widetilde{\boldsymbol{u}}_t(\boldsymbol{\vartheta}), \quad \widetilde{\boldsymbol{u}}_t(\boldsymbol{\vartheta}) = \left(\widetilde{u}_{1t}^2(\boldsymbol{\vartheta}^{(1)}), \ldots, \widetilde{u}_{Mt}^2(\boldsymbol{\vartheta}^{(M)})\right)^\top. \tag{3.8}$$

Since $\sum_{t=1}^n \widetilde{\boldsymbol{u}}_t^\top(\boldsymbol{\vartheta})\widetilde{\boldsymbol{u}}_t(\boldsymbol{\vartheta}) = \sum_{j=1}^M \sum_{t=1}^n \widetilde{u}_{jt}^2(\underline{\boldsymbol{\vartheta}})$, one can see that the collection of the equation-by-equation estimators is nothing else than the global least squares estimator:

$$\widehat{\boldsymbol{\vartheta}}_n = \left(\widehat{\boldsymbol{\vartheta}}_n^{(1)^\top}, \ldots, \widehat{\boldsymbol{\vartheta}}_n^{(M)^\top}\right)^\top.$$

It is, however, clearly easier to compute $\widehat{\boldsymbol{\vartheta}}_n$ by solving the d-dimensional optimisations (3.7), for $j = 1, \ldots, M$, than the Md-dimensional optimisation (3.8).

Theorem 3.1 *Assume the log-GARCH(1,1)-X model (2.1)-(2.3) with (3.1), (3.2), (3.3), (3.4) and (3.5). If $\boldsymbol{\vartheta}_0^{(j)}$ belongs to the compact set Θ and $\phi_{0jj} \neq \beta_{0jj}$, i.e., $\alpha_{0jj} \neq 0$, then $\widehat{\boldsymbol{\vartheta}}_n^{(j)} \to \boldsymbol{\vartheta}_0^{(j)}$ almost surely as $n \to \infty$.*

Remark 3.3 *Note that $\phi_{0jj} \neq \beta_{0jj}$ is an identifiability condition. It appears naturally when considering (2.5) as an ARMA(1,1)-X model for $\ln \epsilon_{jt}^2$, with covariates $\ln \epsilon_{\ell,t-1}^2$ ($\ell \neq j$) and $X_{\ell,t-1}$ ($\ell \in \{1, \ldots K\}$).*

To obtain a consistent estimator of the log-GARCH-X parameters, it remains to find a consistent estimator of $\mathrm{E} \ln \boldsymbol{\eta}_1^2 := \boldsymbol{\tau} = (\tau_1, \ldots, \tau_M)^\top$. Denote by $\widehat{u}_{jt} = \widetilde{u}_{jt}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})$ the residuals of the j-th ARMA-X equation.

Lemma 3.1 *Let the assumptions of Theorem 3.1, $\mathrm{E}\eta_{jt}^4 < \infty$ and $\mathrm{E}\,|\eta_{jt}|^{-s} < \infty$ for some $s > 0$. Almost surely, as $n \to \infty$ we have*

$$\widehat{\tau}_{jn} := -\ln \frac{1}{n}\sum_{t=1}^{n} e^{\widehat{u}_{jt}} \to \tau_j.$$

Let $\boldsymbol{\zeta}_0^{(j)} = (\omega_{0j}, \alpha_{0j1}, \ldots, \alpha_{0jM}, \beta_{0jj}, \lambda_{0j1}, \ldots, \lambda_{0jK})^\top$ be the parameter involved in the j-th equation of the log-GARCH model (2.3). The whole log-GARCH parameter $\boldsymbol{\zeta}_0 =$

$(\boldsymbol{\zeta}_0^{(1)\top}, \ldots, \boldsymbol{\zeta}_0^{(M)\top})^\top$ is a function of the VARMA-X parameter:

$$\boldsymbol{\zeta}_0 = \Psi(\boldsymbol{\varphi}_0) \qquad \text{where} \qquad \boldsymbol{\varphi}_0 = (\boldsymbol{\vartheta}_0^\top, \boldsymbol{\tau}^\top)^\top.$$

The following result is an immediate consequence of Theorem 3.1 and Lemma 3.1.

Corollary 3.1 *Let* $\widehat{\boldsymbol{\zeta}}_n = \Psi(\widehat{\boldsymbol{\varphi}}_n)$, *with* $\widehat{\boldsymbol{\varphi}}_n = \left(\widehat{\boldsymbol{\vartheta}}_n^\top, \widehat{\boldsymbol{\tau}}_n^\top\right)^\top$, *and* $\widehat{\boldsymbol{\tau}}_n = \left(\widehat{\tau}_{1n}^\top, \ldots, \widehat{\tau}_{Mn}^\top\right)^\top$. *Under the assumptions of Lemma 3.1,* $\widehat{\boldsymbol{\zeta}}_n$ *is a strongly consistent estimator of* $\boldsymbol{\zeta}_0$.

3.2 Asymptotic normality

We now show the asymptotic normality of the equation-by-equation estimator $\widehat{\boldsymbol{\vartheta}}_n$.

Theorem 3.2 *Suppose that the assumptions of Lemma 3.1 hold and that* $\boldsymbol{\vartheta}_0^{(j)}$ *belongs to the interior of* Θ *for* $j = 1, \ldots, M$. *As* $n \to \infty$, *we have*

$$\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right),$$

where $\boldsymbol{J}$ *is a block diagonal matrix with* j*-th* $d \times d$ *block* $\boldsymbol{J}^{(j)} = \mathrm{E}\frac{\partial u_{jt}}{\partial \underline{\boldsymbol{\vartheta}}}\frac{\partial u_{jt}}{\partial \underline{\boldsymbol{\vartheta}}^\top}(\boldsymbol{\vartheta}_0^{(j)})$ *and* $\boldsymbol{I}$ *is a matrix whose block* (i,j) *of size* $d \times d$ *is* $\boldsymbol{I}(i,j) = \mathrm{E}u_{it}u_{jt}\mathrm{E}\frac{\partial u_{it}}{\partial \underline{\boldsymbol{\vartheta}}}(\boldsymbol{\vartheta}_0^{(i)})\frac{\partial u_{jt}}{\partial \underline{\boldsymbol{\vartheta}}^\top}(\boldsymbol{\vartheta}_0^{(j)})$.

Note that the theorem implies

$$\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)} - \boldsymbol{\vartheta}_0^{(j)}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left\{\mathbf{0}, \mathrm{E}u_{jt}^2\left(\boldsymbol{J}^{(j)}\right)^{-1}\right\} \tag{3.9}$$

as $n \to \infty$, for $j = 1, \ldots, M$. The matrices $\boldsymbol{J}^{(j)}$ and $\boldsymbol{I}(i,j)$ can be estimated by

$$\widehat{\boldsymbol{J}}^{(j)} = \frac{1}{n}\sum_{t=1}^{n}\widehat{\boldsymbol{\Upsilon}}_{jt}\widehat{\boldsymbol{\Upsilon}}_{jt}^\top, \quad \widehat{\boldsymbol{I}}(i,j) = \frac{1}{n}\sum_{t=1}^{n}\widehat{u}_{it}\widehat{u}_{jt}\widehat{\boldsymbol{\Upsilon}}_{it}\widehat{\boldsymbol{\Upsilon}}_{jt}^\top, \quad \text{with } \widehat{\boldsymbol{\Upsilon}}_{jt} = \frac{\partial \widetilde{u}_{jt}}{\partial \underline{\boldsymbol{\vartheta}}}(\widehat{\boldsymbol{\vartheta}}_n^{(j)}),$$

the $\widehat{\boldsymbol{\Upsilon}}_{jt}$ being recursively computed, for $t = 2, \ldots, n$, by

$$\widehat{\boldsymbol{\Upsilon}}_{jt} = \widetilde{\boldsymbol{d}}_{j,t-1} + \widehat{\beta}_n^{(j)}\widehat{\boldsymbol{\Upsilon}}_{j,t-1}, \tag{3.10}$$

with $\widehat{\beta}_n^{(j)}$ being the estimate of β_{0jj} and

$$\widetilde{\boldsymbol{d}}_{j,t-1} = (-1, -\ln\epsilon_{1,t-1}^2, \ldots, -\ln\epsilon_{M,t-1}^2, \widehat{u}_{j,t-1}, -X_{1,t-1}, \ldots, -X_{K,t-1})^\top$$

and the initial value $\widehat{\boldsymbol{\Upsilon}}_{j1} = \mathbf{0}_d$.

Proposition 3.1 *Under the assumptions of Theorem 3.2, for all* $i, j = 1, \ldots, M$ *we have*

$$\widehat{\boldsymbol{J}}^{(j)} \to \boldsymbol{J} \quad \text{ and } \quad \widehat{\boldsymbol{I}}(i,j) \to \boldsymbol{I}(i,j) \quad \text{ a.s. as } n \to \infty.$$

We now give the asymptotic distribution of all the VARMA-X parameters $\widehat{\boldsymbol{\varphi}}_n = (\widehat{\boldsymbol{\vartheta}}_0, \widehat{\boldsymbol{\tau}})^\top$.

Theorem 3.3 *Under the assumptions of Theorem 3.2 and Lemma 3.1*

$$\sqrt{n}\left(\widehat{\boldsymbol{\varphi}}_n - \boldsymbol{\varphi}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varphi}}\right),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\varphi}}$ *can be estimated by* $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varphi}} = \widehat{\boldsymbol{M}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Upsilon}}\widehat{\boldsymbol{M}}^\top$ *with*

$$\widehat{\boldsymbol{M}} = \begin{pmatrix} -\widehat{\boldsymbol{J}}^{-1} & \mathbf{0}_{Md\times M} \\ \widehat{\boldsymbol{D}}\widehat{\boldsymbol{J}}^{-1} & \widehat{\boldsymbol{E}} \end{pmatrix}, \quad \widehat{\boldsymbol{D}} = \begin{pmatrix} \widehat{\boldsymbol{D}}_1^\top & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \widehat{\boldsymbol{D}}_M^\top \end{pmatrix} \quad \text{with } \widehat{\boldsymbol{D}}_j = \frac{1}{n}\sum_{t=1}^{n} \widehat{\boldsymbol{\Upsilon}}_{jt},$$

$\widehat{\boldsymbol{E}} = -\mathrm{diag}(e^{\widehat{\tau}_1}, \dots, e^{\widehat{\tau}_M})$ *and* $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\Upsilon}} = \frac{1}{n}\sum_{t=1}^{n} \widehat{\boldsymbol{\Upsilon}}_t \widehat{\boldsymbol{\Upsilon}}_t^\top$ *with*

$$\widehat{\boldsymbol{\Upsilon}}_t = \left(\widehat{u}_{1t}\widehat{\boldsymbol{\Upsilon}}_{1t}^\top, \cdots, \widehat{u}_{Mt}\widehat{\boldsymbol{\Upsilon}}_{Mt}^\top, e^{\widehat{u}_{1t}} - e^{-\widehat{\tau}_1}, \cdots, e^{\widehat{u}_{Mt}} - e^{-\widehat{\tau}_M}\right)^\top.$$

