Variance targeting estimation of the BEKK-X model

Thieu, Le Quyen

University Pierre and Marie Curie

1 August 2016

Variance targeting estimation of the BEKK-X model

Le Quyen Thieu,*

Université Pierre et Marie Curie.

Abstract

This paper studies the BEKK model with exogenous variables (BEKK-X), which intends to take into account the influence of explanatory variables on the conditional covariance of the asset returns. Strong consistency and asymptotic normality of a variance targeting estimator (VTE) is proved. Monte Carlo experiments and an application to financial series illustrate the asymptotic results.

Keywords: BEKK model augmented with exogenous variables, BEKK-X model, Variance targeting estimation (VTE),

1 Introduction

Analysing asset return covariances is important since it is a crucial input, in particular, for portfolio selection, asset management and risk assessment. Forecasting sequences of covariance matrices can be done by using multivariate conditional heteroskedastic (GARCH) models, see Bauwens and Rombouts (2006) and Silvennoinen and Terasvirta (2009) for extensive surveys. The first generation of models, for example the VEC model of Bollerslev and Wooldridge (1988) and the BEKK model of Engle and Kroner (1995), are direct extensions of the univariate GARCH model of Bollerslev (1986). These models take into account the information contained in the past of the individual asset returns, but can

*Corresponding author: Le Quyen Thieu, Université Pierre et Marie Curie, France. Telephone: (+33)7 81 95 25 89. E-mail: thieulequyen1411@gmail.com

not exploit external information. In practice, the relations between the volatilities and co-volatilities of several markets can be affected by external influences that we call exogenous variables. In this paper, we employ this term in a wide sens for any external explanatory variable (see Koopmans and Reiersol (1950), Engle et al. (1983), Florens and Mouchart (1982) and Bouissou and Vuong (1986) for other concepts of exogeneity). Engle (2009) provide evidence that economic fundamentals such as inflation and industrial production growth drive stock market volatility. Cakmakli and Dijk (2010) demonstrate that a number of macroeconomic variables can help predicting US stocks volatility between 1980 and 2005. Christiansen and Schrimpf (2012) get similar results for the foreign exchange, the commodity, and the bond market.

Despite the fact that such additional information in financial and macroeconomic variables are widely used to explain and forecast volatility in financial markets, there are, however, relatively few results on the asymptotic behavior of the estimation in presence of exogenous variables. In the univariate case, Han and Kristensen (2014) give conditions for the Consistency and Asymptotic Normality (CAN) of the Gaussian QMLE for the standard GARCH(1,1) augmented by a single covariate. Francq and Thieu (2015) study the asymptotic distribution of the QMLE for a versatile class of model: The Asymmetric Power ARCH(p, q)-X model with an unrestrictive number of the exogenous variables. In the multivariate case, Francq and Sucarrat (2015) provide the proof of the CAN of an estimator of the volatilites for the components of a vectorial log-GARCH model with co-variates. Their framework does not directly specify the conditional covariance. Engle and Kroner (1995) suggest the BEKK model augmented by exogenous variables (BEKK-X). In their model, the covariates can affect all the volatilies and co-volatilities of the returns. However, they only provide the estimation of the model without exogenous influences. Moreover, the CAN of their estimator is not proved. In this paper, the estimation of the BEKK-X model will be presented and its CAN will be established. There are several advantages with the BEKK model. First, although the asymptotic theory for multivariate GARCH has been less investigated than the asymptotic theory for univariate models, several papers have established asymptotic results for different methods of estimation of the BEKK model without covariates. Comte and Lieberman (2003) show the CAN of the QMLE. However, the BEKK model contains a large number of parameters, even for moderate dimensions. This implies that it is difficult to estimate the model by the classical QMLE. Pedersen and Rahbek (2014) consider a simplified estimation method, the variance targeting estimation (VTE), and provide its CAN. The VTE method has been proposed by Engle and Mezrich (1996) to alleviate the numerical difficulties encountered in the maximization of the quasi likelihood. The VTE is numerically more efficient than the QMLE, in particular, in the presence of exogenous variables, because it requires an optimization of lower dimension. This estimator has also the advantage of being relatively robust for long term predictions (see Francq and Zakoïan (2011)). In the present paper, we establish the CAN of the VTE for the BEKK-X model.

The paper is organized as follows. Section 2 introduces the BEKK model augmented with explanatory variables and presents the VTE method. The consistency and asymptotic distribution of the VTE are investigated in Section 3. Numerical illustrations are presented in Section 4. Section 5 concludes the paper. All the proofs are collected in Section 6.

2 The model and variance targeting estimation

2.1 The model

Let $\{\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \cdots, \varepsilon_{mt})'\}$ be a $m-$dimensional process and $\boldsymbol{x}_t = (x_{1t}, \cdots, x_{rt})' \in \mathbb{R}^r$ be a vector of r exogenous variables. Denote $\mathcal{F}_{t-1}$ the $\sigma-$field generated by the past of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{x}_t$; *i.e.* $\mathcal{F}_{t-1} = \sigma\{\boldsymbol{\varepsilon}_u, \boldsymbol{x}_v; u < t, v < t\}$. Assume that

$$E(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = \mathbf{0}, \qquad Var(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = \boldsymbol{H}_t \text{ exists and is positive definite.} \tag{1}$$

The $m \times m$ matrix $\boldsymbol{H}_t$ is specified as a function of the past values of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{x}_t$.

We consider the following BEKK-X(1,1) model

$$\begin{cases} \boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t \\ \boldsymbol{H}_t = \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{H}_{t-1}\boldsymbol{B}' + \boldsymbol{C}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\boldsymbol{C}', \end{cases} \tag{2}$$

where $\boldsymbol{\Omega}, \boldsymbol{A}, \boldsymbol{B}$ are $m \times m$ parameter matrices and $\boldsymbol{C}$ is $m \times r$ parameter matrix. To ensure the positivity of the conditional covariance $\boldsymbol{H}_t$, we assume that the coefficient matrix $\boldsymbol{\Omega} > 0$, where the symbol $>$ denotes the positive definiteness of a matrix.

The attractive property of the BEKK model is that the conditional covariance matrices are positive definite by construction. Introducing the explanatory variables under the form $\boldsymbol{C}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\boldsymbol{C}'$ still guarantees the positive definiteness of $\boldsymbol{H}_t$. Furthermore, they are not restricted to a single variable.

Let $\|A\| = \sqrt{Tr(A'A)}$ be the Euclidean norm of a vector or a matrix A , where $Tr(\cdot)$ is the trace of a square matrix. The following assumptions are made throughout the paper.

A1: $E(\boldsymbol{\eta}_t|\mathcal{F}_{t-1}) = 0$ and $E(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t|\mathcal{F}_{t-1}) = I_m$.

A2: $(\boldsymbol{\varepsilon}_t, \boldsymbol{x}_t)$ is a strictly stationary and ergodic process.

A4: $E\|\boldsymbol{x}_t\|^2 < \infty$ and $E\|\boldsymbol{\varepsilon}_t\|^2 < \infty$.

Remark 1 *Boussama and Stelzer (2011) provide sufficient conditions for the existence of a unique stationary and ergodic solution to BEKK multivariate GARCH models. For the model (2) without covariate, for example, this solution exists if the following assumptions are satisfied*

i) The innovation $\boldsymbol{\eta}_t$ admits a density absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$ and positive in a neighborhood of the point zero

ii) $\rho(\boldsymbol{A}_0 \otimes \boldsymbol{A}_0 + \boldsymbol{B}_0 \otimes \boldsymbol{B}_0) < 1$

where the symbol $\otimes$ stands for the Kronecker product and $\rho(\cdot)$ denotes the spectral radius of a matrix which is the maximum among the absolute values of the eigenvalues of a matrix. Furthermore, under these conditions, the existence $E\|\boldsymbol{\varepsilon}_t\|^2 < \infty$ in **A4** *is satisfied.*

Remark 2 *Under Assumptions* **A2, A4***, the intercept matrix in the volatility of (2) can be represented as a function of the unconditional covariance of the observations and of the unconditional second moment of covariates. That allows us to apply the variance targeting estimation method.*

2.2 Variance targeting estimation

The VTE is a two-step estimation technique whose advantages are to reduce the computational complexity of the optimization produce (see Pedersen and Rahbek (2014), Francq and Zakoïan (2011) and Francq and Zakoïan (2014)) and to guarantee that the implied variance is equal to the sample variance. This method is based on a reparamerization of the volatility equation, in which the intercept is replaced by the returns unconditional variance in case that there is no covariates.

Denote by $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} := Var(\boldsymbol{\varepsilon}_t) = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = E(\boldsymbol{H}_t)$ the variance matrix of the observations and $\boldsymbol{\Sigma}_{\boldsymbol{x}} := E(\boldsymbol{x}_t\boldsymbol{x}_t')$ the second-order moment of the vector of exogenous variables. These matrices are well defined under Assumption **A4**. By taking the expectation of the two hand sides of (2), we get

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{B}' + \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{C}'. \tag{3}$$

Then (2) can be reparameterized by

$$\begin{cases} \boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \\ \boldsymbol{H}_t = (\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} - \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{A}' - \boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{B}' - \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{C}') + \boldsymbol{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{H}_{t-1}\boldsymbol{B}' \\ \qquad +\boldsymbol{C}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\boldsymbol{C}'. \end{cases} \tag{4}$$

Note that in this reparametrization, the constraint of the positive definiteness of the intercept $\boldsymbol{\Omega} > 0$ in (2) becomes

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} - \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{A}' - \boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{B}' - \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{C}' > 0. \tag{5}$$

The generic parameter of the model (4) consists of the elements of the matrices $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ and the ones of the matrices $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$. As mentioned, the parameters of the model will be estimated in the two steps VTE. In the first step, the matrices $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ will be empirically estimated . In the second step, the other parameters will be estimated by QML optimization. The vector of unknown parameters is thus decomposed by $\boldsymbol{\vartheta}_0 = (\boldsymbol{\gamma}_0', \boldsymbol{\theta}_0')' \in \mathbb{R}^d$ with

$$\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{\boldsymbol{x}0}', \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0}')' = (vech'(\boldsymbol{\Sigma}_{\boldsymbol{x}}), vech'(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}))' \in \mathbb{R}^{d_1}, \quad d_1 = \frac{r(r+1)}{2} + \frac{m(m+1)}{2},$$
$$\boldsymbol{\theta}_0 = (vec'(\boldsymbol{A}_0), vec'(\boldsymbol{B}_0), vec'(\boldsymbol{C}_0))' \in \mathbb{R}^{d_2}, \quad d_2 = 2m^2 + mr$$

and $d = d_1 + d_2$ is the total number of unknown parameter of (4), where vec denotes the operator that stacks all columns of a matrix into a column vector, and $vech$ denotes the one that stacks only the lower triangular part including the diagonal of a symmetric matrix into a vector. Likewise, we define the parameter space

$$\boldsymbol{\Theta} := \boldsymbol{\Theta}_{\boldsymbol{\gamma}} \times \boldsymbol{\Theta}_{\boldsymbol{\theta}} \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2},$$

whose a generic parameter vector is denoted by

$$\boldsymbol{\vartheta} = (\boldsymbol{\gamma}', \boldsymbol{\theta}')' = (\boldsymbol{\gamma}'_{\boldsymbol{x}}, \boldsymbol{\gamma}'_{\boldsymbol{\varepsilon}}, \boldsymbol{\theta}')' = \left(vech'(\boldsymbol{\Sigma}_{\boldsymbol{x}}), vech'(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}), vec'(\boldsymbol{A}), vec'(\boldsymbol{B}), vec'(\boldsymbol{C})\right)'.$$

To emphasize that the conditional covariance matrix in (4) depends on the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ and that they are independently estimated, we write $\boldsymbol{H}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$.

