Estimação de ordem em modelos AR, ARCH e BEKK-GARCH usando o critério EDC

por

Paulo Angelo Alves Resende

Brasília – DF

2014

Estimação de ordem em modelos AR, ARCH e BEKK-GARCH usando o critério EDC

Tese apresentada ao Programa de Pós-Graduação em Matemática da Universidade de Brasília (UnB), como requisito parcial para obtenção do grau de DOUTOR EM MATEMÁTICA.

por

Paulo Angelo Alves Resende

Orientadora:
Chang Chung Yu Dorea

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE MATEMÁTICA

Brasília – DF

2014

À minha mãe, Angela Maria.

Resumo

O critério de informação EDC – *Efficient Determination Criterion* – foi proposto originalmente para definir uma classe de estimadores de ordem para cadeias de Markov de espaço de estados finitos. Nesse trabalho, o conceito de modelos parcialmente aninhados é definido e a classe de estimadores EDC é estendida nesse contexto. Esses resultados são aplicados para estabelecer a consistência forte de um novo estimador de ordem para modelos Autoregressivos (AR) e para demonstrar a consistência forte de uma classe de estimadores de ordem para processos Autoregressivos de Heteroscedasticidade Condicional (ARCH) e para o caso multivariado de modelos Autoregressivos de Heteroscedasticidade Condicional Generalizado na Representação BEKK (BEKK-GARCH). Como resultado imediato, a consistência forte dos estimadores de ordem BIC para ARCH e BEKK-GARCH é estabelecida. Também é ilustrado por meio de simulações numéricas que o estimador de ordem EDC proposto para processos AR apresenta melhor performance que suas principais alternativas, os estimadores baseados nos critérios AIC, BIC e HQC.

Abstract

The Efficient Determination Criterion (EDC) was originally stated to define a class of estimators for the order of a Markov chain with finite state space. In this work, we define the concept of partially nested models and extend the class of EDC estimators within this context. This framework is applied to establish the consistency for a new order estimator for Autoregressive process (AR) and to prove the consistency for a class of order estimators for Autoregressive Conditional Heteroskedasticity models (ARCH) and for a multivariate version, the Generalized Autoregressive Conditional Heteroskedasticity in the BEKK representation (BEKK-GARCH). As an immediate consequence, the strong consistency for the BIC order estimators for ARCH and BEKK-GARCH is established. Also, using numerical simulation, we show that the proposed EDC order estimator for AR performs better than the wide-known alternatives based on the criteria AIC, BIC and HQC.

Sumário

Introdução

No contexto de seleção de modelos, a classe de estimadores de ordem de dependência EDC – *Efficient Determination Criterion* – foi proposto por Zhao, Dorea & Gonçalves (2001) como generalização dos critérios BIC e AIC para a estimação de ordem de cadeias de Markov de espaço de estados finitos. Posteriormente, Lopes (2005) estendeu os resultados de Zhao et al. (2001) para cadeias de Markov de espaço de estados enumerável e Dorea (2008) determinou o termo de penalidade ótimo que define o estimador assintoticamente com melhor performance dentro da classe dos estimadores EDC fortemente consistentes.

Com o uso extensivo de simulações numéricas, Resende (2009) verificou a melhor performance do estimador EDC ótimo comparado com as alternativas AIC e BIC. Os resultados empíricos motivaram o presente trabalho, que tem como objetivo generalizar o EDC para a estimação de ordem em modelos markovianos com espaço de estados gerais, que incluem as famílias AR e ARCH, e para a estimação de outros parâmetros, como por exemplo o tamanho do espaço de estados oculto em cadeias de Markov Ocultas.

Uma sequência de variáveis aleatórias $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ é uma cadeia de Markov com espaço de estados finito E e ordem r se, para todo $(a_1,\dots,a_{t+1}) \in E^{t+1}$,

$$P(X_{t+1} = a_{t+1}|X_1 = a_1,\cdots,X_t = a_t) = P(X_{t+1} = a_{t+1}|X_{t-r+1} = a_{t-r+1},\cdots,X_t = a_t)$$

e r é o menor valor com essa propriedade.

Em contextos práticos de modelagem é necessário primeiramente identificar a ordem r da cadeia para posteriormente estimar os parâmetros, que neste caso são as probabilidades de transição.

Esse problema de estimação de ordem foi inicialmente abordado com testes de hipóteses sobre a razão de verossimilhança de Neyman-Pearson por Bartlett (1951), Hoel (1954), Good (1955), Anderson & Goodman (1957) e Billingsley (1961). Neste caso, observou-se que, supondo que k é maior ou igual a ordem verdadeira r,

$$\log\left(\frac{L_n(k+1)}{L_n(k)}\right) \sim \chi^2\left(\gamma(k+1)-\gamma(k)\right)$$

para $L_n(k)$ a função de verossimilhança estimada de uma amostra $x_1,\cdots,x_n$, supondo a ordem da cadeia k e $\gamma(k)$ é o número de parâmetros livres. Com isso basta calcular os intervalos de confiança para se obter uma estimativa de r.

No contexto de seleção de modelos aninhados, Akaike (1974) propõe o uso do critério de informação AIC, com o simples argumento de tornar a seleção do modelo mais objetiva e não depender de avaliação estatística. Esse critério foi utilizado como base por Tong (1975) para a definição do seguinte estimador de ordem em cadeias de Markov.

$$\hat{r}_{aic} = \underset{k\leq K}{\operatorname{argmin}}\,\{\mathrm{AIC}(k)\}$$

onde K é uma cota superior conhecida de r,

$$\mathrm{AIC}(k) = -2\log L_n(k) + 2\gamma(k)$$

e $\gamma(k) = |E|^k(|E|-1)$ (isto é, a cardinalidade do conjunto E). De forma semelhante, Schwarz (1978) propõe o critério de informação BIC e Katz (1981) demonstra a inconsistência do estimador $\hat{r}_{aic}$ e a consistência fraca do seguinte estimador baseado no BIC.

$$\hat{r}_{bic} = \underset{k\leq K}{\operatorname{argmin}}\,\{\mathrm{BIC}(k)\}$$

para

$$\mathrm{BIC}(k) = -2\log L_n(k) + \gamma(k)\log n.$$

A consistência forte do $\hat{r}_{bic}$ foi demonstrada por Finesso (1990) e posteriormente, sem a hipótese de existência de um limitante superior K, por Csiszár & Shields (2000).

Em seus trabalhos, Katz (1981) e Csiszár & Shields (2000), apontam que o estimador BIC possui uma tendência a subestimar a ordem verdadeira de cadeias de Markov em amostras pequenas, o que sugere a necessidade de demonstrar a consistência forte para termos de penalidades menores que $\gamma(k)\log n$. Nesse sentido, Zhao et al. (2001) generalizam os critérios AIC e BIC, criando a classe de estimadores

$$\hat{r}_{edc} = \underset{k \leq K}{\operatorname{argmin}} \{\text{EDC}(k)\}$$

para

$$\text{EDC}(k) = -2\log L_n(k) + \gamma(k)c_n$$

e demonstram a consistência forte do estimador $\hat{r}_{edc}$ para c_n satisfazendo

$$\lim_{n\to\infty} \frac{c_n}{n} = 0 \quad e \quad \liminf_{n\to\infty} \frac{c_n}{\log\log n} = \infty.$$

Posteriormente, Dorea (2008) demonstra a consistência forte para $\hat{r}_{edc}$ quando c_n satisfazendo

$$\lim_{n\to\infty} \frac{c_n}{n} = 0 \quad e \quad \liminf_{n\to\infty} \frac{c_n}{\log\log n} \geq 2\frac{|E|}{|E|-1}$$

e propõe o estimador ótimo $\hat{r}_{opt}$ com o termo de penalidade $\gamma(k)c_n = 2|E|^{k+1}\log\log n$. Usando simulações numéricas, Resende (2009) comprovou a melhor performance do estimador $\hat{r}_{opt}$ e verificou que essa eficiência relativa é substancialmente superior à medida que se considera modelos com maior complexidade (número de parâmetros livres).

Observando que os estimadores citados são baseados no comportamento da verossimilhança, que é o mesmo fundamento utilizado nos testes de hipóteses, Baigorri, Gonçalves & Resende (2014) propõem o estimador GDL, baseado na comparação das distribuições empíricas condicionadas usando a divergência chi-quadrado. Em simulações numéricas, o GDL apresentou performance superior às demais alternativas. Entretanto, o método utiliza diretamente em sua definição a finitude do espaço de estados da cadeia de Markov, o que torna improvável uma generalização a outros processos.

Embora os trabalhos citados usem o conceito de modelos aninhados, os resultados são sempre particularizados para o processo considerado. Dessa forma, nessa tese, os resultados existentes foram aperfeiçoados para tratar a consistência do estimador $\hat{r}_{edc}$ baseada, basicamente, em hipóteses sobre regularidade e comportamento assintótico de funções $\log L_{n,k}$, que possuem propriedades semelhantes à log-verossimilhança.

O ferramental desenvolvido é aplicado para obter a consistência de estimadores de ordem para os processos Autoregressivos (AR), Autoregressivos de Heteroscedasticidade Condicional (ARCH) e ARCH multivariado generalizado na representação BEKK (BEKK-GARCH).

Em simulações numéricas, o estimador obtido para processos Autoregressivos apresentou, no geral, melhor performance quando comparado com os estimadores AIC (Akaike 1974), BIC (Shibata 1976) e HQC (Hannan & Quinn 1979).

Ressalta-se que, para processos ARCH e BEKK-GARCH, até então não é conhecida na literatura a existência de estimadores fortemente consistentes. Assim, esse resultado é inédito e de alta relevância.

Para os processos AR e ARCH, a ordem a ser estimada é um parâmetro univariado, o que permite a definição de sequência de modelos aninhados, onde o EDC é originalmente definido por Zhao et al. (2001). Entretanto, em modelos BEKK-GARCH, a ordem a ser estimada é um parâmetro com duas variáveis (de forma semelhante ocorre com modelos ARMA). Nesse sentido, o conceito de sequência de modelos aninhados é estendido a classe de modelos parcialmente aninhados, permitindo assim a definição do EDC nesse novo contexto para a estimação de ordens multivariadas que tem como caso particular modelos BEKK-GARCH.

A técnica utilizada nos três casos para se obter o comportamento assintótico de $\log L_{n,k}$ é essencialmente a mesma, que basicamente consiste na utilização do comportamento assintótico das derivadas de primeira ordem de $\log L_{n,k}$ para concluir determinado comportamento assintótico dos estimadores dos parâmetros da densidade em questão e, a partir desse comportamento, conclui-se o comportamento assintótico de $\log L_{n,k}$.

Essa técnica foi empregada por Nishii (1988) para a estimação de dimensão em modelos aninhados no caso particular de sequências de variáveis independentes e identicamente distribuídas (i.i.d.) e considerando $\log L_{n,k}$ a log-verossimilhança. Posteriormente, Finesso (1990) baseou-se nessa técnica para demonstrar a consistência forte do estimador de ordem BIC para cadeias de Markov. Vale ressaltar que Basawa & Heyde (1976) utilizam de técnica semelhante para concluir a normalidade assintótica do estimador do parâmetro da densidade a partir do comportamento assintótico das derivadas de primeira ordem da log-verossimilhança.

Geralmente as derivadas de $\log L_{n,k}$ são mais simples, isso motiva o uso desse caminho. Outro fato importante é que várias hipóteses para a aplicação do método são usadas para a obtenção da normalidade assintótica dos estimadores e por isso frequentemente suas demonstrações são encontradas na literatura.

No Capítulo 1 os conceitos de modelos aninhados e ordem são generalizados e formalizados, e alguns resultados gerais relativos à consistência do estimador EDC são apresentados. No Capítulo 2 são introduzidos conceitos e resultados de cadeias de Markov com espaço de estados gerais. Esses resultados são empregados para obter o comportamento assintótico das respectivas funções $\log L_{n,k}$ definidas para os processos AR, ARCH e BEKK-GARCH, para então concluir os casos de consistência forte do EDC. Nos apêndices são encontrados resultados auxiliares, que são utilizados nas demonstrações.