To deduce the asymptotic distribution of $\widehat{\boldsymbol{\zeta}}_n$ from that of $\widehat{\boldsymbol{\varphi}}_n$ we need to compute the $Md \times M(d+1)$ matrix $\frac{\partial \Psi(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top}$. For instance, in the case $K = 1$ and $M = 2$, we have

$$\frac{\partial \Psi(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top} = \begin{pmatrix}
1 & 0 & 0 & \tau_1 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_{011}-1 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \tau_2 & 0 & 0 & \beta_{022}-1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}.$$

More generally, for $j = 1, \dots, M$, the line $d(j-1)+1$ of that matrix is given by

$$\frac{\partial \Psi_{d(j-1)+1}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top} = \begin{pmatrix} \mathbf{0}_{d(j-1)}^\top & 1 & \mathbf{0}_M^\top & \tau_j & \mathbf{0}_K^\top & \mathbf{0}_{d(M-j)}^\top & \mathbf{0}_{j-1}^\top & \beta_{0jj}-1 & \mathbf{0}_{M-j}^\top \end{pmatrix},$$

the line $d(j-1)+1+j$ is given by

$$\frac{\partial \Psi_{d(j-1)+1+j}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top} = \begin{pmatrix} \mathbf{0}_{d(j-1)}^\top & \mathbf{0}_j^\top & 1 & \mathbf{0}_{M-j}^\top & -1 & \mathbf{0}_K^\top & \mathbf{0}_{d(M-j)}^\top & \mathbf{0}_M^\top \end{pmatrix},$$

the line $d(j-1)+1+k$ for $k = 1, \dots, j-1$ or $k = j+1, \dots, M$ is given by

$$\frac{\partial \Psi_{d(j-1)+1+k}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top} = \begin{pmatrix} \mathbf{0}_{d(j-1)}^\top & \mathbf{0}_k^\top & 1 & \mathbf{0}_{M-k}^\top & 0 & \mathbf{0}_K^\top & \mathbf{0}_{d(M-j)}^\top & \mathbf{0}_M^\top \end{pmatrix},$$

the line $d(j-1)+M+2$ is given by

$$\frac{\partial \Psi_{d(j-1)+M+2}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top} = \begin{pmatrix} \mathbf{0}_{d(j-1)}^\top & \mathbf{0}_{M+1}^\top & 1 & \mathbf{0}_K^\top & \mathbf{0}_{d(M-j)}^\top & \mathbf{0}_M^\top \end{pmatrix},$$

and the line $d(j-1)+M+2+k$ for $k=1,\dots,K$ is given by

$$\frac{\partial \Psi_{d(j-1)+M+2+k}(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top} = \begin{pmatrix} \mathbf{0}_{d(j-1)}^\top & \mathbf{0}_{M+k+1}^\top & 1 & \mathbf{0}_{K-k}^\top & \mathbf{0}_{d(M-j)}^\top & \mathbf{0}_M^\top \end{pmatrix}.$$

A direct application of the delta method then gives the following result.

Corollary 3.2 *Under the assumptions of Theorem 3.3*

$$\sqrt{n}\left(\widehat{\boldsymbol{\zeta}}_n - \boldsymbol{\zeta}_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\zeta}} := \frac{\partial \Psi(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}^\top} \boldsymbol{\Sigma}_{\boldsymbol{\varphi}} \frac{\partial \Psi^\top(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}}\right),$$

as $n \to \infty$.

3.3 Constrained models

The asymptotic results of the previous section can readily be used to test the significance of the log-GARCH parameters (see Section 5.1 for an illustration). Such tests may lead to the estimation of new models in which some coefficients are constrained to be zero. Similarly, even if estimation of a large system of dimension M is performed equation-by-equation, a general model of the form (2.3) remains intractable when M is large (in each equation, the number of parameters is $M+K+2$). For these reasons, one could want to impose constraints, such as a diagonal or block-diagonal form for the matrix $\boldsymbol{\alpha}_0$ (see Section 5.2 for an illustration).

A way to introduce very general constraints is to assume that the VARMA-X model (2.4) is parameterised by a vector $\boldsymbol{\vartheta}_0$, that does not necessarily correspond to the parameter defined in Section 3.1, and may be of lower dimension. We thus assume that $\boldsymbol{c}_0 = \boldsymbol{c}(\boldsymbol{\vartheta}_0)$, $\boldsymbol{\phi}_0 = \boldsymbol{\phi}(\boldsymbol{\vartheta}_0)$, $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\vartheta}_0)$ and $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}(\boldsymbol{\vartheta}_0)$, and that $\boldsymbol{\vartheta}_0 = \left(\boldsymbol{\vartheta}_0^{(1)\top}, \dots, \boldsymbol{\vartheta}_0^{(M)\top}\right)^\top$, where $\boldsymbol{\vartheta}_0^{(j)}$ is the vector of the unknown parameters involved in the j-th VARMA-X equation. Parameter $\boldsymbol{\vartheta}_0^{(j)}$ belongs to some compact parameter space $\Theta^{(j)}$, whose generic element $\underline{\boldsymbol{\vartheta}}$ has typically less than $M+K+2$ components. If, for instance, matrix $\boldsymbol{\alpha}_0$ is assumed to be diagonal, then one can set

$$\underline{\boldsymbol{\vartheta}} = (c_j, \alpha_{jj}+\beta_{jj}, \beta_{jj}, \lambda_{j1}, \dots, \lambda_{jK})^\top, \qquad \Theta^{(j)} = \Theta \subset \mathbb{R}^{K+3}.$$

In the general case, it is assumed that the parameterisation satisfies

$$\text{the application } \boldsymbol{\vartheta} \mapsto \{\boldsymbol{c}(\boldsymbol{\vartheta}), \boldsymbol{\phi}(\boldsymbol{\vartheta}), \boldsymbol{\beta}(\boldsymbol{\vartheta}), \boldsymbol{\lambda}(\boldsymbol{\vartheta})\} \text{ is injective} \tag{3.11}$$

and

$$\text{admits continuous third order derivatives in } \Theta^{(1)} \times \dots \times \Theta^{(M)}. \tag{3.12}$$

Assume also that, denoting by $\beta_{jj}(\boldsymbol{\vartheta})$ the j-th diagonal term of $\boldsymbol{\beta}(\boldsymbol{\vartheta})$,

$$|\beta_{jj}(\boldsymbol{\vartheta})| < 1, \qquad \forall \boldsymbol{\vartheta} \in \Theta^{(1)} \times \cdots \times \Theta^{(M)}. \tag{3.13}$$

With this change of notation, the estimator of $\boldsymbol{\vartheta}_0^{(j)}$ can still be defined by (3.7), replacing Θ by $\Theta^{(j)}$ if necessary. The asymptotic behaviour of $\widehat{\boldsymbol{\vartheta}}_n^{(j)}$ is unchanged.

More precisely, the strong consistency of $\widehat{\boldsymbol{\vartheta}}_n^{(j)}$ to $\boldsymbol{\vartheta}_0^{(j)}$ holds true under the previous assumptions (3.1), (3.2), (3.3), (3.4), (3.5), the invertibility condition (3.13) and the identifiability conditions (3.11) and $\alpha_{jj}(\boldsymbol{\vartheta}_0) \neq 0$. The asymptotic normality of Theorem 3.2 continues to hold under the additional assumption that $\boldsymbol{\vartheta}_0^{(j)}$ belongs to the interior of $\Theta^{(j)}$ for all j, and under the smoothness condition (3.12). The output of Proposition 3.1 also remains valid if, in (3.10), the definition of $\widetilde{\boldsymbol{d}}_{j,t-1}$ is changed into

$$\begin{aligned}\widetilde{\boldsymbol{d}}_{j,t-1} = & -\frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} c_j\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}\right) - \sum_{\ell=1}^{M} \ln \epsilon_{\ell,t-1}^2 \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} \phi_{j\ell}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}\right) \\ & - \sum_{\ell=1}^{K} X_{\ell,t-1} \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} \lambda_{j\ell}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}\right) + \widetilde{u}_{j,t-1}(\widehat{\boldsymbol{\vartheta}}_n^{(j)}) \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} \beta_{jj}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}\right).\end{aligned}$$

With this modification, Theorem 3.3 directly applies.

As an illustration, consider for instance the case where $\boldsymbol{\alpha}_0$ is assumed to be diagonal. Moreover, assume that one wants to estimate the model with $K = 2$ covariates under the constraint $\lambda_{j1} = \lambda_{j2}$ for all j. The condition (3.11)-(3.12) is satisfied with $\underline{\boldsymbol{\vartheta}} = (c_j, \phi_{jj}, \beta_{jj}, \lambda_{j1}) = (c, \phi_j, \beta, \lambda_1)$. If $\Theta^{(j)}$ is assumed to be a compact subset of $\mathbb{R}^2 \times (-1,1) \times \mathbb{R}$, the condition (3.13) also holds true. In particular, we have (3.9) where $\boldsymbol{J}^{(j)}$ is a 4×4 matrix which can be consistently estimated as in Proposition 3.1 with

$$\widetilde{\boldsymbol{d}}_{j,t-1} = (-1, -\ln \epsilon_{j,t-1}^2, \widehat{u}_{j,t-1}, -X_{1,t-1} - X_{2,t-1})^\top.$$

4 Proofs

We first state a lemma that will be used to show the identifiability of the parameters in model (2.5).

Lemma 4.1 *Assume (3.2), (3.3) and (3.4). If for some non-random vectors $\boldsymbol{\nu}_1$ of $\mathbb{R}^M$ and $\boldsymbol{\nu}_2$ of $\mathbb{R}^K$, and for some random variable $\nu_{3,t-1} \in \mathcal{H}_{t-1}$ we have*

$$\boldsymbol{\nu}_1^\top \ln \boldsymbol{\epsilon}_t^2 + \boldsymbol{\nu}_2^\top \ln \boldsymbol{X}_t + \nu_{3,t-1} = 0 \qquad a.s. \tag{4.1}$$

then $\boldsymbol{\nu}_1 = \boldsymbol{0}_M$, $\boldsymbol{\nu}_2 = \boldsymbol{0}_K$, and $\nu_{3,t-1} = 0$ almost surely.

