Model identification using the Efficient Determination Criterion

Paulo Angelo Alves RESENDE\textsuperscript{a,*}, Chang Chung Yu DOREA\textsuperscript{a}

\textsuperscript{a}Department of Mathematics, University of Brasília, Brasília 70910-900, Distrito Federal, Brazil

Abstract

In the realm of the model selection context, Akaike’s and Schwarz’s information criteria, AIC and BIC, have been applied successfully for decades for model order identification. The Efficient Determination Criterion (EDC) is a generalization of these criteria, proposed originally to define a strongly consistent class of estimators for the dependency order of a multiple Markov chain. In this work, the EDC is generalized to partially nested models, which encompass many other order identification problems. Based on some assumptions, a class of strongly consistent estimators is established in this general environment. This framework is applied to BEKK multivariate GARCH models and, in particular, the strong consistency of the order estimator based on BIC is established for these models.

Keywords: EDC, BIC, AIC, Order estimation, BEKK-GARCH

2010 MSC: 62M05, 62F12, 60J10

1. Introduction

The order identification problem was initially dealt by using nested hypothesis tests in evaluating the order of multiple Markov chains \cite{1,2,3,4,5}, Autoregressive models \cite{6,7,8,9}, among others. In the selection model context, Akaike \textsuperscript{10} proposed the use of the information criterion AIC, aiming to avoid

\textsuperscript{*}Corresponding author

\textit{Email addresses:} pa@paulangelo.com (Paulo Angelo Alves RESENDE), chang@mat.umb.br (Chang Chung Yu DOREA)
empirical analysis on the estimation process. Subsequently, Schwarz proposed the information criterion BIC. Since then, these criteria have been applied in such contexts as selecting models in Autoregressive (AR) and Autoregressive Moving Average (ARMA) processes, estimating dependency order in multiple Markov chains, detecting change-points in non-homogeneous Markov chains, estimating the length of the hidden state space of a hidden Markov model, estimating order in Autoregressive Conditional Heteroskedasticity process (ARCH), and on estimating dependency order in specific situations.

Zhao et al. introduced the Efficient Determination Criterion (EDC), which allows for adjustments on the penalty term used in the criteria AIC and BIC. Also, a class of strongly consistent estimators was established in the same work. Afterwards, Dorea extended this class and proposed the asymptotic optimal order estimator, which had its better performance verified by the extensive use of numerical simulations.

In this work, the concept of “nested models” is generalized to class of partially nested models and the EDC criterion is extended to this new context. Some results regarding the consistency of EDC order estimators are established based essentially on assumptions about the likelihood function. This approach is applied to state the consistency of the BIC order estimator for BEKK multivariate GARCH models, which encompass the univariate version GARCH as particular case.

Section 2 provides the general results, that may be applied in a variety of models to establish the EDC order estimators. Section 3 presents the approach applied to BEKK multivariate GARCH models. The proofs of the stated results are in the appendices.
2. General framework

The essence of nested models have been used since the pioneer researches using hypothesis tests. However, practically all works focused on particular cases and the formal definition and treatment of the concept of nested models were unused. Nishii [28] firstly proposed a general estimator for the dimension of i.i.d. models. A relevant piece of Nishii’s technique is adapted to our purposes.

For an arbitrary time discrete stochastic process $X = \{X_t\}_{t \in \mathbb{N}}$, $E \subseteq \mathbb{R}^p$ the set of possible values of $X_t$ and $\nu$ a fixed measure on $E$, we define a family of statistical models for $X$ as

$$M = \{f(x^n_1, \theta, n) : \theta \in \Theta, n \geq 1\}$$

where $f(x^n_1, \theta, n)$ represents the set of possible densities for $x^n_1$ with respect to the product measure on $E^n$, which depends on the parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, and $x^n_1 = x_1 x_2 \ldots x_n$ is a realization of $X$. We may denote $f(x^n_1, \theta) = f(x^n_1, \theta, n)$ to simplify the notation.

Two statistical models

$$M_k = \{f(x^n_1, \theta, n) : \theta \in \Theta_k, n \geq 1\} \quad \text{and} \quad M_p = \{f(x^n_1, \theta, n) : \theta \in \Theta_p, n \geq 1\}$$

are nested, denoted by $M_k \subseteq M_p$, if $\Theta_k \subseteq \Theta_p$ and, for all $\theta \in \Theta_k$, $x^n_1 \in E^n$, exists $c \in (0, \infty)$ such as

$$\lim_{n \to \infty} \frac{f_k(x^n_1, \theta)}{f_p(x^n_1, \theta)} = c.$$ 

For $q \in \mathbb{N}$, $p = (p_1, \ldots, p_q) \in \mathbb{N}^q$ and $k = (k_1, \ldots, k_q) \in \mathbb{N}^q$, we define the usual order relation $p \geq k$ iff $p_i \geq k_i$ for $i = 1 \ldots q$, which makes $(\mathbb{N}, \geq)$ a partially ordered set. For $p \not\geq k$ we mean that $p < k$ or $p$ and $k$ are not related. The set $\mathcal{M} = \{M_k\}_{k \in \mathbb{N}^q}$ is a class of partially nested models if

$$M_k \subseteq M_p \iff k \leq p.$$ 

We say that an element $m_r \in \bigcup_{k \in \mathbb{N}^q} M_k$ has order $r \in \mathbb{N}^q$ if $m_r \in M_r$ and $m_r \in M_k$ implies that $M_r \subseteq M_k$. In this context, for a sample $x^n_1$ and $\hat{\theta}_k$ the maximum
likelihood estimator (MLE) of $\theta$ supposing the order $k$, \(\{L_{n,k}(x^n_1, \theta)\}\) is a class of functions \(L_{n,k} : E^n \times \Theta_k \rightarrow \mathbb{R}\) that satisfies
\[
L_{n,k}(x^n_1, \hat{\theta}_k) = \sup_{\theta \in \Theta_k} \{L_{n,k}(x^n_1, \theta)\}  \tag{1}
\]
and for $\theta \in \Theta_k$ and $p \geq k$,
\[
L_{n,p}(x^n_1, \theta) \geq L_{n,k}(x^n_1, \theta) \quad \text{and} \quad \lim_{n \to \infty} \frac{L_{n,p}(x^n_1, \theta)}{L_{n,k}(x^n_1, \theta)} < \infty.  \tag{2}
\]
To simplify notation, we shall denote $L_{n,k}(\theta) = L_{n,k}(x^n_1, \theta)$. In most situations, the $L_{n,k}$ functions are merely the likelihood for each $n$ and $k$. Now, we define the EDC estimator for a class of partially nested models.

**Definition 1.** Let $\mathcal{M}$ be a class of partially nested models, $m_r \in \bigcup_{k \in \mathbb{N}} \mathcal{M}_k$ of order $r$ and $K \geq r$. The EDC estimator is defined by
\[
\hat{r} = \arg\min_{k \leq K} \{\text{EDC}(k)\}  \tag{3}
\]
for
\[
\text{EDC}(k) = -\log L_{n,k}(\hat{\theta}_k) + c_n \gamma(k),
\]
c$_n$ a sequence of positive numbers and $\gamma(k) = \dim(\Theta_k)$.

We need the following assumptions to conclude consistency for the EDC estimator based on the asymptotic behaviour of the $c_n$ sequence. In what follows, $r$ is the order of $\mathcal{X}$, $\theta_r$ is the true parameter, i.e. the one that gives the density for the process $\mathcal{X}$ and $\hat{\theta}_k$ is the MLE of $\theta_r$ supposing the order $k$.

**Assumption A1.** For all $k \geq r$, $\theta_r$ is an interior point of $\Theta_k$ and
\[
\hat{\theta}_k \xrightarrow{a.s.} \theta_r.
\]

**Assumption A2.** For all $k, n \in \mathbb{N}$, $\log L_{n,k}(x^n_1, \theta)$ and its derivatives
\[
D^1_{\theta}(\log L_{n,k}(x^n_1, \theta)), \ D^2_{\theta}(\log L_{n,k}(x^n_1, \theta)) \text{ and } D^3_{\theta}(\log L_{n,k}(x^n_1, \theta))
\]
are measurable with respect to $x^n_1$ and continuous with respect to $\theta$. 