1 Estimação de ordem em modelos aninhados

Os estudos pioneiros em estimação de ordem para diversos processos estocásticos clássicos foram focados na definição de condições para a aplicação de testes de hipóteses no apoio ao trabalho de análise e decisão (Bartlett 1951, Whittle 1951). Em um segundo momento, ainda com uma visão estatística para abordar o problema, tiveram os trabalhos no desenvolvimento de critérios de informação no contexto de seleção de modelos aninhados (Akaike 1974, Schwarz 1978). Com a necessidade de se obter um melhor entendimento, alguns autores utilizaram com sucesso uma abordagem analítica na avaliação do comportamento assintótico dos objetos envolvidos para concluir aspectos de consistência dos estimadores (Nishii 1988, Finesso 1990). Nesse cenário, Zhao et al. (2001) propõem o EDC como uma generalização dos critérios de informação que usam o método da log-verossimilhança penalizada.

Ao longo das últimas quatro décadas, os critérios de informação vêm sendo implementados nos mais diversos processos e situações (Tong 1975, Shibata 1976, Ozaki 1977, Ogata 1980, Zhao, Dorea & Gonçalves 2001, Lopes 2005, Polansky 2007). Entretanto, não houve uma proposta efetiva de ferramental que facilite o estudo da consistência de estimadores de ordem em novos casos. Nesse sentido, na Seção 1.1 os conceitos de modelos aninhados e ordem são apresentados de forma mais geral. A Seção 1.2 contém a definição do estimador de ordem EDC para esse contexto e o aperfeiçoamento e generalização dos resultados de Nishii (1988) e Dorea (2008). Com isso é possível obter os casos de consistência do estimador de ordem EDC baseado, basicamente, em condições sobre as funções $\log L_{n,k}$, que em muitas situações podem ser as funções log-verossimilhança. A última seção contém a generalização de sequência de modelos aninhados para classe de modelos parcialmente aninhados e a reapresentação dos resultados da Seção 1.2 nesse cenário ainda mais geral.

1.1 Definições

A maioria dos autores citados que aplicam algum critério de informação não necessitam tratar diretamente em caso geral de seleção de modelos, com isso não definem modelos aninhados. Por outro lado, Akaike (1974) e Schwarz (1978) usam livremente a palavra “modelo” sem definição explícita do objeto. Na demonstração da consistência forte do estimador BIC para cadeias de Markov, Csiszár & Shields (2000) definem conjunto de modelos estatísticos de índice k como a classe de processos que satisfaz a propriedade de Markov de ordem k.

Com o objetivo de generalizar os critérios de informação para o caso i.i.d., Nishii (1988) define um conjunto de modelos como uma família de densidades, mas não se depara com o problema de definição de aninhamento para sequências de variáveis aleatórias que possuem dependência. A definição abaixo é a generalização da definição de Nishii. Observa-se que, como as densidades são em relação a uma medida fixada ν, pode-se valer disso para simplificar a expressão da log-verossimilhança e possivelmente das funções $\log L_{n,k}$ na aplicação dos resultados subsequentes.

Definição 1.1. *Para um processo estocástico $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$, $E \subseteq \mathbb{R}^p$ o conjunto de possíveis valores de X_t, ν uma medida fixada em E, assumimos uma família de modelos estatísticos para $\mathbb{X}$ como*

$$M = \left(f(x_1^n, \theta, n), \Theta \right),\ \theta \in \Theta,\ \forall\, n \geq 1,$$

onde $f(x_1^n, \theta, n)$ representa o conjunto de possíveis densidades de x_1^n com respeito à medida produto em E^n, que depende do parâmetro $\theta \in \Theta \subseteq \mathbb{R}^d$ e $x_1^n = x_1 x_2 \ldots x_n$ é uma possível realização de $\mathbb{X}$. Para simplificar a notação, denotamos $f(x_1^n, \theta, n)$ por $f(x_1^n, \theta)$. Os elementos de M podem ser denotados apenas pelas suas densidades, exemplo $m = f(x_1^n, \theta_0) \in M$.

O fundamento do conceito de “aninhamento” foi utilizado em hipóteses aninhadas para a aplicação dos testes na razão da log-verossimilhança de Neyman-Pearson, como exemplo em Hoel (1954). A seguir apresentamos uma generalização desse conceito e a definição de ordem para sequência de modelos aninhados.

Definição 1.2. *Para*

$$M_i = \left(f_i(x_1^n,\theta),\Theta_i \right),\ \theta \in \Theta_i,\ n \geq 1,$$

$$M_j = \left(f_j(x_1^n,\theta),\Theta_j \right),\ \theta \in \Theta_j,\ n \geq 1.$$

(i) *Dizemos que* $M_i \subseteq M_j$ *se e somente se* $\Theta_i \subseteq \Theta_j$ *e, para todo* $\theta \in \Theta_i$*,* $n \in \mathbb{N}$ *e* $x_1^n \in E^n$ *existe* $c_1 \in (0,\infty)$*, que não depende de n, tal que, se, para* c_2 *suficientemente grande,* $n > c_2 \in \mathbb{N}$*,*

$$\frac{f_i(x_1^n,\theta)}{f_j(x_1^n,\theta)} = c_1. \tag{1.1}$$

Neste caso denotamos por $f_i(\cdot,\theta) \simeq f_j(\cdot,\theta)$*.*

(ii) $\mathbb{M} = \{M_k\}_{k\in\mathbb{N}}$ *é uma sequência de modelos aninhados se* $M_k \subseteq M_{k+1}$ *para todo* $k \in \mathbb{N}$*.*

(iii) *Se* $\mathbb{M}$ *é sequência de modelos aninhados e* $m \in \bigcup_{k=0}^{\infty} M_k$*, dizemos que m é de ordem* 0 *se* $m \in M_0$ *e é de ordem* $r > 0$ *se* $m \in M_r$ *e* $m \notin M_{r-1}$*.*

(iv) *Se* $\mathbb{M} = \{M_k\}_{k\in\mathbb{N}}$ *é sequência de modelos aninhados, denotamos* $\gamma(k) = dim(\Theta_k)$*.*

Observa-se que em (1.1), bastaria

$$\lim_{n\to\infty} \frac{f_i(x_1^n,\theta)}{f_j(x_1^n,\theta)} = c_1.$$

Entretanto, a forma como foi definido é suficiente e mais simples. Essencialmente, o objetivo é definir equivalência das densidades conjuntas em função do comportamento assintótico quase certo, em outras palavras $f_i(n) = O(f_j(n))$ para quase toda realização do processo.

Mesmo no contexto de seleção de modelos, o termo "verossimilhança" (ou "verossimilhança estimada") comumente se refere a uma função. Entretanto, implicitamente trata-se de uma classe de funções que depende da hipótese assumida como verdadeira e do tamanho da amostra. Caso contrário, encontramos dificuldades na definição explícita do domínio dessa função. Para ilustrar essa situação, suponha que temos uma cadeia de Markov de ordem r com espaço de estados finito E. Considere $\mathbb{M}$ sua respectiva sequência de modelos aninhados, $\theta_k \in \Theta_k$ uma matriz de transição e $p_{i,j}(\theta_k)$ as respectivas probabilidades de transição considerando $i \in E^k$. Para uma amostra $x_1^n = x_1 x_2 \ldots x_n$, fixando n e k, a verossimilhança é dada por

$$C(x_1^k)\prod_{i=k+1}^{n} p_{x_{i-k}^{i-1},x_i}(\theta_k).$$

Se por outro lado supormos a ordem $k+1$ temos

$$C(x_1^{k+1})\prod_{i=k+2}^{n} p_{x_{i-k-1}^{i-1},x_i}(\theta_{k+1}).$$

Mesmo desprezando os termos $C(\cdot)$, se ambas expressões são referentes à imagem de uma única função, teríamos que permitir amostras de tamanhos arbitrários e inserir n e k no domínio da função. Neste caso, uma alternativa de definição para a função verossimilhança seria

$$L : E^{\infty}\times\Theta\times\mathbb{N}\times\mathbb{N}\to\mathbb{R}$$
$$L(x_1^n,\theta_k,n,k) = C(x_1^k)\prod_{i=k+1}^{n} p_{x_{i-k}^{i-1},x_i}(\theta_k)$$

para $\Theta = \cup_{i=0}^{\infty}\Theta_k$ e considerando as devidas imersões de $\Theta_k\subset\Theta$ e $E^n\subset E^{\infty}$. Entretanto, dentre possíveis outras, teríamos dificuldades em manipular as derivadas de L em relação a θ. Como exemplo, manipular a matriz Hessiana $D_{\theta}^2 L(\hat{\theta}_k)$ para um estimador de máxima verossimilhança $\hat{\theta}_k\in\Theta_k\subset\Theta$.

Além disso, em várias implementações de estimadores de ordem encontradas na literatura, é comum "desprezar" alguns fatores da verossimilhança ou até mesmo "aproximar" a verossimilhança por outra função. Esse fato evidencia que a estimação da ordem, embora tenha suas raízes no uso da verossimilhança, não necessita que a função a ser penalizada seja precisamente a log-verossimilhança.

Nesse sentido, definimos abaixo a classe de funções $\{L_{n,k}\}$, ou em uma nomenclatura mais simples, *funções* $L_{n,k}$, que possui as propriedades adequadas para a definição do estimador EDC e estabelecimento dos resultados subsequentes. No geral, as funções $L_{n,k}$ simplesmente se referem à função verossimilhança para seus respectivos n e k.

Definição 1.3. *Para x_1^n uma realização de $\mathbb{X}$, $\mathbb{M}=\{M_k\}_{k\in\mathbb{N}}$ uma sequência de modelos aninhados e $\hat{\theta}_k$ o estimador de máxima verossimilhança considerando k a ordem verdadeira, definimos a família de funções $\{L_{n,k}\}$, com $L_{n,k} : E^n\times\Theta_k\to\mathbb{R}$, que satisfaça*

$$L_{n,k}(x_1^n,\hat{\theta}_k) = \sup_{\theta\in\Theta_k}\left\{L_{n,k}(x_1^n,\theta)\right\} \tag{1.2}$$

e, para todo k, $\theta \in \Theta_k$, e $l \geq k$,

$$L_{n,l}(\theta) \geq L_{n,k}(\theta) \quad e \quad \lim_{n\to\infty} \frac{L_{n,l}(\theta)}{L_{n,k}(\theta)} < \infty. \tag{1.3}$$

Quando não houver dúvidas, utilizamos $L_n(\hat{\theta}_k) = L_{n,k}(x_1^n, \hat{\theta}_k)$ ou $L_{n,k}(\theta) = L_{n,k}(x_1^n, \theta)$.

Usando (1.3) temos que, para todo $l \geq k$, $\theta \in \Theta_k$ e sequência h_n tal que $\lim_{n\to\infty} h_n = \infty$,

$$0 \leq \frac{\log L_{n,l}(\theta) - \log L_{n,k}(\theta)}{h_n} \to 0. \tag{1.4}$$

Essa relação é utilizada para comparar as diferenças $\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)$ no que segue.

Exemplo 1.1. *Considere a família de modelos "Weibull", M_W, com*

$$\begin{aligned} E &= \mathbb{R}^+, \\ f(x_1^n, \theta) &= \prod_{k=1}^{n} \frac{\theta_1}{\theta_2} x_k^{\theta_1 - 1} e^{-\frac{x_k^{\theta_1}}{\theta_2}}, \\ \forall\, \theta &= (\theta_1, \theta_2) \in \mathbb{R}^+ \times \mathbb{R}^+. \end{aligned}$$

Com a restrição abaixo, temos a família de modelos "exponencial", M_e, e $M_e \subseteq M_W$.

$$\forall\, \theta = (1, \theta_2),\ \theta_2 \in \mathbb{R}^+.$$

No contexto apresentado, além dos casos i.i.d. como ilustrado no exemplo acima, é possível inserir os problemas de identificação de ordem de dependência para cadeias de Markov, cadeias de Markov ocultas [ordem de dependência oculta], processos Autoregressivos (AR), processos Autoregressivos de Heteroscedasticidade Condicional (ARCH), além de outros casos particulares pouco difundidos como Raftery (1985), Logan (1981) e Pegram (1980). O ferramental também pode ser utilizado além dos casos de ordem de dependência, como no problema da identificação do tamanho do espaço de estados oculto de cadeias de Markov ocultas.