Proof of Lemma 4.1. Subtracting the mean, conditionally to $\mathcal{H}_{t-1}$, on both sides of equality (4.1), we obtain $\boldsymbol{\nu}_1^\top \boldsymbol{u}_t + \boldsymbol{\nu}_2^\top \boldsymbol{v}_t = 0$ a.s., which entails $\boldsymbol{\nu}_1 = \boldsymbol{0}_M$ and $\boldsymbol{\nu}_2 = \boldsymbol{0}_K$ by

(3.4), and then that $\nu_{3,t-1} = 0$ a.s. $\square$

Proof of Theorem 3.1. We have

$$\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} |u_{jt}(\underline{\boldsymbol{\vartheta}}) - \widetilde{u}_{jt}(\underline{\boldsymbol{\vartheta}})| = \sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} \beta^{t-1} |u_{j1}(\underline{\boldsymbol{\vartheta}}) - \widetilde{u}_{j1}(\underline{\boldsymbol{\vartheta}})| \leq K\rho^t, \tag{4.2}$$

where, here and in the sequel of the paper, K denotes a generic positive random variable which is $\mathcal{F}_0$-measurable, and ρ denotes a generic constant belonging to $[0,1)$. Note that $K\rho^t$ tends almost surely to zero as $t\to\infty$ wihout any moment assumption on K because $\Pr\{\omega\in\Omega : \lim_{t\to\infty} K(\omega)\rho^t = 0\} = 1$ for any real random variable defined on some probability space $(\Omega, \mathcal{A}, \Pr)$. Letting $Q_n^{(j)}(\underline{\boldsymbol{\vartheta}}) = n^{-1}\sum_{t=1}^n u_{jt}^2(\underline{\boldsymbol{\vartheta}})$, we then have

$$\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} \left| Q_n^{(j)}(\underline{\boldsymbol{\vartheta}}) - \widetilde{Q}_n^{(j)}(\underline{\boldsymbol{\vartheta}}) \right| \leq \frac{K}{n}\sum_{t=1}^{\infty} \rho^t \left(2\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} |u_{jt}(\underline{\boldsymbol{\vartheta}})| + K\rho^t \right) = O\left(\frac{1}{n}\right) \text{ a.s.} \tag{4.3}$$

For the last equality, we use the fact that, under the moments conditions and the compactness assumption, $\left\|\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} |u_{jt}(\underline{\boldsymbol{\vartheta}})|\right\|_2 < \infty$, and thus the L^2-norm of the sum is finite, which entails that the sum is finite almost surely.

By the ergodic theorem we have almost surely

$$\lim_{n\to\infty} Q_n^{(j)}(\underline{\boldsymbol{\vartheta}}) = \mathrm{E}u_{jt}^2(\underline{\boldsymbol{\vartheta}}) = \mathrm{E}u_{jt}^2 + \mathrm{E}\left\{u_{jt}(\underline{\boldsymbol{\vartheta}}) - u_{jt}\right\}^2,$$

because u_{jt} is uncorrelated with $u_{jt}(\underline{\boldsymbol{\vartheta}}) - u_{jt} \in \mathcal{F}_{t-1}$, under assumption (3.5). In view of (4.3), we thus have

$$\lim_{n\to\infty} \widetilde{Q}_n^{(j)}(\underline{\boldsymbol{\vartheta}}) = \lim_{n\to\infty} Q_n^{(j)}(\underline{\boldsymbol{\vartheta}}) \geq \lim_{n\to\infty} Q_n^{(j)}(\boldsymbol{\vartheta}_0^{(j)}) = \lim_{n\to\infty} \widetilde{Q}_n^{(j)}(\boldsymbol{\vartheta}_0^{(j)}) \quad \text{a.s.}$$

where the inequality is an equality if and only if $\Pr\{u_{jt}(\underline{\boldsymbol{\vartheta}}) = u_{jt}\} = 1$. The last equality is equivalent to

$$\left(\frac{1-\phi_{0jj}B}{1-\beta_{0jj}B} - \frac{1-\phi_j B}{1-\beta B}\right)\ln\epsilon_{jt}^2 - \left(\frac{c_{0j}}{1-\beta_{0jj}} - \frac{c}{1-\beta}\right)$$
$$-\sum_{\ell\neq j}\left(\frac{\phi_{0j\ell}B}{1-\beta_{0jj}B} - \frac{\phi_\ell B}{1-\beta B}\right)\ln\epsilon_{\ell t}^2 - \sum_{\ell=1}^{K}\left(\frac{\lambda_{0j\ell}}{1-\beta_{0jj}B} - \frac{\lambda_\ell}{1-\beta B}\right)X_{\ell,t-1} = 0 \quad \text{a.s.}$$

By Lemma 4.1, this entails that the four terms displayed in brackets are equal to zero. Under the condition $\phi_{0jl} \neq \beta_{0jj}$, this implies that $\underline{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta}_0^{(j)}$. We thus have shown that

$$\lim_{n\to\infty} \widetilde{Q}_n^{(j)}(\underline{\boldsymbol{\vartheta}}) > \lim_{n\to\infty} \widetilde{Q}_n^{(j)}(\boldsymbol{\vartheta}_0^{(j)}),$$

when $\underline{\boldsymbol{\vartheta}} \neq \boldsymbol{\vartheta}_0^{(j)}$. Using standard arguments (used, for instance, to show (d) on Page 157 in Francq and Zakoïan (2010)) the result can be extended to show that for any $\underline{\boldsymbol{\vartheta}} \neq \boldsymbol{\vartheta}_0^{(j)}$

there exists a neighborhood $V(\underline{\boldsymbol{\vartheta}})$ of $\underline{\boldsymbol{\vartheta}}$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\vartheta}^*\in V(\underline{\boldsymbol{\vartheta}})\cap\Theta} Q_n^{(j)}(\boldsymbol{\vartheta}^*) > \lim_{n\to\infty} Q_n^{(j)}(\boldsymbol{\vartheta}_0^{(j)}).$$

The conclusion then follows from a compactness argument. □

Proof of Lemma 3.1. We first study the effect of the initial values. A Taylor expansion and (4.2) show that

$$\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} \left| e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} - e^{\widetilde{u}_{jt}(\underline{\boldsymbol{\vartheta}})} \right| \le K\rho^t \sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} e^{u_{jt}(\underline{\boldsymbol{\vartheta}})}.$$

Since $\mathrm{E}\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} |u_{jt}(\underline{\boldsymbol{\vartheta}})| < \infty$, we have $\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} |u_{jt}(\underline{\boldsymbol{\vartheta}})|/t \to 0$ a.s. as $t\to\infty$. Therefore,

$$\limsup_{t\to\infty} \frac{1}{t} \ln \sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} \left| e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} - e^{\widetilde{u}_{jt}(\underline{\boldsymbol{\vartheta}})} \right| \le \ln\rho < 0,$$

and thus

$$\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} \left| e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} - e^{\widetilde{u}_{jt}(\underline{\boldsymbol{\vartheta}})} \right| < K\rho^t,$$

from which we deduce

$$\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} \left| \frac{1}{n}\sum_{t=1}^n e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} - \frac{1}{n}\sum_{t=1}^n e^{\widetilde{u}_{jt}(\underline{\boldsymbol{\vartheta}})} \right| = O\left(\frac{1}{n}\right) \text{ a.s.} \tag{4.4}$$

Now note that

$$\frac{\partial e^{u_{jt}(\underline{\boldsymbol{\vartheta}})}}{\partial\underline{\boldsymbol{\vartheta}}} = e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} \frac{\partial u_{jt}(\underline{\boldsymbol{\vartheta}})}{\partial\underline{\boldsymbol{\vartheta}}}, \qquad \frac{\partial u_{jt}(\underline{\boldsymbol{\vartheta}})}{\partial\underline{\boldsymbol{\vartheta}}} = \boldsymbol{d}_{t-1}(\underline{\boldsymbol{\vartheta}}) + \beta\frac{\partial u_{j,t-1}(\underline{\boldsymbol{\vartheta}})}{\partial\underline{\boldsymbol{\vartheta}}}, \tag{4.5}$$

with

$$\boldsymbol{d}_{t-1}(\underline{\boldsymbol{\vartheta}}) = (-1, -\ln\epsilon_{1,t-1}^2, \ldots, -\ln\epsilon_{M,t-1}^2, u_{j,t-1}(\underline{\boldsymbol{\vartheta}}), -X_{1,t-1}, \ldots, -X_{K,t-1})^\top. \tag{4.6}$$

Moreover, we have

$$\left\| \frac{\partial e^{u_{jt}(\underline{\boldsymbol{\vartheta}})}}{\partial\underline{\boldsymbol{\vartheta}}} - \frac{\partial e^{u_{jt}(\boldsymbol{\vartheta}_0^{(j)})}}{\partial\underline{\boldsymbol{\vartheta}}} \right\| \le \left| e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} - e^{u_{jt}(\boldsymbol{\vartheta}_0^{(j)})} \right| \left\| \frac{\partial u_{jt}(\underline{\boldsymbol{\vartheta}})}{\partial\underline{\boldsymbol{\vartheta}}} \right\| + e^{u_{jt}(\boldsymbol{\vartheta}_0^{(j)})} \left\| \frac{\partial u_{jt}(\underline{\boldsymbol{\vartheta}})}{\partial\underline{\boldsymbol{\vartheta}}} - \frac{\partial u_{jt}(\boldsymbol{\vartheta}_0^{(j)})}{\partial\underline{\boldsymbol{\vartheta}}} \right\|.$$

The compactness of Θ, the fact that $\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta}|\beta| < 1$ and (3.3) entail that

$$\mathrm{E}\sup_{\underline{\boldsymbol{\vartheta}}\in\Theta} \left\| \frac{\partial u_{jt}(\underline{\boldsymbol{\vartheta}})}{\partial\underline{\boldsymbol{\vartheta}}} \right\|^2 < \infty. \tag{4.7}$$

By Lemma 2.1 and Remark 2.1 in Francq and Sucarrat (2013), it can be shown that there exists a neighbourhood $V(\boldsymbol{\vartheta}_0^{(j)})$ of $\boldsymbol{\vartheta}_0^{(j)}$ such that

$$\mathrm{E} \sup_{\underline{\boldsymbol{\vartheta}} \in V(\boldsymbol{\vartheta}_0^{(j)})} \left| e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} \right|^2 < \infty. \tag{4.8}$$

Let $V_k(\boldsymbol{\vartheta}_0^{(j)})$ be the ball of centre $\boldsymbol{\vartheta}_0^{(j)}$ and radius $1/k$. The dominated convergence theorem and (4.8) imply

$$\lim_{k\to\infty} \mathrm{E} \sup_{\underline{\boldsymbol{\vartheta}} \in V_k(\boldsymbol{\vartheta}_0^{(j)}) \cap \Theta} \left| e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} - e^{u_{jt}(\boldsymbol{\vartheta}_0^{(j)})} \right|^2 = 0. \tag{4.9}$$

The Lebesgue dominated convergence theorem also shows that

$$\lim_{k\to\infty} \mathrm{E} \sup_{\underline{\boldsymbol{\vartheta}} \in V_k(\boldsymbol{\vartheta}_0^{(j)}) \cap \Theta} \left\| \frac{\partial u_{jt}(\underline{\boldsymbol{\vartheta}})}{\partial \underline{\boldsymbol{\vartheta}}} - \frac{\partial u_{jt}(\boldsymbol{\vartheta}_0^{(j)})}{\partial \underline{\boldsymbol{\vartheta}}} \right\| = 0. \tag{4.10}$$

Note that (2.1)-(2.2) entail that $\mathrm{E}(\eta_{jt}^2) = 1$, and thus $\mathrm{E}e^{u_{jt}} = e^{-\tau_j}$. Using (4.7) and (4.9) with the Cauchy-Schwarz inequality, and (4.10) with (3.5) and $\mathrm{E}e^{u_{jt}} = e^{-\tau_j}$, we finally obtain

$$\lim_{k\to\infty} \mathrm{E} \sup_{\underline{\boldsymbol{\vartheta}} \in V_k(\boldsymbol{\vartheta}_0^{(j)}) \cap \Theta} \left\| \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} e^{u_{jt}(\boldsymbol{\vartheta}_0^{(j)})} - \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} e^{u_{jt}(\underline{\boldsymbol{\vartheta}})} \right\| = 0. \tag{4.11}$$