---

4
Assumption A3. For \( k \geq r \), there exists \( c < \infty \) and a symmetric and positive definite matrix \( A_2 \), such as, for all \( \hat{\theta} = (1 - s)\hat{\theta}_k + s\theta_r, \ s \in (0, 1), (i, j, l) \in \{1, \ldots, \gamma(k)\}^3 \),
\[
\lim_{n \to \infty} \frac{D^3_\hat{\theta}(\log L_{n,k}(\hat{\theta}))}{n} < c \quad \text{a.s. and}
\]
\[
\lim_{n \to \infty} \frac{D^2_\hat{\theta}(\log L_{n,k}(\hat{\theta}))}{n} = A_2 \quad \text{a.s.}
\]

Assumption A4. If \( k \geq r \),
\[
\limsup_{n \to \infty} \left\| \frac{D_1^4 \log L_{n,k}(\theta_r)}{\sqrt{2n \log \log n}} \right\| < \infty \quad \text{a.s.} \tag{4}
\]

Assumption A5. If \( k \not\geq r \),
\[
0 < \lim_{n \to \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,p}(\hat{\theta}_k)}{n} \quad \text{a.s.}
\]

We note that Basawa & Heyde [29] propose basically the use of assumptions A1-A3 and an analogous of A4 to conclude the asymptotically normality of the parameter estimator \( \hat{\theta}_k \). The approach used here is quite similar to their. The next result establishes the class of consistent EDC order estimators based on the assumptions and on the asymptotic behaviour of the sequence \( c_n \). The proof is in the Appendix A. The Corollary concludes the strong consistency for the BIC order estimator defined in this general context.

Theorem 1. Let \( \mathcal{X} \) be a discrete time stochastic process taking values in \( \mathbb{R}^m \), \( \mathcal{M} \) its respective class of partially nested models, \( m_r = \bigcup_{k \in \mathbb{N}} M_k \) of order \( r \), \( \hat{r} \) as defined in (3), and the assumptions A1-A5 are satisfied. Then \( \hat{r} \to r \) if
\[
\liminf_{n \to \infty} \frac{c_n}{\log \log n} = \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{c_n}{n} = 0.
\]

Corollary 1. Supposing the same hypothesis of Theorem 7, the BIC order estimator bellow is strongly consistent.
\[
\hat{r}_{bic} = \argmin_{k \leq K} \left\{ -\log L_{n,k}(\hat{\theta}_k) + \frac{\log n}{2} \gamma(k) \right\}
\]
for a known \( K \geq r \).
The proof consists in the determination of the asymptotic behaviour of the differences below for arbitrary \( p \geq r \) and \( k > r \).

\[
\frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,p}(\hat{\theta}_p)}{n}
\]

(5)

\[
\frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r)}{\log \log n}
\]

(6)

Assumption A5 is precisely (5). For (6) we use assumptions A1 and A2 to enable the use of Taylor series and state

\[
\limsup_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\theta_r)}{\log \log n} \leq \limsup_{n \to \infty} \frac{\sqrt{n}(\theta_r - \hat{\theta}_k) - D^2_\theta \left( \log L_{n,k}(\hat{\theta}_k) \right) \left( \sqrt{n}(\theta_r - \hat{\theta}_k) \right)'}{\sqrt{2 \log \log n}}
\]

and

\[
\frac{1}{\sqrt{n \log \log n}} D^1_\theta \log L_{n,k}(\theta_r) \left( -\frac{D^2_\theta \left( \log L_{n,k}(\hat{\theta}_k) \right)}{n} \right)^{-1} = \frac{\sqrt{n}}{\sqrt{\log \log n}} (\hat{\theta}_k - \theta_r).
\]

Using A3 and A4 we define a upper bound for (6).

Defining the asymptotic behaviour of \( D^1_\theta \log L_{n,k}(\theta_r) \) is generally easier when compared to the effort in manipulating directly the equation (6) to state its asymptotic behaviour. Besides that, assumptions A1-A3 are commonly used to establish the asymptotic normality for the MLE and usually are available in the literature.

3. BEKK-GARCH order estimation

Engle [30] originally proposed the use of ARCH models on modelling time series in economy. His work has been hugely influential in the area and motivated many generalizations and/or adaptations such as GARCH [31], NGARCH [32], EGARCH [33] and the multivariate generalizations BEKK-GARCH [34], VEC-GARCH [35], CCC-GARCH [36], Factor-GARCH [37], among others.

The multivariate models have special applications in portfolio selection and asset pricing. In this family, the BEKK-GARCH models has particular relevance due to its generality and the amount of research available in the literature.
Among the mentioned, only the VEC-GARCH is more general than the BEKK-GARCH model. However, the VEC-GARCH cases that can not be represented in the BEKK-GARCH parametrization are somewhat degenerated \[38, 39\].

Boussama \[40\] immersed the BEKK-GARCH models into the framework of general state space Markov chains and used algebraic topology to conclude the geometric ergodicity of such models under regularity conditions. This work is also published in \[38\] with minor changes.

Comte & Lieberman \[41\] used Boussama’s results to prove the consistency conditions proposed by Jeantheau \[42\] and the conditions proposed by Basawa & Heyde \[29\] and conclude the strong consistency and asymptotic normality for the MLE of the parameter \(\theta_r\).

As with the particular case of ARCH models, until now, there is no results regarding consistency of order estimators for BEKK-GARCH models. However, the AIC and BIC information criteria have been used without further formalization. In what follows, we present some preliminary results, which are used to prove the assumptions A1-A5 and conclude the consistency of the EDC order estimator \(\hat{r}\) for such models, which encompass the consistency of the BIC order estimator as a particular case.

For \(k = (k_1, k_2) \in \mathbb{N}^2\), a random sequence \(X = \{X_t\}_{t \in \mathbb{N}}\) taking values in \(\mathbb{R}^m\) is a BEKK-GARCH(\(k\)) model if for all \(t \in \mathbb{N}\),

\[
X_t = (H_t)^{1/2} \varepsilon_t,
\]

where, for \(m \times m\) matrices \(C, \{A_{ls}\}\) and \(\{B_{ls}\}\), \(C\) positive definite and \(N \in \mathbb{N}\),

\[
H_t = C + \sum_{l=1}^{k_2} \left( \sum_{s=1}^{N} A_{ls} X_{t-l} X_{t-l}^\prime A_{ls}^\prime \right) + \sum_{l=1}^{k_1} \left( \sum_{s=1}^{N} B_{ls} H_{t-l} B_{ls}^\prime \right),
\]

\(\{\varepsilon_t\}_{t \in \mathbb{N}} \sim \mathcal{N}(0, I_m)\), and \(I_m\) is the \(m \times m\) identity matrix. The process \(X\) can be represented as a Markov chain \(Y = \{Y_t\}_{t \in \mathbb{N}}\) defined by

\[
Y_t = (vech(H_{t+1})^\prime, vech(H_t)^\prime, \ldots, vech(H_{t-k_1+2})^\prime, X_t^\prime, X_{t-1}^\prime, \ldots, X_{t-k_2+1}^\prime)^\prime,
\]

where \(vech\) is the operator that stacks the lower triangular portion of a matrix. Boussama proved that \(Y\) is a positive Harris and geometric ergodic Markov
chain if
\[ \rho \left( \sum_{l=1}^{k_2} \tilde{A}_l + \sum_{l=1}^{k_1} \tilde{B}_l \right) < 1, \]  
(8)
where \( \rho \) is the spectral radius,
\[ \tilde{A}_l = D_m^+ \sum_{s=1}^{N} (A_{ls} \otimes A_{ls}) D_m \quad \text{and} \quad \tilde{B}_l = D_m^+ \sum_{s=1}^{N} (B_{ls} \otimes B_{ls}) D_m, \]
\( \otimes \) is the Kronecker product, \( D_m \) and \( D_m^+ \) are defined by the operators that satisfy
\[ \vec{\text{vec}}(A) = D_m \text{vech}(A) \quad \text{and} \quad \text{vech}(A) = D_m^+ \text{vec}(A). \]