1.2 Consistência de estimadores baseados no critério EDC

No que segue, definimos o estimador de ordem EDC em sequências de modelos aninhados, que não apresenta diferença significativa em relação à definição de Zhao et al. (2001).

Definição 1.4. *Para $\mathbb{M}$ uma sequência de modelos aninhados, $m_r \in \bigcup_{k=0}^{\infty} M_k$ de ordem r e $K \geq r$, definimos o estimador* EDC *de r por*

$$\hat{r}_{edc} = \underset{k\leq K}{\operatorname{argmin}} \{\mathrm{EDC}(k)\} \tag{1.5}$$

onde

$$\mathrm{EDC}(k) = -\log L_n(\hat{\theta}_k) + c_n\gamma(k),$$

$\{c_n\}_{n\in\mathbb{N}}$ é sequência de números positivos e $L_n(\hat{\theta}_k)$ é como na Definição 1.3.

A rigor, $\gamma(k)$ pode ser qualquer função crescente em k. Na literatura é comum adotar c_n como sequência que não depende de x_1^n e/ou k. Entretanto essa restrição não é necessária.

O objetivo do Teorema abaixo é concluir os casos de consistência do estimador $\hat{r}_{edc}$ baseado no comportamento assintótico das funções $\log L_{n,k}$ e da sequência c_n. Esse resultado foi estabelecido em contextos particulares, diretamente ou indiretamente, por diversos pesquisadores, dentre eles Nishii (1988), Finesso (1990) e Dorea (2008).

Teorema 1.5. *Seja $\mathbb{X}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva sequência de modelos aninhados, $m_r \in \bigcup_{k=0}^{\infty} M_k$ de ordem r, $\hat{r}_{edc}$ como definido em (1.5), $\hat{\theta}_k$ e $\log L_n(\hat{\theta}_k)$ como na Definição 1.3.*

(i) $\hat{r}_{edc}$ é fortemente consistente ($\hat{r}_{edc} \underset{q.c.}{\rightarrow} r$) se

(H1) para $k < r$, existe $c_1 \in (0,\infty)$ tal que

$$\lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_r) - \log L_n(\hat{\theta}_k)}{n} \geq c_1 \; q.c.,$$

(H2) para $k > r$, existe $c_2 \in (0,\infty)$ tal que

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \leq c_2(\gamma(k) - \gamma(r)) \; q.c.$$

e c_n*, dado em* (1.5)*, satisfaz*

$$\lim_{n\to\infty} \frac{c_n}{n} = 0 \quad e \quad \liminf_{n\to\infty} \frac{c_n}{\log\log n} \geq c_2. \tag{1.6}$$

(ii) $\hat{r}_{edc}$ *é consistente (*$\hat{r}_{edc} \underset{P}{\to} r$*) se H1 é satisfeita,* $\lim_{n\to\infty} c_n = \infty$, $\lim_{n\to\infty} \frac{c_n}{n} = 0$ *e*

(H3) para $k > r$

$$\lim_{n\to\infty} \mathrm{P}\left(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > (\gamma(k) - \gamma(r))c_n \right) = 0.$$

(iii) $\hat{r}_{edc}$ *é inconsistente se existe* $c_3 \in (0,\infty)$ *tal que* $\limsup_{n\to\infty} c_n \leq c_3 < \infty$ *e*

(H4) para $k \geq r$

$$\lim_{n\to\infty} \mathrm{P}\left(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > c_3(\gamma(k) - \gamma(r)) \right) > 0.$$

Demonstração. Temos que

$$\begin{aligned} &\left(-\log L_n(\hat{\theta}_p) + \gamma(p)c_n\right) - \left(-\log L_n(\hat{\theta}_l) + \gamma(l)c_n\right) \\ &\qquad = \left(\log L_n(\hat{\theta}_l) - \log L_n(\hat{\theta}_p)\right) - c_n\left(\gamma(l) - \gamma(p)\right). \end{aligned} \tag{1.7}$$

(i) Usando a hipótese H2 e (1.6) , substituindo $p = r$ e $l = k > r$ em (1.7), temos

$$\begin{aligned} \limsup_{n\to\infty} \frac{\left(-\log L_n(\hat{\theta}_r) + \gamma(r)c_n\right) - \left(-\log L_n(\hat{\theta}_k) + \gamma(k)c_n\right)}{\log\log n} &\leq c_2(\gamma(k) - \gamma(r)) - \liminf_{n\to\infty} \left(\frac{c_n}{\log\log n}\right)(\gamma(k) - \gamma(r)) \quad q.c. \\ &\leq c_2(\gamma(k) - \gamma(r)) - c_2(\gamma(k) - \gamma(r)) \quad q.c. \\ &= 0. \end{aligned}$$

Da mesma forma, se $l = r$ e $p = k < r$, em (1.7), usando a hipótese H1 e (1.6), temos q.c.

$$\begin{aligned} \liminf_{n\to\infty} \frac{\left(-\log L_n(\hat{\theta}_k) + \gamma(k)c_n\right) - \left(-\log L_n(\hat{\theta}_r) + \gamma(r)c_n\right)}{n} &\geq c_1 - \limsup_{n\to\infty} \frac{c_n}{n}\left(\gamma(r) - \gamma(k)\right) \quad q.c. \\ &> 0 \quad q.c. \end{aligned} \tag{1.8}$$

Então $\hat{r}_{edc} \underset{q.c.}{\to} r$.

(ii) Usando H1 concluímos (1.8) e portanto $\lim_{n\to\infty} \hat{r}_{edc} \geq r$ q.c. Por outro lado temos que

$$\mathrm{P}(\hat{r}_{edc} > r) \leq \sum_{k=r+1}^{K} \mathrm{P}(\mathrm{EDC}(k) < \mathrm{EDC}(r)) \tag{1.9}$$

e para $k > r$, usando (1.7) com $p = r$ e $l = k$,

$$\begin{aligned} \mathrm{P}(\mathrm{EDC}(k) < \mathrm{EDC}(r)) &= \mathrm{P}\left(-\log L_n(\hat{\theta}_k) + \gamma(k)c_n < -\log L_n(\hat{\theta}_r) + \gamma(r)c_n\right) \\ &= \mathrm{P}\left(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > c_n(\gamma(k) - \gamma(r))\right) \end{aligned}$$

Usando H3 e (1.9) temos que $\hat{r}_{edc} \underset{P}{\rightarrow} r$.

(iii) Para $k > r$,

$$\lim_{n\to\infty} \mathrm{P}(\hat{r}_{edc} > r) > \lim_{n\to\infty} \mathrm{P}(\mathrm{EDC}(r) - \mathrm{EDC}(k) > 0)$$

e tomando $l = k > r$ e $p = r$ em (1.7) temos

$$\begin{aligned} \lim_{n\to\infty} \mathrm{P}(\mathrm{EDC}(r) - \mathrm{EDC}(k) > 0) &= \lim_{n\to\infty} \mathrm{P}\left(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > c_n(\gamma(k) - \gamma(r))\right) \\ &\geq \lim_{n\to\infty} \mathrm{P}\left(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > c_3(\gamma(k) - \gamma(r))\right) \\ &> 0. \end{aligned}$$

Portanto $\hat{r}_{edc}$ possui uma probabilidade positiva a superestimar r.

□

Se $\log L_{n,k}$ é a log-verossimilhança, temos que a diferença $\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)$ é o logaritmo da razão da verossimilhança de Neyman-Pearson, que foi bastante utilizada nos trabalhos focados em testes de hipóteses, anteriores a Akaike (1974). Portanto, é comum se obter na literatura o comportamento assintótico dessa diferença para cada caso, que geralmente converge para uma distribuição qui-quadrado (Whittle 1951, Anderson & Goodman 1957, van der Vaart 2000, Shao 2007). Com esse resultado temos condições de provar as hipóteses H3 e H4. A hipótese H1 geralmente pode ser demonstrada, sem muitas dificuldades, usando a Lei Forte dos Grandes Números e/ou divergência de Kullback-Leibler (Kullback 1959). A questão maior é a demonstração da hipótese H2, que geralmente depende da Lei do Logaritmo Iterado, que nem sempre está disponível de forma utilizável para o processo a que se pretende estender o EDC.

Nesse sentido, o Teorema 1.6 conclui H2 usando hipóteses sobre o comportamento assintótico do estimador $\hat{\theta}_k$ ou sobre o comportamento das derivadas de primeira ordem das funções $\log L_{n,k}$. Para isso, são necessárias as condições de regularidade abaixo. Observa-se que Nishii (1988) usa condições e resultado semelhante para o caso i.i.d.

Condição 1.1 (Regularidade)**.** *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva sequência de modelos aninhados, $L_{n,k}$ como na Definição 1.3, $m_r = f(x_1^n, \theta_r) \in M_r$ as densidades de dimensão finita de $\mathbb{X}$, $k \geq r$ e $\hat{\theta}_k$ o estimador de máxima verossimilhança de θ_r. São satisfeitos os seguintes.*

(i) *θ_r é ponto interior de Θ_k e*

$$\hat{\theta}_k \underset{q.c.}{\rightarrow} \theta_r. \tag{1.10}$$

(ii) *Para todo $k,n \in \mathbb{N}$, $\log L_{n,k}(\theta)$ e suas derivadas, $D^1_\theta(\log L_{n,k}(\theta))$, $D^2_\theta(\log L_{n,k}(\theta))$ e $D^3_\theta(\log L_{n,k}(\theta))$, são mensuráveis com respeito a x_1^n e contínuas com respeito a θ.*

(iii) *Para $\dot{\theta} = (1-s)\hat{\theta}_k + s\theta_r$, $s \in (0,1)$, $(i,j,l) \in \{1,\ldots,\gamma(k)\}^3$,*

$$\lim_{n\to\infty} \frac{\left(D^3_\theta(\log L_{n,k}(\dot{\theta}))\right)_{i,j,l}}{n} < \infty \quad q.c. \tag{1.11}$$

Teorema 1.6. *Seja $\mathbb{X}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva sequência de modelos aninhados, $\log L_n(\hat{\theta}_k)$ como na Definição 1.3, $\theta_r = (\alpha_1,\ldots,\alpha_{\gamma(k)}) \in \Theta_k$, $\hat{\theta}_k = (\hat{\alpha}_1,\ldots,\hat{\alpha}_{\gamma(k)}) \in \Theta_k$, $\hat{\theta}_k$ o estimador de máxima verossimilhança de θ_r e valem as Condições de Regularidade (Condição 1.1).*

(i) *Se existe matriz A_2 positiva definida tal que*

$$\lim_{n\to\infty} -\frac{D^2_\theta\left(\log L_n(\hat{\theta}_k)\right)}{n} = A_2 \quad q.c. \tag{1.12}$$

e existe $c_4 \in (0,\infty)$ tal que

$$\limsup_{n\to\infty} \left\| \frac{\sqrt{n}(\theta_r - \hat{\theta}_k)}{\sqrt{2\log\log n}} \right\| \leq c_4 \quad q.c. \tag{1.13}$$

então

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \leq 2c_4^2\lambda_1 \quad q.c.$$

onde λ_1 é o maior autovalor de A_2.

(ii) *Se existe matriz A_2 positiva definida tal que, para $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$ e $s \in [0,1]$,*

$$\lim_{n\to\infty} -\frac{D^2_\theta\left(\log L_n(\dot{\theta})\right)}{n} = A_2 \quad q.c., \tag{1.14}$$

e existe $c_5 \in (0,\infty)$ tal que

$$\limsup_{n\to\infty} \left\| \frac{D_\theta^1 \log L_n(\theta_r)}{\sqrt{2n\log\log n}} \right\| \le c_5 \quad q.c. \tag{1.15}$$

então, para todo $i \in \{1,\ldots,\gamma(k)\}$,

$$\limsup_{n\to\infty} \left| \frac{\sqrt{n}(\hat{\alpha}_i - \alpha_i)}{\sqrt{2\log\log n}} \right| < \infty \quad q.c. \tag{1.16}$$

e

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \le \frac{2c_5^2}{\lambda_{\gamma(k)}} \quad q.c.$$

onde $\lambda_{\gamma(k)}$ é o menor autovalor de A_2.