Note also that, by (3.5), $u_{jt} = u_{jt}(\boldsymbol{\vartheta}_0^{(j)})$ is independent of $\mathcal{F}_{t-1}$. In view of the strong consistency established in Theorem 3.1, it follows from (4.11) that for any sequence $\boldsymbol{\vartheta}_n$ between $\widehat{\boldsymbol{\vartheta}}_n^{(j)}$ and $\boldsymbol{\vartheta}_0^{(j)}$, almost surely

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} e^{u_{jt}(\boldsymbol{\vartheta}_n)} = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} e^{u_{jt}(\boldsymbol{\vartheta}_0^{(j)})} = e^{-\tau_j} \mathrm{E} \frac{\partial u_{jt}(\boldsymbol{\vartheta}_0^{(j)})}{\partial \underline{\boldsymbol{\vartheta}}}. \tag{4.12}$$

Now it suffices to point out that (4.4) and a Taylor expansion entail

$$\frac{1}{n} \sum_{t=1}^{n} e^{\widehat{u}_{jt}} = \frac{1}{n} \sum_{t=1}^{n} e^{u_{jt}} + \left(\widehat{\boldsymbol{\vartheta}}_n^{(j)} - \boldsymbol{\vartheta}_0^{(j)} \right)^\top \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} e^{u_{jt}(\boldsymbol{\vartheta}_n)} + O\left(\frac{1}{n}\right) = e^{-\tau_1} + o(1).$$

□

Proof of Theorem 3.2. Similarly to (4.3), it can be seen that

$$\sup_{\underline{\boldsymbol{\vartheta}} \in \Theta} \left\| \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} Q_n^{(j)}(\underline{\boldsymbol{\vartheta}}) - \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} \widetilde{Q}_n^{(j)}(\underline{\boldsymbol{\vartheta}}) \right\| = O\left(\frac{1}{n}\right) \quad \text{a.s.} \tag{4.13}$$

By arguments that are similar to those used to prove (e) on Page 174 in Francq and Zakoïan (2010), one can show that for any sequence $\boldsymbol{\vartheta}_n$ between $\widehat{\boldsymbol{\vartheta}}_n^{(j)}$ and $\boldsymbol{\vartheta}_0^{(j)}$, we have

$$\lim_{n\to\infty} \frac{\partial^2}{\partial \underline{\boldsymbol{\vartheta}} \partial \underline{\boldsymbol{\vartheta}}^\top} Q_n^{(j)}(\boldsymbol{\vartheta}_n) = 2\boldsymbol{J}^{(j)} \quad \text{a.s.}$$

The existence of $\boldsymbol{J}^{(j)}$ comes from (4.5)-(4.6) and (3.2)-(3.3). Therefore, by using a Taylor expansion, we obtain

$$o(1) = \sqrt{n}\frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} Q_n^{(j)}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}\right) = \sqrt{n}\frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} Q_n^{(j)}\left(\boldsymbol{\vartheta}_0^{(j)}\right) + \sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)} - \boldsymbol{\vartheta}_0^{(j)}\right)\left\{2\boldsymbol{J}^{(j)} + o(1)\right\}.$$

The central limit theorem of Billingsley (1961) for stationary and square integrable martingale differences entails that

$$\sqrt{n}\begin{pmatrix} \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} Q_n^{(1)}\left(\boldsymbol{\vartheta}_0^{(1)}\right) \\ \vdots \\ \frac{\partial}{\partial \underline{\boldsymbol{\vartheta}}} Q_n^{(M)}\left(\boldsymbol{\vartheta}_0^{(M)}\right) \end{pmatrix} = \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix} u_{1t}\frac{\partial u_{1t}}{\partial \underline{\boldsymbol{\vartheta}}} \\ \vdots \\ u_{Mt}\frac{\partial u_{Mt}}{\partial \underline{\boldsymbol{\vartheta}}} \end{pmatrix}_{\boldsymbol{\vartheta}_0} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, 4\boldsymbol{I}\right)$$

as $n \to \infty$. It remains to show that the matrices $\boldsymbol{J}^{(j)}$ are nonsingular. If $\boldsymbol{J}^{(j)}$ is singular, then there exists a non-zero vector $\boldsymbol{\nu}$ of $\mathbb{R}^d$, such that $\boldsymbol{\nu}^\top \frac{\partial u_{jt}}{\partial \underline{\boldsymbol{\vartheta}}} = 0$ a.s. By (4.5), this entails $\boldsymbol{\nu}^\top \boldsymbol{d}_t(\boldsymbol{\vartheta}_0^{(j)}) = 0$ a.s. In view of Lemma 4.1, this is in contradiction with (3.4). The conclusion follows. □

Proof of Proposition 3.1. Using already-used arguments, it can be shown that the initial values are unimportant. More precisely, we have $\widehat{\boldsymbol{I}}(i,j) = \boldsymbol{I}_n(\widehat{\boldsymbol{\vartheta}}_n) + o(1)$ a.s., with

$$\boldsymbol{I}_n(\boldsymbol{\vartheta}) = \boldsymbol{I}_n^{i,j}(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{t=1}^{n} u_{it}(\boldsymbol{\vartheta}^{(i)})u_{jt}(\boldsymbol{\vartheta}^{(j)})\boldsymbol{\Upsilon}_{it}(\boldsymbol{\vartheta}^{(i)})\boldsymbol{\Upsilon}_{jt}^\top(\boldsymbol{\vartheta}^{(j)}), \quad \boldsymbol{\Upsilon}_{jt}(\boldsymbol{\vartheta}^{(j)}) = \frac{\partial u_{jt}}{\partial \underline{\boldsymbol{\vartheta}}}(\boldsymbol{\vartheta}^{(j)}).$$

By the ergodic theorem $\boldsymbol{I}_n(\boldsymbol{\vartheta}_0) \to \boldsymbol{I}(i,j)$ a.s. as $n \to \infty$. Since $\widehat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0$ a.s., it remains to show that for all $\varepsilon > 0$, there exists a neighbourhood $V(\boldsymbol{\vartheta}_0)$ of $\boldsymbol{\vartheta}_0$, such that

$$\lim_{n\to\infty} \sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)} \left\|\boldsymbol{I}_n(\boldsymbol{\vartheta}) - \boldsymbol{I}_n^{i,j}(\boldsymbol{\vartheta}_0)\right\| < \varepsilon. \tag{4.14}$$

Let $V_k(\boldsymbol{\vartheta}_0)$ be the ball of centre $\boldsymbol{\vartheta}_0$ and radius $1/k$. Note that

$$\sup_{\boldsymbol{\vartheta}\in V_k(\boldsymbol{\vartheta}_0)} \|\boldsymbol{I}_n(\boldsymbol{\vartheta}) - \boldsymbol{I}_n(\boldsymbol{\vartheta}_0)\| \leq \frac{1}{n}\sum_{t=1}^{n} x_t(k)$$

where

$$x_t(k) = \sup_{\boldsymbol{\vartheta}\in V_k(\boldsymbol{\vartheta}_0)} \|\boldsymbol{Y}_t(\boldsymbol{\vartheta}) - \boldsymbol{Y}_t(\boldsymbol{\vartheta}_0)\|, \quad \boldsymbol{Y}_t(\boldsymbol{\vartheta}) = u_{it}(\boldsymbol{\vartheta}^{(i)})u_{jt}(\boldsymbol{\vartheta}^{(j)})\boldsymbol{\Upsilon}_{it}(\boldsymbol{\vartheta}^{(i)})\boldsymbol{\Upsilon}_{jt}^\top(\boldsymbol{\vartheta}^{(j)}).$$

Since $u_{it}(\boldsymbol{\vartheta}^{(i)})$ and the components of $\boldsymbol{\Upsilon}_{it}(\boldsymbol{\vartheta}^{(i)})$ admit moments of order 2, uniformly in Θ (see (4.7)), we have

$$\mathrm{E}\sup_{\boldsymbol{\vartheta}\in\Theta^M} \|\boldsymbol{Y}_t(\boldsymbol{\vartheta})\| < \infty. \tag{4.15}$$

The process $\{x_t(k)\}_t$ being stationary and ergodic, the left-hand side of (4.14) is a.s. bounded by $\mathrm{E}x_t(k)$. Noting that $x_t(k) \to 0$ a.s. as $k \to \infty$, we obtain (4.14) by the dominated convergence theorem and (4.15). The consistency of $\widehat{\boldsymbol{I}}(i,j)$ is shown. That of $\widehat{\boldsymbol{J}}^{(j)}$ is obtained similarly. $\square$

Proof of Theorem 3.3. We first show that the initial values are asymptotically negligible. In view of (4.4) and the almost sure convergence of $\frac{1}{n}\sum_{t=1}^n e^{\widehat{u}_{jt}}$ to $e^{-\tau_1}>0$, a Taylor expansion shows that

$$\left|\widehat{\tau}_{jn}+\ln\frac{1}{n}\sum_{t=1}^n e^{u_{jt}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})}\right|\leq K\left|\frac{1}{n}\sum_{t=1}^n e^{u_{jt}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})}-\frac{1}{n}\sum_{t=1}^n e^{\widetilde{u}_{jt}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})}\right|=o\left(n^{-1/2}\right)\quad\text{a.s.}$$

Doing again a Taylor expansion, we obtain

$$\ln\frac{1}{n}\sum_{t=1}^n e^{u_{jt}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})}=\ln\frac{1}{n}\sum_{t=1}^n e^{u_{jt}(\boldsymbol{\vartheta}_0^{(j)})}+\frac{1}{\frac{1}{n}\sum_{t=1}^n e^{u_{jt}(\boldsymbol{\vartheta}_n)}}\frac{1}{n}\sum_{t=1}^n\frac{\partial}{\partial\underline{\boldsymbol{\vartheta}}^\top}e^{u_{jt}(\boldsymbol{\vartheta}_n)}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}-\boldsymbol{\vartheta}_0^{(j)}\right)$$

for some sequence $\boldsymbol{\vartheta}_n$ between $\widehat{\boldsymbol{\vartheta}}_n^{(j)}$ and $\boldsymbol{\vartheta}_0^{(j)}$. In view of (4.12) and Theorem 3.2, it follows that

$$\sqrt{n}\left(\widehat{\tau}_{jn}-\tau_j\right)=\sqrt{n}\left(-\ln\frac{1}{n}\sum_{t=1}^n e^{u_{jt}}-\tau_j\right)-\boldsymbol{D}_j^\top\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}-\boldsymbol{\vartheta}_0^{(j)}\right)+o_{\mathrm{Pr}}(1),$$

where $\boldsymbol{D}_j=\mathrm{E}\frac{\partial u_{j1}(\boldsymbol{\vartheta}_0^{(j)})}{\partial\underline{\boldsymbol{\vartheta}}}$. Now, using Lemma 3.1 we have

$$\ln\frac{1}{n}\sum_{t=1}^n e^{u_{jt}}=-\tau_j+\frac{1}{e^{-\tau_j}+o(1)}\frac{1}{n}\sum_{t=1}^n\left(e^{u_{jt}}-e^{-\tau_j}\right).$$