For \( k = (k_1, k_2) \) and a fixed \( N \geq 1 \), the BEKK-GARCH(k) model can be immersed in a class of partially ordered models considering, for \( p > k \) and \( \theta_k \in \Theta_k, \Theta_k \subset \Theta_p \), the matrices \( \{A_{ls}(\theta_k)\}, \{B_{ls}(\theta_k)\} \) and \( C(\theta_k) \) for \( l = 1 \cdots k_2 \), \( l' = 1 \cdots k_1 \) and \( s = 1 \cdots N \). Denoting \( \bar{k} = \max(k_1, k_2) \), we have
\[ f(x^n_t, \theta_k) = C_1(x^n_{\bar{k}}) \prod_{t=1+\bar{k}}^{n} \frac{1}{\sqrt{(2\pi)^{m/2} \det(H_{t, \theta_k})}} \exp \left( -\frac{1}{2} x_{t}' H_{t, \theta_k}^{-1} x_t \right) \]
for
\[ H_{t, \theta_k} = C + \sum_{l=1}^{k_2} A_l X_{t-l} X'_{t-l} A_l' + \sum_{l=1}^{k_1} B_l H_{t-l, \theta_k} B_l'. \]

Also, the following definition for the functions \( \log L_{n,k} \) satisfies (11) and (2).
\[ \log L_{n,k}(\theta_k) = \sum_{1+\bar{k}}^{n} l_t(\theta_k) \]
for
\[ l_t(\theta_k) = -\frac{1}{2} X_{t}' H_{t, \theta_k}^{-1} X_t - \frac{1}{2} \log \det(H_{t, \theta_k}). \]

The nesting relation \( M_k \subset M_p \) can be observed taking \( A_{ls} \) and \( B_{ls} \) as null matrices for \( l > k_2 \) and \( l' > k_1 \).

In particular, if \( N = 1 \) and \( \theta_k \) is the columns of the matrices \( \{A_l\}, \{B_{l'}\} \) and \( C \), we may construct \( \Theta_k \subset \mathbb{R}^{m^2(2k+1)} \),
\[ \Theta_k = \Omega_0 \times \Omega_1 \times \cdots \times \Omega_{2k}, \]
for $\Omega_i = \{0\}^m$ if $i/2 > k_2$ and $i$ is odd or $i/2 > k_1$ and $i$ is even, for the remaining cases, $\Omega_i \subseteq \mathbb{R}^m$ has non-empty interior. Assuming $A_i = 0$ if $i > k_2$, $B_i = 0$ if $i > k_2$,

$$\theta_k = (\text{vec}(C), \text{vec}(A_1), \text{vec}(B_1), \ldots, \text{vec}(A_{\bar{k}}), \text{vec}(B_{\bar{k}})) \in \Theta_k.$$  

In this case,

$$\gamma(k) = m^2(1 + k_1 + k_2).$$

For the order, $r$, of a BEKK-GARCH process $X$, we consider the lowest $k$ such as $X$ can be represented by (7). Assuming the following conditions (B1-B5), we establish the Theorem 2 that concludes assumptions A1-A5 and states the class of strong consistent EDC order estimators. For $k \geq r$,

**B1.** $\Theta_k$ is compact and $\theta_r$ is an interior point of $\Theta_k$.

**B2.** There exists a $c > 0$ such as $\inf_{\theta \in \Theta_k} \det C(\theta) > c$.

**B3.** The model is identifiable, i.e. $H_t(\theta) = H_t(\theta')$ a.s. if and only if $\theta = \theta'$.

**B4.** $C(\theta)$, $\tilde{A}_i(\theta)$ and $\tilde{B}_i(\theta)$ and their derivatives, with respect to $\theta$, until order 3 are continuous.

**B5.** $X_t$ admits bounded moments of order 16.

Comte & Lieberman use B1-B4 and finite moments of order 8 in B5 to conclude the asymptotic normality of $\hat{\theta}_k$. We need the finiteness for moments of order 16 to conclude assumption A4.

**Theorem 2.** Let $X$ be a BEKK-GARCH($r$) of order $r$, satisfying (8) and conditions B1-B5. Then the EDC order estimator defined at (3) is strongly consistent if

$$\lim_{n \to \infty} \frac{c_n}{\log \log n} = \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{c_n}{n} = 0.$$  

Boussama [40] concluded the geometric ergodicity for the associated Markov chain of $X$ and, in particular, enabled the use of the Strong Law of Large Numbers (SLLN), which can be found in Meyn & Tweedie [43]. However, the geometric ergodicity is not sufficient to conclude the Law of Iterated Logarithm.
(LIL), needed to prove assumption A4. To overcome this, we use the LIL below, stated for square integrable Martingales, which can be found in [44]. Also, some auxiliary results stated by Comte & Lieberman [41] are used to conclude assumptions A1-A5.

**Theorem 3** (Hall & Heyde (1980)). Let \( \{S_n, F_{t-1}\} \) be a martingale, \( S_n = \sum_{t=1}^{n} U_t, E(S_n) = 0, E(S_n^2) < \infty, \) \( \{Z_t\}_{t\in\mathbb{N}} \) and \( \{W_n\}_{n\in\mathbb{N}} \) non-negative random variables such as \( Z_t \) and \( W_t \) are \( F_{t-1} \) measurable. If

\[
\lim_{n \to \infty} \frac{\sum_{t=1}^{n} U_t I(|U_t| > Z_t) - E[U_t I(|U_t| > Z_t)|F_{t-1}]}{\sqrt{2W_n^2 \log \log W_n^2}} = 0 \quad \text{a.s.}, \quad (L1)
\]

\[
\lim_{n \to \infty} \frac{\sum_{t=1}^{n} E[U_t^2 I(|U_t| \leq Z_t)|F_{t-1}] - E[U_t I(|U_t| \leq Z_t)|F_{t-1}]}{W_n^2} = 1 \quad \text{a.s.}, \quad (L2)
\]

\[
\lim_{n \to \infty} \sum_{t=1}^{n} \frac{E[U_t^4 I(|U_t| \leq Z_t)|F_{t-1}]}{W_n^4} < \infty \quad \text{a.s.}, \quad (L3)
\]

\[
\lim_{n \to \infty} \frac{W_n}{W_{n+1}} = 1 \quad \text{a.s. and} \quad \lim_{n \to \infty} W_n = \infty \quad \text{a.s.} \quad (L4)
\]

Then

\[
\limsup_{n \to \infty} \frac{S_n}{\sqrt{2W_n^2 \log \log W_n^2}} = 1 \quad \text{a.s.}
\]

and

\[
\liminf_{n \to \infty} \frac{S_n}{\sqrt{2W_n^2 \log \log W_n^2}} = -1 \quad \text{a.s.}
\]

**4. Conclusion**

The Efficient Determination Criterion (EDC) raises as a promising approach in the context of partially nested models. Mainly because the assumptions A1-A5 simplify the establishment of strongly consistent order estimators in a variety of models. Some of these assumptions, for each case, can be found in the literature on defining the asymptotic normality for the respective MLE.
Hafner & Preminger [45] state some results for VEC-GARCH(1,1) models. If it is possible to generalize these results for arbitrary $k \in \mathbb{N}^2$, the EDC order estimator can be easily defined for VEC-GARCH models.

As future works, we suggest to weaken the hypothesis B5 and state the consistency of the EDC estimator for $c_n = O(\log \log n)$.

Acknowledgement

The authors gratefully acknowledge the attention of Dr. Farid Boussama in sending his thesis and indirectly encouraging the establishment of the EDC for BEKK-GARCH models.

Appendix A. Proof of Theorem 1

The following Lemma is an adaptation of results that can be found at [26].