(iii) Se para todo $(i,j) \in \{1,\ldots,\gamma(k)\}^2$, existem c_6, c_7 tais que

$$\limsup_{n\to\infty} \left| \frac{\sqrt{n}(\hat{\alpha}_i - \alpha_i)}{\sqrt{2\log\log n}} \right| \le c_6 \quad q.c. \tag{1.17}$$

e

$$\lim_{n\to\infty} \left(\frac{D_\theta^2 \left(\log L_n(\hat{\theta}_k)\right)}{n} \right)_{i,j} \le c_7 \quad q.c., \tag{1.18}$$

então

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \le 2\gamma(k)^2 c_6^2 c_7 \quad q.c.$$

Demonstração. Usando as condições de regularidade e tomando a expansão em série de Taylor para $\log L_{n,k}(\theta_r)$ no ponto $\hat{\theta}_k$ temos

$$\begin{aligned} \log L_{n,k}(\theta_r) &= \log L_{n,k}(\hat{\theta}_k) + (\theta_r - \hat{\theta}_k) D_\theta^1 \left(\log L_{n,k}(\hat{\theta}_k)\right) \\ &\quad + \frac{1}{2}(\theta_r - \hat{\theta}_k) D_\theta^2 \left(\log L_{n,k}(\hat{\theta}_k)\right)(\theta_r - \hat{\theta}_k)^T + r_n(\theta_r - \hat{\theta}_k) \end{aligned} \tag{1.19}$$

onde

$$r_n(\theta_r - \hat{\theta}_k) = \frac{1}{3!} \sum_{i,j,l} \left(D_\theta^3 \log L_{n,k}(\dot{\theta})\right)_{i,j,l} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j)(\alpha_l - \hat{\alpha}_l),$$

$\theta_r = (\alpha_i,\ldots,\alpha_{\gamma(k)})$, $\hat{\theta}_k = (\hat{\alpha}_i,\ldots,\hat{\alpha}_{\gamma(k)})$ e $\dot{\theta} = (1-s)\hat{\theta}_k + s\theta_r$, $s \in (0,1)$. Como $\hat{\theta}_k$ maximiza

$L_{n,k}$, temos que $D^1_\theta\left(\log L_{n,k}(\hat\theta_k)\right)=0$. Organizando (1.19) e dividindo por $\log\log n$, temos

$$\begin{aligned}\limsup_{n\to\infty}\frac{\log L_{n,k}(\hat\theta_k)-\log L_{n,k}(\theta_r)}{\log\log n} &\le \limsup_{n\to\infty}\frac{\sqrt{n}(\theta_r-\hat\theta_k)}{\sqrt{2\log\log n}}\frac{-D^2_\theta\left(\log L_{n,k}(\hat\theta_k)\right)}{n}\frac{\left(\sqrt{n}(\theta_r-\hat\theta_k)\right)^T}{\sqrt{2\log\log n}}\\ &\quad+\limsup_{n\to\infty}\frac{|r_n(\theta_r-\hat\theta_k)|}{\log\log n} \qquad (1.20)\end{aligned}$$

Usando (1.13) ou (1.26) ou (1.17) temos que

$$\lim_{n\to\infty}\left|\frac{\sqrt{n}(\alpha_i-\hat\alpha_i)}{\sqrt{\log\log n}}\right|<\infty \quad q.c.$$

Portanto, usando (1.11) e (1.10),

$$\begin{aligned}\limsup_{n\to\infty}\frac{|r_n(\theta_r-\hat\theta_k)|}{\log\log n} &\le \limsup_{n\to\infty}\frac{1}{3!}\sum_{i,j,l}\left|\frac{\left(D^3_\theta\log L_{n,k}(\dot\theta)\right)_{i,j,l}}{n}\right|\left|\frac{\sqrt{n}(\alpha_i-\hat\alpha_i)}{\sqrt{\log\log n}}\right|\left|\frac{\sqrt{n}(\alpha_j-\hat\alpha_j)}{\sqrt{\log\log n}}\right||\alpha_l-\hat\alpha_l|\\ &\le c\sum_l\limsup_{n\to\infty}|\alpha_l-\hat\alpha_l|=0 \quad q.c. \qquad (1.21)\end{aligned}$$

(i) Como A_2 é positiva definida, temos que $\forall x\in\mathbb{R}^{\gamma(k)}$, $xA_2x^T\le\lambda_1||x||^2$, para λ_1 o maior autovalor de A_2 (Zhang 2011). Usando (1.20), (1.12), (1.13) e (1.21) temos

$$\begin{aligned}\limsup_{n\to\infty}\frac{\log L_{n,k}(\hat\theta_k)-\log L_{n,k}(\theta_r)}{\log\log n} &\le \limsup_{n\to\infty}\frac{\sqrt{n}(\theta_r-\hat\theta_k)}{\sqrt{2\log\log n}}\frac{-D^2_\theta\left(\log L_{n,k}(\hat\theta_k)\right)}{n}\frac{\left(\sqrt{n}(\theta_r-\hat\theta_k)\right)^T}{\sqrt{2\log\log n}}\\ &\le \limsup_{n\to\infty}\lambda_1\left\|\frac{\sqrt{n}(\theta_r-\hat\theta_k)}{\sqrt{2\log\log n}}\right\|^2\\ &\le \lambda_1c_4^2 \quad q.c. \qquad (1.22)\end{aligned}$$

Portanto, aplicando (1.22) duas vezes e usando (1.4), obtemos

$$\begin{aligned}\limsup_{n\to\infty}\frac{\log L_{n,k}(\hat\theta_k)-\log L_{n,r}(\hat\theta_r)}{\log\log n} &= \limsup_{n\to\infty}\frac{\log L_{n,k}(\hat\theta_k)-\log L_{n,k}(\theta_r)-(\log L_{n,r}(\hat\theta_r)-\log L_{n,r}(\theta_r))+(\log L_{n,k}(\theta_r)-\log L_{n,r}(\theta_r))}{\log\log n}\\ &\le \limsup_{n\to\infty}\frac{\log L_{n,k}(\hat\theta_k)-\log L_{n,k}(\theta_r)}{\log\log n}+\limsup_{n\to\infty}\frac{\log L_{n,r}(\hat\theta_r)-\log L_{n,r}(\theta_r)}{\log\log n} \qquad (1.23)\\ &\le 2\lambda_1c_4^2.\end{aligned}$$

Usamos em (1.23) que, pela definição de $L_{n,k}$, $\log L_{n,k}(\hat\theta_k)\ge\log L_{n,k}(\theta_r)$ e $\log L_{n,r}(\hat\theta_r)\ge\log L_{n,r}(\theta_r)$.

(ii) Tomando a série de Taylor de $D^1_\theta\log L_{n,k}(\hat\theta_k)$ no ponto θ_r,

$$\begin{aligned}(0,\cdots,0) &= D^1_\theta\log L_{n,k}(\hat\theta_k)\\ &= D^1_\theta\log L_{n,k}(\theta_r)+(\hat\theta_k-\theta_r)D^2_\theta\log L_{n,k}(\dot\theta),\end{aligned}$$

onde $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$ e $s \in (0,1)$. Organizando,

$$\begin{aligned}
\frac{1}{\sqrt{n\log\log n}} D_\theta^1 \log L_{n,k}(\theta_r) &= \frac{\sqrt{n}}{n\sqrt{\log\log n}}\left\{(\hat{\theta}_k - \theta_r) D_\theta^2 \log L_{n,k}(\dot{\theta})\right\} \\
&= -\frac{\sqrt{n}}{\sqrt{\log\log n}}(\hat{\theta}_k - \theta_r)\left[\frac{D_\theta^2 \log L_{n,k}(\dot{\theta})}{n}\right].
\end{aligned}$$

Como A_2 é positiva definida, é invertível, e para n suficientemente grande,

$$A_n := -\left[\frac{D_\theta^2 \log L_{n,k}(\dot{\theta})}{n}\right]$$

possui inversa A_n^{-1}, então

$$\frac{1}{\sqrt{n\log\log n}} D_\theta^1 \log L_{n,k}(\theta_r) A_n^{-1} = \frac{\sqrt{n}}{\sqrt{\log\log n}}(\hat{\theta}_k - \theta_r). \tag{1.24}$$

Aplicando (1.24), (1.21), (1.14), (1.15) e que $A_n \underset{q.c.}{\rightarrow} A_2$, obtemos

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\theta_r)}{\log\log n} &\leq \limsup_{n\to\infty} \frac{\left(D_\theta^1 \log L_{n,k}(\theta_r)\right) A_n^{-1}}{\sqrt{2n\log\log n}} \frac{-D_\theta^2\left(\log L_{n,k}(\hat{\theta}_k)\right)}{n}\left(\frac{\left(D_\theta^1 \log L_{n,k}(\theta_r)\right) A_n^{-1}}{\sqrt{2n\log\log n}}\right)^T \\
&= \limsup_{n\to\infty} \frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n\log\log n}} A_2^{-1} A_2 \left(\frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n\log\log n}} A_2^{-1}\right)^T \\
&= \limsup_{n\to\infty} \frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n\log\log n}} A_2^{-1} \left(\frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n\log\log n}}\right)^T \\
&\leq \frac{1}{\lambda_{\gamma(k)}} \limsup_{n\to\infty} \left\|\frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n\log\log n}}\right\|^2 \\
&\leq \frac{c_5^2}{\lambda_{\gamma(k)}} \quad q.c.
\end{aligned} \tag{1.25}$$

Usamos o fato de matrizes positivas definidas serem simétricas ($A_2 = A_2^T$) [1] e possuírem inversas positivas definidas. Aplicando o mesmo argumento de (1.23), concluímos de (1.25) a segunda inequação do resultado. Para a primeira, basta utilizar

[1] Nesse trabalho consideramos simetria na definição de matrizes positiva definidas.

(1.24), considerar $P_i : \mathbb{R}^{\gamma(k)} \to \mathbb{R}$ como a projeção da coordenada i e observar que

$$\begin{aligned}
\limsup_{n\to\infty}\left|\frac{\sqrt{n}(\hat{\alpha}_i-\alpha_i)}{\sqrt{2\log\log n}}\right| &= \limsup_{n\to\infty}\left|P_i\left(\frac{\sqrt{n}}{\sqrt{2\log\log n}}(\hat{\theta}_k-\theta_r)\right)\right| \\
&= \limsup_{n\to\infty}\left|P_i\left(\frac{1}{\sqrt{2n\log\log n}}D^1_\theta\log L_{n,k}(\theta_r)A_n^{-1}\right)\right| \\
&= \limsup_{n\to\infty}\left|P_i\left(\frac{1}{\sqrt{2n\log\log n}}D^1_\theta\log L_{n,k}(\theta_r)A_2^{-1}\right)\right| \\
&< \infty. \qquad (1.26)
\end{aligned}$$

(iii) Aplicando (1.17), (1.18), (1.21) e (1.20), obtemos

$$\limsup_{n\to\infty}\frac{\log L_n(\hat{\theta}_k)-\log L_n(\theta_r)}{\log\log n} \leq \gamma(k)^2c_6^2c_7 \quad q.c.$$

Com o mesmo raciocínio usado em (1.23), obtemos o resultado.

□

O resultado dos itens (i)-(iii) do Teorema 1.6 é basicamente o mesmo – uma cota superior em H2. Entretanto, quanto menor a cota superior encontrada maior é a classe de estimadores $\hat{r}_{edc}$ fortemente consistentes. Assim, as três opções visam simplificar o trabalho de obtenção de uma cota pequena.

Geralmente, logo após a definição de determinado processo, a primeira questão abordada são as condições de ergodicidade para ser possível a aplicação dos equivalentes a Lei Forte dos Grandes Números (LFGN), Teorema do Limite Central (TLC) e Lei do Logaritmo Iterado (LLI). Na sequência, buscam-se condições para a consistência forte do estimador do parâmetro θ_r e a normalidade assintótica desse estimador. Para a obtenção desse último passo, geralmente são necessárias algumas das hipóteses exigidas pelo Teorema 1.6, que podem ser utilizadas para o estabelecimento dos estimadores EDC.

O item (ii) do Teorema 1.6, essencialmente, conclui H2 baseado na LLI para as derivadas de primeira ordem de $\log L_{n,k}$. Em resultado análogo, e usando técnica semelhante, Basawa & Heyde (1976) conclui a normalidade assintótica de $\hat{\theta}_r$ a partir de normalidade assintótica das derivadas de primeira ordem da log-verossimilhança. Para isso são utilizadas condições semelhantes às exigidas pelo Teorema 1.6.