Putting the results together, we obtain

$$\sqrt{n}\left(\widehat{\tau}_{jn}-\tau_j\right)=-e^{\tau_j}\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(e^{u_{jt}}-e^{-\tau_j}\right)-\boldsymbol{D}_j^\top\sqrt{n}\left(\widehat{\boldsymbol{\vartheta}}_n^{(j)}-\boldsymbol{\vartheta}_0^{(j)}\right)+o_{\mathrm{Pr}}(1).$$

From the proof of Theorem 3.2, it follows that

$$\sqrt{n}\left(\widehat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}_0\right)=\boldsymbol{M}\frac{1}{\sqrt{n}}\sum_{t=1}^n\boldsymbol{\Upsilon}_t+o_{\mathrm{Pr}}(1),$$

where

$$\boldsymbol{\Upsilon}_t=\begin{pmatrix}u_{1t}\frac{\partial u_{1t}}{\partial\underline{\boldsymbol{\vartheta}}}(\boldsymbol{\vartheta}^{(1)})\\ \vdots\\ u_{Mt}\frac{\partial u_{Mt}}{\partial\underline{\boldsymbol{\vartheta}}}(\boldsymbol{\vartheta}^{(M)})\\ e^{u_{1t}}-e^{-\tau_1}\\ \vdots\\ e^{u_{Mt}}-e^{-\tau_M}\end{pmatrix},\quad\boldsymbol{M}=\begin{pmatrix}-\boldsymbol{J}^{-1}&\boldsymbol{0}_{Md\times M}\\ \boldsymbol{D}\boldsymbol{J}^{-1}&\boldsymbol{E}\end{pmatrix},\quad\boldsymbol{D}=\begin{pmatrix}\boldsymbol{D}_1^\top&&\boldsymbol{0}\\&\ddots&\\ \boldsymbol{0}&&\boldsymbol{D}_M^\top\end{pmatrix}$$

and $\boldsymbol{E} = -\mathrm{diag}(e^{\tau_1}, \ldots, e^{\tau_M})$. By already-given arguments, $\frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{\Upsilon}_t \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}})$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_{\boldsymbol{\vartheta},\tau} \\ \boldsymbol{\Sigma}_{\boldsymbol{\vartheta},\tau}^\top & \boldsymbol{\Sigma}_\tau \end{pmatrix}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\vartheta},\tau} = \begin{pmatrix} \mathrm{cov}(u_{1t}, e^{u_{1t}})\boldsymbol{D}_1 & \cdots & \mathrm{cov}(u_{1t}, e^{u_{Mt}})\boldsymbol{D}_1 \\ & \vdots & \\ \mathrm{cov}(u_{Mt}, e^{u_{1t}})\boldsymbol{D}_M & \cdots & \mathrm{cov}(u_{Mt}, e^{u_{Mt}})\boldsymbol{D}_M \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}_\tau = \left(\ \mathrm{cov}(e^{u_{it}}, e^{u_{jt}}) \ \right).$$

□

5 Empirical applications

5.1 Volatility spillover

How, and to what extent, volatility in one financial market may spill over to others is of importance for both policymakers and investors. In ordinary multivariate GARCH models, the study of volatility spillover rests upon complicated conditions on the parameters, and on the restrictive assumption that the conditional correlations are constant, see Conrad and Weber (2013), and Pedersen (2015). By contrast, in our model, complicated assumptions on the parameters are not required (due to the exponential volatility specification), and our tests are valid under time-varying conditional correlations of unknown form. Indeed, a variety of tests can be undertaken either equation-by-equation, or jointly for the whole system. Furthermore, we can also test whether covariates, e.g., volatility proxies, provide additional – or alternative – channels of volatility spillover.

We illustrate the flexibility and generality of the results from the previous section in a study of how the stock market volatilities of Europe and the US affect each other. For illustration purposes, we restrict our attention to only two indices, the FTSE 100 and the Standard and Poor's 100 (SP100), and initially we only include two variables in the X-part: The log of a range-based volatility proxy for FTSE 100 and SP100, respectively. The range, i.e., the difference between the maximum and minimum prices, is often available, and constitutes a volatility proxy, see, e.g., Parkinson (1980) and Garman and Klass (1980). The log of our range-based volatility-proxy is computed as $\ln\left[(high_t - low_t)\cdot 100\right]^2$, where $high_t$ is the natural log of the maximum value during day t, and where low_t is the natural log of the minimum value during day t. The source of our data is Bloomberg and goes from 2 January 1998 to 1 June 2015, a total 4297 observations before differencing and lagging. Initially, before we add more variables to the X-part, all of our estimated models will be contained in the two-dimensional log-GARCH(1,1)-X model

$$\begin{aligned} \ln\sigma_{1,t}^2 &= \omega_{01} + \alpha_{011}\ln\epsilon_{1,t-1}^2 + \alpha_{012}\ln\epsilon_{2,t-1}^2 + \beta_{01}\ln\sigma_{1,t-1}^2 \\ &\quad +\lambda_{011}X_{1,t-1} + \lambda_{012}X_{2,t-1}, \qquad (5.1) \\ \ln\sigma_{2,t}^2 &= \omega_{02} + \alpha_{021}\ln\epsilon_{1,t-1}^2 + \alpha_{022}\ln\epsilon_{2,t-1}^2 + \beta_{02}\ln\sigma_{2,t-1}^2 \\ &\quad +\lambda_{021}X_{1,t-1} + \lambda_{022}X_{2,t-1}, \qquad (5.2) \end{aligned}$$

where $\epsilon_{1,t}$ and $\epsilon_{2,t}$ denote daily European and US return (in percent), respectively, at day t, and $X_{1,t}$ and $X_{2,t}$ are the logs of the volatility proxies at day t. The VARMA-X representation of this model is

$$\begin{aligned}
\ln \epsilon_{1,t}^2 &= c_{01} + \phi_{011} \ln \epsilon_{1,t-1}^2 + \alpha_{012} \ln \epsilon_{2,t-1}^2 - \beta_{01} u_{1,t-1} \\
&\quad + \lambda_{011} X_{1,t-1} + \lambda_{012} X_{2,t-1} + u_{1,t}, \qquad (5.3) \\
\ln \epsilon_{2,t}^2 &= c_{02} + \alpha_{021} \ln \epsilon_{1,t-1}^2 + \phi_{022} \ln \epsilon_{2,t-1}^2 - \beta_{02} u_{2,t-1} \\
&\quad + \lambda_{021} X_{1,t-1} + \lambda_{022} X_{2,t-1} + u_{2,t}, \qquad (5.4)
\end{aligned}$$

where, for equation $j = 1, 2$,

$$c_{0j} = \omega_{0j} + (1 - \beta_{0j}) \mathrm{E}(\ln \eta_{j,t}^2), \quad \phi_{0jj} = \alpha_{0jj} + \beta_{0j} \quad \text{and} \quad u_{j,t} = \ln \eta_{j,t}^2 - \mathrm{E}(\ln \eta_{j,t}^2).$$

We start by estimating two univariate log-GARCH(1,1) models for comparison purposes:

$$\ln \widehat{\sigma}_{1,t}^2 = 0.066 + \underset{(0.006)}{0.047} \ln \epsilon_{1,t-1}^2 + \underset{(0.008)}{0.943} \ln \widehat{\sigma}_{1,t-1}^2, \qquad (5.5)$$

$$\ln \widehat{\sigma}_{2,t}^2 = 0.070 + \underset{(0.006)}{0.046} \ln \epsilon_{2,t-1}^2 + \underset{(0.007)}{0.947} \ln \widehat{\sigma}_{2,t-1}^2. \qquad (5.6)$$

The numbers in parentheses are standard errors of the estimates (we explain how they are computed below). It should be noted, though, that they cannot be used to test whether the ARCH parameter α_{0jj} is equal to zero (under the null), since $\alpha_{0jj} \neq 0$ is required in Theorem 3.1. Confidence intervals can, however, be derived (under the assumption that this parameter is nonzero). The ARCH and GARCH estimates are in the usual range: The ARCH parameters are close to 0.05, the GARCH parameters are close to 0.95 and their sum in each equation is close to 1. Next, we estimate the two-dimensional log-GARCH(1,1)-X model using our EBEE, which gives

$$\begin{aligned}
\ln \widehat{\sigma}_{1,t}^2 &= -0.191 - \underset{(0.014)}{0.020} \ln \epsilon_{1,t-1}^2 - \underset{(0.014)}{0.012} \ln \epsilon_{2,t-1}^2 + \underset{(0.048)}{0.674} \ln \widehat{\sigma}_{1,t-1}^2 \\
&\quad + \underset{(0.032)}{0.158} X_{1,t-1} + \underset{(0.042)}{0.172} X_{2,t-1}, \qquad (5.7) \\
\ln \widehat{\sigma}_{2,t}^2 &= -0.232 - \underset{(0.013)}{0.010} \ln \epsilon_{1,t-1}^2 - \underset{(0.014)}{0.056} \ln \epsilon_{2,t-1}^2 + \underset{(0.032)}{0.751} \ln \widehat{\sigma}_{2,t-1}^2 \\
&\quad + \underset{(0.030)}{0.081} X_{1,t-1} + \underset{(0.035)}{0.258} X_{2,t-1}. \qquad (5.8)
\end{aligned}$$

It is noteworthy that all ARCH effects become negative – currently there are no QMLE results for ordinary, i.e., non-exponential, GARCH models in the presence of negative ARCH effects (Pedersen (2015)), and that only one of the ARCH effects – that of SP100 on its own log-volatility – is significant according to usual significance levels. It is also noteworthy that the GARCH effects fall to 0.674 and 0.751, respectively. This is in line with the findings of Lamoureux and Lastrapes (1990). Both volatility proxies are significant in both equations according to t-tests at usual significance levels, hence overall, the t-tests suggest the volatility spill-over is via the volatility proxies, and not the ARCH effects.

The standard errors can be computed in at least two ways. The first, which is usually the simplest in practice, exploits the fact that numerical software often provides utility functions for the numerical computation of the Hessian of the criterion function. In numerical software, the least squares estimator of an ARMA-X specification is typically implemented by minimising $\sum_{t=1}^{n}\widetilde{u}_{jt}^{2}(\underline{\boldsymbol{\vartheta}})$ (rather than the average $Q_n^{(j)}(\underline{\boldsymbol{\vartheta}}) = n^{-1}\sum_{t=1}^{n}\widetilde{u}_{jt}^{2}(\underline{\boldsymbol{\vartheta}})$). Let $\widehat{\boldsymbol{H}}_n^{(j)}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})$ denote a numerical estimate of the Hessian of the criterion function $\sum_{t=1}^{n}\widetilde{u}_{jt}^{2}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})$ about the least squares estimate $\widehat{\boldsymbol{\vartheta}}_n^{(j)}$. This means that $n^{-1}\widehat{\boldsymbol{H}}_n^{(j)}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})$ provides an estimate of $2\boldsymbol{J}^{(j)}$, since (see the proof of Theorem 3.2)

$$\lim_{n\to\infty}\frac{\partial^2}{\partial\underline{\boldsymbol{\vartheta}}\partial\underline{\boldsymbol{\vartheta}}^\top}Q_n^{(j)}(\boldsymbol{\vartheta}_n) = 2\boldsymbol{J}^{(j)} \quad \text{a.s.}$$

The expression

$$\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_{jt}^{2}\right)\cdot 2n\cdot\left(\widehat{\boldsymbol{H}}_n^{(j)}(\widehat{\boldsymbol{\vartheta}}_n^{(j)})\right)^{-1}$$

thus provides an estimate of the asymptotic variance-covariance matrix $\mathrm{E}u_{jt}^2\left(J^{(j)}\right)^{-1}$ for equation j. Finally, to obtain the empirical variance-covariance matrix of the log-GARCH-X parameters in equation j, the relationships between the ARMA-X and log-GARCH-X parameters are used. The standard errors in all of the equations above have been computed in this way.[4] The second way the standard errors can be computed is by using the formulas following Theorem 3.2, and whose strong consistency is ensured by Proposition 3.1. This is necessary if the joint variance-covariance matrix of all the parameters in the M-dimensional VARMA-X system of equations is needed. When studying volatility spillover effects, we are indeed interested in the joint variance-covariance of all the parameters, hence we now use these formulas instead. The null-hypothesis of no spillover effects between European and US markets corresponds to

$$H_0 : \alpha_{012} = \lambda_{012} = \alpha_{021} = \lambda_{021} = 0. \tag{5.9}$$

These are linear restrictions on a subset of the parameters of the VARMA-X representation. The associated Wald-statistic is distributed as a Chi-squared distribution with 4 degrees of freedom, and turns out to be huge: 7601.2. Therefore, the null of no spill-over is resoundingly rejected at common significance levels.