Let $(\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n)$ be a realization of length n of the stationary ergodic process $(\boldsymbol{\varepsilon}_t)$ and $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ be n observations of the exogenous variables $(\boldsymbol{x}_t)$. Conditionally on the initial values $\boldsymbol{\varepsilon}_0, \boldsymbol{x}_0$ and $\widetilde{\boldsymbol{H}}_0$, the conditional covariance matrix can be recursively defined, for $t \geq 1$, as follows

$$\begin{aligned} \widetilde{\boldsymbol{H}}_t(\boldsymbol{\gamma}, \boldsymbol{\theta}) =& (\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} - \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{A}' - \boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{B}' - \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{C}') + \boldsymbol{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\boldsymbol{A}' + \boldsymbol{B}\widetilde{\boldsymbol{H}}_{t-1}(\boldsymbol{\gamma}, \boldsymbol{\theta})\boldsymbol{B}' \\ &+ \boldsymbol{C}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\boldsymbol{C}'. \end{aligned} \tag{6}$$

Let us define the functions

$$\widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta}), \qquad \widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \boldsymbol{\varepsilon}'_t \widetilde{\boldsymbol{H}}_t^{-1}(\boldsymbol{\gamma}, \boldsymbol{\theta})\boldsymbol{\varepsilon}_t + \log\det\left(\widetilde{\boldsymbol{H}}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})\right). \tag{7}$$

As mentioned, in the first stage of VT estimation method, the parameter $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_{\boldsymbol{x}}, \boldsymbol{\gamma}'_{\boldsymbol{\varepsilon}})'$ is pre-estimated directly from the sample by the method of moments:

$$\widehat{\boldsymbol{\gamma}}_n = (\widehat{\boldsymbol{\gamma}}'_{\boldsymbol{x}n}, \widehat{\boldsymbol{\gamma}}'_{\boldsymbol{\varepsilon}n})' = \left(vech'(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n}), vech'(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n})\right)', \tag{8}$$

where $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}'_t$ are the empirical estimators of the covariance matrix of $\boldsymbol{\varepsilon}_t$ and the second-order moment of $\boldsymbol{x}_t$, respectively. Once $\widehat{\boldsymbol{\gamma}}_n$ is obtained, the parameter $\boldsymbol{\theta}$ is estimated by using the quasi likelihood, conditioning on the parameters estimated in the first stage

$$\widetilde{Q}_n(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\ell}_t(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\varepsilon}'_t \widetilde{\boldsymbol{H}}_t^{-1}(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta})\boldsymbol{\varepsilon}_t + \log\det\left(\widetilde{\boldsymbol{H}}_t(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta})\right), \tag{9}$$

where the covariance process $\widetilde{\boldsymbol{H}}_t\left(\widehat{\gamma}_n, \boldsymbol{\theta}\right)$ can be recursively calculated by replacing $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ in (6) by $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n}$ respectively. The estimator of the parameter $\boldsymbol{\theta}$ is thus defined as any measurable solution $\widehat{\boldsymbol{\theta}}_n$ of

$$\widehat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}{\arg\min}\, \widetilde{Q}_n\left(\widehat{\gamma}_n, \boldsymbol{\theta}\right). \tag{10}$$

The VTE of $\boldsymbol{\vartheta}_0$ is then given by $\widehat{\boldsymbol{\vartheta}}_n = \left(\widehat{\gamma}_n', \widehat{\boldsymbol{\theta}}_n'\right)'$.

The estimation of $\boldsymbol{\Omega}$ in the original BEKK-X model (2) can be obtained by

$$\widehat{\boldsymbol{\Omega}}_n = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} - \widehat{\boldsymbol{A}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}\widehat{\boldsymbol{A}}_n' - \widehat{\boldsymbol{B}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}\widehat{\boldsymbol{B}}_n' - \widehat{\boldsymbol{C}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n}\widehat{\boldsymbol{C}}_n', \tag{11}$$

where $\widehat{\boldsymbol{A}}_n, \widehat{\boldsymbol{B}}_n$ and $\widehat{\boldsymbol{C}}_n$ are the QML estimators of $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ respectively. Then the estimator of original parameter vector, denoted by $\boldsymbol{\xi}_0 = (vech(\boldsymbol{\Omega}_0)', \boldsymbol{\theta}_0')'$, of (2) can be given by $\widehat{\boldsymbol{\xi}}_n = \left(vech'(\widehat{\boldsymbol{\Omega}}_n), \widehat{\boldsymbol{\theta}}_n'\right)'$.

3 VTE inference

In this section, the asymptotic properties of the VTE will be established. The computation of the asymptotic covariance matrix will be aslo given.

3.1 Consistency and asymptotic normality

For the strong consistency of the VTE, the following assumptions are required

A5: The true parameter $\boldsymbol{\vartheta}_0 \in \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$ is compact.

A6: $\rho(\boldsymbol{B}) < 1$ and $\rho(\boldsymbol{A}\otimes\boldsymbol{A} + \boldsymbol{B}\otimes\boldsymbol{B}) < 1$ for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$.

A7: If for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\boldsymbol{\theta}}$, $\boldsymbol{H}_t(\gamma_0, \boldsymbol{\theta}) = \boldsymbol{H}_t(\gamma_0, \boldsymbol{\theta}_0)$ a.s., then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

A8: If $\boldsymbol{\pi}$ is a non zero vector of $\mathbb{R}^r$ then $\boldsymbol{\pi}'\boldsymbol{x}_1$ is non-degenerate.

Remark 3 *Assumption* **A7** *is a condition for the identifiability of the model. Note that, Comte and Lieberman (2003), Hafner and Preminger (2009) and Pedersen and Rahbek (2014) give an identification condition equivalent to Assumption* **A7** *for BEKK models without covariates.*

Remark 4 *Assumption* **A8** *is an identifiability condition which is obviously necessary to avoid multicollinearity of the exogenous variables.*

Theorem 1 *Under Assumptions* **A1** - **A8**,

$$\widehat{\boldsymbol{\vartheta}}_n \to \boldsymbol{\vartheta}_0, \ \ a.s. \ \ as \ n \to \infty. \tag{12}$$

To establish the asymptotic normality of VTE the following additional assumptions are needed.

A9: The true parameter $\boldsymbol{\theta}_0$ belongs to the interior of $\boldsymbol{\Theta}_{\boldsymbol{\theta}}$.

A10: $E\left\|\boldsymbol{\varepsilon}_t\right\|^6 < \infty$ and $E\left\|\boldsymbol{x}_t\right\|^6 < \infty$.

We denote by $\alpha_X(h)$ the strong mixing coefficient of a stationary process $X = (X_t)$

$$\alpha_X(h) = \sup_{A\in\sigma(X_u, u\leq t), B\in\sigma(X_u, u\geq t+h)} |P(A\cap B) - P(A)P(B)|.$$

A11: $\boldsymbol{z}_t = (\boldsymbol{x}_t', \boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t')'$ is a $\alpha-$mixing process such that, for some $\nu > 0$ and $\delta > 0$,

$$E\|\boldsymbol{\varepsilon}_t\|^{(4+2\nu)(1+1/\delta)} < \infty, \quad E\|\boldsymbol{x}_t\|^{(4+2\nu)(1+1/\delta)} < \infty, \quad E\|\boldsymbol{\eta}_t\|^{(4+2\nu)(1+\delta)} < \infty$$

and $\sum_{h=0}^{\infty}\{\alpha_{\boldsymbol{z}}(h)\}^{\nu/(2+\nu)} < \infty$.

Let $\underline{\boldsymbol{H}}_{t,s}(\boldsymbol{\vartheta})$ be such that, for $s > 0$,

$$vec(\underline{\boldsymbol{H}}_{t,s}(\boldsymbol{\vartheta})) = \sum_{k=0}^{s}(\boldsymbol{B}^{\otimes 2})^k \left(vec(\boldsymbol{\Omega}) + \boldsymbol{A}^{\otimes 2}vec(\boldsymbol{\varepsilon}_{t-k-1}\boldsymbol{\varepsilon}_{t-k-1}') + \boldsymbol{C}^{\otimes 2}vec(\boldsymbol{x}_{t-k-1}\boldsymbol{x}_{t-k-1}')\right),$$

where $A^{\otimes 2}$ denotes the Kronecker product of a matrix A and itself. Let also $\mathcal{S}$ be a subspace such that for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, $\boldsymbol{H}_t(\boldsymbol{\vartheta}) \in \mathcal{S}$ and for all $s > 0$, $\underline{\boldsymbol{H}}_{t,s}(\boldsymbol{\vartheta}) \in \mathcal{S}$.

A12: There exists $K > 0$ such that

$$\left\|\boldsymbol{H}_t^{1/2}(\boldsymbol{\vartheta}) - \boldsymbol{H}_t^{*1/2}(\boldsymbol{\vartheta})\right\| \leq K\left\|\boldsymbol{H}_t(\boldsymbol{\vartheta}) - \boldsymbol{H}_t^{*}(\boldsymbol{\vartheta})\right\| \quad \text{for all } \boldsymbol{H}_t(\boldsymbol{\vartheta}), \boldsymbol{H}_t^{*}(\boldsymbol{\vartheta}) \in \mathcal{S}.$$

Remark 5 *The condition that the observations $\boldsymbol{\varepsilon}_t$ admit finite moment of order 6 is also found in the existing body of literature on asymptotic normality of the QMLE (see Hafner and Preminger (2009)) or the one of the VTE (see Pedersen and Rahbek (2014)) of the models without covariates. Assumption* **A10** *is needed to show the existence of moment of second-order derivatives of the log-likelihood function and its uniform convergence on the parameter space.*

Remark 6 *Under Assumption* **A1**, $(\boldsymbol{\eta}_t, \mathcal{F}_t)$ *is a conditionally homoscedastic martingale difference and* (2) *becomes a semi-strong model. The exogenous variables need not to be independent on the innovations* $\boldsymbol{\eta}_t$*. The mixing assumption in* **A11** *is used to apply the central limit theorem (CLT) of Herrndorf (1984) to specify the limiting distribution in Theorem 2. When* (2) *is a strong model,* i.e. *when* $\boldsymbol{\eta}_t$ *is iid, the moment conditions in* **A11** *can be weakened as follows*

A11*: $\boldsymbol{z}_t = (\boldsymbol{x}_t', \boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t')'$ *is a* $\alpha-$*mixing process such that, for some* $\nu > 0$, $E\|\boldsymbol{z}_t\|^{4+2\nu} < \infty$ *and* $\sum_{h=0}^{\infty}\{\alpha_{\boldsymbol{z}}(h)\}^{\nu/(2+\nu)} < \infty$.

Remark 7 *In univariate case, Assumption* **A12** *is always satisfied. Indeed, for simplicity, we consider the univariate GARCH*$(1,1)$ *model* $\varepsilon_t = \sigma_t^2(\theta)\eta_t$*, where* $\sigma_t^2(\theta) = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2(\theta)$*, with* $\omega \geq \underline{\omega} > 0$*. For any* $\sigma_t^2(\theta)$ *and* $\sigma_t^{*2}(\theta)$*, we have*

$$|\sigma_t(\theta) - \sigma_t^*(\theta)| = \left|\sqrt{\sigma_t^{*2}(\theta)} + (\sigma_t^2(\theta) - \sigma_t^{*2}(\theta))\frac{1}{2\sqrt{\bar{\sigma}_t^2(\theta)}} - \sqrt{\sigma_t^{*2}(\theta)}\right| \leq K\left|\sigma_t^2(\theta) - \sigma_t^{*2}(\theta)\right|,$$

where $\bar{\sigma}_t^2(\theta)$ *is between* $\sigma_t^2(\theta)$ *and* $\sigma_t^{*2}(\theta)$ *and the inequality follows from* $\bar{\sigma}_t^2(\theta) \geq \underline{\omega}$ *for all* θ.

Let $Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and $\ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ be obtained by replacing $\widetilde{\boldsymbol{H}}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ by $\boldsymbol{H}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ in $\widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and $\widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$. We define the following matrices.

$$\boldsymbol{J} = E\left(\frac{\partial^2\ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right), \quad \boldsymbol{K}_{\boldsymbol{\varepsilon}} = E\left(\frac{\partial^2\ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}'}\right), \quad \boldsymbol{K}_{\boldsymbol{x}} = E\left(\frac{\partial^2\ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{x}}'}\right) \tag{13}$$

and

$$\boldsymbol{\Sigma}_{11} = \sum_{h=-\infty}^{\infty} cov\left(vech(\boldsymbol{x}_t\boldsymbol{x}_t'), vech(\boldsymbol{x}_{t-h}\boldsymbol{x}_{t-h}')\right), \tag{14}$$

$$\boldsymbol{\Sigma}_{22} = \sum_{h=-\infty}^{\infty} cov\left(\boldsymbol{\Upsilon}_{0t}vec\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right), \boldsymbol{\Upsilon}_{0,t-h}vec\left(\boldsymbol{\eta}_{t-h}\boldsymbol{\eta}_{t-h}'\right)\right), \tag{15}$$

$$\boldsymbol{\Sigma}_{12} = \sum_{h=-\infty}^{\infty} cov\left(vech(\boldsymbol{x}_t\boldsymbol{x}_t'), \boldsymbol{\Upsilon}_{0,t-h}vec\left(\boldsymbol{\eta}_{t-h}\boldsymbol{\eta}_{t-h}'\right)\right), \tag{16}$$

where

$$\boldsymbol{\Upsilon}_{0t} = \begin{pmatrix} \boldsymbol{H}_{0t}^{1/2} \otimes \boldsymbol{H}_{0t}^{1/2} \\ -\frac{\partial vec'(\boldsymbol{H}_{0t})}{\partial\boldsymbol{\theta}}\left(\boldsymbol{H}_{0t}^{-1/2} \otimes \boldsymbol{H}_{0t}^{-1/2}\right)' \end{pmatrix}. \tag{17}$$

Denote by D_m and L_m the duplication matrix and elimination matrix defined such that, for any symmetric $(m \times m)$ matrix A, $vec(A) = D_m vech(A)$ and $vech(A) = L_m vec(A)$.

The following theorem gives the asymptotic distribution of VTE estimators.