Como $\hat{\theta}_k$ maximiza $\log L_{n,k}$, temos que a Hessiana no ponto $\hat{\theta}_k$ é negativa. Neste caso, $-D^2_\theta \log L_n(\hat{\theta}_k)$ é positiva para todo n. Suponha positiva definida para melhor entendimento da equação (1.12). A equação requer que $-D^2_\theta \log L_n(\hat{\theta}_k)/n$ possua limite quase certo a matriz A_2, positiva definida. Em outro ponto de vista, suponha que o processo seja markoviano e que L_n seja a função verossimilhança, neste caso, suponha ainda que $-D^2_\theta \log L_n(\hat{\theta})/n = -\sum D^2_\theta \log f_{\hat{\theta}}(x_i|x_{i-r},\dots,x_{i-1})\,/n$ para alguma densidade condicional $f_{\hat{\theta}}(.|.)$. Se vale a LFGN para esse limite, temos que $-\sum D^2_\theta \log f_{\hat{\theta}}(x_i|x_{i-r},\dots,x_{i-1})\,/n$ converge quase certamente para a matriz de Informação de Fisher para valores próximos a θ_r e, que por (1.12), deve ser positiva definida.

A normalidade assintótica do estimador $\hat{\theta}_k$ pode ser utilizada para concluir H3. O próximo resultado provê condições para isso.

Teorema 1.7. *Seja $\mathbb{X}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva sequência de modelos aninhados, $L_n(\hat{\theta}_k)$ como na Definição 1.3, $\theta_r = (\alpha_1,\dots,\alpha_{\gamma(k)})$ $\in \Theta_k$, $\hat{\theta}_k = (\hat{\alpha}_1,\dots,\hat{\alpha}_{\gamma(k)}) \in \Theta_k$, $\hat{\theta}_k$ o estimador de máxima verossimilhança de θ_r e valem as Condições de Regularidade (Condição 1.1). Se existe matriz A_2 positiva definida tal que*

$$\lim_{n\to\infty} -\frac{D^2_\theta\left(\log L_n(\hat{\theta}_k)\right)}{n} = A_2 \quad q.c., \tag{1.27}$$

para todo $i \in \{1,\dots,\gamma(k)\}$,

$$\limsup_{n\to\infty}\left|\frac{\sqrt{n}(\hat{\alpha}_i-\alpha_i)}{\sqrt{\log\log n}}\right| < \infty \quad q.c. \tag{1.28}$$

e, para todo h_n, tal que $\lim_{n\to\infty} h_n = \infty$,

$$\lim_{n\to\infty} P(\left\|\sqrt{n}(\theta_r-\hat{\theta}_k)\right\| > h_n) = 0$$

então

$$\lim_{n\to\infty} \mathrm{P}(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > h_n) = 0. \tag{1.29}$$

para todo h_n, tal que $\lim_{n\to\infty} h_n = \infty$.

Demonstração. Usando as condições de regularidade e tomando a expansão em série de

Taylor para $\log L_{n,k}(\theta_r)$ no ponto $\hat{\theta}_k$ temos

$$\begin{aligned}\log L_{n,k}(\theta_r) &= \log L_{n,k}(\hat{\theta}_k)+(\theta_r-\hat{\theta}_k)D^1_\theta\left(\log L_{n,k}(\hat{\theta}_k)\right)\\ &+\frac{1}{2}(\theta_r-\hat{\theta}_k)D^2_\theta\left(\log L_{n,k}(\hat{\theta}_k)\right)(\theta_r-\hat{\theta}_k)^T+r_n(\theta_r-\hat{\theta}_k)\end{aligned}$$

onde

$$r_n(\theta_r-\hat{\theta}_k) = \frac{1}{3!}\sum_{i,j,l}\left(D^3_\theta \log L_{n,k}(\dot{\theta})\right)_{i,j,l}(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)(\alpha_l-\hat{\alpha}_l)$$

Para $\dot{\theta}=(1-s)\theta_k+s\theta_r$, $s\in(0,1)$. Usando (1.28) e (1.11), obtemos

$$\begin{aligned}\limsup_{n\to\infty}|r_n(\theta_r-\hat{\theta}_k)| &\leq \limsup_{n\to\infty}\frac{1}{3!}\sum_{i,j,l}\left|\frac{(D^3_\theta\log L_{n,k}(\dot{\theta}))_{i,j,l}}{n}\right|\left|\frac{\sqrt{n}(\alpha_i-\hat{\alpha}_i)}{\sqrt{\log\log n}}\right|\left|\frac{\sqrt{n}(\alpha_j-\hat{\alpha}_j)}{\sqrt{\log\log n}}\right|\left|\frac{\sqrt{n}(\alpha_l-\hat{\alpha}_l)}{\sqrt{\log\log n}}\right|\frac{(\log\log n)^{3/2}}{\sqrt{n}}\\ &\leq c\limsup_{n\to\infty}\frac{(\log\log n)^{3/2}}{\sqrt{n}}=0\quad q.c.\end{aligned}$$

Usando (1.27), para n suficientemente grande, $A_n=-\frac{D^2_\theta(\log L_{n,k}(\hat{\theta}_k))}{n}$ é positiva definida, e portanto possui maior autovalor, λ_n, tal que $\lambda_n \underset{q.c.}{\to} \lambda_1$, onde λ_1 é o maior autovalor de A_2. Como $\forall x$, $xA_nx^T\leq\lambda_n||x||^2$, temos

$$\begin{aligned}\lim_{n\to\infty}\mathrm{P}(\log L_{n,k}(\hat{\theta}_k)-\log L_{n,k}(\theta_r)>h_n) &= \lim_{n\to\infty}\mathrm{P}\left(\frac{1}{2}\sqrt{n}(\theta_r-\hat{\theta}_k)\frac{-D^2_\theta(\log L_n(\hat{\theta}_k))}{n}(\sqrt{n}(\theta_r-\hat{\theta}_k))^T>h_n\right)\\ &\leq \lim_{n\to\infty}\mathrm{P}\left(\frac{\lambda_n}{2}\left\|\sqrt{n}(\theta_r-\hat{\theta}_k)\right\|^2>h_n\right)\\ &= \lim_{n\to\infty}\mathrm{P}\left(\left\|\sqrt{n}(\theta_r-\hat{\theta}_k)\right\|>\frac{\sqrt{2h_n}}{\sqrt{\lambda_n}}\right)=0\end{aligned}$$

onde concluímos (1.29), considerando (1.4). □

Em resumo, se as condições abaixo são atendidas, o Teorema 1.6 pode ser utilizado para obter H2. Adicionalmente, se a hipótese H1 é verdadeira, as condições para a definição de uma classe de estimadores EDC fortemente consistentes são atendidas.

(i) θ_r é ponto interior de Θ_k e

$$\hat{\theta}_k \underset{q.c.}{\to} \theta_r.$$

(ii) Para todo $k,n\in\mathbb{N}$, $\log L_{n,k}(\theta)$ e suas derivadas, $D^1_\theta(\log L_{n,k}(\theta))$, $D^2_\theta(\log L_{n,k}(\theta))$ e $D^3_\theta(\log L_{n,k}(\theta))$, são mensuráveis com respeito a x_1^n e contínuas com respeito a θ.

(iii) Para $\dot{\theta} = (1-s)\hat{\theta}_k + s\theta_r$, $s \in (0,1)$, $(i,j,l) \in \{1,\ldots,\gamma(k)\}^3$,

$$\lim_{n\to\infty} \frac{\left(D_\theta^3(\log L_{n,k}(\dot{\theta}))\right)_{i,j,l}}{n} < \infty \quad q.c.$$

(iv) Existe matriz A_2 positiva definida tal que, para $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$ e $s \in [0,1]$,

$$\lim_{n\to\infty} -\frac{D_\theta^2\left(\log L_{n,k}(\dot{\theta})\right)}{n} = A_2 \quad q.c.$$

(v)

$$\limsup_{n\to\infty} \left\| \frac{D_\theta^1 \log L_{n,k}(\theta_r)}{\sqrt{2n\log\log n}} \right\| \leq c_5 \quad q.c.$$

1.3 Generalização: ordem em modelos parcialmente aninhados

Sob algumas hipóteses, um processo AR(r) pode ser especificado com r parâmetros e um processo AR($r-1$) pode ser considerado como um processo AR(r) com um parâmetro pré-fixado. Nesse sentido temos o conceito de "aninhamento" definido na Seção anterior e podemos considerar a sequência de processos $\{AR(0), AR(1), \cdots\}$ como uma sequência de modelos aninhados. De forma semelhante, temos os processos ARMA(r_1,r_2), que são especificados com r_1+r_2 parâmetros. Quando $k_1 \geq r_1$ e $k_2 \geq r_2$ podemos fixar alguns parâmetros e considerar um modelo ARMA(r_1,r_2) imerso em um modelo ARMA(k_1,k_2). Entretanto quando $k_1 < r_1$ e $k_2 \geq r_2$ o processo de imersão não é aplicável.

Mesmo nesses casos, os critérios de informação (AIC, BIC, HQC, dentre outros) são utilizados na determinação da ordem, como exemplo cita-se Hannan (1980). Entretanto, como os problemas de determinação de ordem são tratados na literatura de forma particularizada, a questão de aninhamento não é abordada diretamente. Por outro lado, para se propor uma generalização, há a necessidade de se definir bem esses conceitos no contexto de seleção de modelos.

A alternativa aparentemente viável seria definir o aninhamento diagonal, como exemplo, para o caso mencionado teríamos a sequência de modelos aninhados $\{ARMA(0,0),$ $ARMA(1,1), \cdots\}$, e tratar o caso geral fixando uma dimensão e considerando sequências de modelos aninhados na outra dimensão. Entretanto, para implementar essa alternativa seria necessário dividir o processo de estimação da ordem em etapas, o que não é desejável.

Para o caso unidimensional, a relação de aninhamento na sequência de modelos aninhado coincide com a relação de ordem em $\mathbb{N}$. Assim, a definição proposta nesse trabalho para tratar casos com ordem de dimensão p é relacionar a ordem parcial em $\mathbb{N}^p$ com a relação de aninhamento. As próximas definições tratam essa generalização. Em seguida são apresentados os resultados da Seção anterior nesse novo contexto.

Definição 1.8. *Para $p \in \mathbb{N}$, seja $k = (k_1, \ldots, k_p) \in \mathbb{N}^p$ e $r = (r_1, \ldots, r_p) \in \mathbb{N}^p$. Define-se os seguintes.*

(i) A relação de ordem: $k \geq r \Leftrightarrow k_i \geq r_i$ para $i = 1 \ldots p$. [Com isso $(\mathbb{N}^p, \geq)$ é um conjunto parcialmente ordenado.]

(ii) $r \leq k$ quando $k \geq r$ e $k < r$ quando $r \geq k$ e $k \neq r$.

(iii) $k \ngeq r$ se $r < k$ ou quando k e r não estão relacionados.

Definição 1.9. *(i) Para $p \in \mathbb{N}$, $\mathbb{M} = \{M_k\}_{k \in \mathbb{N}^p}$ é uma classe de modelos parcialmente aninhados se vale: $k \leq r$ se e somente se $M_k \subseteq M_r$.*

(ii) Se $\mathbb{M}$ é classe de modelos parcialmente aninhados e $m_r \in \bigcup\limits_{k \in \mathbb{N}^p} M_k$, dizemos que m_r é de ordem r se $m_r \in M_r$ e, se $m_r \in M_j$, então $M_r \subseteq M_j$.

(iii) Se $\mathbb{M} = \{M_k\}_{k \in \mathbb{N}^p}$ é classe de modelos parcialmente aninhados, denotamos $\gamma(k) = dim(\Theta_k)$.

Com pequenas adaptações nos Teoremas 1.5, 1.6 e 1.7 obtemos os seguintes resultados.