Lemma 1. Let $\mathbb{X}$ be a discrete time stochastic process with values in $\mathbb{R}^m$, $\mathbb{M}$ its respective class of partially nested models, $m_r \in \bigcup_{k=0}^{\infty} M_k$ of order $r$ and $\hat{r}$ as defined in (3). Then $\hat{r}$ is strongly consistent ($\hat{r} \to r$) if, for $k \not\geq r$, exists $c_1 \in (0, \infty)$ such as

$$\lim_{n \to \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,k}(\hat{\theta}_k)}{n} \geq c_1 \text{ a.s.}, \quad (A.1)$$

for $k > r$, exists $c_2 \in (0, \infty)$ such as

$$\limsup_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r)}{\log \log n} \leq c_2(\gamma(k) - \gamma(r)) \text{ a.s.} \quad (A.2)$$

and $c_n$ satisfies

$$\lim_{n \to \infty} \frac{c_n}{n} = 0 \quad \text{and} \quad \liminf_{n \to \infty} \frac{c_n}{\log \log n} \geq c_2. \quad (A.3)$$

Proof. We have that

$$\left( - \log L_{n,p}(\hat{\theta}_p) + \gamma(p)c_n \right) - \left( - \log L_{n,l}(\hat{\theta}_l) + \gamma(l)c_n \right)
= \left( \log L_{n,l}(\hat{\theta}_l) - \log L_{n,p}(\hat{\theta}_p) \right) - c_n(\gamma(l) - \gamma(p)). \quad (A.4)$$
Taking \( p = r \) and \( l = k \) in (A.4) and using (A.2) we get

\[
\limsup_{n \to \infty} \left( -\log L_{n,k}(\hat{\theta}_r) + \gamma(r)c_n \right) - \left( -\log L_{n,k}(\hat{\theta}_k) + \gamma(k)c_n \right) \log \log n
\]

\[
\leq c_2(\gamma(k) - \gamma(r)) - \liminf_{n \to \infty} \left( \frac{c_n}{\log \log n} \right) (\gamma(k) - \gamma(r)) \quad a.s.
\]

\[
\leq c_2(\gamma(k) - \gamma(r)) - c_2(\gamma(k) - \gamma(r)) \quad a.s.
\]

\[= 0.\]

In the same manner, but taking \( l = r \) and \( p = k < r \) in (A.4), and using (A.1) and (A.3), we have

\[
\liminf_{n \to \infty} \left( -\log L_{n,k}(\hat{\theta}_k) + \gamma(k)c_n \right) - \left( -\log L_{n,k}(\hat{\theta}_r) + \gamma(r)c_n \right) \log \log n
\]

\[
\geq c_1 - \limsup_{n \to \infty} \frac{c_n}{n} (\gamma(r) - \gamma(k)) \quad a.s.
\]

\[> 0 \quad a.s.
\]

Then, using its definition, we conclude that \( \hat{r} \to r \). \( \square \)

**Lemma 2.** Let \( \mathcal{X} \) be a discrete time stochastic process with values in \( \mathbb{R}^m \), \( r \) its order, \( \mathcal{M} \) its respective class of partially nested models, the \( \log L_{n,k} \) functions as defined above and \( \hat{\theta}_k \in \Theta_k \) the MLE of the true parameter \( \theta_r \in \Theta_k \). If assumptions A1-A5 are true, then

\[
\limsup_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r)}{\log \log n} \leq \frac{2c^2}{\lambda_{\gamma(k)}} \quad a.s.,
\]

where \( \lambda_{\gamma(k)} \) is the lowest eigenvalue of \( A_2 \) and \( c \) is an upper bound for (4).

**Proof.** Using A1-A2, for large enough \( n \), we can take the Taylor expansion of \( \log L_{n,k}(\theta_r) \) at \( \hat{\theta}_k \), which gives

\[
\log L_{n,k}(\theta_r) = \log L_{n,k}(\hat{\theta}_k) + (\theta_r - \hat{\theta}_k)D_{\theta}^1 \left( \log L_{n,k}(\hat{\theta}_k) \right)
\]

\[
+ \frac{1}{2}(\theta_r - \hat{\theta}_k)D_{\theta}^2 \left( \log L_{n,k}(\hat{\theta}_k) \right) (\theta_r - \hat{\theta}_k)^T + r_n(\theta_r - \hat{\theta}_k) \quad (A.5)
\]

where, for \( \theta_r = (\alpha_1, \ldots, \alpha_{\gamma(k)}) \), \( \hat{\theta}_k = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{\gamma(k)}) \) and \( \hat{\theta} = (1 - s)\theta_k + s\theta_r \), \( s \in (0, 1) \),

\[
r_n(\theta_r - \hat{\theta}_k) = \frac{1}{3!} \sum_{i,j,l} \left( D_{\theta}^3 \log L_{n,k}(\hat{\theta}) \right)_{i,j,l} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)(\alpha_l - \hat{\alpha}_l).
\]
By definition, \( \hat{\theta}_k \) maximizes \( L_{n,k} \), which gives \( D_\hat{\theta}^1 \left( \log L_{n,k}(\hat{\theta}_k) \right) = 0 \). Organizing (A.5) and dividing by \( \log \log n \), we have

\[
\limsup_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\theta_r)}{\log \log n} \leq \limsup_{n \to \infty} \frac{\sqrt{n}(\theta_r - \hat{\theta}_k) - D_\hat{\theta}^2 \left( \log L_{n,k}(\hat{\theta}_k) \right) \left( \sqrt{n}(\theta_r - \hat{\theta}_k) \right)^T}{2 \log \log n} + \limsup_{n \to \infty} \frac{|r_n(\theta_r - \hat{\theta}_k)|}{\log \log n} \quad (A.6)
\]

Now, taking the Taylor expansion of \( D_\theta^1 \log L_{n,k}(\hat{\theta}_k) \) at \( \theta_r \),

\[
(0, \cdots, 0) = D_\theta^1 \log L_{n,k}(\hat{\theta}_k) = D_\theta^1 \log L_{n,k}(\theta_r) + (\hat{\theta}_k - \theta_r) D_\theta^2 \log L_{n,k}(\hat{\theta}_k),
\]

where \( \hat{\theta} = s\theta_r + (1-s)\hat{\theta}_k \) and \( s \in (0,1) \). Organizing, we have

\[
\frac{1}{\sqrt{n \log \log n}} D_\theta^1 \log L_{n,k}(\theta_r) = \frac{\sqrt{n}}{n \sqrt{\log \log n}} \left\{ (\hat{\theta}_k - \theta_r) D_\theta^2 \log L_{n,k}(\hat{\theta}_k) \right\}
\]

\[
= -\frac{\sqrt{n}}{\sqrt{\log \log n}} (\hat{\theta}_k - \theta_r) \left[ D_\theta^2 \log L_{n,k}(\hat{\theta}_k) \right]_n.
\]

Using that \( A_2 \) is positive definite we conclude that it is invertible and, for large enough \( n \),

\[
A_n := -\left[ D_\theta^2 \log L_{n,k}(\hat{\theta}_k) \right]_n
\]

has inverse \( A_{n}^{-1} \), then

\[
\frac{1}{\sqrt{n \log \log n}} D_\theta^1 \log L_{n,k}(\theta_r) A_{n}^{-1} = \frac{\sqrt{n}}{\sqrt{\log \log n}} (\hat{\theta}_k - \theta_r). \quad (A.7)
\]

Using A4 and A41, considering \( P_i : \mathbb{R}^{\gamma(k)} \to \mathbb{R} \) as the projection of coordinate \( i \), we have

\[
\limsup_{n \to \infty} \frac{\sqrt{n}(\hat{\alpha}_i - \alpha_i)}{\sqrt{2 \log \log n}} = \limsup_{n \to \infty} \left| P_i \left( \frac{\sqrt{n}}{\sqrt{2 \log \log n}} (\hat{\theta}_k - \theta_r) \right) \right| = \limsup_{n \to \infty} \left| P_i \left( \frac{1}{\sqrt{2n \log \log n}} D_\theta^1 \log L_{n,k}(\theta_r) A_{n}^{-1} \right) \right| \leq \limsup_{n \to \infty} \left| P_i \left( \frac{1}{\sqrt{2n \log \log n}} D_\theta^1 \log L_{n,k}(\theta_r) A_{n_2}^{-1} \right) \right| < \infty,
\]