Theorem 2 *Under Assumptions* **A1** - **A12**, *as* $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n} - \boldsymbol{\gamma}_{\boldsymbol{x}0} \\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0} \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Gamma\Phi\Sigma\Phi}'\boldsymbol{\Gamma}'\right), \tag{18}$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \qquad \boldsymbol{\Gamma} = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{m(m+1)/2} & \mathbf{0} \\ -\boldsymbol{J}^{-1}\boldsymbol{K}_{\boldsymbol{x}} & -\boldsymbol{J}^{-1}\boldsymbol{K}_{\boldsymbol{\varepsilon}} & -\boldsymbol{J}^{-1} \end{pmatrix} \tag{19}$$

and

$$\boldsymbol{\Phi} = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0} \\ L_m(I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1}\boldsymbol{C}_0^{\otimes 2} D_r & L_m(I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1}(I_{m^2} - \boldsymbol{B}_0^{\otimes 2}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -I_{2m^2+mr} \end{pmatrix}. \tag{20}$$

The asymptotic normality of the estimation of the original parameters is given in the following corollary.

Corollary 1 *Under the assumptions of Theorem 2, the VTE of* $\boldsymbol{\xi}_0$ *satisfies*

$$\sqrt{n}\left(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Delta\boldsymbol{\Gamma\Phi\Sigma\Phi}'\boldsymbol{\Gamma}'\Delta'\right), \tag{21}$$

where

$$\Delta = \begin{pmatrix} \Delta_1 & \Delta_2 \\ \mathbf{0}_{(2m^2+mr)\times(m^2+r^2)} & I_{(2m^2+mr)\times(2m^2+mr)} \end{pmatrix} \tag{22}$$

with

$$\Delta_1 = \begin{pmatrix} -L_m\left(\boldsymbol{C}_0 \otimes \boldsymbol{C}_0\right) D_d & L_m\left(I_{m^2} - \boldsymbol{A}_0 \otimes \boldsymbol{A}_0 - \boldsymbol{B}_0 \otimes \boldsymbol{B}_0\right) D_m \end{pmatrix},$$

$$\Delta_2 = -L_m(I_{m^2} + M_{mm})\begin{pmatrix} (\boldsymbol{A}_0\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}0} \otimes I_m) & (\boldsymbol{B}_0\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}0} \otimes I_m) & (\boldsymbol{C}_0\boldsymbol{\Sigma}_{\boldsymbol{x}0} \otimes I_m) \end{pmatrix}$$

and M_{pq} *denotes the commutation matrix such that, for any* $(p \times q)$ *matrix* A, $M_{pq}vec(A) = vec(A')$.

3.2 Estimating the asymptotic covariance matrix

In the econometric literature the nonparametric kernel estimator, also called heteroscekastic autocorrelation consistent (HAC) estimator (see Newey and West (1987), Andrews (1991) and Phillips and Jin (2003)) is widely used to estimate covariance matrix of the form $\boldsymbol{\Sigma}_{11}$. The consistent estimators $\widehat{\boldsymbol{\Sigma}}_{11n}$, $\widehat{\boldsymbol{\Sigma}}_{22n}$ and $\widehat{\boldsymbol{\Sigma}}_{12n}$ of $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ and $\boldsymbol{\Sigma}_{12}$, respectively, can thus be given by

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_{11n} &= \frac{1}{n}\sum_{t,s=1}^{n} w_{|t-s|} vech(\boldsymbol{x}_t\boldsymbol{x}_t') vech'(\boldsymbol{x}_s\boldsymbol{x}_s'),\\
\widehat{\boldsymbol{\Sigma}}_{22n} &= \frac{1}{n}\sum_{t,s=1}^{n} w_{|t-s|} \widehat{\boldsymbol{\Upsilon}}_t vec\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right) vec'\left(\boldsymbol{\eta}_s\boldsymbol{\eta}_s'\right) \widehat{\boldsymbol{\Upsilon}}_t',\\
\widehat{\boldsymbol{\Sigma}}_{12n} &= \frac{1}{n}\sum_{t,s=1}^{n} w_{|t-s|} vech(\boldsymbol{x}_t\boldsymbol{x}_t') vec'\left(\boldsymbol{\eta}_s\boldsymbol{\eta}_s'\right) \widehat{\boldsymbol{\Upsilon}}_s',
\end{aligned}$$

where $w_0, \dots, w_{n-1}$ is a sequence of weights (see Newey and West (1987), Andrews (1991) and Phillips and Jin (2003) for the problem of the choice of weights) and

$$\widehat{\boldsymbol{\Upsilon}}_t = \begin{pmatrix} \widetilde{\boldsymbol{H}}_t^{1/2}(\widehat{\boldsymbol{\vartheta}}_n) \otimes \widetilde{\boldsymbol{H}}_t^{1/2}(\widehat{\boldsymbol{\vartheta}}_n) \\ -\dfrac{\partial vec'\left(\widetilde{\boldsymbol{H}}_t(\widehat{\boldsymbol{\vartheta}}_n)\right)}{\partial \boldsymbol{\theta}} \left(\widetilde{\boldsymbol{H}}_t^{-1/2}(\widehat{\boldsymbol{\vartheta}}_n) \otimes \widetilde{\boldsymbol{H}}_t^{-1/2}(\widehat{\boldsymbol{\vartheta}}_n)\right)' \end{pmatrix}.$$

Let

$$\widehat{\boldsymbol{J}}_n = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \widetilde{\ell}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}, \quad \widehat{\boldsymbol{K}}_{\boldsymbol{x}n} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \widetilde{\ell}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{x}}'}, \quad \widehat{\boldsymbol{K}}_{\boldsymbol{\varepsilon}n} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \widetilde{\ell}_t(\widehat{\boldsymbol{\vartheta}}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}'}. \tag{23}$$

Then under the assumptions of Theorem 2, strongly consistent estimators of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are given by

$$\widehat{\boldsymbol{\Gamma}}_n = \begin{pmatrix} I_{r(r+1)/2} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & I_{m(m+1)/2} & \boldsymbol{0} \\ -\widehat{\boldsymbol{J}}_n^{-1}\widehat{\boldsymbol{K}}_{\boldsymbol{x}n} & -\widehat{\boldsymbol{J}}_n^{-1}\widehat{\boldsymbol{K}}_{\boldsymbol{\varepsilon}n} & -\widehat{\boldsymbol{J}}_n^{-1} \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_n = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11n} & \widehat{\boldsymbol{\Sigma}}_{12n} \\ \widehat{\boldsymbol{\Sigma}}_{12n}' & \widehat{\boldsymbol{\Sigma}}_{22n} \end{pmatrix} \tag{24}$$

Note that, the computation of the matrices $\widehat{\boldsymbol{\Gamma}}_n$ and $\widehat{\boldsymbol{\Sigma}}_n$ requires the evaluation of complicated first and second-order derivatives. More precisely, for $\widehat{\boldsymbol{\Sigma}}_{22n}$ and $\widehat{\boldsymbol{\Sigma}}_{12n}$ one needs to compute $\partial vec'\left(\widetilde{\boldsymbol{H}}_t(\widehat{\boldsymbol{\vartheta}}_n)\right)/\partial\boldsymbol{\theta}$. Francq and Zakoïan (2014) show that these n vectors of derivatives cannot be numerically calculated within a reasonable amount of time. They provide thus recursive formulas for a rapid computation of these derivatives.

4 Numerical illustration

In this section, we illustrate our asymptotic results of Section 3 on Monte Carlo simulations and on US stock series.

4.1 A Monte Carlo experiment

This subsection presents the results from a series of Monte Carlo experiments that allow us to evaluate the performance of the BEKK framework when the exogenous variables are introduced.

In order to reduce the computation burden of the simulations, we consider a simplified version of the bivariate BEKK-X(1,1) model (2) with $\boldsymbol{B}$ a diagonal matrix. The vector of the exogenous variables is $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t})'$ where $\boldsymbol{x}_{1t}$ and $\boldsymbol{x}_{2t}$ are two lagged values of an APARCH(1,1)

$$\begin{cases} \boldsymbol{z}_t = \sigma_t e_t, \\ \sigma_t = 0.046 + 0.027\boldsymbol{z}_{t-1}^{+} + 0.092\boldsymbol{z}_{t-1}^{-} + 0.843\sigma_{t-1}, \end{cases} \tag{25}$$

where $\sqrt{2}e_t$ is i.i.d and follows a Student distribution with 4 degrees of freedom. Two components of $\boldsymbol{\eta}_t$ are independent and normally distributed $\mathcal{N}(0,1)$. The true the parameter matrix are taken as follows

$$\boldsymbol{\Omega}_0 = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \boldsymbol{A}_0 = \begin{pmatrix} 0.15 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \boldsymbol{B}_0 = \begin{pmatrix} 0.8 & 0.0 \\ 0.0 & 0.9 \end{pmatrix}, \boldsymbol{C}_0 = \begin{pmatrix} 0.15 & 0.05 \\ 0.1 & 0.2 \end{pmatrix}. \tag{26}$$

We investigate samples with $n = 1000$ and $n = 5000$ observations. All simulations are repeated 500 times. For each data series, we simulated $(n + 500)$ observations of $\boldsymbol{\varepsilon}_t$ and then the first 500 observations are discarded in each simulation to minimize the effect of the initial values. In order to assess the statistical properties of the estimates we have computed the bias, the root mean squared error (RMSE) and the quartiles of the estimated parameters $\widehat{\boldsymbol{\xi}}_n$

$$bias\left(\widehat{\boldsymbol{\xi}}_n\right) = \frac{1}{500}\sum_{i=1}^{500}\left(\widehat{\boldsymbol{\xi}}_n^{(i)} - \boldsymbol{\xi}_0\right)$$

$$RMSE\left(\widehat{\boldsymbol{\xi}}_n\right) = \left(\frac{1}{500}\sum_{i=1}^{500}\left(\widehat{\boldsymbol{\xi}}_n^{(i)} - \bar{\boldsymbol{\xi}}\right)^2\right)^{1/2}$$

where $\widehat{\boldsymbol{\xi}}_n^{(i)}$ is the estimator at the i^{th} replication and $\bar{\boldsymbol{\xi}}$ is their empirical mean. The results of the simulation study are presented in Table 1. They are in accordance with the consistency of the VTE, in particular the medians of the estimated parameters are close to the true values. As expected, the accuracy of the estimation increases as the sample size increases from $n = 1000$ to $n = 5000$.

4.2 An application to stocks US

Is the intraday realized volatility useful for predicting the the volatility of the financial returns? In the univariate case, Francq and Thieu (2015) demonstrate that, for the capitalization stocks of American stock exchanges, yesterday's realized volatility often helps in predicting today's squared returns. Another question that we would like to investigate is whether the realized volatilities of some series returns affect their co-volatilities. The aim of this subsection is to apply the BEKK-X model in order to answer this question. For illustration purposes we restrict our attention to only 3 indices, the MSFT (Microsoft Corporation), the AAPL (Apple) and the DELL, and initially we only include one exogenous variable that is the yesterday's realized volatility of the MSFT.

The data come from Section 4.2 of Laurent et al. (2014), covering the period from January 4, 1999 to December 31, 2008 (2,489 trading days). In the end of each trading day t, the log-return in percentage ε_{kt} and the realized volatility rv_{kt} (computed as the sum of intraday squared 5-minute log-returns) are available.

With obvious notations (in particular the estimated standard deviations, obtained from the empirical estimator (24) in Section 3, are into brackets), the estimated parameters can be written as

$$\widehat{\Omega}_n^{VTE} = \begin{pmatrix} \underset{(0.0418)}{0.0218} & & \\ \underset{(0.0186)}{0.0118} & \underset{(0.0218)}{0.0257} & \\ \underset{(0.0065)}{0.0041} & \underset{(0.0047)}{-0.0070} & \underset{(0.0070)}{0.0052} \end{pmatrix}, \quad \widehat{A}_n^{VTE} = \begin{pmatrix} \underset{(0.0320)}{0.1888} & \underset{(0.0113)}{-0.0032} & \underset{(0.0184)}{-0.0015} \\ \underset{(0.0673)}{0.0062} & \underset{(0.0457)}{0.1378} & \underset{(0.0143)}{-0.0157} \\ \underset{(0.0330)}{-0.0049} & \underset{(0.0138)}{0.0455} & \underset{(0.0204)}{0.2014} \end{pmatrix}$$

$$\widehat{B}_n^{VTE} = \begin{pmatrix} \underset{(0.0142)}{0.9721} & 0 & 0 \\ 0 & \underset{(0.0006)}{0.9888} & 0 \\ 0 & 0 & \underset{(0.0056)}{0.9731} \end{pmatrix}, \quad \widehat{C}_n^{VTE} = \begin{pmatrix} \underset{(0.0523)}{0.0390} \\ \underset{(0.0126)}{0.0181} \\ \underset{(0.0264)}{0.0167} \end{pmatrix}$$