To further illustrate the computational attractiveness of our estimator, we add more variables to the X-part. It is often the case that additional explanatory information is readily available, for example, volume and leverage. For ordinary or non-exponential GARCH models, strong non-negative restrictions on parameters are needed if one were to include this additional information. In the multivariate log-GARCH(1,1)-X model, by

[4] The R package `lgarch` version 0.6, see Sucarrat (2015), was used.

contrast, we readily obtain the following estimates by means of our EBEE:

$$\begin{aligned}\ln\widehat{\sigma}^2_{1,t} &= -0.280 - \underset{(0.011)}{0.010}\ln\epsilon^2_{1,t-1} - \underset{(0.011)}{0.005}\ln\epsilon^2_{2,t-1} + \underset{(0.034)}{0.807}\ln\widehat{\sigma}^2_{1,t-1} + \underset{(0.025)}{0.117}X_{1,t-1} \\ &\quad + \underset{(0.028)}{0.067}X_{2,t-1} + \underset{(0.048)}{0.143}X_{3,t-1} + \underset{(0.049)}{0.212}X_{4,t-1} - \underset{(0.134)}{0.055}X_{5,t-1} + \underset{(0.165)}{0.262}X_{6,t-1} \qquad (5.10)\\ \ln\widehat{\sigma}^2_{1,t} &= -0.428 - \underset{(0.012)}{0.004}\ln\epsilon^2_{1,t-1} - \underset{(0.013)}{0.050}\ln\epsilon^2_{2,t-1} + \underset{(0.031)}{0.776}\ln\widehat{\sigma}^2_{1,t-1} + \underset{(0.027)}{0.068}X_{1,t-1} \\ &\quad + \underset{(0.034)}{0.214}X_{2,t-1} + \underset{(0.052)}{0.241}X_{3,t-1} + \underset{(0.052)}{0.248}X_{4,t-1} - \underset{(0.136)}{0.061}X_{5,t-1} - \underset{(0.177)}{0.111}X_{6,t-1}. \qquad (5.11)\end{aligned}$$

The $X_{3,t}$ and $X_{4,t}$ are leverage terms defined as $I_{\eta_{1,t}<0}$ and $I_{\eta_{2,t}<0}$, and $X_{5,t}$ and $X_{6,t}$ are the first-difference of log-volume. The leverage terms are highly significant at usual significant levels in both equations according to t-tests, but the volume variables are not.

5.2 A 50-dimensional log-GARCH-X model with time-varying correlations

Estimation of large multivariate GARCH models is plagued by the curse of dimensionality: As the dimension grows, it becomes computationally unfeasible to reliably estimate the parameters jointly. Our estimator sidesteps this problem by estimating the system equation-by-equation. To illustrate we estimate a constrained version (see Section 3.3) of an M-dimensional log-GARCH-X model of the returns of the stocks that make up the EURO STOXX 50 index. The source of our individual stock price data is Bloomberg and goes from 2 January 1998 to 11 June 2015, a total of up to 4474 observations before differencing, lagging and data-cleaning (e.g., removal of missing X-values, etc.). Estimation of the 50 equations is fast, since it takes just over one minute on a laptop computer.[5] The jth. equation is given by

$$\begin{aligned}\ln\sigma^2_{jt} &= \omega_j + \alpha_{j1}\ln\epsilon^2_{j,t-1} + \beta_j\ln\sigma^2_{j,t-1} + \lambda_{j1}X_{j1,t-1} \\ &\quad +\lambda_{j2}X_{2,t-1} + \lambda_{j3}X_{3,t-1} + \lambda_{j4}X_{4,t-1} + \lambda_{j5}X_{5,t-1}, \qquad (5.12)\end{aligned}$$

where $X_{j1,t}$ is a leverage-term for return $\epsilon_{j,t}$ computed as $X_{j1,t} = I_{\eta_{1,t}<0}$, $X_{2,t}$ is the log of squared EURO STOXX 50 return, $X_{3,t}$ is a leverage-term for the EURO STOXX 50 return, $X_{4,t}$ is the difference of the log of EURO STOXX 50 volume and $X_{5,t}$ is a range-based volatility proxy of EURO STOXX 50 return (computed in the same way as earlier). The constraint we impose on the multivariate system is thus that $\boldsymbol{\alpha}_0$ is diagonal.[6] Table 1 contains the results, and an asterisk ($*$) indicates significance at the 10% level. The ARCH (α_{j1}), GARCH (β_j) and leverage (λ_{j1}) parameters behave as expected. All the ARCH parameters are positive and in the 0.01 to 0.06 range, all the GARCH parameters are positive, in the 0.81 to 0.97 range and significant at 10%, and all the leverage parameters are positive (41 of them significant at 10%). The first unexpected result is the impact of $X_{2,t}$, which is defined as $\ln\epsilon^2_t$ where ϵ_t is the daily log-return (in percent) of the EURO

[5] The model is estimated with the R package `lgarch` version 0.6 (Sucarrat (2015)) under R version 3.2.2 running on Windows 7 (64-bit) on a Lenovo X250 with an Intel CORE i7 processor and 16GB RAM.

[6] Otherwise each of the j equations would have $M + K + 2 = 57$ parameters.

STOXX 50 index. The twelve terms that are significant all have a negative effect, while one would maybe have expected a positive one. The effect is small, however, since the largest – in absolute value – is 0.023 (for $j = 8$). An economic explanation for the negative impact is, possibly, that higher EURO STOXX 50 volatility reduces volatility for certain types of stocks. Further investigation is needed in order to shed light on this hypothesis. The leverage term of the EURO STOXX 50 index, $X_{3,t}$, is positive and significant for all but four (when $j = 5, 23, 27$ and 47) out of the 50 stocks, and economically quite substantial when compared with the ARCH effects, i.e., α_{j1} and λ_{0j2}. Thirteen of the stocks are significantly affected by EURO STOXX 50 volume, $X_{4,t}$, and in all these cases the effect is positive. In other words, higher EURO STOXX 50 volume increases volatility – on average – on the next day for these stocks. Finally, the EURO STOXX 50 volatility proxy, $X_{5,t}$, has a positive and significant impact in thirty-seven out of fifty cases.

To obtain estimates of the conditional correlations we fit the corrected DCC (cDCC) model of Aielli (2013), which is a modified version of Engle's (2002) DCC. The cDCC model is given by

$$\boldsymbol{R}_t = \boldsymbol{Q}_t^{*-1/2}\boldsymbol{Q}_t\boldsymbol{Q}_t^{*-1/2}, \quad \boldsymbol{Q}_t = (1-\gamma_0-\delta_0)\boldsymbol{S}_0 + \gamma_0\boldsymbol{Q}_{t-1}^{*1/2}\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}_{t-1}^\top\boldsymbol{Q}_{t-1}^{*1/2} + \delta_0\boldsymbol{Q}_{t-1} \tag{5.13}$$

where $\gamma_0, \delta_0 \geq 0$, $\gamma_0 + \delta_0 < 1$, $\boldsymbol{S}_0$ is a correlation matrix, $\boldsymbol{Q}_t^*$ is a diagonal matrix with the elements from the diagonal of $\boldsymbol{Q}_t$ and $\boldsymbol{\eta}_t = \boldsymbol{D}_t^{-1}\boldsymbol{\epsilon}_t$. Estimation of γ_0 and δ_0 by Gaussian QML yields the estimator

$$(\widehat{\gamma}, \widehat{\delta}) = \arg\min_{(\gamma,\delta)} \sum_{t=1}^{n} \left(\ln|\widehat{\boldsymbol{R}}_t| + \widehat{\boldsymbol{\eta}}_t^\top \widehat{\boldsymbol{R}}_t^{-1} \widehat{\boldsymbol{\eta}}_t \right), \tag{5.14}$$

where $\widehat{\boldsymbol{\eta}}_t$ are the standardised residuals from the 50 log-GARCH models,

$$\begin{aligned}
\widehat{\boldsymbol{R}}_t &= \widehat{\boldsymbol{Q}}_t^{*-1/2}\widehat{\boldsymbol{Q}}_t\widehat{\boldsymbol{Q}}_t^{*-1/2}, \quad \widehat{\boldsymbol{Q}}_t = (1-\widehat{\gamma}-\widehat{\delta})\boldsymbol{S}_n + \widehat{\gamma}\widehat{\boldsymbol{Q}}_{t-1}^{*1/2}\widehat{\boldsymbol{\eta}}_{t-1}\widehat{\boldsymbol{\eta}}_{t-1}^\top\widehat{\boldsymbol{Q}}_{t-1}^{*1/2} + \widehat{\delta}\widehat{\boldsymbol{Q}}_{t-1} \\
\boldsymbol{S}_n &= \frac{1}{n}\sum_{t=1}^{n}\widehat{\boldsymbol{Q}}_t^{*1/2}\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t^\top\widehat{\boldsymbol{Q}}_t^{*1/2}, \qquad \widehat{\boldsymbol{Q}}_t^{*} = \mathrm{diag}(\widehat{q}_{11t}, \ldots, \widehat{q}_{mmt}) \\
\widehat{q}_{iit} &= (1-\widehat{\gamma}-\widehat{\delta}) + \widehat{\gamma}\widehat{\eta}_{i,t-1}^2 + \widehat{\delta}\widehat{q}_{i,t-1} \quad \text{for } i = 1, \ldots, m.
\end{aligned}$$

The estimates of γ_0 and δ_0 are 0.003 and 0.963, respectively, which suggests the correlations are very persistent. Figure 1 contains histograms of selected descriptive statistics of the $50 \cdot 49/2 = 1225$ fitted correlations paths. Graphically, the empirical distribution of the unconditional correlations is bell-shaped and symmetric about 0.49 (the mean and median are virtually identical, 0.487 and 0.485, respectively), and the maximum and minimum unconditional correlations are 0.89 and 0.16. The distributions of the minima and maxima of the conditional correlations are also unimodal, but somewhat skewed to the left and right, respectively. Moreover, the distribution of the minima reveals that some paths cross the zero line.