13
which gives, using A1 and A2,

\[
\limsup_{n \to \infty} \frac{|r_n(\theta_r - \hat{\theta}_k)|}{\log \log n} \\
\leq \limsup_{n \to \infty} \frac{1}{3^i} \sum_{i,j,l} \left| \left| \frac{D^3 \log L_{n,k}(\theta)}{n} \right|_{i,j,l} \right| \frac{\sqrt{\log \log n}}{\sqrt{\log \log n}} |\alpha_i - \hat{\alpha}_i| \\
\leq c \sum_{t} \limsup_{n \to \infty} |\alpha_t - \hat{\alpha}_t| = 0 \quad a.s. \quad (A.8)
\]

Using A3, (A.6), (A.7), (A.8) and that \( A_n \to A_2 \), we have

\[
\limsup_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\theta_r)}{\log \log n} \\
\leq \limsup_{n \to \infty} \frac{\left( D^1 \log L_{n,k}(\theta) \right) A_{n,k}^{-1} - D^2 \left( \log L_{n,k}(\theta) \right) A_{n,k}^{-1}}{\sqrt{2n \log \log n}} \frac{T}{\sqrt{2n \log \log n}} \\
= \limsup_{n \to \infty} \frac{D^1 \log L_{n,k}(\theta)}{\sqrt{2n \log \log n}} A_{n,k}^{-1} A_{n,k} \left( D^2 \log L_{n,k}(\theta) \right) \frac{T}{\sqrt{2n \log \log n}} \\
= \limsup_{n \to \infty} \frac{D^1 \log L_{n,k}(\theta)}{\sqrt{2n \log \log n}} \frac{T}{\sqrt{2n \log \log n}} \\
\leq \frac{1}{\lambda_{\gamma(k)}} \limsup_{n \to \infty} \left\| \frac{D^1 \log L_{n,k}(\theta)}{\sqrt{2n \log \log n}} \right\|^2 \\
\leq \frac{c^2}{\lambda_{\gamma(k)}} \quad a.s. \quad (A.9)
\]

We used that \( A_2 \) is symmetric positive definite so is its inverse. Applying (A.9) twice, we conclude the proof. \( \square \)

Using Assumption A5 and Lemma 2 we have (A.1) and (A.2). The Theorem follows from Lemma 4.

Appendix B. Proof of Theorem 2

Lemma 3 (Comte & Lieberman (2003)). Let \( \mathcal{X} = \{X_t\}_{t \in \mathbb{N}} \) be a BEKK-GARCH\( (k) \), \( \theta_r = (\alpha_1, \ldots, \alpha_{\gamma(k)}) \) its true parameter, \( \hat{\theta}_k = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{\gamma(k)}) \) the MLE of \( \theta_r \). If conditions (iii) and B1-B5 are true, then

(i) \( \lim_{n \to \infty} \frac{D^2 \log L_{n,k}(\theta_r)}{n} = A_2 \quad a.s. \),

where

\[
A_2 = -E \left( \frac{\partial^2 l_t(\theta_r)}{\partial \theta \partial \theta'} \right). \quad (B.1)
\]
(ii) $A_2$ is positive definite.

(iii) For all $i, j, l \in \{1, \ldots, \gamma(k)\}$,
\[ E \left( \sup_{\|\theta - \theta_\ast\| \leq \delta} \left| \frac{\partial^3 l_i(\theta)}{\partial \alpha_i \alpha_j \alpha_l} \right| \right) < c(\delta). \]

(iv) For all $i \in \{1, \ldots, \gamma(k)\}$, $\frac{\partial \log L_{n,k}(\theta)}{\partial \alpha_i}$ is a square-integrable Martingale.

(v) The MLE $\hat{\theta}_k$ is strongly consistent.

(vi) Exists $c \in (0, \infty)$, which does not depend on $t$ or $\theta$, such as
\[ \|H_t^{-1}\| \leq c. \]

(vii) $E \left( \log \left( \det(\hat{H}_t(\theta_\ast)) \right) \right) < \infty.$

The following Lemma adapts some results from Comte & Lieberman to our purposes.

Lemma 4. Let $X = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH($k$), $\theta_\ast = (\alpha_1, \ldots, \alpha_{\gamma(k)})$ its true parameter, $\log L_{n,k}$ as defined early, $\hat{\theta}_k \in \Theta_k$ the MLE of $\theta_\ast$, $\hat{\theta} = s\theta_\ast + (1-s)\hat{\theta}_k$ and $s \in [0, 1]$ and $B_3(\theta_\ast) \subset \Theta_k$ a neighborhood of $\theta_\ast$. If conditions (S) and B1-B5 are true, then

(i) Exists $c \in (0, \infty)$, such as, for all $i, j, l \in \{1, \ldots, \gamma(k)\}$,
\[ \limsup_{n \to \infty} \left\| \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in B_3(\theta_\ast)} \frac{\partial^3 l_i(\theta)}{\partial \alpha_i \alpha_j \alpha_l} \right\| \leq c. \]

(ii) \[ \lim_{n \to \infty} \frac{-D^2_0 \log L_{n,k}(\hat{\theta})}{n} = A_2 \quad a.s. \]

for $A_2$ as defined in (B.1).

(iii) $E \left( \left| \log \left( \det(\hat{H}_t(\theta_\ast)) \right) + X_t^\prime H_t^{-1} X_t \right| \right) < \infty.$
Proof. (i) Using item (iii) of Lemma 3 and the Boussama’s results, we just apply the SLLN that can be found at [43].

(ii) Analogous to the technique used in Lemma 5 of [46], using that \(D_2^2l_t(\theta)\) and \(D_3^2l_t(\theta)\) are continuous with respect to \(\theta\), that \(\hat{\theta}_k\) is strongly consistent (Lemma 3) and the mean value Theorem, we have

\[
\left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\hat{\theta})}{\partial \alpha_i \alpha_j} - \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t(\theta_\tau)}{\partial \alpha_i \alpha_j} \right\| \leq \sup_{\theta \in B_\delta(\theta_\tau)} \left\{ \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta'} \left( \frac{\partial^2 l_t(\theta)}{\partial \alpha_i \alpha_j} \right) \right\| \right\}.
\]

Using item (i) and the strong consistency of \(\hat{\theta}\) we conclude the result.

(iii)

\[
E \left( |\log [\det (H_t(\theta_\tau))] + X_t' H_t^{-1} X_t| \right) \leq E (|\log [\det (H_t(\theta_\tau))]|) + E (|X_t' H_t^{-1} X_t|)
\]

\[
\leq E (|\log [\det (H_t(\theta_\tau))]|) + E \left( |X_t|^2 \right) E (||H_t^{-1}||)
\]

which is bounded by items (vi) and (vii) of Lemma 3 and by B3.

\[\square\]

**Lemma 5.** Let \(X = \{X_t\}_{t \in \mathbb{N}}\) be a BEKK-GARCH\((k)\), \(\theta_\tau = (\alpha_1, \ldots, \alpha_{\gamma(k)})\) its true parameter, if conditions (8) and B1-B5 are true, thus, for all \(i \in \{1, \ldots, \gamma(k)\}\),

\[
\limsup_{n \to \infty} \frac{\partial \log L_{n,k}(\theta_\tau)}{\partial \alpha_i} / \sqrt{2n \log \log n} = E \left( \frac{\partial l_1(\theta_\tau)}{\partial \alpha_i}^2 \right)^{1/2} \ a.s.,
\]

\[
\liminf_{n \to \infty} \frac{\partial \log L_{n,k}(\theta_\tau)}{\partial \alpha_i} / \sqrt{2n \log \log n} = -E \left( \frac{\partial l_1(\theta_\tau)}{\partial \alpha_i}^2 \right)^{1/2} \ a.s. \quad \text{and}
\]

\[
\limsup_{n \to \infty} \frac{D_3^2 \log L_{n,k}(\theta_\tau)}{\sqrt{2 \log \log n}} \leq c \ a.s.
\]

for \(c \in (0, \infty)\).