Table 1: Sampling distribution of the VTE of $\boldsymbol{\vartheta}_0$ over 500 replications for the BEKK-X(1,1) model

parameter	true val.	bias	RMSE	min	Q_1	Q_2	Q_3	max
				$n = 1,000$				
$vec(\boldsymbol{\Omega})$	0.30	-0.0024	0.0942	0.0376	0.2411	0.2952	0.3544	0.6045
	0.20	0.0047	0.1053	-0.0810	0.1409	0.1998	0.2692	0.5817
	0.40	0.0101	0.2037	0.0005	0.2842	0.4037	0.5079	2.2127
$\boldsymbol{A}$	0.15	-0.0101	0.0790	0.0000	0.0841	0.1420	0.1960	0.3428
	0.10	-0.0124	0.1404	-0.3823	0.0185	0.1094	0.2027	0.5215
	0.10	-0.0004	0.0404	-0.0447	0.0734	0.1000	0.1261	0.2087
	0.20	-0.0138	0.0770	0.0000	0.1400	0.1908	0.2394	0.4012
$diag(\boldsymbol{B})$	0.80	-0.0011	0.0353	0.6854	0.7801	0.7984	0.8224	0.9135
	0.90	-0.0030	0.0227	0.6665	0.8857	0.8998	0.9112	0.9519
$\boldsymbol{C}$	0.15	-0.0010	0.0154	0.1033	0.1390	0.1492	0.1600	0.2123
	0.10	-0.0001	0.0265	0.0079	0.0813	0.1005	0.1174	0.1814
	0.05	0.0007	0.0164	0.0014	0.0405	0.0510	0.0616	0.1082
	0.20	-0.0006	0.0251	0.1275	0.1834	0.1990	0.2147	0.2854
				$n = 5,000$				
$vec(\boldsymbol{\Omega})$	0.30	0.0015	0.0385	0.0777	0.2752	0.3008	0.3270	0.4342
	0.20	0.0028	0.0393	-0.0660	0.1770	0.2012	0.2284	0.3717
	0.40	0.0046	0.0687	0.0560	0.3600	0.4010	0.4469	0.6909
$\boldsymbol{A}$	0.15	-0.0027	0.0324	0.0392	0.1295	0.1481	0.1683	0.2403
	0.10	0.0013	0.0543	-0.0714	0.0651	0.0999	0.1352	0.2885
	0.10	0.0011	0.0168	0.0377	0.0905	0.1011	0.1118	0.1557
	0.20	-0.0019	0.0307	0.0871	0.1799	0.1992	0.2172	0.4138
$diag(\boldsymbol{B})$	0.80	-0.0013	0.0146	0.7575	0.7898	0.8000	0.8084	0.8487
	0.90	-0.0008	0.0075	0.8785	0.8943	0.8996	0.9044	0.9309
$\boldsymbol{C}$	0.15	0.0001	0.0071	0.1305	0.1456	0.1498	0.1547	0.1752
	0.10	0.0003	0.0118	0.0684	0.0927	0.0998	0.1086	0.1404
	0.05	0.0002	0.0072	0.0254	0.0457	0.0504	0.0547	0.0721
	0.20	0.0002	0.0112	0.1561	0.1928	0.2004	0.2075	0.2563

RMSE is the Root Mean Square Error, Q_i, $i = 1, 3$, denote the quartiles.

5 Conclusion

In this paper we establish the asymptotic behavior of the variance-targeting estimator of the parameters for the multivariate BEKK augmented by exogenous variables. We do not restrict the number of covariates that we want to investigate. They are introduced in the conditional covariance equation such that the positivity of the conditional covariance matrix is still assured. The model BEKK-X is reparameterized such that the unconditional covariane matrix of the observed process and the second moment matrix of the explanatory variables appear explicitly in the model equation. We demonstrate the strong consistency of the VTE under the existence of the second-order moments of the observations and the covariates. We also establish the asymptotic normality under the conditions that the process and the exogenous variables have finite sixth-order moments and that the exogenous variables follow an α-mixing process. We also provide the asymptotic distribution of the original parameters. One Monte-Carlo simulation and one empirical application illustrate the usefulness of the results.

6 Proofs

6.1 Proof of the consistency of VTE in Theorem 1

Proof of Theorem 1. The strong convergence of $\widehat{\boldsymbol{\gamma}}_n$ to $\boldsymbol{\gamma}_0$ is a direct consequence of the ergodic theorem and Assumption **A4**. To show the strong consistency of $\widehat{\boldsymbol{\theta}}_n$, it suffices to establish the following results:

i) $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}} \left| Q_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}) - \widetilde{Q}_n(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}) \right| = 0$ a.s.

ii) $E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})| \right) < \infty$ and if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $E(\ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta})) > E(\ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0))$.

iii) For any $\bar{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$, there exists a neighborhood $V(\bar{\boldsymbol{\theta}})$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in V(\bar{\boldsymbol{\theta}})} \widetilde{Q}_n(\widehat{\boldsymbol{\gamma}}_n, \boldsymbol{\theta}) > E\ell_1(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0) \text{ a.s.}$$

For notation simplicity, we denote $f_t := f_t(\boldsymbol{\gamma}, \boldsymbol{\theta})$ for any function f_t depending on parameters $(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and denote f_{0t} when $(\boldsymbol{\gamma}, \boldsymbol{\theta}) = (\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$. In the sequel, K and ϱ denote generic constants such that $K > 0$ and $\varrho \in (0, 1)$ whose exact values are unimportant.

Assumption **A6** implies that $\rho(\boldsymbol{B}^{\otimes 2}) < 1$ for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. By the compactness of $\boldsymbol{\Theta}$, we even have

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \rho(\boldsymbol{B}^{\otimes 2}) < 1. \tag{27}$$

Using the relation $vec(ABC) = (C' \otimes A)vec(B)$, the vec representation of (4) is given by

$$\begin{aligned} vec\left(\boldsymbol{H}_t\right) =& (I_{m^2} - \boldsymbol{A}^{\otimes 2} - \boldsymbol{B}^{\otimes 2})vec(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) - \boldsymbol{C}^{\otimes 2}vec(\boldsymbol{\Sigma}_{\boldsymbol{x}}) + \boldsymbol{A}^{\otimes 2}vec(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}) \\ &+ \boldsymbol{B}^{\otimes 2}vec(\boldsymbol{H}_{t-1}) + \boldsymbol{C}^{\otimes 2}vec(\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}). \end{aligned} \tag{28}$$

Iteratively using equation (28), we deduce that almost surely

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \widetilde{\boldsymbol{H}}_t - \boldsymbol{H}_t \right\| \leq K\varrho^t, \qquad \forall t. \tag{29}$$

Observe that K is a random variable that depends on the past values $\{\boldsymbol{\varepsilon}_s, \boldsymbol{x}_s; s \leq 0\}$ but does not depend on t. It can thus be considered as a constant, such as ρ. Applying the inequality $det(A+B) \geq det(A)$, for $A > 0$ and $B \geq 0$, where $\geq$ denotes that the matrix is positive semi definite, we have $\det\left(\boldsymbol{H}_t\right) > 0$, for all t and for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. It implies that $\boldsymbol{H}_t$ is invertible. Moreover, by using the equality $0 < tr\left((A+B)^{-1}\right) \leq tr\left(B^{-1}\right)$, we have

$$\|\boldsymbol{H}_t^{-1}\| \leq \|\boldsymbol{H}_t^{-1/2}\|^2 = tr\left(\boldsymbol{H}_t^{-1}\right) \leq tr\left(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} - \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{A}' - \boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{B}' - \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{C}'\right)^{-1} \leq K.$$

Hence, it yields

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \|\boldsymbol{H}_t^{-1}\| < K. \tag{30}$$

By the same arguments, $\widetilde{\boldsymbol{H}}_t$ is also invertible and, for some constant K

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \|\widetilde{\boldsymbol{H}}_t^{-1}\| < K. \tag{31}$$

We thus have almost surely,

$$\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \boldsymbol{H}_t^{-1} - \widetilde{\boldsymbol{H}}_t^{-1} \right\| \leq \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \widetilde{\boldsymbol{H}}_t^{-1} \right\| \left\| \widetilde{\boldsymbol{H}}_t - \boldsymbol{H}_t \right\| \|\boldsymbol{H}_t^{-1}\| \leq K\varrho^t. \tag{32}$$

Now

$$\begin{aligned} &\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left| Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta}) - \widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}) \right| \\ &\quad \leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left| tr\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t \left(\boldsymbol{H}_t^{-1} - \widetilde{\boldsymbol{H}}_t^{-1} \right) \right) \right| + \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left| \log \frac{\det\left(\boldsymbol{H}_t\right)}{\det\left(\widetilde{\boldsymbol{H}}_t\right)} \right| \\ &\quad \leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left(\|\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t\| \left\| \boldsymbol{H}_t^{-1} - \widetilde{\boldsymbol{H}}_t^{-1} \right\| \right) + \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left| \log \frac{\det\left(\boldsymbol{H}_t\right)}{\det\left(\widetilde{\boldsymbol{H}}_t\right)} \right|. \end{aligned} \tag{33}$$

The first sum converges to zero almost surely by using Assumption **A4**, (32) and the same arguments to show Theorem 11.7(a) in Francq and Zakoïan (2010). The convergence to zero of the second sum is also showed as on page $297-298$ of the previous reference. Therefore we obtain

$$\lim_{n\to\infty}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left|Q_n\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)-\widetilde{Q}_n\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)\right|=0\text{ a.s.}\tag{34}$$

Now we have

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\left|Q_n\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)-\widetilde{Q}_n\left(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}\right)\right|\\
&\qquad\leq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\left|Q_n\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)-Q_n\left(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}\right)\right|+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\left|Q_n\left(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}\right)-\widetilde{Q}_n\left(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}\right)\right|\\
&\qquad\leq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\left|Q_n\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)-Q_n\left(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}\right)\right|+\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left|Q_n\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)-\widetilde{Q}_n\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)\right|.
\end{aligned}\tag{35}$$

To show point i), it thus remains to show that the first term in (35) also almost surely converges to zero. For m large enough, let $V_m(\boldsymbol{\gamma}_0)$ be the open ball of center $\boldsymbol{\gamma}_0$ and radius $1/m$. Because of the consistency of $\widehat{\boldsymbol{\gamma}}_n$, for n large enough, we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\left|Q_n\left(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}\right)-Q_n\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)\right|\leq\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\gamma}\in V_m(\boldsymbol{\gamma}_0)}\left|\ell_t\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)-\ell_t\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)\right|.$$

Then

$$\begin{aligned}
\lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\left|Q_n\left(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}\right)-Q_n\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)\right|&\leq\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\sup_{\boldsymbol{\gamma}\in V_m(\boldsymbol{\gamma}_0)}\left|\ell_t\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)-\ell_t\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)\right|\\
&=E\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}}\sup_{\boldsymbol{\gamma}\in V_m(\boldsymbol{\gamma}_0)}\left|\ell_t\left(\boldsymbol{\gamma},\boldsymbol{\theta}\right)-\ell_t\left(\boldsymbol{\gamma}_0,\boldsymbol{\theta}\right)\right|
\end{aligned}$$

where the last equality follows by the ergodicity and the existence of the expectation of the term under the summation symbol. By Lebesgue's dominated convergence theorem, the latter expectation tends to zero when the neighborhood $V_m(\boldsymbol{\gamma}_0)$ shrinks to the singleton $\boldsymbol{\gamma}_0$. The point i) is proved.

We turn now to prove ii). Iteratively using equation (28) and then using (27) and Assumption **A4**, we easily get

$$E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\|vec(\boldsymbol{H}_t)\|\right)\leq\sum_{k=0}^{\infty}K\varrho^k\left(1+E\|\boldsymbol{\varepsilon}_{t-k-1}\|^2+E\|\boldsymbol{x}_{t-k-1}\|^2\right)<\infty.\tag{36}$$

Then using the inequalities $|tr(AB)| \leq \|A\|\|B\|$ and $log|A| \leq Tr(A)$, for matrix $A > 0$, we have

$$\begin{aligned}
E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\gamma},\boldsymbol{\theta})|\right) &\leq E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} \|\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\|\|\boldsymbol{H}_t^{-1}\|\right) + E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} |tr(\boldsymbol{H}_t)|\right) \\
&\leq E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} \|\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\|\|\boldsymbol{H}_t^{-1}\|\right) + \sqrt{m}E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} \|\boldsymbol{H}_t\|\right) \\
&< \infty.
\end{aligned}$$

Let $\lambda_{kt}, k = 1,\ldots,m$ be the eigenvalues of matrix $\boldsymbol{H}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)\boldsymbol{H}_t^{-1}(\boldsymbol{\gamma}_0,\boldsymbol{\theta})$. Using the same arguments used to show the point (c) in the proof of Theorem 11.7 in Francq and Zakoïan (2010), we can obtain

$$E\left(\ell_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta})\right) - E\left(\ell_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)\right) = \sum_{k=0}^{m} E(\lambda_{kt} - 1 - \log(\lambda_{kt})) \geq 0.$$

The inequality is strict unless if, for all k, $\lambda_{kt} = 1$ a.s., that is, if $\boldsymbol{H}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0) = \boldsymbol{H}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta})$ a.s. which implies that $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$. The second inequality of the point ii) is thus obtained.