6 Conclusions

We derive an equation-by-equation estimator (EBEE) of a multivariate log-GARCH-X model that admits feedback effects between the equations, and Dynamic Conditional Correlations (DCCs). Our least squares EBEE does not rely on financial returns being distributed according to a specific conditional distribution, e.g., the multivariate normal. Equation-by-equation estimation is particularly attractive when the dimensionality of the system is large, or when the number of covariates is large, or both, since then estimation often becomes numerically unfeasible due to the "curse of dimensionality". The vector of covariates ("X") is assumed to be stationary and ergodic, which means many of the variables that are believed to have an impact on volatility, and which are readily available, e.g., leverage, volume and volatility proxies, can be included as conditioning variables. Both strong consistency and asymptotic normality of our EBEE is proved under mild assumptions, and consistency of the estimators of the terms involved in the asymptotic variance-covariance matrices is also proved.

Two empirical applications illustrate the usefulness of the results. In the first, we show how the volatility-spillover hypothesis can be tested when it involves restrictions in several equations – the null of no spillover is resoundingly rejected. In the second application, we illustrate how a high-dimensional multivariate log-GARCH-X model can readily be estimated in minutes. The model concerns the constituent returns of the EURO STOXX 50 index, i.e., the model is 50-dimensional, and the X-part contains a leverage term, past values of EURO STOXX 50 variability together with its leverage, volume and a range-based volatility proxy. One or more of these are found to be significant in most of the equations.

Among the potential developments of the present work, let us mention that, in particular for the applications, it would be sensible to allow covariates obtained from mixed data sampling (MIDAS). The linear structure of the log-GARCH-X model should help to define an operational log-GARCH-MIDAS model.

References

Aielli, G. P. (2013). Dynamic Conditional Correlations: On Properties and Estimation. *Journal of Business and Economic Statistics 31*, 282–299. http://dx.doi.org/10.1080/07350015.2013.771027.

Apergis, N. and A. Rezitis (2011). Food Price Volatility and Macroeconomic Factors: Evidence from GARCH and GARCH-X Estimates. *Journal of Agricultural and Applied Economics 43*, 95–110.

Bauwens, L., D. Rime, and G. Sucarrat (2006). Exchange Rate Volatility and the Mixture of Distribution Hypothesis. *Empirical Economics 30*, 889–911.

Billingsley, P. (1961). Statistical methods in Markov chains. *The Annals of Mathematical Statistics 32*, 12–40.

Bollerslev, T. and M. Melvin (1994). Bid-Ask spreads and the volatility in the foreign exchange market: An empirical analysis. *Journal of International Economics 36*, 355–372.

Boussama, F., F. Fuchs, and R. Stelzer (2011). Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models. *Stochastic Processes and Their Applications 121*, 2331–2360.

Brenner, R., R. Harjes, and K. Kroner (1996). Another look at models of the short-term interest rate. *Journal of Financial and Quantitative Analysis 31*, 85–107.

Chen, M. and Q. Song (2015). Semi-parametric estimation and forecasting for exogenous log-GARCH models. *TEST*. DOI: http://dx.doi.org/10.1007/s11749-015-0442-6.

Clark, P. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica 41*, 135–155.

Conrad, C. and E. Weber (2013). Measuring persistence in volatility spillovers. Discussion Paper 543, Department of Economics, University of Heidelberg. http://archiv.ub.uni-heidelberg.de/volltextserver/14865/.

Dominguez, K. M. (1998). Central bank interventions and exchange rate volatility. *Journal of International Money and Finance 17*, 161–190.

Duan, J.-C. (1995). The GARCH option pricing model. *Mathematical finance 5*(1), 13–32.

Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics 20*, 339–350.

Engle, R. and G. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*, 3–27.

Francq, C. and G. Sucarrat (2013). An Exponential Chi-Squared QMLE for Log-GARCH Models Via the ARMA Representation. http://mpra.ub.uni-muenchen.de/51783/.

Francq, C. and L. Q. Thieu (2015). Qml inference for volatility models with covariates. http://mpra.ub.uni-muenchen.de/63198/.

Francq, C. and J.-M. Zakoïan (2010). *GARCH Models*. New York: Marcel Dekker.

Francq, C. and J.-M. Zakoïan (2015). Estimating multivariate GARCH and stochastic correlation models equation by equation. Forthcoming in *The Journal of the Royal Statistical Society. Series B*. Working paper version: MPRA Paper No. 54250. Online at http://mpra.ub.uni-muenchen.de/54250/.

Garman, M. B. and M. J. Klass (1980). On the Estimation of Securities Price Volatilities from Historical Data. *Journal of Business 53*, 67–78.

Hagiwara, M. and M. Herce (1999). Endogenous exchange rate volatility, trading volume and interest rate differentials in a model of portfolio selection. *Review of International Economics 7*, 202–218.

Han, H. and D. Kristensen (2014). Asymptotic theory for the QMLE in GARCH-X models with stationary and non-stationary covariates. *Journal of Business and Economic Statistics*, ??–??

Hansen, P. R., Z. Huan, and H. H. Shek (2012). Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. *Journal of Applied Econometrics 27*, 877–906.

Hwang, S. and S. Satchell (2005). GARCH model with cross-sectional volatility: GARCHX models. *Applied Financial Economics 15*, 203–216.

Lamoureux, C. G. and W. D. Lastrapes (1990). Heteroscedasticity in Stock Return Data: Volume versus GARCH Effects. *The Journal of Finance*, 221–229.

McNeil, A., F. R. and P. Embrechts (2005). *Quantitative risk management.* Princeton: Princeton Series in Finance.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica 59*, 347–370.

Parkinson, M. (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business 53*, 61–65.

Pedersen, R. S. (2015). Inference and testing on the boundary in extended constant correlation GARCH models. Working Paper.

Santos, A. and G. V. Moura (2014). Dynamic factor multivariate GARCH model. *Computational Statistics & Data Analysis 76*, 606–617.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high frequency based volatility (HEAVY). *Journal of Applied Econometrics 25*, 197–231.

Straumann, D. and T. Mikosch (2006). Quasi-Maximum-Likelihood Estimation in Conditionally Heteroscedastic Time Series: A Stochastic Recurrence Equations Approach. *The Annals of Statistics 34*, 2449–2495.

Sucarrat, G. (2015). *lgarch: Simulation and estimation of log-GARCH models*. R package version 0.6. http://cran.r-project.org/web/packages/lgarch/.

Sucarrat, G., G. Bjø nnes, and S. Grø nneberg (2015). Models of Financial Return With Time-Varying Zero Probability. Work in progress.

Sucarrat, G. and Á. Escribano (2014). Unbiased Estimation of Log-GARCH Models in the Presence of Zero Returns. MPRA Paper No. 59040. http://mpra.ub.uni-muenchen.de/59040/.

Sucarrat, G., S. Grønneberg, and Á. Escribano (2015). Estimation and Inference in Univariate and Multivariate Log-GARCH-X Models When the Conditional Density is Unknown. MPRA Paper No. 62352. Online at http://mpra.ub.uni-muenchen.de/62352/. An earlier version circulated as "The Power Log-GARCH Model", Universidad Carlos III de Madrid Working Paper 10-13 in the Economic Series, June 2010.

Tauchen, G. and M. Pitts (1983). The Price Variability-Volume Relationship on Speculative Markets. *Econometrica 51*, 485–505.

Wintenberger, O. (2013). Continuous Invertibility and Stable QML Estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics 40*, 846–867.

Table 1: A multivariate Log-GARCH(1,1)-X model of the EURO STOXX 50 constituent returns (see Section 5.2)