Proof. Consider item (iv) of Lemma 3 and assume \(F_{t-1} = \sigma(X_1, \ldots, X_t), Z_t = t^\delta, \delta > 1,\)

\[U_t = \frac{\partial l_t(\theta_\tau)}{\partial \alpha_i} \quad \text{and} \quad W_n = \left[ nE \left( \frac{\partial l_t(\theta_\tau)}{\partial \alpha_i} \right)^2 \right]^{1/2}, \]

16
where, by (9),
\[ \frac{\partial l_t(\theta_r)}{\partial \alpha_i} = \frac{1}{2} \text{Tr} \left( X_t X'_t H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} H_t^{-1} - H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} \right). \]

To apply Theorem 3, we need to prove conditions L1-L4 below.

(L1) By the Chebyshev’s inequality, we have
\[ P(|U_t| > Z_t) = P \left( \left| \frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right| > t^\delta \right) \leq \frac{1}{t^{2\delta}} E \left( \left( \frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right)^2 \right). \]

Using item (iii) of Lemma 3
\[ \sum_{t=1}^{\infty} P(|U_t| > Z_t) \leq E \left( \left( \frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right)^2 \right) \sum_{t=1}^{\infty} \frac{1}{t^{2\delta}} < \infty. \]

By the Borel-Cantelli Lemma,
\[ P \left( \{ \omega : \|U_t\| > t^\delta \text{ i.o.} \} \right) = P \left( \{ \omega : |U_t| > t^\delta \text{ i.o.} \} \right) = 0 \]
and thus
\[ \lim_{n \to \infty} \frac{\sum_{t=1}^{n} U_t \mathbb{1}(|U_t| > Z_t) - E[U_t \mathbb{1}(|U_t| > Z_t)|\mathcal{F}_{t-1}]}{\sqrt{2W_n^2 \log \log W_n^2}} = 0 \text{ a.s.} \]

(L2)
\[ E(U_t | \mathcal{F}_{t-1}) = \frac{1}{2} \text{Tr} \left[ E(X_t X'_t | \mathcal{F}_{t-1}) H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} H_t^{-1} - H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} \right] \]
\[ = \frac{1}{2} \text{Tr} \left[ H_t H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} H_t^{-1} - H_t^{-1} \frac{\partial H_t}{\partial \alpha_i} \right] \]
\[ = 0. \]

Using item (iv) of Lemma 3
\[ E(E(U_t^2 | \mathcal{F}_{t-1})) = E(U_t^2) < \infty \]
and then, by the SLLN, we have
\[ \lim_{n \to \infty} \frac{\sum_{t=1}^{n} E[U_t^2 | \mathcal{F}_{t-1}]}{nE \left( \left( \frac{\partial l_t(\theta_r)}{\partial \alpha_i} \right)^2 \right)} = 1 \text{ a.s.} \]
By the Dominated Convergence Theorem,

$$\lim_{t \to \infty} E(U_t I(|U_t| \leq n) | F_{t-1}) = 0 \quad a.s.$$ 

Considering an arbitrary $\varepsilon > 0$, it is required to find a $t$-summable upper bound for

$$P \left[ |E(U_t^2 | F_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | F_{t-1})| > \varepsilon \right]$$

and apply the Borel-Cantelli Lemma to conclude

$$\lim_{t \to \infty} \left[ E(U_t^2 I(|U_t| \leq t^\delta) | F_{t-1}) - E(U_t^2 | F_{t-1}) \right] = 0 \quad a.s.$$ 

and apply the Cesàro’s Mean Theorem to conclude

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E(U_t^2 I(|U_t| \leq Z_t) | F_{t-1}) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} E(U_t^2 | F_{t-1})$$

By the generalized Chebyshev’s inequality,

$$P \left[ |E(U_t^2 | F_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | F_{t-1})| > \varepsilon \right] \leq \frac{1}{\varepsilon} \left[ E(U_t^2 | F_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | F_{t-1}) \right]$$

and

$$E \left[ |E(U_t^2 | F_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | F_{t-1})| \right]$$

Using (B.2), (B.7) and (B.3), for suitable $c > 0,$

$$E \left[ |E(U_t^2 | F_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta) | F_{t-1})| \right] \leq \frac{1}{t^\delta}$$

which is $t$-summable.
(L3) Using the notation $\dot{H}_t := \frac{\partial H}{\partial t}$,

$$0 \leq E[U_t^4 I([U_t] \leq Z_t)|\mathcal{F}_{t-1}]$$

$$\leq E[U_t^4 |\mathcal{F}_{t-1}]$$

$$\leq E \left\{ \text{Tr} \left[ X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} - H_t^{-1} \dot{H}_t \right]^4 |\mathcal{F}_{t-1} \right\}$$

$$= E \left\{ \left[ \text{Tr} \left( X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right) - \text{Tr} \left( H_t^{-1} \dot{H}_t \right) \right]^4 |\mathcal{F}_{t-1} \right\}$$

$$= E \left\{ \text{Tr} \left( X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right)^4 \right. \right.$$  

$$- 4 \text{Tr} \left( X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right)^3 \text{Tr} \left( H_t^{-1} \dot{H}_t \right)$$

$$+ 6 \text{Tr} \left( X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right)^2 \text{Tr} \left( H_t^{-1} \dot{H}_t \right)^2$$

$$- 4 \text{Tr} \left( X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right) \text{Tr} \left( H_t^{-1} \dot{H}_t \right)^3$$

$$+ \text{Tr} \left( H_t^{-1} \dot{H}_t \right)^4 |\mathcal{F}_{t-1} \right\}. \quad (B.4)$$

Also, by Lemma 3 (vi), for suitable $c \in (0, \infty)$,

$$\left| \text{Tr} \left( X_t X_t' H_t^{-1} \dot{H}_t H_t^{-1} \right) \right| = \left| \text{Tr} \left( H_t^{1/2} \varepsilon_t \left( H_t^{1/2} \varepsilon_t' \right)' \dot{H}_t H_t^{-1} \right) \right|$$

$$= \left| \text{Tr} \left( H_t^{1/2} \varepsilon_t \varepsilon_t' H_t^{1/2} \dot{H}_t H_t^{-1} \right) \right|$$

$$= \left| \text{Tr} \left( \varepsilon_t \varepsilon_t' H_t^{-1/2} \dot{H}_t H_t^{-1/2} \right) \right|$$

$$\leq \| \varepsilon_t \varepsilon_t' \| \| H_t^{-1/2} \| \| \dot{H}_t \|$$

$$\leq c \| \varepsilon_t \varepsilon_t' \| \| \dot{H}_t \|$$ \quad (B.5)

and

$$\left| \text{Tr} \left( H_t^{-1/2} \dot{H}_t H_t^{-1/2} \right) \right| \leq c \| \dot{H}_t \|. \quad (B.6)$$

We used the relation $|\text{Tr}(ABC)| \leq \|A\| \|B\| \|C\|$. Thus,

$$E[U_t^4 I([U_t] \leq Z_t)|\mathcal{F}_{t-1}] \leq c^4 E \left( \| \varepsilon_t \varepsilon_t' \|^4 |\mathcal{F}_{t-1} \right) \| \dot{H}_t \|^4$$

$$+ 4c^4 E \left( \| \varepsilon_t \varepsilon_t' \|^4 |\mathcal{F}_{t-1} \right) \| \dot{H}_t \|^4$$

$$+ 6c^4 E \left( \| \varepsilon_t \varepsilon_t' \|^2 |\mathcal{F}_{t-1} \right) \| \dot{H}_t \|^4$$

$$+ 3c^4 \| \dot{H}_t \|^4$$

$$\leq c_1 E \left[ \| \varepsilon_t \varepsilon_t' \|^2 + \| \varepsilon_t \varepsilon_t' \|^3 + \| \varepsilon_t \varepsilon_t' \|^4 + 1 |\mathcal{F}_{t-1} \right] \| \dot{H}_t \|^4. \quad (B.7)
Adapting the proof of Lemma A.2 of Comte & Lieberman \footnote{41} and using B5, we conclude that

\[
E \sup_{\theta \in \Theta_k} \left[ \left\| \frac{\partial H_t}{\partial \alpha_i} (\theta) \right\|^2 \right] < \infty. \tag{B.8}
\]