It now remains to show the point iii). For any $\bar{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$, let $V_k(\bar{\boldsymbol{\theta}})$ be the open ball with center $\bar{\boldsymbol{\theta}}$ and radius $1/k$. By properties of the supremum and infimum of a function and using successively i), the ergodic theorem, the monotone convergence theorem and ii), we obtain almost surely

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in V_k(\bar{\boldsymbol{\theta}})\cap\boldsymbol{\Theta}_{\boldsymbol{\theta}}} \widetilde{Q}_n(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}) \\
&\quad\geq \liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in V_k(\bar{\boldsymbol{\theta}})\cap\boldsymbol{\Theta}_{\boldsymbol{\theta}}} Q_n(\boldsymbol{\gamma}_0,\boldsymbol{\theta}) - \limsup_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\boldsymbol{\theta}}} \left|\widetilde{Q}_n(\widehat{\boldsymbol{\gamma}}_n,\boldsymbol{\theta}) - Q_n(\boldsymbol{\gamma}_0,\boldsymbol{\theta})\right| \\
&\quad\geq \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \inf_{\boldsymbol{\theta}\in V_k(\bar{\boldsymbol{\theta}})\cap\boldsymbol{\Theta}_{\boldsymbol{\theta}}} \ell_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}) \\
&\quad= E \inf_{\boldsymbol{\theta}\in V_k(\bar{\boldsymbol{\theta}})\cap\boldsymbol{\Theta}_{\boldsymbol{\theta}}} \ell_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}) \\
&\quad> E\ell_1(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)
\end{aligned}$$

for k large enough. □

6.2 Proof of the asymptotic normality in Theorem 2

For the proof of the asymptotic distribution we need a few elementary results on the differentiation of expressions involving matrices.

If $X \in \mathcal{M}_n(\mathbb{R})$ is a symmetric matrix then

$$\frac{\partial vec(AXB)}{\partial vech(X)'} = (B' \otimes A)D_m. \tag{37}$$

If $X \in \mathcal{M}_{m\times n}(\mathbb{R})$ and $A \in \mathcal{M}_n(\mathbb{R})$ is a symmetric matrix then

$$\frac{\partial vec(XAX')}{\partial vec(X)'} = (I_{m^2} + M_{mm})(XA \otimes I_m). \tag{38}$$

Let x be a vector

$$\frac{\partial vec(Y^k)}{\partial x'} = \sum_{i=0}^{k} \left((Y')^{k-i-1} \otimes Y^i\right) \frac{\partial vec(Y)}{\partial x'}. \tag{39}$$

The proof is based on several technical lemmas.

Lemma 1 *Under Assumptions* **A1-A11***,*

$$E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\gamma},\boldsymbol{\theta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} \right\|\right) < \infty \tag{40}$$

for all $i, j = 1, \ldots, d$*.*

Proof of Lemma 1. We have

$$\frac{\partial \ell_t}{\partial \boldsymbol{\vartheta}_i} = Tr\left(\left(\boldsymbol{H}_t^{-1} - \boldsymbol{H}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{H}_t^{-1}\right)\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i}\right), \tag{41}$$

$$\begin{aligned}\frac{\partial^2 \ell_t}{\partial \boldsymbol{\vartheta}_i\partial \boldsymbol{\vartheta}_j} =& - Tr\left(\boldsymbol{H}_t^{-1}\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_j}\boldsymbol{H}_t^{-1}\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i}\right) + Tr\left(\boldsymbol{H}_t^{-1}\frac{\partial^2 \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i\partial \boldsymbol{\vartheta}_j}\right)\\ &+ 2Tr\left(\boldsymbol{H}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{H}_t^{-1}\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_j}\boldsymbol{H}_t^{-1}\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i}\right) - Tr\left(\boldsymbol{H}_t^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{H}_t^{-1}\frac{\partial^2 \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i\partial \boldsymbol{\vartheta}_j}\right).\end{aligned} \tag{42}$$

The triangle inequality and $|Tr(AB)| \leq \|A\|\|B\|$ give

$$\begin{aligned}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left|\frac{\partial^2 \ell_t}{\partial \boldsymbol{\vartheta}_i\partial \boldsymbol{\vartheta}_j}\right| \leq& \sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{H}_t^{-1}\right\|^2\left\|\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i}\right\|\left\|\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_j}\right\| + \sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{H}_t^{-1}\right\|\left\|\frac{\partial^2 \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i\partial \boldsymbol{\vartheta}_j}\right\|\\ &+ \sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} 2\left\|\boldsymbol{H}_t^{-1}\right\|^3\|\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\|\left\|\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i}\right\|\left\|\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_j}\right\| + \sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{H}_t^{-1}\right\|^2\|\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\|\left\|\frac{\partial^2 \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i\partial \boldsymbol{\vartheta}_j}\right\|.\end{aligned}$$

Note that $\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{H}_t^{-1}\right\|^3 < K$ follows from (30). Then by Hölder's inequality and assumption **A10**, the existence of the second-order moment of the score will be proved by showing

$$E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i}\right\|^3\right) = E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial vec(\boldsymbol{H}_t)}{\partial \boldsymbol{\vartheta}_i}\right\|^3\right) < \infty \tag{43}$$

and

$$E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2\boldsymbol{H}_t}{\partial\boldsymbol{\vartheta}_i\partial\boldsymbol{\vartheta}_j}\right\|^2\right) = E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 vec(\boldsymbol{H}_t)}{\partial\boldsymbol{\vartheta}_i\partial\boldsymbol{\vartheta}_j}\right\|^2\right) < \infty, \tag{44}$$

for any $i, j = 1, \ldots, d$.

Denote $\boldsymbol{a} = vec(\boldsymbol{A}), \boldsymbol{b} = vec(\boldsymbol{B})$ and $\boldsymbol{c} = vec(\boldsymbol{C})$. Using (37), we can calculate the components of the first derivation of $vec(\boldsymbol{H}_t)$ as the following

$$\frac{\partial vec(\boldsymbol{H}_t)}{\partial\boldsymbol{\gamma}'_{\boldsymbol{\varepsilon}}} = \sum_{k=0}^{\infty}\left(\boldsymbol{B}^{\otimes 2}\right)^k\left(I_{m^2} - \boldsymbol{A}^{\otimes 2} - \boldsymbol{B}^{\otimes 2}\right)D_m, \tag{45}$$

$$\frac{\partial vec(\boldsymbol{H}_t)}{\partial\boldsymbol{\gamma}'_{\boldsymbol{x}}} = -\sum_{k=0}^{\infty}\left(\boldsymbol{B}^{\otimes 2}\right)^k\boldsymbol{C}^{\otimes 2}D_m, \tag{46}$$

$$\begin{aligned}\frac{\partial vec(\boldsymbol{H}_t)}{\partial\boldsymbol{a}'} &= (I_{m^2} + M_{mm})\left(\boldsymbol{A}(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\otimes I_m\right) + \boldsymbol{B}^{\otimes 2}\frac{\partial vec(\boldsymbol{H}_{t-1})}{\partial\boldsymbol{a}'},\\ &= \sum_{k=0}^{\infty}\left(\boldsymbol{B}^{\otimes 2}\right)^k(I_{m^2} + M_{mm})\left(\boldsymbol{A}(\boldsymbol{\varepsilon}_{t-k-1}\boldsymbol{\varepsilon}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\otimes I_m\right)\end{aligned} \tag{47}$$

$$\begin{aligned}\frac{\partial vec(\boldsymbol{H}_t)}{\partial\boldsymbol{c}'} &= (I_{m^2} + M_{mm})\left(\boldsymbol{C}(\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1} - \boldsymbol{\Sigma}_{\boldsymbol{x}})\otimes I_m\right) + \boldsymbol{B}^{\otimes 2}\frac{\partial vec(\boldsymbol{H}_{t-1})}{\partial\boldsymbol{c}'},\\ &= \sum_{k=0}^{\infty}\left(\boldsymbol{B}^{\otimes 2}\right)^k(I_{m^2} + M_{mm})\left(\boldsymbol{C}(\boldsymbol{x}_{t-k-1}\boldsymbol{x}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{x}})\otimes I_m\right),\end{aligned} \tag{48}$$

and

$$\begin{aligned}\frac{\partial vec(\boldsymbol{H}_t)}{\partial\boldsymbol{b}'} = \sum_{k=1}^{\infty}&\Big(\big(\boldsymbol{A}^{\otimes 2}vec(\boldsymbol{\varepsilon}_{t-k-1}\boldsymbol{\varepsilon}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\\ &+\boldsymbol{C}^{\otimes 2}vec(\boldsymbol{x}_{t-k-1}\boldsymbol{x}'_{t-k-1} - \boldsymbol{\Sigma}_{\boldsymbol{x}})\big)'\otimes I_{m^2}\Big)\frac{\partial vec(\left(\boldsymbol{B}^{\otimes 2}\right)^k)}{\partial\boldsymbol{b}'},\end{aligned} \tag{49}$$

where, using (39),

$$\frac{\partial vec(\left(\boldsymbol{B}^{\otimes 2}\right)^k)}{\partial\boldsymbol{b}'} = \sum_{i=0}^{k}\left(\left(\boldsymbol{B}^{\otimes 2'}\right)^{k-i-1}\otimes\left(\boldsymbol{B}^{\otimes 2}\right)^i\right)\frac{\partial vec(\boldsymbol{B}^{\otimes 2})}{\partial\boldsymbol{b}'} \tag{50}$$

To show (43), it suffices to prove that $E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial\boldsymbol{H}_t}{\partial\boldsymbol{d}'}\right\|^3\right) = E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial vec(\boldsymbol{H}_t)}{\partial\boldsymbol{d}'}\right\|^3\right) < \infty$, where $\boldsymbol{d} = \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}, \boldsymbol{\gamma}_{\boldsymbol{x}}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. We immediately see from (45) and (46) that the derivatives with respect to the elements of $\boldsymbol{\gamma}$ are obviously bounded. Using the inequalities $\|A\otimes B\|_{sp} = \|A\|_{sp}\|B\|_{sp}$, $\|A\|_{sp} \leq \|A\| \leq \sqrt{m}\|A\|_{sp}$, $E\left(\sum a_i\right)^3 \leq \left(\sum (Ea_i^3)^{1/3}\right)^3$, (27) and assumption **A10**, we get

$$E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial vec(\boldsymbol{H}_t)}{\partial \boldsymbol{a}'}\right\|^3\right) \leq K\left(\sum_{k=0}^{\infty}\left(E\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\|\boldsymbol{B}^{\otimes 2}\|^{3k}\|vec(\boldsymbol{\varepsilon}_{t-k-1}\boldsymbol{\varepsilon}'_{t-k-1}-\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\|^3\right)^{1/3}\right)^3$$
$$\leq K\left(\sum_{k=0}^{\infty}\varrho^k\left(E\|\boldsymbol{\varepsilon}_{t-k-1}\boldsymbol{\varepsilon}'_{t-k-1}-\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\|^3\right)^{1/3}\right)^3 < \infty. \qquad (51)$$

Similary we also obtain

$$E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial vec(\boldsymbol{H}_t)}{\partial \boldsymbol{c}'}\right\|^3\right) < \infty. \qquad (52)$$

From (27)

$$\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\sum_{i=0}^{k}\left\|\left(\boldsymbol{B}^{\otimes 2'}\right)^{k-i-1}\otimes\left(\boldsymbol{B}^{\otimes 2}\right)^i\right\| \leq \sum_{i=0}^{k}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{B}^{\otimes 2'}\right\|^{k-i-1}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\left(\boldsymbol{B}^{\otimes 2}\right)^i\right\| < Kk\varrho^k.$$

Using the same arguments to show (51) and (52), the following result is also obtained

$$E\left(\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial vec(\boldsymbol{H}_t)}{\partial \boldsymbol{b}'}\right\|^3\right) < \infty.$$

(43) is thus shown. The second moment condition (44) is obtained by doing similar developments for the second order derivatives. □

Lemma 2 *Under Assumptions* **A1-A11**, $\boldsymbol{J}$ *is non-singular.*

Proof of Lemma 2. We apply the approach of Comte and Lieberman (2003) to prove the invertibility of the matrix $\boldsymbol{J}$. Starting by writing $\boldsymbol{J}$ as a function of $\boldsymbol{H}_t$ and of its derivatives. From (42), we have

$$E\left(\frac{\partial^2\ell_t(\gamma_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\,|\mathcal{F}_{t-1}\right) = Tr\left(\boldsymbol{H}_{0t}^{-1}\frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_i}\boldsymbol{H}_{0t}^{-1}\frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_j}\right)$$
$$= \left(\left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}vec\left(\frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_i}\right)\right)'\left(\left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}vec\left(\frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_j}\right)\right).$$