Teorema 1.10. *Seja $\mathbb{X}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva classe de modelos parcialmente aninhados, $m_r \in \bigcup\limits_{k \in \mathbb{N}^p} M_k$ de ordem r e $\hat{r}_{edc}$ como definido em (1.5).*

(i) $\hat{r}_{edc}$ é fortemente consistente ($\hat{r}_{edc} \underset{q.c.}{\rightarrow} r$) se

(H1) para $k \ngeq r$, existe $c_1 \in (0,\infty)$ tal que

$$\lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_r) - \log L_n(\hat{\theta}_k)}{n} \geq c_1 \; q.c.,$$

(H2) para $k > r$, existe $c_2 \in (0,\infty)$ tal que

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \leq c_2(\gamma(k) - \gamma(r)) \; q.c.$$

e c_n satisfaz

$$\lim_{n\to\infty} \frac{c_n}{n} = 0 \quad e \quad \liminf_{n\to\infty} \frac{c_n}{\log\log n} \geq c_2.$$

(ii) $\hat{r}_{edc}$ é consistente ($\hat{r}_{edc} \underset{P}{\to} r$) se H1 é satisfeita, $\lim_{n\to\infty} c_n = \infty$, $\lim_{n\to\infty} \frac{c_n}{n} = 0$ e

(H3) para $k > r$

$$\lim_{n\to\infty} \mathrm{P}(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > (\gamma(k) - \gamma(r))c_n) = 0.$$

(iii) $\hat{r}_{edc}$ é inconsistente se $\limsup_{n\to\infty} c_n \leq c_3 < \infty$ e

(H4) para $k \geq r$

$$\lim_{n\to\infty} \mathrm{P}(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > c_3(\gamma(k) - \gamma(r))) > 0.$$

Condição 1.2 (Regularidade)**.** *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva classe de modelos parcialmente aninhados, $L_{n,k}$ como na Definição 1.3, $m_r = f(x_1^n, \theta_r) \in M_r$ as densidades de dimensão finita de $\mathbb{X}$, $k \geq r$ e $\hat{\theta}_k$ o estimador de máxima verossimilhança de θ_r. São satisfeitos os seguintes.*

(i) θ_r é ponto interior de Θ_k e

$$\hat{\theta}_k \underset{q.c.}{\to} \theta_r.$$

(ii) Para todo $k, n \in \mathbb{N}$, $\log L_{n,k}(\theta)$ e suas derivadas, $D^1_\theta(\log L_{n,k}(\theta))$, $D^2_\theta(\log L_{n,k}(\theta))$ e $D^3_\theta(\log L_{n,k}(\theta))$, são mensuráveis com respeito a x_1^n e contínuas com respeito a θ.

(iii) Para $\dot{\theta} = (1-s)\hat{\theta}_k + s\theta_r$, $s \in (0,1)$, $(i,j,l) \in \{1,\ldots,\gamma(k)\}^3$,

$$\lim_{n\to\infty} \frac{\left(D^3_\theta(\log L_{n,k}(\dot{\theta}))\right)_{i,j,l}}{n} < \infty \quad q.c.$$

Teorema 1.11. *Seja $\mathbb{X}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva classe de modelos parcialmente aninhados, $L_n(\hat{\theta}_k)$ como na Definição 1.3, $\theta_r = (\alpha_1, \ldots, \alpha_{\gamma(k)}) \in \Theta_k$, $\hat{\theta}_k = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{\gamma(k)}) \in \Theta_k$, $\hat{\theta}_k$ o estimador de máxima verossimilhança de θ_r e valem as Condições de Regularidade (Condição 1.2).*

(i) Se existe matriz A_2 positiva definida tal que

$$\lim_{n\to\infty} -\frac{D^2_\theta\left(\log L_n(\hat{\theta}_k)\right)}{n} = A_2 \quad q.c.$$

e

$$\limsup_{n\to\infty} \left\| \frac{\sqrt{n}(\theta_r - \hat{\theta}_k)}{\sqrt{2\log\log n}} \right\| \leq c_4 \quad q.c.,$$

então

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \leq 2c_4^2\lambda_1 \quad q.c.,$$

onde λ_1 é o maior autovalor de A_2.

(ii) Se existe matriz A_2 positiva definida tal que, para $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$ e $s \in [0,1]$,

$$\lim_{n\to\infty} -\frac{D^2_\theta\left(\log L_n(\dot{\theta})\right)}{n} = A_2 \quad q.c.,$$

e

$$\limsup_{n\to\infty} \left\| \frac{D^1_\theta \log L_n(\theta_r)}{\sqrt{2n\log\log n}} \right\| \leq c_5 \quad q.c.$$

então, para todo $i \in \{1, \ldots, \gamma(k)\}$,

$$\limsup_{n\to\infty} \left| \frac{\sqrt{n}(\hat{\alpha}_i - \alpha_i)}{\sqrt{2\log\log n}} \right| < \infty \quad q.c.$$

e

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \leq \frac{2c_5^2}{\lambda_{\gamma(k)}} \quad q.c.,$$

onde $\lambda_{\gamma(k)}$ é o menor autovalor de A_2.

(iii) Se para todo $(i,j) \in \{1, \ldots, \gamma(k)\}^2$,

$$\limsup_{n\to\infty} \left| \frac{\sqrt{n}(\hat{\alpha}_i - \alpha_i)}{\sqrt{2\log\log n}} \right| \leq c_6 \quad q.c.$$

e

$$\lim_{n\to\infty}\left(\frac{D^2_\theta\left(\log L_n(\hat{\theta}_k)\right)}{n}\right)_{i,j}\leq c_7\quad q.c.,$$

então

$$\limsup_{n\to\infty}\frac{\log L_n(\hat{\theta}_k)-\log L_n(\hat{\theta}_r)}{\log\log n}\leq 2\gamma(k)^2c_6^2c_7\quad q.c.$$

Teorema 1.12. *Seja $\mathbb{X}$ um processo estocástico a tempo discreto com valores em $\mathbb{R}^m$, $\mathbb{M}$ sua respectiva classe de modelos parcialmente aninhados, $k\geq r$, $L_n(\hat{\theta}_k)$ como na Definição 1.3, $\theta_r=(\alpha_1,\ldots,\alpha_{\gamma(k)})\in\Theta_k$, $\hat{\theta}_k=(\hat{\alpha}_1,\ldots,\hat{\alpha}_{\gamma(k)})\in\Theta_k$, $\hat{\theta}_k$ o estimador de máxima verossimilhança de θ_r e valem as Condições de Regularidade (Condição 1.2). Se existe matriz A_2 positiva definida tal que*

$$\lim_{n\to\infty}-\frac{D^2_\theta\left(\log L_n(\hat{\theta}_k)\right)}{n}=A_2\quad q.c.,$$

para todo $(i,j)\in\{1,\ldots,\gamma(k)\}^2$,

$$\limsup_{n\to\infty}\left|\frac{\sqrt{n}(\hat{\alpha}_i-\alpha_i)}{\sqrt{\log\log n}}\right|<\infty\quad q.c.$$

e, para todo h_n, tal que $\lim\limits_{n\to\infty}h_n=\infty$,

$$\lim_{n\to\infty}P(\left\|\sqrt{n}(\theta_r-\hat{\theta}_k)\right\|>h_n)=0,$$

então

$$\lim_{n\to\infty}\mathrm{P}(\log L_n(\hat{\theta}_k)-\log L_n(\hat{\theta}_r)>h_n)=0$$

para todo h_n, tal que $\lim\limits_{n\to\infty}h_n=\infty$.

2 Cadeias de Markov de espaço de estados gerais

Para se obter resultados sobre a consistência do estimador EDC é necessário garantir o comportamento assintótico das funções $\log L_{n,k}$, que no fundo se resume em versões aplicadas da Lei Forte dos Grandes Números (LFGN), Lei do Logaritmo Iterado (LLI) e Teorema do Limite Central (TLC).

Para a classe dos processos que possuem dependência finita, o ambiente mais geral é o de cadeias de Markov com espaço de estados gerais. Nesse contexto podem ser inseridos os seguintes.

- Modelo Linear em Espaço de Estados (LSS), que tem como caso particular o processo Autoregressivo (AR).
- Processos Autoregressivos com Médias Móveis (ARMA), que tem o casos particulares o processo de Médias Móveis (MA) e os Autoregressivos (AR).
- A família de processos ARCH, que incluem GARCH (Bollerslev 1986), NGARCH (Engle & Ng 1993), EGARCH (Nelson 1991) e variações multivariadas como BEKK-GARCH (Engle & Kroner 1995), VEC-GARCH (Bollerslev, Engle & Wooldridge 1988), CCC-GARCH (Bollerslev 1990), dentre outros.
- Modelos Não-Lineares em Espaço de Estados, que incluem modelos Multilineares em Espaço de Estados.

Todavia, ser inserido no ambiente de cadeias de Markov não é suficiente para a conclusão dos resultados assintóticos. Ainda é necessário o estabelecimento de condições para obter a ergodicidade geométrica ou ergodicidade V-uniforme para ser possível o uso dos resultados assintóticos. Mesmo assim, há um grande trabalho desenvolvido para a definição dessas condições em cada caso particular.

Nas últimas décadas, cadeias de Markov teve um grande desenvolvimento, dentre muitos outros, citamos os trabalhos de Kolmogorov (1936), Doeblin (1937, 1940), Foster (1953), Harris (1956), Rosenblatt (1964, 1974), Doob (1966), Feller (1968), Orey (1959, 1971), Cogburn (1972), Tweedie (1974, 1975, 1976), Athreya & Ney (1978, 1980), Nummelin (1978, 1984) e Niemi & Nummelin (1982). A parte significativa desse desenvolvimento, acrescida de novos resultados, pode ser encontrada em Meyn & Tweedie (1993), que é a referência mais importante na área.

Nesse capítulo aplicamos os resultados do capítulo anterior na estimação de ordem de processos Autoregressivos (AR), Autoregressivos de Heteroscedasticidade Condicional (ARCH) e Autoregressivos de Heteroscedasticidade Condicional Generalizado na Modelagem BEKK (BEKK-GARCH). Para processos AR foi possível encontrar a ergodicidade V-uniforme suficiente para a aplicação da LLI, necessária para a conclusão da hipótese H2. Porém, para os outros casos a ergodicidade V-uniforme ainda não foi estabelecida de forma suficiente para concluir H2. Entretanto, as condições para aplicação da LFGN no contexto de cadeias de Markov estão atendidas e são utilizadas. Em razão disso, foi necessário a aplicação da LLI estabelecida para Martingales por Hall & Heyde (1980).

Na Seção 2.1 são apresentados resumidamente os principais resultados de Meyn&Tweedie. Nas Seções subsequentes, esses resultados são utilizados para a definição de estimadores de ordem EDC para os processos AR, ARCH e BEKK-GARCH.

2.1 Definições e principais resultados

A Definição abaixo expõe os conceitos relacionados à existência e estacionariedade de cadeias de Markov. As Definições 2.2 e 2.3 versam sobre conceitos relacionados a recorrência e ergodicidade, respectivamente.

Definição 2.1. *Seja $\mathscr{Y}$ um espaço topológico e $\mathscr{B}(\mathscr{Y})$ sua σ-álgebra de Borel.*

(i) $\mathrm{P} : \mathscr{Y} \times \mathscr{B}(\mathscr{Y}) \to [0,1]$ *é núcleo de transição se*

(I) *Para todo* $A \in \mathscr{B}(\mathscr{Y})$, $\mathrm{P}(\cdot, A)$ *é não-negativa e mensurável em* $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ *;*

(II) *Para todo* $y \in \mathscr{Y}$, $\mathrm{P}(y, \cdot)$ *é medida de probabilidade em* $\mathscr{B}(\mathscr{Y})$.