For \( \delta_1 \in (\frac{1}{2}, 1) \) and suitable \( c_2, c_3 \in (0, \infty) \), using the Chebyshev’s and Jensen’s inequalities, (B.7) and (B.8),

\[
P \left[ E[U_t^4 \mathbb{1}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] > t^{\delta_1} \right] 
\leq P \left[ E \left[ \left( \| \varepsilon_t \varepsilon_t' \|^2 + \| \varepsilon_t \varepsilon_t'' \|^3 + \| \varepsilon_t \varepsilon_t'' \|^4 + 1 \right) \left\| \frac{\partial H_t}{\partial \alpha_i} \right\|^2 \right] \right] 
\leq \frac{c_2^2}{t^{2\delta_1}} E \left[ \left( \| \varepsilon_t \varepsilon_t' \|^2 + \| \varepsilon_t \varepsilon_t'' \|^3 + \| \varepsilon_t \varepsilon_t'' \|^4 + 1 \right) \left\| \frac{\partial H_t}{\partial \alpha_i} \right\|^2 \right] 
\leq \frac{c_2^2}{t^{2\delta_1}} E \left[ \left( \| \varepsilon_t \varepsilon_t' \|^2 + \| \varepsilon_t \varepsilon_t'' \|^3 + \| \varepsilon_t \varepsilon_t'' \|^4 + 1 \right)^2 \left\| \frac{\partial H_t}{\partial \alpha_i} \right\|^2 \right] 
\leq \frac{c_3}{t^{2\delta_1}} E \left[ \left\| \frac{\partial H_t}{\partial \alpha_i} \right\|^2 \right]
\]

and

\[
\sum_{t=1}^{\infty} P \left[ E[U_t^4 \mathbb{1}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] > t^{\delta_1} \right] 
\leq \sum_{t=1}^{\infty} \frac{c_3}{t^{2\delta_1}} E \left[ \left\| \frac{\partial H_t}{\partial \alpha_i} \right\|^2 \right] < \infty.
\]

By the Borel-Cantelli Lemma,

\[
P \left[ E[U_t^4 \mathbb{1}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}] > t^{\delta_1} \text{ i.o.} \right] = 0
\]

and then

\[
\lim_{n \to \infty} \sum_{t=1}^{n} \frac{E[U_t^4 \mathbb{1}(|U_t| \leq Z_t) | \mathcal{F}_{t-1}]}{W_t^2} \leq \lim_{n \to \infty} \sum_{t=1}^{n} \frac{1}{t^{2-\delta_1}} < \infty \text{ a.s.}
\]
Lemma 6. Let $X = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH($k$), where conditions (8) and B1-B5 are true. Then,

$$E \sup_{\theta \in \Theta_k} \left[ \left| \text{Tr}(\dot{H}_t(\theta)H_t^{-1}(\theta) - X_tX_t'H_t^{-1}(\theta)\dot{H}_t(\theta)H_t^{-1}(\theta)) \right| \right] < \infty,$$

where $\dot{H}_t := D_\theta H_t$.

Proof. Using Lemma 3 and (B.8), for a suitable $c \in (0, \infty)$,

$$E \sup_{\theta \in \Theta_k} \left[ \left| \text{Tr}(\dot{H}_tH_t^{-1} - X_tX_t'H_t^{-1}\dot{H}_tH_t^{-1}) \right| \right] \leq E \sup_{\theta \in \Theta_k} \left[ \left| \dot{H}_t \right| + \left| X_tX_t'H_t^{-1} \dot{H}_t \right| \right] < \infty.$$

Lemma 7. Let $X = \{X_t\}_{t \in \mathbb{N}}$ be a BEKK-GARCH($r$), of order $r$, $k \geq r$, $\theta_r$ its true parameter, $\hat{\theta}_k$ the MLE of $\theta_r$. If conditions (8) and B1-B5 are true, then

$$\lim_{n \to \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,k}(\hat{\theta}_k)}{n} > 0 \ a.s.$$

Proof. Assuming $p$ such as $p \geq k$ and $p \geq r$,

$$\lim_{n \to \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,k}(\hat{\theta}_k)}{n} = \lim_{n \to \infty} \frac{\log L_{n,p}(\hat{\theta}_p) - \log L_{n,k}(\hat{\theta}_k)}{n}.$$

Applying Lemma 2 using the results above, we have

$$\lim_{n \to \infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,p}(\hat{\theta}_p)}{n} = 0 \ a.s.$$

By Lemma 4, item (iii),

$$E \left[ |l_t(\theta_r)| \right] < \infty$$
and then, using the SLLN,

\[
\lim_{n \to \infty} \frac{\log L_{n,p}(\theta_r)}{n} = \lim_{n \to \infty} \frac{\sum_{t=1+k}^{n} l_t(\theta_r)}{n} = E(l_1(\theta_r)) = c_1 < \infty \quad \text{a.s.}
\]

By the Mean Value Theorem, for \( \dot{\theta} = s\theta_r + (1-s)\hat{\theta}_p \), \( s \in (0,1) \), sufficiently large \( n \) and \( B_\delta(\theta_r) \) a sufficiently small neighborhood of \( \theta_r \),

\[
\left| \frac{\sum_{t=1+k}^{n} l_t(\theta_r)}{n} - \frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_p)}{n} \right| = \left| \frac{\sum_{t=1+k}^{n} D_{\theta}^1 l_t(\dot{\theta})}{n} (\hat{\theta}_p - \theta_r) \right|
\]

\[
\leq \sup_{\theta \in B_\delta(\theta_r)} \left\| \frac{\sum_{t=1+k}^{n} D_{\theta}^1 l_t(\dot{\theta})}{n} \right\| \left\| (\hat{\theta}_p - \theta_r) \right\|
\]

Using the SLLN, Lemma 6 and the strong consistency of \( \hat{\theta}_p \),

\[
\left| \frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_p)}{n} - \frac{\sum_{t=1+k}^{n} l_t(\theta_r)}{n} \right| \to 0. \quad \text{(B.9)}
\]

And thus,

\[
\lim_{n \to \infty} \frac{\log L_{n,p}(\hat{\theta}_p)}{n} = \lim_{n \to \infty} \frac{\log L_{n,p}(\theta_r)}{n} = c_1. \quad \text{(B.10)}
\]

Besides that, \( \Theta_k \subset \Theta_p \) and \( \hat{\theta}_k \) is the MLE of \( \theta_r \), thus

\[
\lim_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \leq \lim_{n \to \infty} \frac{\log L_{n,p}(\hat{\theta}_p)}{n} = c_1 \quad \text{a.s.}
\]

and then,

\[
\lim_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \leq \lim_{n \to \infty} \frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_k)}{n} = c_2 \leq c_1 \quad \text{a.s.}
\]

Let \( n_1 \) be a subsequence of \( n \) such as

\[
\lim_{n \to \infty} \frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_k)}{n_1} = c_2 \quad \text{a.s.}
\]
Using that $\Theta_k$ is compact, assume $n_j$ a subsequence of $n_i$ such as

$$\hat{\theta}_k(n_j) \to \bar{\theta}_k \in \Theta_k \text{ a.s.}$$

And then,

$$\lim_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \leq \limsup_{n \to \infty} \frac{\log L_{n,k}(\hat{\theta}_k)}{n} = \lim_{n_j \to \infty} \frac{\log L_{n_j}(\hat{\theta}_k(n_j))}{n_j} \text{ a.s.}$$