Let $\boldsymbol{u}_{ti} = \left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}\boldsymbol{v}_{ti}, \boldsymbol{v}_{ti} = vec\left(\frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_i}\right)$ and the matrices $\boldsymbol{u}_t = (\boldsymbol{u}_{t1}|\cdots|\boldsymbol{u}_{td_2})$ and $\boldsymbol{v}_t = (\boldsymbol{v}_{t1}|\cdots|\boldsymbol{v}_{td_2})$. Then $\boldsymbol{u}_t = \left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}\boldsymbol{v}_t$ and $\boldsymbol{J} = E(\boldsymbol{u}_t'\boldsymbol{u}_t)$. If $\boldsymbol{J}$ is singular, there exists $\boldsymbol{c} = (c_1,\ldots,c_{d_2})' \in \mathbb{R}^{d_2}$ such that $\boldsymbol{c} \neq 0$ and

$$\boldsymbol{c}'\boldsymbol{J}\boldsymbol{c} = \boldsymbol{c}'E(\boldsymbol{u}_t'\boldsymbol{u}_t)\boldsymbol{c} = E\left((\boldsymbol{u}_t\boldsymbol{c})'(\boldsymbol{u}_t\boldsymbol{c})\right) = 0 \text{ a.s.}$$

Due to the positivity of $(\boldsymbol{u}_t\boldsymbol{c})'(\boldsymbol{u}_t\boldsymbol{c})$, then almost surely

$$(\boldsymbol{u}_t\boldsymbol{c})'(\boldsymbol{u}_t\boldsymbol{c}) = \boldsymbol{c}'\boldsymbol{u}_t'\boldsymbol{u}_t\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{v}_t'\left(\left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}\right)'\left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}\boldsymbol{v}_t\boldsymbol{c} = 0 \text{ a.s.}$$

Because $\left(\left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}\right)'\left(\boldsymbol{H}_{0t}^{-1/2}\right)^{\otimes 2}$ is strictly positive definite with probability one, it follows that

$$\boldsymbol{v}_t\boldsymbol{c} = \sum_{i=1}^{d_2} c_i vec\left(\frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i}\right) = 0 \text{ a.s.}$$

Let $\boldsymbol{\Omega}_0^* = (I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})vec(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) - \boldsymbol{C}_0^{\otimes 2}vec(\boldsymbol{\Sigma}_{\boldsymbol{x}})$. Then we have

$$0 = \overline{\boldsymbol{\Omega}}_0 + \overline{\boldsymbol{A}}_0 vec(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}') + \overline{\boldsymbol{B}}_0 vec(\boldsymbol{H}_{0t-1}) + \overline{\boldsymbol{C}}_0 vec(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'),$$

where $\overline{\boldsymbol{\Omega}}_0 = \sum_{i=1}^{d_2} \boldsymbol{c}_i \frac{\partial \boldsymbol{\Omega}_0^*}{\partial \boldsymbol{\theta}_i}$, $\overline{\boldsymbol{A}}_0 = \sum_{i=1}^{d_2} \boldsymbol{c}_i \frac{\partial \boldsymbol{A}_0^{\otimes 2}}{\partial \boldsymbol{\theta}_i}$, $\overline{\boldsymbol{B}}_0 = \sum_{i=1}^{d_2} \boldsymbol{c}_i \frac{\partial \boldsymbol{B}_0^{\otimes 2}}{\partial \boldsymbol{\theta}_i}$, $\overline{\boldsymbol{C}}_0 = \sum_{i=1}^{d_2} \boldsymbol{c}_i \frac{\partial \boldsymbol{C}_0^{\otimes 2}}{\partial \boldsymbol{\theta}_i}$. Then $vec(\boldsymbol{H}_{0t})$ can be represented by

$$\begin{aligned} vec(\boldsymbol{H}_{0t}) = &\left(\boldsymbol{\Omega}_0^* - \overline{\boldsymbol{\Omega}}_0\right) + \left(\boldsymbol{A}_0^{\otimes 2} - \overline{\boldsymbol{A}}_0\right) vec(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}') + \left(\boldsymbol{B}_0^{\otimes 2} - \overline{\boldsymbol{B}}_0\right) vec(\boldsymbol{H}_{0t-1}) \\ &+ \left(\boldsymbol{C}_0^{\otimes 2} - \overline{\boldsymbol{C}}_0\right) vec(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'). \end{aligned}$$

Because $\boldsymbol{c} \neq 0$, we have found another representation of $vec(\boldsymbol{H}_{0t})$, which contradicts Assumption **A7**. Hence, $\boldsymbol{J}$ must be non-singular. □

Lemma 3 *Under Assumptions* **A1-A12**,

$$\sqrt{n}\left\|\frac{\partial \widetilde{Q}_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial Q_n(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\| \to 0 \textit{ in probability when } n \to \infty. \tag{53}$$

$$\sqrt{n}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} - \frac{\partial^2 Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\| \to 0 \textit{ in probability when } n \to \infty. \tag{54}$$

Proof of Lemma 3. This lemma means that the effect of the initial values on the derivatives of the criterion vanishes asymptotically. By the definition of $Q_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$ and $\widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\theta})$, (53) and (54) are entailed by showing that

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\left\|\frac{\partial \widetilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\| \to 0 \text{ in probability when } n \to \infty, \tag{55}$$

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial^2 \widetilde{\ell}_t(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'} - \frac{\partial^2 \ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\right\| \to 0 \text{ in probability when } n \to \infty. \tag{56}$$

For any $i = 1, \ldots, 3m^2$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left| \frac{\partial \widetilde{\ell}_t (\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i} - \frac{\partial \ell_t (\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i} \right| \\
&\leq \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left| Tr \left(\widetilde{\boldsymbol{H}}_{0t}^{-1} \frac{\partial \widetilde{\boldsymbol{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \boldsymbol{H}_{0t}^{-1} \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right| \\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left| Tr \left(\widetilde{\boldsymbol{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \widetilde{\boldsymbol{H}}_{0t}^{-1} \frac{\partial \widetilde{\boldsymbol{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \boldsymbol{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \boldsymbol{H}_{0t}^{-1} \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right|. \qquad (57)
\end{aligned}$$

We will show that two terms in the right-hand side of the last inequality (57) tend to zero as $n \to \infty$. For the first term, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left| Tr \left(\widetilde{\boldsymbol{H}}_{0t}^{-1} \frac{\partial \widetilde{\boldsymbol{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \boldsymbol{H}_{0t}^{-1} \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right| \\
&\leq \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left\| \widetilde{\boldsymbol{H}}_{0t}^{-1} - \boldsymbol{H}_{0t}^{-1} \right\| \left\| \frac{\partial \widetilde{\boldsymbol{H}}_{0t}}{\partial \boldsymbol{\theta}_i} \right\| + \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left\| \boldsymbol{H}_{0t}^{-1} \right\| \left\| \frac{\partial \widetilde{\boldsymbol{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right\| \\
&\leq \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \widetilde{\boldsymbol{H}}_{t}^{-1} - \boldsymbol{H}_{t}^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \frac{\partial \widetilde{\boldsymbol{H}}_{t}}{\partial \boldsymbol{\vartheta}_i} \right\| + \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \boldsymbol{H}_{t}^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \frac{\partial \widetilde{\boldsymbol{H}}_{t}}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \boldsymbol{H}_{t}}{\partial \boldsymbol{\vartheta}_i} \right\|.
\end{aligned}$$

The *vec* representation of $\widetilde{\boldsymbol{H}}_t$ is obtained by replacing $\boldsymbol{H}_t$ in (28) with $\widetilde{\boldsymbol{H}}_t$. By simple differentiation and using Assumption **A4** and (27), we can obtain

$$E \left(\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \frac{\partial \widetilde{\boldsymbol{H}}_t}{\partial \boldsymbol{\vartheta}_i} \right\| \right) < \infty \text{ and } E \left(\sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| vec \left(\frac{\partial \widetilde{\boldsymbol{H}}_t}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{\vartheta}_i} \right) \right\| \right) = O(t \varrho^t). \qquad (58)$$

Using Markov's inequality, (32) and (58), we can show that, for any $\epsilon > 0$,

$$P \left(\frac{1}{\sqrt{n}} \sum_{t=0}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \widetilde{\boldsymbol{H}}_{t}^{-1} - \boldsymbol{H}_{t}^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \frac{\partial \widetilde{\boldsymbol{H}}_{t}}{\partial \boldsymbol{\vartheta}_i} \right\| > \epsilon \right) \to 0$$

and

$$P \left(\frac{1}{\sqrt{n}} \sum_{t=0}^{n} \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \boldsymbol{H}_{t}^{-1} \right\| \sup_{\boldsymbol{\vartheta} \in \boldsymbol{\Theta}} \left\| \frac{\partial \widetilde{\boldsymbol{H}}_{t}}{\partial \boldsymbol{\vartheta}_i} - \frac{\partial \boldsymbol{H}_{t}}{\partial \boldsymbol{\vartheta}_i} \right\| > \epsilon \right) \to 0.$$

For the second term, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left| Tr \left(\widetilde{\boldsymbol{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \widetilde{\boldsymbol{H}}_{0t}^{-1} \frac{\partial \widetilde{\boldsymbol{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \boldsymbol{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \boldsymbol{H}_{0t}^{-1} \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right| \\
&\qquad \leq \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left| Tr \left(\widetilde{\boldsymbol{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \widetilde{\boldsymbol{H}}_{0t}^{-1} \left(\frac{\partial \widetilde{\boldsymbol{H}}_{0t}}{\partial \boldsymbol{\theta}_i} - \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right) \right| \\
&\qquad + \frac{1}{\sqrt{n}} \sum_{t=0}^{n} \left| Tr \left(\left(\widetilde{\boldsymbol{H}}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \widetilde{\boldsymbol{H}}_{0t}^{-1} - \boldsymbol{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \boldsymbol{H}_{0t}^{-1} \right) \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \right|.
\end{aligned}$$

Using the inequality $|Tr(AB)| \leq \|A\|\|B\|$ and $Tr(AB) = Tr(BA)$ for matrices of appropriate sizes, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\left|Tr\left(\widetilde{\boldsymbol{H}}_{0t}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\widetilde{\boldsymbol{H}}_{0t}^{-1}\left(\frac{\partial\widetilde{\boldsymbol{H}}_{0t}}{\partial\boldsymbol{\theta}_i} - \frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_i}\right)\right)\right| \\
&\qquad \leq \frac{1}{\sqrt{n}}\sum_{t=0}^{n}\left\|\widetilde{\boldsymbol{H}}_{0t}^{-1}\right\|^2\|\boldsymbol{H}_{0t}\|\|\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\|\left\|\frac{\partial\widetilde{\boldsymbol{H}}_{0t}}{\partial\boldsymbol{\theta}_i} - \frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_i}\right\| \\
&\qquad \leq K\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\|\boldsymbol{H}_t\|\|\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial\widetilde{\boldsymbol{H}}_t}{\partial\boldsymbol{\vartheta}_i} - \frac{\partial\boldsymbol{H}_t}{\partial\boldsymbol{\vartheta}_i}\right\|.
\end{aligned}$$

Then by Holdër's inequality, we can get

$$P\left(K\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\|\boldsymbol{H}_t\|\|\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\|\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}}\left\|\frac{\partial\widetilde{\boldsymbol{H}}_t}{\partial\boldsymbol{\vartheta}_i} - \frac{\partial\boldsymbol{H}_t}{\partial\boldsymbol{\vartheta}_i}\right\| > \epsilon\right) \to 0.$$

It implies that almost surely, as $n \to \infty$

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\left|Tr\left(\widetilde{\boldsymbol{H}}_{0t}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\widetilde{\boldsymbol{H}}_{0t}^{-1}\left(\frac{\partial\widetilde{\boldsymbol{H}}_{0t}}{\partial\boldsymbol{\theta}_i} - \frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_i}\right)\right)\right| \to 0. \tag{59}$$

Applying the same arguments to show (59), we also get

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{n}\left|Tr\left(\left(\widetilde{\boldsymbol{H}}_{0t}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\widetilde{\boldsymbol{H}}_{0t}^{-1} - \boldsymbol{H}_{0t}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\boldsymbol{H}_{0t}^{-1}\right)\frac{\partial\boldsymbol{H}_{0t}}{\partial\boldsymbol{\theta}_i}\right)\right| \to 0 \text{ a.s.}$$

It follows that the second term in the right-hand of (57) almost surely tends to zero. The proof of $i)$ is thus obtained. By the same arguments, (56) can be also showed. □

Lemma 4 *Under Assumptions* **A1-A11**,

$$\frac{\partial^2\widetilde{Q}_n(\boldsymbol{\vartheta}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'} \to E\left(\frac{\partial^2\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'}\right) \quad \textit{in probability when } \boldsymbol{\vartheta}_n \to \boldsymbol{\vartheta}_0 \textit{ in probability.} \tag{60}$$