(ii) *Um processo estocástico* $\mathbb{Y} = \{Y_t\}_{t \in \mathbb{N}}$ *é cadeia de Markov (homogênea no tempo) com núcleo de transição* $\mathrm{P}(y, A)$ *e distribuição inicial* π *se a distribuição de dimensão finita de* $\mathbb{Y}$ *satisfaz, para todo n e para todo* $A_i \in \mathscr{B}(\mathscr{Y})$, $i = 1, \cdots, n$,

$$\mathrm{P}_\pi(Y_1 \in A_1, \cdots, Y_n \in A_n) = \int\limits_{y_1 \in A_1} \cdots \int\limits_{y_{n-1} \in A_{n-1}} \pi(dy_1)\mathrm{P}(y_1, dy_2) \cdots \mathrm{P}(y_{n-1}, A_n).$$

(iii) $\mathrm{P}^n(y, A)$ *é definida indutivamente por*

$$\mathrm{P}^n(y, A) = \int\limits_{z \in \mathscr{Y}} \mathrm{P}(y, dz)\mathrm{P}^{n-1}(z, A) \quad y \in \mathscr{Y} \ \text{ e } \ A \in \mathscr{B}(\mathscr{Y})$$

e P_y *é a medida estendida à* σ*-álgebra produto* $\vee_{i=1}^{\infty} \mathscr{B}(\mathscr{Y})$ *usando o sistema de distribuições*

$$\mathrm{P}_y^1(A) = \mathrm{P}(x, A)$$

e

$$\mathrm{P}_y^n(A_1 \times \cdots \times A_n) = \int\limits_{A_1} \mathrm{P}(y, dy_1) \int\limits_{A_2} \mathrm{P}(y_1, dy_2) \cdots \mathrm{P}(y_{n-1}, A_n).$$

(iv) *Uma distribuição* σ*-finita* π *em* $\mathscr{B}(\mathscr{Y})$ *é invariante se para todo* $A \in \mathscr{B}(\mathscr{Y})$

$$\pi(A) = \int\limits_{y} \pi(dy)\mathrm{P}(y, A).$$

(v) *O tempo de ocupação de A ou número de visitas a A é definido por*

$$\eta_A = \sum_{t=1}^{\infty} \mathbb{I}(Y_t \in A).$$

(vi) *O tempo de primeira visita a* $A \in \mathscr{B}(\mathscr{Y})$ *é definido por*

$$\tau_A = \min\{t \in \mathbb{N} \,:\, Y_t \in A\}.$$

Definição 2.2. *Seja* $\mathbb{Y}$ *uma cadeia de Markov.*

(i) $\mathbb{Y}$ *é* φ*-irredutível se existe uma medida* φ *em* $\mathscr{B}(\mathscr{Y})$ *tal que, se* $A \in \mathscr{B}(\mathscr{Y})$ *e* $\varphi(A) > 0$, *então,* $\forall y \in \mathscr{Y}$*,*

$$L(y,A) > 0,$$

onde $L(y,A) := P_y(\tau_A < \infty)$*.*

(ii) $\mathbb{Y}$ *é* ψ*-irredutível se é* φ*-irredutível para alguma medida* φ *e* ψ *é maximal, isto é,* $\mathbb{Y}$ *é* φ'*-irredutível se e somente se* $\psi \succ \varphi'$*.*

(iii) Um conjunto $A \in \mathscr{B}(\mathscr{Y})$ *é Harris recorrente se, para todo* $y \in A$*,*

$$Q(y,A) := \mathrm{P}_y(\eta_A = \infty) = 1.$$

(iv) $\mathbb{Y}$ *é Harris recorrente se é* ψ*-irredutível e todo conjunto em* $\{A \in \mathscr{B}(\mathscr{Y}) : \psi(A) > 0\}$ *é Harris recorrente.*

(v) $\mathbb{Y}$*,* ψ*-irredutível, é positiva se admite distribuição invariante.*

(vi) $\mathbb{Y}$ *é Harris positiva se é Harris recorrente e positiva.*

Definição 2.3. *Para* $\mathbb{Y}$ *uma cadeia de Markov.*

(i) $\mathbb{Y}$ *é ergódica se,* $\forall y \in \mathscr{Y}$*,*

$$\lim_{n\to\infty} \sup_{A\in\mathscr{B}(\mathscr{Y})} |\mathrm{P}^n(y,A) - \pi(A)| = 0.$$

(ii) Para $V : E \to [1,\infty)$ *e* ν *uma medida em* $\mathscr{B}(\mathscr{Y})$*,*

$$\|\nu\|_V = \sup_{g:|g|\le V} \left\{ \int_{y\in\mathscr{Y}} g(y)\nu(dy) \right\}.$$

(iii) $\mathbb{Y}$ *é V-uniformemente ergódica se*

$$\lim_{n\to\infty} \sup_{y\in\mathbb{Y}} \frac{\|\mathrm{P}^n(y,\cdot) - \pi\|_V}{V(y)} = 0.$$

No que segue, são apresentadas as chamadas condições do tipo “*drift*”, que têm como objetivo concluir a ergodicidade geométrica ou ergodicidade V-uniforme baseado no comportamento de uma única transição do processo.

Definição 2.4. *Um conjunto $B \in \mathscr{B}(\mathscr{Y})$ é chamado "pequeno" se existe um $m > 0$ e uma medida não-trivial ν_m em $\mathscr{B}(\mathscr{Y})$, tais que, para todo $y \in B$ e $A \in \mathscr{B}(\mathscr{Y})$,*

$$\mathrm{P}^m(y,A) \geq \nu_m(A).$$

Condição 2.1. *Para $f : \mathscr{Y} \to [1,\infty)$, $B \in \mathscr{B}(\mathscr{Y})$ conjunto pequeno, $b < \infty$ e $V : \mathscr{Y} \to [0,\infty)$,*

$$E\left[V(Y_{t+1})|Y_t\right] - V(Y_t) \leq -f(Y_t) + b\mathbb{I}_B(Y_t). \tag{2.1}$$

Condição 2.2 (Foster-Lyapunov)**.** *Existe $V : \mathscr{Y} \to [1,\infty)$, $B \in \mathscr{B}(\mathscr{Y})$ conjunto pequeno, $\beta > 0$ e $b_1 < \infty$, tais que*

$$E\left[V(Y_{t+1})|Y_t\right] - V(Y_t) \leq -\beta V(Y_t) + b_1\mathbb{I}_B(Y_t). \tag{2.2}$$

Observa-se que a Condição 2.2 é a Condição 2.1 no caso particular $f = \beta V$. O Teorema 2.5 é uma adaptação dos Teoremas 17.0.1, 17.3.6 e 17.5.3 de Meyn & Tweedie e é peça fundamental no estabelecimento do comportamento assintótico das funções $\log L_{n,k}$.

Teorema 2.5 (Meyn & Tweedie (1993))**.** *Seja $\mathbb{Y} = \{Y_t\}_{t\in\mathbb{N}}$ uma cadeia de Markov a tempo discreto com espaço de estados $\mathscr{Y}$, Harris recorrente com distribuição invariante π, $g : \mathscr{Y} \to \mathbb{R}$ e $S_n(g) = \sum\limits_{t=1}^{n} g(Y_t)$.*

(i) Se $E_\pi(|g|) < \infty$, então

$$\lim_{n\to\infty} \frac{S_n(g)}{n} = E_\pi(g).$$

(ii) Se $\mathbb{Y}$ é V-uniformemente ergódica, $g^2 \leq V$ e $\int g d\pi = 0$, então $\phi_g^2 := E_\pi\left[g^2(X_1)\right] + 2\sum_{k=2}^{\infty} E_\pi\left[g(X_1)g(X_k)\right]$ está bem definido.

(iii) Se $\mathbb{Y}$ é ergódica, e existe uma função $f : \mathscr{Y} \to [1,\infty)$, um conjunto pequeno C, $b_0 < \infty$ satisfazendo (2.1), $\pi(V^2) < \infty$, $|g| \leq f$ e $\int g d\pi = 0$, então $\phi_g^2 := E\left[g^2(Y_1)\right] + 2\sum_{k=2}^{\infty} E\left[g(Y_1)g(Y_k)\right]$ está bem definido.

(iv) Nas hipóteses de (ii) ou nas hipóteses de (iii), se $\phi_g^2 > 0$, então

$$\lim_{n\to\infty} \frac{S_n(g)}{\sqrt{n\phi_g^2}} \xrightarrow{d} \mathscr{N}(0,1),$$

$$\limsup_{n\to\infty} \frac{S_n(g)}{\sqrt{2n\phi_g^2 \log\log(n)}} = 1 \quad q.c. \quad e$$

$$\liminf_{n\to\infty} \frac{S_n(g)}{\sqrt{2n\phi_g^2 \log\log(n)}} = -1 \quad q.c.$$

A Lei do Logaritmo Iterado para Martingales, transcrita abaixo, pode ser encontrada em Hall & Heyde (1980), Teoremas 4.7 e 4.8.

Teorema 2.6 (Hall & Heyde (1980))**.** *Seja* $\{S_n, \mathscr{F}_{t-1}\}$ *uma martingale,* $S_n = \sum_{t=1}^n U_t$, $E(S_n) = 0$, $E(S_n^2) < \infty$, $\{Z_t\}_{t\in\mathbb{N}}$ *e* $\{W_t\}_{t\in\mathbb{N}}$ *sequências de variáveis aleatórias não negativas tais que* Z_t *e* W_t *são* $\mathscr{F}_{t-1}$ *mensuráveis. Se*

$$\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} U_t\mathbb{I}(|U_t| > Z_t) - E[U_t\mathbb{I}(|U_t| > Z_t)|\mathscr{F}_{t-1}]}{\sqrt{2W_n^2 \log\log W_n^2}} = 0 \quad q.c., \tag{L1}$$

$$\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} E[U_t^2\mathbb{I}(|U_t| \le Z_t)|\mathscr{F}_{t-1}] - E[U_t\mathbb{I}(|U_t| \le Z_t)|\mathscr{F}_{t-1}]^2}{W_n^2} = 1 \quad q.c., \tag{L2}$$

$$\lim_{n\to\infty} \sum_{t=1}^{n} \frac{E[U_t^4\mathbb{I}(|U_t| \le Z_t)|\mathscr{F}_{t-1}]}{W_t^4} < \infty \quad q.c., \tag{L3}$$

$$\lim_{n\to\infty} \frac{W_n}{W_{n+1}} = 1 \quad q.c. \ e \quad \lim_{n\to\infty} W_n = \infty \quad q.c. \tag{L4}$$

Então

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2W_n^2 \log\log W_n^2}} = 1 \quad q.c.$$

e

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2W_n^2 \log\log W_n^2}} = -1 \quad q.c.$$

2.2 Modelo Autoregressivo (AR)

O processo Autoregressivo (AR) foi proposto inicialmente por Yule (1921) para a modelagem de séries temporais e vem sendo utilizado em diversas áreas, dentre outras citamos econometria (Maddala & Lahiri 2009), engenharia (Schneider 1988) e genética (Carvalheira, Blake, Pollak, Quaas & Duran-Castro 1998).

O problema de estimação de ordem em modelos AR inicialmente foi abordado usando testes de hipóteses por Quenouille (1947), Whittle (1951, 1954) e Bartlett & Rajalakshman (1953). Akaike (1969) propôs o método de minimização do erro final de predição (FPE – *Final Prediction Error*) para a estimação da ordem em modelos AR, que foi o precursor do método AIC, proposto pelo mesmo autor em 1974. Após isso, foram desenvolvidos outros métodos de estimação, que se destacam o BIC (Akaike 1979) e HQC (Hannan & Quinn 1979). Vale ressaltar que o problema de estimação de ordem em modelos AR motivaram a criação do AIC e influenciou o desenvolvimento de estimadores de ordem em outros processos, como exemplo em cadeias de Markov. Para uma visão mais detalhada sobre o desenvolvimento das técnicas de estimação de ordem em processo ARMA (que tem o AR como caso particular), veja Choi (1992).

No que segue, o ferramental desenvolvido nesse trabalho é aplicado para a estimação da ordem em modelos AR e um novo estimador de ordem é proposto. Também são apresentadas simulações numéricas que indicam, no geral, uma melhor performance do estimador proposto quando comparado com as alternativas AIC, BIC e HQC.

2.2.1 Definições

Abaixo o processo AR é definido para inovações Gaussianas e em seguida é apresentada a Condição 2.3, que é requisito clássico para a obtenção de resultados assintóticos.

Definição 2.7. *Uma sequência de variáveis aleatórias $\mathbb{X} = \{X_t\}_{t\in\mathbb{Z}}$ com valores em $\mathbb{R}$ é um processo Autoregressivo de ordem $r \in \mathbb{N}$, denotado por AR(r), se satisfaz as condições abaixo.*

(i) Para todo $t \in \mathbb{Z}$,

$$X_t = h_t + \varepsilon_t.$$

(ii) $h_t = \alpha_1 X_{t-1} + \cdots + \alpha_r X_{t-r}$, $\theta_r = (\alpha_1, \ldots, \alpha_r) \in \mathbb{R}^r$ e $\alpha_r \neq 0$.

(iii) $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ é uma sequência de variáveis aleatórias i.i.d. $\mathcal{N}(0, \sigma^2)$.