Applying the same argument used in (B.9), we have

$$\lim_{n_i \to \infty} \frac{\log L_{n_j}(\bar{\theta}_k)}{n_j} = E \left( l_1(\bar{\theta}) \right) \text{ and}$$

$$\left| \frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_k)}{n} - \frac{\sum_{t=1+k}^{n} l_t(\bar{\theta})}{n} \right| \to_{a.s.} 0.$$ 

Additionally,

$$\lim_{n \to \infty} \left[ \frac{\log L_{n,p}(\hat{\theta}_p)}{n} - \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \right] \geq E \left[ \log \left( \frac{f(\bar{\theta}_k)}{f(\theta_r)} \right) \right].$$

By other hand,

$$E \left[ \log \left( \frac{f(\bar{\theta}_k)}{f(\theta_r)} \right) \right]$$

is the Kullback-Leibler divergence, which is positive if $f(\bar{\theta}_k) \neq f(\theta_r)$, and, as $\theta_r \notin \Theta_k \subseteq \mathbb{R}^{\gamma(k)}$, we have $\theta_r \neq \hat{\theta}_k$ and then, by B3, $f(\bar{\theta}_k) \neq f(\theta_r)$. Thus, we conclude

$$\lim_{n \to \infty} \left[ \frac{\log L_{n,k}(\hat{\theta}_r)}{n} - \frac{\log L_{n,k}(\hat{\theta}_k)}{n} \right] \geq E \left[ \log \left( \frac{f(\bar{\theta}_k)}{f(\theta_r)} \right) \right] = c > 0.$$

Lemmas 3 and 4 provides A1-A3 and Lemmas 5 and 7 provides, respectively, A4 and A5. The EDC estimator’s consistency is established using Theorem 1.
References

[1] M. S. Bartlett, The frequency goodness of fit test for probability chains, Proceedings of the Cambridge Philosophical Society 47 (1951) 86–95.

[2] P. G. Hoel, A test for Markoff chains, Biometrika 41 (3/4) (1954) 430–433.

[3] I. J. Good, The likelihood ratio test for Markoff chains, Biometrika 42 (3/4) (1955) 531–533.

[4] P. Billingsley, Statistical methods in Markov chains, The Annals of Mathematical Statistics 32 (1) (1961) 12–40.

[5] T. W. Anderson, L. A. Goodman, Statistical inference about Markov chains, The Annals of Mathematical Statistics 28 (1) (1957) 89–110.

[6] M. H. Quenouille, A large-sample test for the goodness of fit of autoregressive schemes, Journal of the Royal Statistical Society 110 (2) (1947) 123–129.

[7] P. Whittle, Hypothesis Testing in Time-Series Analysis, Almqvist and Wiksells, Uppsala, 1951.

[8] P. Whittle, Some recent contributions to the theory of stationary processes: A Study in the analysis of stationary time series., Almqvist and Wiksells, Uppsala, 1954.

[9] M. S. Bartlett, D. V. Rajalakshman, Goodness of fit tests for simultaneous autoregressive series, Journal of the Royal Statistical Society 15 (1) (1953) 107–124.

[10] H. Akaike, A new look at the statistical model identification, Automatic Control, IEEE Transactions on 19 (6) (1974) 716–723.

[11] G. Schwarz, Estimating the dimension of a model, The Annals of Statistics 6 (2) (1978) 461–464.
[12] H. Akaike, A bayesian extension of the minimum AIC procedure of autoregressive model fitting, Biometrika 66 (1979) 237–242.

[13] E. J. Hannan, B. G. Quinn, The determination of the order of an autoregression, Journal of the Royal Statistical Society 41 (2) (1979) 190–195.

[14] E. J. Hannan, The estimation of the order of an arma process, The Annals of Statistics 8 (1980) 1071–1081.

[15] T. Ozaki, On the order determination of ARIMA models, Journal of Applied Statistics 26 (1977) 290–301.

[16] Y. Ogata, Maximum likelihood estimates of incorrect Markov models for time series and the derivation of AIC, Journal of Applied Probability 17 (1980) 59–72.

[17] B. Choi, ARMA Model Identification, Springer, New York, 1992.

[18] H. Tong, Determination of the order of a Markov chain by Akaike’s information criterion, Journal of Applied Probability 12 (3) (1975) 488–497.

[19] R. W. Katz, On some criteria for estimating the order of a Markov chain, Technometrics 23 (3) (1981) 243–249.

[20] I. Csiszár, P. C. Shields, The consistency of the BIC Markov order estimator, The Annals of Statistics 28 (6) (2000) 1601–1619.

[21] A. M. Polansky, Detecting change-points in Markov chains, Computational Statistics & Data Analysis 51 (2007) 6013–6026.

[22] L. Finesso, Consistent estimation of the order for Markov and hidden Markov chains, Ph.D. thesis, University of Maryland (1990).

[23] A. W. Hughes, M. L. King, K. T. Kwek, Selecting the order of an ARCH model, Economics Letters 83 (2) (2004) 269–275.

[24] A. E. Raftery, A model for high-order Markov chains, Journal of the Royal Statistical Society 47 (3) (1985) 528–539.
[25] L. Zhao, C. Dorea, C. Gonçalves, On determination of the order of a Markov chain, Statistical Inference for Stochastic Processes 4 (3) (2001) 273–282.

[26] C. C. Y. Dorea, Optimal penalty term for EDC Markov chain order estimator, Annales de l’Institut de Statistique de l’Universite de Paris 52 (3) (2008) 15–26.

[27] P. A. A. Resende, Análise comparativa de estimadores da ordem de cadeias de Markov, Master’s thesis, Universidade de Brasília (2009).

[28] R. Nishii, Maximum likelihood principle and model selection when the true model is unspecified, Journal of Multivariate Analysis 27 (2) (1988) 392–403.

[29] I. V. Basawa, C. C. Heyde, Asymptotic properties of maximum likelihood estimators for stochastic processes, The Indian Journal of Statistics 38 (1976) 259–270.

[30] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica 50 (4) (1982) 987–1007.

[31] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (3) (1986) 307–327.

[32] R. F. Engle, V. K. Ng, Measuring and testing the impact of news on volatility, The Journal of Finance 48 (5) (1993) 1749–1778.

[33] D. B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59 (2) (1991) 349–370.

[34] R. F. Engle, K. F. Kroner, Multivariate simultaneous generalized ARCH, Econometric Theory 11 (1) (1995) 122–150.

[35] T. Bollerslev, R. F. Engle, J. M. Wooldridge, A Capital Asset Pricing Model with Time-Varying Covariances, Journal of Political Economy 96 (1) (1988) 116–131.
[36] T. Bollerslev, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, Review of Economics and Statistics 72 (3) (1990) 498–505.

[37] R. F. Engle, V. K. Ng, M. Rothschild, Asset pricing with a Factor-ARCH covariance structure: Empirical estimates for Treasury bills, Journal of Econometrics 45 (1990) 213–237.

[38] F. Boussama, F. Fuchs, R. Stelzer, Stationarity and geometric ergodicity of BEKK multivariate GARCH models, Stochastic Processes and their Applications 121 (10) (2011) 2331 – 2360.

[39] R. Stelzer, On the relation between the VEC and BEKK multivariate GARCH models, Econometric Theory 24 (04) (2008) 1131–1136.

[40] F. Boussama, Ergodicité, mélange et estimation dans les modèles garch, Ph.D. thesis, Université 7 Paris (1998).

[41] F. Comte, O. Lieberman, Asymptotic theory for multivariate GARCH processes, Journal of Multivariate Analysis 84 (2003) 61–84.

[42] T. Jeantheau, Strong consistency of estimators for multivariate ARCH models, Econometric Theory 14 (1) (1998) 70–86.

[43] S. P. Meyn, R. L. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, London, 1993.

[44] P. Hall, C. C. Heyde, Martingale limit theory and its application, Academic Press, New York, 1980.

[45] C. M. Hafner, A. Preminger, On asymptotic theory for multivariate GARCH models, Journal of Multivariate Analysis 100 (2009) 2044–2054.

[46] C. M. Hafner, A. Preminger, Asymptotic theory for a FACTOR GARCH model, Econometric Theory 25 (2009) 336–363.