Proof of Lemma 4. First note that

$$P\left(\left\|\frac{\partial^2\widetilde{Q}_n(\boldsymbol{\vartheta}_n)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'} - E\left(\frac{\partial^2\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'}\right)\right\| \geq \epsilon\right) \leq p_1 + p_2 + p_3 + p_4,$$

where

$$\begin{aligned}
p_1 &= P\left(\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta}\left\|\frac{\partial^2\widetilde{Q}_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'} - \frac{\partial^2 Q_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'}\right\| \geq \frac{\epsilon}{3}\right), \\
p_2 &= P\left(\sup_{\boldsymbol{\vartheta}\in V(\boldsymbol{\vartheta}_0)\cap\Theta}\left\|\frac{\partial^2 Q_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'} - \frac{\partial^2 Q_n(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'}\right\| \geq \frac{\epsilon}{3}\right), \\
p_3 &= P\left(\left\|\frac{\partial^2 Q_n(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'} - E\left(\frac{\partial^2\ell_t(\boldsymbol{\vartheta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\vartheta}'}\right)\right\| \geq \frac{\epsilon}{3}\right), \qquad p_4 = P\left\{\boldsymbol{\vartheta}_n \notin V(\boldsymbol{\vartheta}_0)\right\}
\end{aligned}$$

for any $\epsilon > 0$ and any neighborhood $V(\boldsymbol{\vartheta}_0)$. By the assumption that $\boldsymbol{\vartheta}_n \to \boldsymbol{\vartheta}_0$ in probability, we have $p_4 \to 0$ as $n \to \infty$. By (54), for any $\epsilon > 0$ and when $V(\boldsymbol{\vartheta}_0)$ is sufficiently small, $p_1 \to 0$. Because $\dfrac{\partial^2 \ell_t(\boldsymbol{\gamma}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\vartheta}'}$ is a function of $(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \dots)$ and $(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \dots)$, under Assumption **A2** it is strictly stationary and ergodic. The uniform law of large numbers for stationary ergodic processes and Lemma 1 thus imply that $p_3 \to 0$ for any $\epsilon > 0$. To prove that $p_2 \to 0$, it suffices to show that, for all $\epsilon > 0$, there exists $V(\boldsymbol{\vartheta}_0)$ satisfying

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\vartheta} \in V(\boldsymbol{\vartheta}_0) \cap \Theta} \left\| \frac{\partial^2 \ell_t(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} - \frac{\partial^2 \ell_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right\| \leq \epsilon \quad \text{a.s.}$$

The result follows from the ergodic theorem, the dominated convergence theorem, the uniform continuity of the second order derivatives of $\ell_t(\boldsymbol{\vartheta})$, and by Lemma 1. □

Lemma 5 *Under Assumptions* **A1 - A11**,

$$\begin{pmatrix} \sqrt{n}\left(\widehat{\gamma}_{\boldsymbol{x},n} - \gamma_{\boldsymbol{x},0}\right) \\ \sqrt{n}\left(\widehat{\gamma}_{\boldsymbol{\varepsilon},n} - \gamma_{\boldsymbol{\varepsilon},0}\right) \\ \dfrac{1}{\sqrt{n}} \sum_{t=1}^{n} \dfrac{\partial \ell_t(\gamma_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \end{pmatrix} = \boldsymbol{\Phi} \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^{n} vech(\boldsymbol{x}_t \boldsymbol{x}_t' - E\boldsymbol{x}_1 \boldsymbol{x}_1') \\ \sum_{t=1}^{n} \boldsymbol{\Upsilon}_{0t} vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \end{pmatrix} + o_p(1), \tag{61}$$

where $\boldsymbol{\Upsilon}_{0t}$ *and* $\boldsymbol{\Phi}$ *are given in* (17) *and* (20), *respectively.*

Proof of Lemma 5. Introduce the martingale difference

$$\boldsymbol{\nu}_t = vec(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') - vec(\boldsymbol{H}_{0t}) = \left(\boldsymbol{H}_{0t}^{1/2}\right)^{\otimes 2} vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m).$$

In the representation of $vec(\boldsymbol{H}_{0t})$ obtained from (2), we replace $vec(\boldsymbol{H}_{0t})$ by $vec(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') - \boldsymbol{\nu}_t$. Then, we get

$$\begin{aligned} vec\left(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' - E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')\right) = {} & \left(\boldsymbol{A}_0^{\otimes 2} + \boldsymbol{B}_0^{\otimes 2}\right) vec\left(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' - E(\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}')\right) \\ & + \boldsymbol{C}_0^{\otimes 2} vec\left(\boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}' - E(\boldsymbol{x}_{t-1} \boldsymbol{x}_{t-1}')\right) + \left(\boldsymbol{\nu}_t - \boldsymbol{B}_0^{\otimes 2} \boldsymbol{\nu}_{t-1}\right). \end{aligned}$$

Note that under assumption **A6**, the matrix $I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2}$ is inversible. Taking the average of the two side of the equality for $t = 1, \dots, n$ gives

$$\begin{aligned} \widehat{\gamma}_{\boldsymbol{\varepsilon},n} - \gamma_{\boldsymbol{\varepsilon},0} = {} & L_m (I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1} (I_{m^2} - \boldsymbol{B}_0^{\otimes 2}) \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{\nu}_t \\ & + L_m (I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1} \boldsymbol{C}_0^{\otimes 2} D_r \left(\widehat{\gamma}_{\boldsymbol{x},n} - \gamma_{\boldsymbol{x},0}\right) + o_p(1), \qquad a.s. \end{aligned}$$

Using the relation $Tr(A'B) = vec'(A)vec(B)$, we have, for $i = 1, \dots, d_2$

$$\begin{aligned}
\frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_i} &= Tr\left(\left(\boldsymbol{H}_{0t}^{-1} - \boldsymbol{H}_{0t}^{-1} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \boldsymbol{H}_{0t}^{-1} \right) \frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \\
&= vec'\left(\frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) vec\left(\left(\boldsymbol{H}_{0t}^{-1/2} \right)' (I_m - \boldsymbol{\eta}_t \boldsymbol{\eta}_t') \left(\boldsymbol{H}_{0t}^{-1/2} \right) \right) \\
&= -vec'\left(\frac{\partial \boldsymbol{H}_{0t}}{\partial \boldsymbol{\theta}_i} \right) \left(\left(\boldsymbol{H}_{0t}^{-1/2} \right)^{\otimes 2} \right)' vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m).
\end{aligned}$$

It follows that

$$\frac{\partial \ell_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = -\frac{\partial vec'(\boldsymbol{H}_{0t})}{\partial \boldsymbol{\theta}} \left(\left(\boldsymbol{H}_{0t}^{-1/2} \right)^{\otimes 2} \right)' vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \tag{62}$$

and (61) is thus obtained. □

Lemma 6 *Under Assumptions* **A1-A11***,*

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^n vech(\boldsymbol{x}_t \boldsymbol{x}_t' - E\boldsymbol{x}_1 \boldsymbol{x}_1') \\ \sum_{t=1}^n \boldsymbol{\Upsilon}_{0t} vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

where $\boldsymbol{\Sigma}$ is given in Theorem 2.

Proof of Lemma 6. From (2), we write, for $s > 0$, $\boldsymbol{H}_{0t} = \underline{\boldsymbol{H}}_{0t,s} + \overline{\boldsymbol{H}}_{0t,s}$ such that

$$vec(\overline{\boldsymbol{H}}_{0t,s}) = \sum_{k=s+1}^{\infty} (\boldsymbol{B}_0^{\otimes 2})^k \left(vec(\boldsymbol{\Omega}_0) + \boldsymbol{A}_0^{\otimes 2} vec(\boldsymbol{\varepsilon}_{t-k-1} \boldsymbol{\varepsilon}_{t-k-1}') + \boldsymbol{C}_0^{\otimes 2} vec(\boldsymbol{x}_{t-k-1} \boldsymbol{x}_{t-k-1}') \right).$$

Let $\boldsymbol{H}_{0t}^{1/2} = \underline{\boldsymbol{H}}_{0t,s}^{1/2} + R_{t,s}$. Note that $\underline{\boldsymbol{H}}_{0t,s}^{1/2}$ is invertible. Then $\boldsymbol{H}_{0t}^{-1/2} = \underline{\boldsymbol{H}}_{0t,s}^{-1/2} + R_{t,s}^*$, where $R_{t,s}^* = -\underline{\boldsymbol{H}}_{0t,s}^{-1/2} R_{t,s} \left(I + \underline{\boldsymbol{H}}_{0t,s}^{-1/2} R_{t,s} \right)^{-1} \underline{\boldsymbol{H}}_{0t,s}^{-1/2}$ (see Miller (1981) for the inverse of the sum of two matrices). Using the elementary relation $(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D$, we can write $\boldsymbol{\Upsilon}_{0t} vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) = \underline{\boldsymbol{Y}}_{t,s} + \boldsymbol{R}_{t,s}$, where

$$\underline{\boldsymbol{Y}}_{t,s} = \begin{pmatrix} \underline{\boldsymbol{H}}_{0t,s}^{1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{1/2} \\ -\frac{\partial vec'(\underline{\boldsymbol{H}}_{0t,s})}{\partial \boldsymbol{\theta}} \left(\underline{\boldsymbol{H}}_{0t,s}^{-1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{-1/2} \right)' \end{pmatrix} vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m)$$

and $\boldsymbol{R}_{t,s}$ is the rest of the development. Note that the processes $(\underline{\boldsymbol{Y}}_{t,s})_t$ and $(\boldsymbol{R}_{t,s})_t$ are stationary and centered. Using the relations $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, $(A \otimes B)' = A' \otimes B'$ and $Tr(A \otimes B) = Tr(A)Tr(B)$, we have

$$E \left\| \underline{\boldsymbol{H}}_{0t,s}^{1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right\|^2 = E\left(Tr\left(\underline{\boldsymbol{H}}_{0t,s} \otimes \underline{\boldsymbol{H}}_{0t,s} \right) \right) = E\left(Tr(\underline{\boldsymbol{H}}_{0t,s}) \right)^2 \leq KE \| \underline{\boldsymbol{H}}_{0t,s} \|^2.$$

Then using the Holder inequality and Assumption **A11**, we have, for some ν and $\delta > 0$

$$\left\| \left(\underline{\boldsymbol{H}}_{0t,s}^{1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right) vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{2+\nu} \leq \left\| \underline{\boldsymbol{H}}_{0t,s}^{1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right\|_{(2+\nu)(1+1/\delta)} \| vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \|_{(2+\nu)(1+\delta)}$$
$$\leq \left\| \underline{\boldsymbol{H}}_{0t,s} \right\|_{(2+\nu)(1+1/\delta)} \| vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \|_{(2+\nu)(1+\delta)} < \infty.$$

Similarly,

$$\left\| \frac{\partial vec'(\underline{\boldsymbol{H}}_{0t,s})}{\partial \boldsymbol{\theta}} \left(\underline{\boldsymbol{H}}_{0t,s}^{-1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{-1/2} \right)' vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{2+\nu} \leq K \left\| \frac{\partial vec'(\underline{\boldsymbol{H}}_{0t,s})}{\partial \boldsymbol{\theta}} vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \right\|_{2+\nu} < \infty.$$

It entails that $\|\underline{\boldsymbol{Y}}_{t,s}\|_{2+\nu} < \infty$. Therefore, under Assumption **A11** and s fixed, the process $(\underline{\boldsymbol{Y}}_{t,s})_t$ is strongly mixing, with mixing coefficients $\alpha_{\underline{\boldsymbol{Y}}}(h) \leq \alpha_{\boldsymbol{z}}(\max\{0, h-s\})$. Applying the CLT of Herrndorf (1984) for mixing processes, we directly obtain

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \underline{\boldsymbol{Y}}_{t,s} \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_{22,s}), \quad \boldsymbol{\Sigma}_{22,s} = \sum_{h=-\infty}^{\infty} \mathrm{cov}(\underline{\boldsymbol{Y}}_{\mathrm{t,s}}, \underline{\boldsymbol{Y}}_{\mathrm{t-h,s}}).$$

Let $\mathbf{x}_t = vech(\boldsymbol{x}_t \boldsymbol{x}_t' - E\boldsymbol{x}_1 \boldsymbol{x}_1')$ and $\boldsymbol{\Sigma}_{12,s} = \sum_{h=-\infty}^{\infty} \mathrm{cov}(\mathbf{x}_{\mathrm{t}}, \underline{\boldsymbol{Y}}_{\mathrm{t-h,s}})$. As in Francq and Zakoïan (1998) Lemma 3, we can show that the matrices $\boldsymbol{\Sigma}_{22} = \lim_{s\to\infty} \boldsymbol{\Sigma}_{22,s}$ and $\boldsymbol{\Sigma}_{12} = \lim_{s\to\infty} \boldsymbol{\Sigma}_{12,s}$ exist. Using Assumption **A12** and the arguments given in the proof of Lemma 4 in the precedent reference , one can show that

$$\lim_{s\to\infty} \limsup_{n\to\infty} P\left(\left\| n^{-1/2} \sum_{t=1}^{n} \boldsymbol{R}_{t,s} \right\| > \epsilon \right) = 0$$

for any $\epsilon > 0$. Then using Assumption **A11** and the CLT of Herrndorf (1984), we get

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^{n} vech(\boldsymbol{x}_t \boldsymbol{x}_t' - E\boldsymbol{x}_1 \boldsymbol{x}_1') \\ \sum_{t=1}^{n} \boldsymbol{\Upsilon}_{0t} vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^{n} \mathbf{x}_t \\ \sum_{t=1}^{n} \underline{\boldsymbol{Y}}_{t,s} \end{pmatrix} + \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{0} \\ \sum_{t=1}^{n} \boldsymbol{R}_{t,s} \end{pmatrix}$$
$$\xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}).$$

□

Proof of Theorem 2.