j	$\widehat{\omega}_{0j}$ $(s.e.)$	$\widehat{\alpha}_{0jj}$ $(s.e.)$	$\widehat{\beta}_{0jj}$ $(s.e.)$	$\widehat{\lambda}_{0j1}$ $(s.e.)$	$\widehat{\lambda}_{0j2}$ $(s.e.)$	$\widehat{\lambda}_{0j3}$ $(s.e.)$	$\widehat{\lambda}_{0j4}$ $(s.e.)$	$\widehat{\lambda}_{0j5}$ $(s.e.)$
1	−0.040	0.047 (0.007)*	0.900 (0.023)*	0.110 (0.036)*	−0.007 (0.007)	0.122 (0.035)*	−0.025 (0.089)	0.034 (0.015)*
2	0.011	0.038 (0.008)*	0.932 (0.017)*	0.059 (0.028)*	−0.005 (0.006)	0.076 (0.033)*	0.328 (0.091)*	0.021 (0.010)*
3	−0.002	0.043 (0.006)*	0.911 (0.013)*	0.117 (0.032)*	0.003 (0.007)	0.078 (0.034)*	0.126 (0.089)	0.034 (0.012)*
4	−0.011	0.033 (0.006)*	0.946 (0.013)*	0.048 (0.030)	−0.001 (0.006)	0.104 (0.030)*	0.082 (0.092)	0.010 (0.009)
5	−0.007	0.028 (0.005)*	0.967 (0.006)*	0.070 (0.019)*	−0.005 (0.004)	0.019 (0.021)	0.182 (0.084)*	0.007 (0.005)
6	−0.016	0.043 (0.006)*	0.939 (0.011)*	0.076 (0.024)*	−0.007 (0.005)	0.063 (0.026)*	0.132 (0.082)	0.016 (0.008)*
7	−0.052	0.027 (0.006)*	0.919 (0.013)*	0.109 (0.030)*	−0.004 (0.006)	0.146 (0.034)*	0.092 (0.085)	0.048 (0.013)*
8	−0.075	0.034 (0.007)*	0.860 (0.021)*	0.179 (0.040)*	−0.023 (0.009)*	0.096 (0.040)*	−0.017 (0.089)	0.091 (0.020)*
9	0.034	0.027 (0.007)*	0.876 (0.027)*	0.081 (0.037)*	0.005 (0.008)	0.082 (0.037)*	0.144 (0.097)	0.054 (0.020)*
10	−0.050	0.038 (0.006)*	0.899 (0.017)*	0.132 (0.033)*	−0.017 (0.007)*	0.083 (0.036)*	0.058 (0.090)	0.069 (0.018)*
11	−0.066	0.033 (0.008)*	0.891 (0.019)*	0.191 (0.037)*	−0.018 (0.008)*	0.072 (0.038)*	0.062 (0.088)	0.080 (0.019)*
12	−0.005	0.032 (0.006)*	0.942 (0.012)*	0.042 (0.025)*	−0.006 (0.005)	0.079 (0.027)*	−0.006 (0.090)	0.023 (0.010)*
13	−0.016	0.046 (0.008)*	0.915 (0.017)*	0.115 (0.033)*	−0.012 (0.007)*	0.071 (0.034)*	0.104 (0.091)	0.037 (0.013)*
14	0.008	0.039 (0.006)*	0.937 (0.012)*	0.039 (0.026)	0.000 (0.006)	0.091 (0.028)*	0.213 (0.086)*	0.014 (0.009)*
15	−0.005	0.035 (0.007)*	0.897 (0.019)*	0.063 (0.032)*	−0.012 (0.008)	0.111 (0.037)*	0.022 (0.093)	0.060 (0.017)*
16	0.010	0.046 (0.007)*	0.933 (0.014)*	0.064 (0.028)*	−0.004 (0.006)	0.058 (0.027)*	0.050 (0.087)	0.011 (0.008)
17	−0.010	0.033 (0.007)*	0.907 (0.021)*	0.083 (0.031)*	−0.004 (0.007)	0.115 (0.037)*	−0.042 (0.093)	0.055 (0.019)*
18	−0.008	0.033 (0.007)*	0.908 (0.024)*	0.069 (0.034)*	−0.004 (0.007)	0.095 (0.039)*	0.048 (0.093)	0.042 (0.017)*
19	0.007	0.047 (0.006)*	0.946 (0.007)*	0.039 (0.026)	−0.003 (0.005)	0.090 (0.026)*	0.152 (0.084)*	0.005 (0.006)
20	−0.028	0.059 (0.008)*	0.886 (0.020)*	0.133 (0.036)*	−0.021 (0.008)*	0.089 (0.035)*	0.041 (0.089)	0.051 (0.016)*
21	−0.031	0.055 (0.007)*	0.926 (0.012)*	0.097 (0.030)*	−0.016 (0.006)*	0.088 (0.032)*	0.130 (0.094)	0.022 (0.009)*
22	0.022	0.012 (0.012)	0.817 (0.059)*	0.065 (0.056)	−0.007 (0.012)	0.118 (0.061)*	−0.004 (0.108)	0.121 (0.046)*
23	−0.006	0.036 (0.006)*	0.927 (0.012)*	0.114 (0.029)*	0.003 (0.006)	0.042 (0.029)	0.165 (0.090)*	0.015 (0.009)*
24	0.014	0.045 (0.006)*	0.939 (0.010)*	0.044 (0.028)	0.001 (0.005)	0.087 (0.027)*	0.074 (0.087)	0.003 (0.007)
25	−0.005	0.041 (0.006)*	0.935 (0.012)*	0.063 (0.026)*	−0.007 (0.006)	0.089 (0.030)*	0.233 (0.088)*	0.024 (0.010)*
26	−0.019	0.052 (0.007)*	0.931 (0.010)*	0.108 (0.025)*	−0.011 (0.006)*	0.060 (0.027)*	0.076 (0.084)	0.018 (0.007)*
27	0.001	0.035 (0.007)*	0.924 (0.018)*	0.088 (0.033)*	−0.008 (0.006)	0.034 (0.034)	0.129 (0.090)	0.033 (0.012)*
28	−0.048	0.049 (0.006)*	0.929 (0.010)*	0.100 (0.030)*	−0.009 (0.006)	0.142 (0.031)*	0.310 (0.087)*	0.025 (0.009)*
29	−0.012	0.046 (0.006)*	0.936 (0.010)*	0.078 (0.025)*	−0.011 (0.006)*	0.072 (0.027)*	0.034 (0.091)	0.020 (0.008)*
30	0.019	0.027 (0.005)*	0.967 (0.008)*	0.012 (0.021)	0.005 (0.004)	0.051 (0.021)*	0.135 (0.083)	−0.005 (0.005)
31	−0.011	0.036 (0.006)*	0.923 (0.016)*	0.077 (0.031)*	−0.004 (0.006)	0.086 (0.031)*	0.104 (0.087)	0.031 (0.012)*
32	0.003	0.040 (0.006)*	0.941 (0.009)*	0.062 (0.025)*	−0.003 (0.006)	0.060 (0.027)*	−0.010 (0.090)	0.017 (0.008)*
33	0.021	0.033 (0.006)*	0.952 (0.009)*	0.029 (0.023)	−0.007 (0.005)	0.056 (0.026)*	0.027 (0.083)	0.016 (0.007)*
34	−0.005	0.045 (0.006)*	0.950 (0.007)*	0.044 (0.023)*	−0.007 (0.005)	0.082 (0.025)*	0.096 (0.085)	0.006 (0.006)
35	−0.020	0.040 (0.005)*	0.944 (0.009)*	0.087 (0.025)*	−0.012 (0.006)*	0.051 (0.027)*	0.179 (0.085)*	0.024 (0.009)*
36	−0.027	0.047 (0.007)*	0.923 (0.014)*	0.079 (0.027)*	−0.014 (0.006)*	0.103 (0.031)*	0.050 (0.087)	0.026 (0.010)*
37	0.033	0.051 (0.007)*	0.914 (0.015)*	0.036 (0.029)	−0.010 (0.007)	0.053 (0.031)*	0.111 (0.089)	0.029 (0.011)*
38	−0.051	0.029 (0.005)*	0.947 (0.011)*	0.137 (0.028)*	−0.009 (0.006)*	0.053 (0.026)*	0.272 (0.088)*	0.026 (0.010)*
39	−0.002	0.042 (0.006)*	0.941 (0.011)*	0.053 (0.027)*	−0.002 (0.005)	0.087 (0.025)*	0.138 (0.082)*	0.009 (0.007)
40	−0.013	0.037 (0.005)*	0.956 (0.007)*	0.096 (0.025)*	−0.001 (0.005)	0.057 (0.025)*	0.306 (0.088)*	0.004 (0.006)
41	−0.006	0.037 (0.006)*	0.929 (0.016)*	0.075 (0.029)*	0.000 (0.006)	0.105 (0.030)*	0.045 (0.092)	0.020 (0.011)*
42	−0.020	0.040 (0.005)*	0.941 (0.009)*	0.073 (0.030)*	−0.010 (0.006)	0.081 (0.031)*	0.124 (0.089)	0.022 (0.009)*
43	−0.024	0.052 (0.007)*	0.929 (0.011)*	0.083 (0.029)*	−0.014 (0.006)*	0.074 (0.029)*	0.023 (0.082)	0.022 (0.009)*
44	−0.022	0.040 (0.006)*	0.940 (0.011)*	0.094 (0.028)*	−0.005 (0.006)	0.065 (0.027)*	0.029 (0.086)	0.014 (0.009)
45	−0.105	0.040 (0.009)*	0.920 (0.020)*	0.187 (0.050)*	−0.013 (0.009)	0.131 (0.048)*	0.121 (0.105)	0.039 (0.018)*
46	−0.002	0.051 (0.006)*	0.942 (0.007)*	0.079 (0.021)*	−0.002 (0.005)	0.078 (0.026)*	0.140 (0.084)*	0.001 (0.006)
47	0.018	0.044 (0.006)*	0.927 (0.014)*	0.083 (0.027)*	0.006 (0.006)	0.036 (0.027)	0.107 (0.085)	0.009 (0.010)
48	−0.006	0.049 (0.009)*	0.923 (0.022)*	0.117 (0.033)*	−0.006 (0.007)	0.066 (0.032)*	0.224 (0.089)*	0.014 (0.011)
49	−0.003	0.046 (0.006)*	0.931 (0.011)*	0.079 (0.029)*	−0.002 (0.006)	0.088 (0.030)*	0.025 (0.086)	0.015 (0.009)*
50	0.085	0.045 (0.008)*	0.902 (0.021)*	0.009 (0.029)	0.002 (0.007)	0.076 (0.033)*	0.014 (0.094)	0.032 (0.015)*

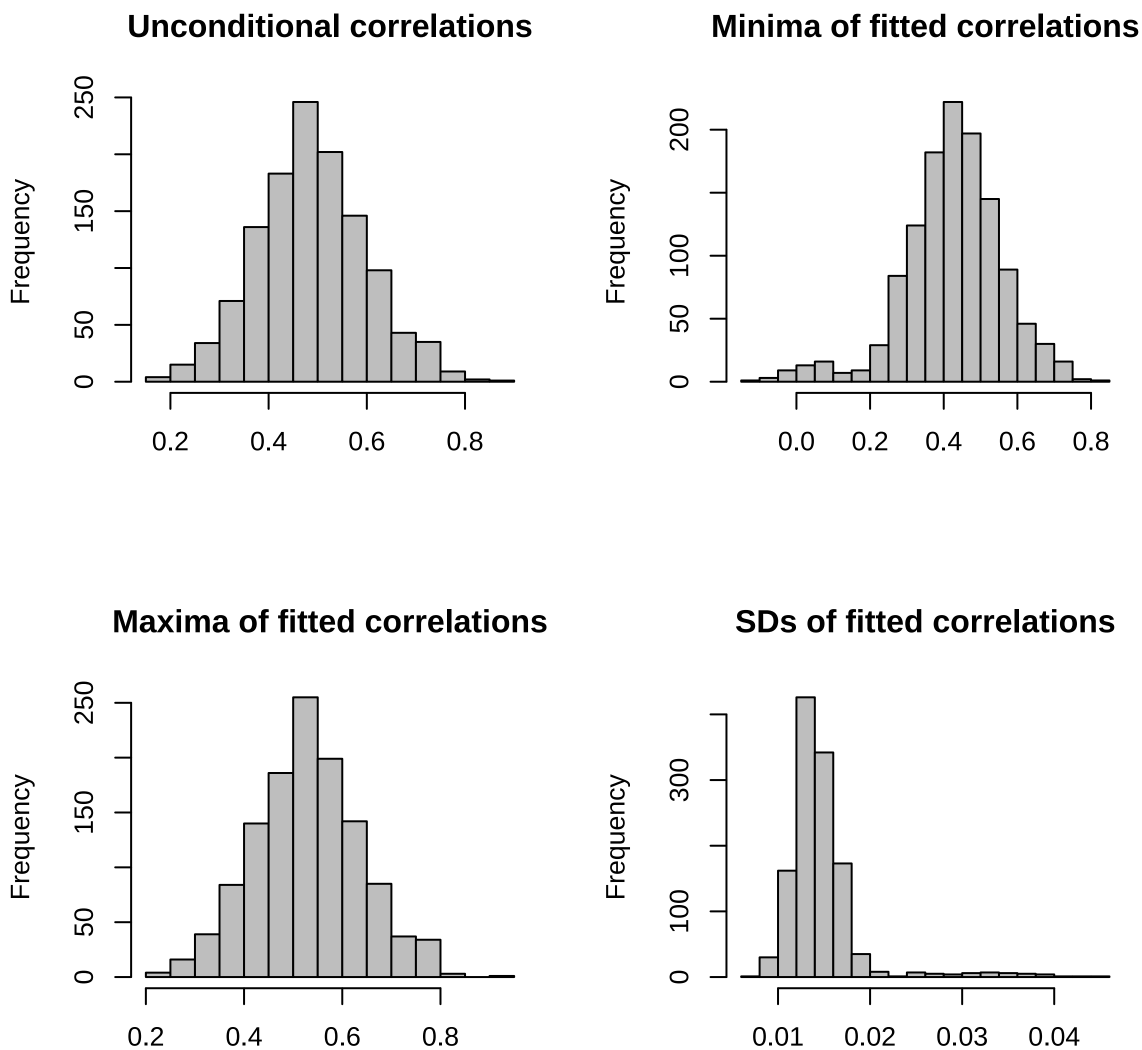

Figure 1: Histograms of the estimated unconditional correlations, the minima and maxima of the fitted conditional correlations, and the sample standard deviations (SDs) of the 1225 fitted conditional correlations of the 50-dimensional model of the EURO STOXX 50 index (see Section 5.2)