By the strong consistency, assumption **A5** and the definition of $\widehat{\boldsymbol{\theta}}_n$, for n large enough, $\widehat{\boldsymbol{\theta}}_n$ is contained $a.s.$ in an arbitrary small neighborhood of $\boldsymbol{\theta}_0$ that belongs to interior of the parameter set $\boldsymbol{\Theta}_{\boldsymbol{\theta}}$. The first-order condition

$$0 = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \widetilde{\ell}_t \left(\widehat{\gamma}_n, \widehat{\boldsymbol{\theta}}_n \right)}{\partial \boldsymbol{\theta}} \tag{63}$$

is thus satisfied.

Let us define the following matrices

$$\boldsymbol{K}_{\boldsymbol{\varepsilon}n}(\boldsymbol{\vartheta}) = \frac{\partial^2 \widetilde{Q}_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}'}. \tag{64}$$

$$\boldsymbol{K}_{\boldsymbol{x}n}(\boldsymbol{\vartheta}) = \frac{\partial^2 \widetilde{Q}_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{x}}'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{x}}'}. \tag{65}$$

$$\boldsymbol{J}_{n}(\boldsymbol{\vartheta}) = \frac{\partial^2 \widetilde{Q}_n(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\boldsymbol{\vartheta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}. \tag{66}$$

The mean-value theorem applied to each element of the right-hand side of the first-order condition gives

$$\begin{aligned}
0 =& \frac{1}{n}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\boldsymbol{\vartheta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}'}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n}-\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0}\right) + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\boldsymbol{\vartheta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}_{\boldsymbol{x}}'}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n}-\boldsymbol{\gamma}_{\boldsymbol{x}0}\right)\\
&+\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2\widetilde{\ell}_t(\boldsymbol{\vartheta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)\\
=&\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \boldsymbol{K}_{\boldsymbol{\varepsilon}n}(\boldsymbol{\vartheta}^*)\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n}-\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0}\right) + \boldsymbol{K}_{\boldsymbol{x}n}(\boldsymbol{\vartheta}^*)\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n}-\boldsymbol{\gamma}_{\boldsymbol{x}0}\right) + \boldsymbol{J}_n(\boldsymbol{\vartheta}^*)\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)
\end{aligned}$$

where $\boldsymbol{\vartheta}^*$ is between $\widehat{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}_0$. By Lemma 2, Lemma 3, Lemma 4 and the consistency of $\widehat{\boldsymbol{\vartheta}}_n$, the matrix $\boldsymbol{J}_n(\boldsymbol{\vartheta}^*)$ is a.s. invertible for sufficiently large n. Hence multiplying by $\sqrt{n}$ and solving for $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right)$ gives

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right) =& -\boldsymbol{J}_n^{-1}(\boldsymbol{\vartheta}^*)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right) - \boldsymbol{J}_n^{-1}(\boldsymbol{\vartheta}^*)\boldsymbol{K}_{\boldsymbol{\varepsilon}n}(\boldsymbol{\vartheta}^*)\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n}-\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0}\right)\\
&-\boldsymbol{J}_n^{-1}(\boldsymbol{\vartheta}^*)\boldsymbol{K}_{\boldsymbol{x}n}(\boldsymbol{\vartheta}^*)\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n}-\boldsymbol{\gamma}_{\boldsymbol{x}0}\right).
\end{aligned}$$

Hence, notice that $\dfrac{\partial\widetilde{Q}_n(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \dfrac{1}{n}\sum_{t=1}^{n}\dfrac{\partial\widetilde{\ell}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$, we have the following representation

$$\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n}-\boldsymbol{\gamma}_{\boldsymbol{x}0}\\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n}-\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0}\\ \widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\end{pmatrix} = \boldsymbol{\Gamma}_n\begin{pmatrix}\sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n}-\boldsymbol{\gamma}_{\boldsymbol{x}0}\right)\\ \sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n}-\boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0}\right)\\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\widetilde{\ell}_t(\boldsymbol{\gamma}_0,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\end{pmatrix}$$

where

$$\boldsymbol{\Gamma}_n = \begin{pmatrix} I_{r(r+1)/2} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & I_{m(m+1)/2} & \mathbf{0}\\ -\boldsymbol{J}_n^{-1}(\boldsymbol{\vartheta}^*)\boldsymbol{K}_{\boldsymbol{x}n}(\boldsymbol{\vartheta}^*) & -\boldsymbol{J}_n^{-1}(\boldsymbol{\vartheta}^*)\boldsymbol{K}_{\boldsymbol{\varepsilon}n}(\boldsymbol{\vartheta}^*) & -\boldsymbol{J}_n^{-1}(\boldsymbol{\vartheta}^*)\end{pmatrix}.$$

By Lemma 4 and the consistency of $\widehat{\boldsymbol{\vartheta}}_n$, we get

$$\boldsymbol{\Gamma}_n \to \boldsymbol{\Gamma} \text{ in probability as } n \to \infty$$

and by Lemma 3, Lemma 5 and Lemma 6,

$$\begin{pmatrix} \sqrt{n}\left(\widehat{\gamma}_{\boldsymbol{x}n} - \boldsymbol{\gamma}_{\boldsymbol{x}0}\right) \\ \sqrt{n}\left(\widehat{\gamma}_{\varepsilon n} - \boldsymbol{\gamma}_{\varepsilon 0}\right) \\ \dfrac{1}{\sqrt{n}} \sum_{t=1}^{n} \dfrac{\partial \widetilde{\ell}_t(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}'\right).$$

The asymptotic normality of the VTE now follows the Slutzky theorem. □

Proof of Corollary 1. The proof of this corollary can be obtained by applying directly the delta method (see Theorem 3.1 in van der Vaart, 1998). Indeed, let ϕ be the map which transforms $\boldsymbol{\vartheta}_0$ into $\boldsymbol{\xi}_0$. This linear map is differentiable at $\boldsymbol{\vartheta}_0$, and is described by the Jacobian matrix

$$\frac{\partial \phi}{\partial \boldsymbol{\vartheta}_0'} = \begin{pmatrix} \dfrac{\partial vec(\boldsymbol{\Omega}_0)}{\partial \boldsymbol{\gamma}_{\boldsymbol{x}0}'} & \dfrac{\partial vec(\boldsymbol{\Omega}_0)}{\partial \boldsymbol{\gamma}_{\varepsilon 0}'} & \dfrac{\partial vec(\boldsymbol{\Omega}_0)}{\partial \boldsymbol{a}_0'} & \dfrac{\partial vec(\boldsymbol{\Omega}_0)}{\partial \boldsymbol{b}_0'} & \dfrac{\partial vec(\boldsymbol{\Omega}_0)}{\partial \boldsymbol{c}_0'} \\ \dfrac{\partial \boldsymbol{a}_0}{\partial \boldsymbol{\gamma}_{\boldsymbol{x}0}'} & \dfrac{\partial \boldsymbol{a}_0}{\partial \boldsymbol{\gamma}_{\varepsilon 0}'} & \dfrac{\partial \boldsymbol{a}_0}{\partial \boldsymbol{a}_0'} & \dfrac{\partial \boldsymbol{a}_0}{\partial \boldsymbol{b}_0'} & \dfrac{\partial \boldsymbol{a}_0}{\partial \boldsymbol{c}_0'} \\ \dfrac{\partial \boldsymbol{b}_0}{\partial \boldsymbol{\gamma}_{\boldsymbol{x}0}'} & \dfrac{\partial \boldsymbol{b}_0}{\partial \boldsymbol{\gamma}_{\varepsilon 0}'} & \dfrac{\partial \boldsymbol{b}_0}{\partial \boldsymbol{a}_0'} & \dfrac{\partial \boldsymbol{b}_0}{\partial \boldsymbol{b}_0'} & \dfrac{\partial \boldsymbol{b}_0}{\partial \boldsymbol{c}_0'} \\ \dfrac{\partial \boldsymbol{c}_0}{\partial \boldsymbol{\gamma}_{\boldsymbol{x}0}'} & \dfrac{\partial \boldsymbol{c}_0}{\partial \boldsymbol{\gamma}_{\varepsilon 0}'} & \dfrac{\partial \boldsymbol{c}_0}{\partial \boldsymbol{a}_0'} & \dfrac{\partial \boldsymbol{c}_0}{\partial \boldsymbol{b}_0'} & \dfrac{\partial \boldsymbol{c}_0}{\partial \boldsymbol{c}_0'} \end{pmatrix} \tag{67}$$

where $\boldsymbol{a}_0 = vec(\boldsymbol{A}_0), \boldsymbol{b}_0 = vec(\boldsymbol{B}_0)$ and $\boldsymbol{c}_0 = vec(\boldsymbol{C}_0)$. Using (37) and (38), the components of the Jacobian matrix can be easily calculed and we get $\dfrac{\partial \phi}{\partial \boldsymbol{\vartheta}_0'} = \Delta$.

□

References

Andrews, Donald WK. (1991) Heteroskedasticity and autocorrelation consis- tent covariance matrix estimation. *Econometrica* 59, 817–858.

Boubacar Maïnassara, Y. and Francq, C. (2011) Estimating structural VARMA models with uncorrelated but non-independent error terms. *Journal of Multivariate Analysis* 102, 496–505.

Billingsley, P. (1961) The Lindeberg-Levy theorem for martingales.

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

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge (1988) A capital asset pricing model with time-varying covariances. *The Journal of Political Economy*, 116–131.

Bouissou, Michel B., Jean-Jacques Laffont, and Quan H. Vuong (1986) Tests of noncausality under Markov assumptions for qualitative panel data.*Econometrica: Journal of the Econometric Society*, 395–414.

Boussama, Farid, Florian Fuchs, and Robert Stelzer (2011) Stationarity and geometric ergodicity of BEKK multivariate GARCH models.*Stochastic Processes and their Applications* 121, 2331–2360.

Cakmakli, Cem, and Dick JC Van Dijk (2010) Getting the most out of macroeconomic information for predicting stock returns and volatility.

Christiansen, Charlotte, Maik Schmeling, and Andreas Schrimpf (2012) A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics* 27, 956–977.

Comte, Fabienne, and Offer Lieberman (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.

Engle, R.F., Hendry, D.F. and J.F. Richard (1983) Exogeneity. *Econometrica* 51, 277–304.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard (1983) Exogeneity.*Econometrica: Journal of the Econometric Society*, 277–304.

Engle, Robert F., and Kenneth F. Kroner (1995) Multivariate simultaneous generalized ARCH. *Econometric theory* 11, 122–150.

Engle, Robert (2009) *Anticipating correlations: a new paradigm for risk management.* Princeton University Press.

Florens, Jean-Pierre, and Michel Mouchart (1982) A note on noncausality.*Econometrica: Journal of the Econometric Society*, 583–591.

Francq, Christian, and Jean-Michel Zakoian (1998) Estimating linear representations of nonlinear processes.*Journal of Statistical Planning and Inference* 68, 145–165.

Francq, Christian, and Jean-Michel Zakoian (2000) Covariance matrix estimation for estimators of mixing weak ARMA models.*Journal of Statistical Planning and Inference* 83, 369–394.

Francq, Christian, Roch Roy, and Jean-Michel Zakoïan (2005) Diagnostic checking in ARMA models with uncorrelated errors.*Journal of the American Statistical Association* 100, 532–544.

Francq, C. and J-M. Zakoïan (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

Francq, Christian, Lajos Horvath, and Jean-Michel Zakoïan (2011) Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 9, 619–656.

Francq, Christian, Lajos Horváth, and Jean-Michel Zakoïan (2014) Variance targeting estimation of multivariate GARCH models.*Journal of Financial Econometrics*, https://mpra.ub.uni-muenchen.de/57794/1/MPRA_paper_57794.pdf

Francq, Christian, and Thieu, Le Quyen (2015) Qml inference for volatility models with covariates. *University Library of Munich, Germany*, No. 63198.

Francq, Christian, and Genaro Sucarrat (2015) Equation-by-Equation Estimation of a Multivariate Log-GARCH-X Model of Financial Returns.

Hafner, Christian M., and Arie Preminger (2009) On asymptotic theory for multivariate GARCH models.*Journal of Multivariate Analysis* 100, 2044–2054.

Han, H. and D. Kristensen (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.

Herrndorf, Norbert (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 141–153.

Kenneth S. Miller (1981) On the Inverse of the Sum of Matrices. *Mathematics Magazine*. http://www.jstor.org/stable/2690437, 54, 67–72

Koopmans, Tjalling C., and Olav Reiersol (1950) The identification of structural characteristics.*The Annals of Mathematical Statistics* 21, 165–181.

Mezrich, J., and R. F. Engle (1996) GARCH for groups.*Risk* 9, 36–40.

Newey, Whitney K., and Kenneth D. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica* 55, 703–708.

Pedersen, Rasmus S., and Anders Rahbek (2014) Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal* 17, 24–55.

Phillips, Peter CB, Yixiao Sun, and Sainan Jin (2003) Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. *Discussion paper, Yale University.*

Silvennoinen, Annastiina, and Timo Teräsvirta (2009) Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373–411.