Condição 2.3 (Ergodicidade)**.** *Para $\mathbb{X}$ um AR(r), e $\theta_r = (\alpha_1, \ldots, \alpha_r) \in \mathbb{R}^r$ seus respectivos parâmetros, assumimos que*

$$1 - \sum_{i=1}^{r} \alpha_i z^i \neq 0 \quad \forall z \in \mathbb{C},\ |z| \leq 1. \tag{2.3}$$

Principais propriedades do processo autoregressivo

Algumas propriedade básicas para processos AR, que são utilizadas no desenvolvimento dessa Seção, são apresentadas abaixo.

(a) $E(X_i|\mathscr{F}_{i-1}) = h_i \quad$ q.c.

(b) Um AR(r), $\mathbb{X}$, pode ser imerso em uma sequência de modelos aninhados $\mathbb{M} = \bigcup_{k=0}^{\infty} M_k$, tomando $\Theta_k = \mathbb{R}^k$ e, para $\theta_k = (\alpha_1, \cdots, \alpha_k) \in \Theta_k$,

$$f(x_1^n, \theta_k) = C(x_1^k) \prod_{t=1+k}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_t - (\alpha_1 x_{t-1} + \cdots + \alpha_k x_{t-k}))^2}{2\sigma^2}} \tag{2.4}$$

(c) Usando que $\mathbb{X}$ é estacionário e (2.3), temos que

$$0 < |m(k)| < \infty. \tag{2.5}$$

onde $m(k) = E(X_t X_{t+k})$ e $t, k \in \mathbb{Z}$ (Anderson (1994), página 173).

Funções $\log L_{n,k}$ e suas derivadas

A log-verossimilhança de $m_k = f(x_1^n, \theta_k) \in M_k$ é dada por

$$\begin{aligned} \log\left(\prod_{t=1+k}^{n} f(X_t|X_{t-k-1}^{t-1})\right) + C_1(X_1^k) &= \sum_{t=1+k}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_t - h_t)^2}{2\sigma^2}}\right) + C_1(X_1^k) \\ &= \sum_{t=1+k}^{n} -\frac{(X_t - h_t)^2}{2\sigma^2} - (n-k)\log\sqrt{2\pi\sigma^2} + C_1(X_1^k). \end{aligned}$$

Definimos abaixo o logaritmo de $L_{n,k}$, que satisfaz a Definição 1.3.

$$\log L_{n,k}(\theta) = \sum_{t=1+k}^{n} -\frac{(X_t - h_t)^2}{2}. \tag{2.6}$$

Dessa forma, para $i, j, l \in \{1, \cdots, k\}$,

$$\frac{\partial \log L_{n,k}(\theta)}{\partial \alpha_i} = \sum_{t=1+k}^{n} (X_t - h_t) X_{t-i}, \tag{2.7}$$

$$\frac{\partial^2 \log L_{n,k}(\theta)}{\partial \alpha_i \alpha_j} = \sum_{t=1+k}^{n} -X_{t-i} X_{t-j} \;\; e \tag{2.8}$$

$$\frac{\partial^3 \log L_{n,k}(\theta)}{\partial \alpha_i \alpha_j \alpha_l} = 0. \tag{2.9}$$

Alguns resultados disponíveis na literatura são estabelecidos para modelos Lineares em Espaço de Estados (LSS), que possui como caso particular os modelos AR. Nesse sentido, definimos abaixo modelos LSS e apresentamos condições para a existência de ergodicidade V-uniforme. Esses resultados podem ser encontrados de forma mais geral em Meyn & Tweedie (1993).

Definição 2.8. *Uma sequência de vetores aleatórios $\mathbb{Y} = \{Y_t\}_{t \in \mathbb{Z}}$ com valores em $\mathbb{R}^r$ é um modelo Linear em Espaço de Estados (LSS) se*

(i) Existe matriz $(r \times r)$ F e matriz $(r \times p)$ G tal que para todo $t \in \mathbb{N}$, Y_t e W_t com respectivos valores em $\mathbb{R}^r$ e $\mathbb{R}^p$ satisfazem, para todo $t \in \mathbb{N}$,

$$Y_t = FY_{t-1} + GW_t \tag{2.10}$$

e Y_0 é arbitrário.

(ii) Os vetores aleatórios $\{W_t\}_{t \in \mathbb{N}}$ são i.i.d., são independentes de Y_0 e possuem distribuição com média e variância finitas.

Condição 2.4. *Se $\mathbb{Y}$ é um modelo LSS, então*

(i) W possui distribuição Gaussiana em $\mathbb{R}^p$ com média zero e variância 1. Isto é, $W \sim \mathcal{N}(0, I)$, I a matriz identidade $(p \times p)$.

(ii) Se $\lambda \in \mathbb{C}$ é autovalor de F, então $|\lambda| < 1$.

Se $\mathbb{X}$ é um AR(r), podemos representá-lo como um modelo Linear de Espaço de Estados

$\mathbb{Y}$ tomando $p = 1$, $Y_t = (X_t, \ldots, X_{t-r+1})$ com valores em $\mathbb{R}^r$, $W_t = \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0,1)$,

$$F = \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_r \\ 1 & 0 & \ldots & 0 \\ 0 & \ddots & & \vdots \\ 0 & \ldots & 1 & 0 \end{bmatrix} \quad e \quad G = \begin{bmatrix} \sigma \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.11}$$

Usando isso, podemos usar o ferramental desenvolvido para cadeias de Markov, no que se refere a estabilidade assintótica. Em outras palavras, temos o seguinte resultado.

Lema 2.9 (Meyn & Tweedie (1993) e Anderson (1994))**.** *Se um AR(r)* $\mathbb{X}$ *satisfaz* (2.3)*, então* $\mathbb{Y}$*, o modelo Linear de Espaço de Estados correspondente conforme definido em* (2.11)*, satisfaz a Condição 2.4, é Harris positivo e V-uniformemente ergódico para* $V(x) = |x|^2 + 1$*. Além disso,* $\mathbb{Y}$ *satisfaz a Condição 2.2 para* $\beta \in (0, \infty)$*,* $b_1 \in (-\infty, \infty)$ *e B conjunto pequeno e compacto.*

Demonstração. Temos que $\mathbb{Y}$ satisfaz a Condição 2.4 (Anderson (1994) página 180). Usando a Proposição 4.4.3 de Meyn & Tweedie (1993) temos que as condições dos Teoremas 12.5.1 e 17.6.2 de Meyn & Tweedie (1993) são satisfeitas e portanto $\mathbb{Y}$ é Harris positiva e é V-uniformemente ergódico para $V(x) = |x|^2 + 1$. A demonstração do Teorema 17.6.2 de Meyn&Tweedie mostra que $\mathbb{Y}$ satisfaz a Condição 2.2. □

2.2.2 Consistência do estimador de ordem de dependência

Abaixo é definido estimador de ordem EDC para o caso particular de processos AR. Em seguida estão uma série de resultados com o objetivo de demonstrar as hipóteses H1-H4 do Teorema 1.5 e concluir os casos de convergência do estimador proposto, que estão sumarizados no Teorema 2.18. Como mencionado anteriormente, a técnica de utilizar o comportamento assintótico das primeiras derivadas de $\log L_{n,k}$ para concluir o comportamento assintótico de $\log L_{n,k}$ foi utilizada por Nishii (1988) em caso particular de estimação de dimensão de modelos i.i.d. e considerando $\log L_{n,k}$ como a log-verossimilhança. Ressalta-se também que, de forma semelhante, Basawa & Heyde (1976) utiliza o comportamento assintótico das primeiras derivadas da log-verossimilhança para concluir o comportamento assintótico do estimador de máxima verossimilhança. A essência dessa técnica é a utilização da expansão em séries de Taylor em determinados pontos, para isso são exigidas as condições de regularidade das derivadas de $\log L_{n,k}$.

Definição 2.10. *Para* $\mathbb{X}$ *um AR(r) e* $K \geq r$*, definimos o estimador* EDC *de* r *por*

$$\hat{r}_{edc} = \underset{k \in \{0,\ldots,K\}}{\operatorname{argmin}} \left\{ -\log L_n(\hat{\theta}_k) + kc_n \right\}$$

para $\log L_n(\hat{\theta}_k)$ *como definida em* (2.6) *e* $\{c_n\}_{n\in\mathbb{N}}$ *uma sequência de números positivos.*

Lema 2.11. *Seja* $\mathbb{X} = \{X_t\}_{t\in\mathbb{Z}}$ *um* $AR(r)$, $k \geq r$, $\log L_n(\hat{\theta}_k)$ *como definida* (2.6), $\hat{\theta}_k \in \Theta_k \subseteq \mathbb{R}^k$ *o estimador de máxima verossimilhança, então*

(i) *Para* $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$, $s \in [0,1]$,

$$\lim_{n\to\infty} -\frac{D^2_\theta \log L_n(\dot{\theta})}{n} = A_2 \quad q.c.,$$

onde

$$A_2 = \begin{bmatrix} m(0) & \ldots & m(k) \\ \vdots & \ddots & \vdots \\ m(k) & \ldots & m(0) \end{bmatrix}$$

e $m(k) = E(X_t X_{t+k})$.

(ii) A_2 *é positiva definida.*

Demonstração. (i) Usando a desigualdade de Hölder

$$E\left(|X_{t-i}X_{t-j}|\right) \leq E(X^2_{t-i})^{\frac{1}{2}} E(X^2_{t-j})^{\frac{1}{2}} < \infty.$$

Pelo Teorema 2.5 e expressão (2.8),

$$\lim_{n\to\infty} \frac{\frac{\partial^2 \log L_n(\dot{\theta})}{\partial \alpha_i \alpha_j}}{n} = -E\left(X_{t-i}X_{t-j}\right) = -m(i-j).$$

(ii) Demonstração pode ser encontrada em Anderson (1994), Lema 5.5.4, página 196.

□

Lema 2.12. *Seja* $\mathbb{X} = \{X_t\}_{t\in\mathbb{Z}}$ *um* $AR(r)$, $k \geq r$, $\log L_{n,k}(\theta)$ *como definida* (2.6), $i \in \{1, \cdots, k\}$, *então*

$$\limsup_{n\to\infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n\log\log n}} = \sqrt{\sigma^2 E(X_1^2)} \quad q.c.,$$

$$\liminf_{n\to\infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n\log\log n}} = -\sqrt{\sigma^2 E(X_1^2)} \quad q.c. \quad e \tag{2.12}$$

$$\frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{n}} \underset{D}{\rightarrow} \mathcal{N}(0, \sigma^2 E(X_1^2)) \quad q.c. \tag{2.13}$$

Demonstração. Assumindo $\mathscr{F}_t$ a σ-álgebra gerada por $\{X_l\}_{l\leq t}$ e considerando o Teorema 2.5, $\{Y_t\}_{t\in\mathbb{Z}}$ o modelo LSS associado, $g(Y_t) = (X_t - h_t)X_{t-i}$, $E(g) = E(X_{t-i}E(\varepsilon_t|\mathscr{F}_t)) = 0$ e

$$\phi_g^2 = E\left[((X_1 - h_1)X_{1-i})^2\right] + 2\sum_{t=k+2}^{\infty} E\left[((X_1 - h_1)X_{1-i})((X_t - h_t)X_{t-i})\right].$$

Para a primeira parcela

$$\begin{aligned} E\left[((X_1 - h_1)X_{1-i})^2\right] &= E\left[X_{1-i}^2 E(\varepsilon_1^2|\mathscr{F}_0)\right] \\ &= \sigma^2 E(X_{1-i}^2). \end{aligned}$$

Usando (2.5), concluímos

$$0 < E\left[((X_1 - h_1)X_{1-i})^2\right] < \infty.$$

Para a segunda parcela

$$\begin{aligned} E\left[((X_1 - h_1)X_{1-i})((X_t - h_t)X_{t-i})\right] &= E\left[(X_1 - h_1)X_{1-i}X_{t-i}E\left((X_t - h_t)|\mathscr{F}_{t-1}\right)\right] \\ &= 0. \end{aligned}$$

Então

$$\phi_g^2 = \sigma^2 E(X_{1-i}^2).$$

Também temos que

$$\begin{aligned}|g(Y_t)| &= |(X_t - h_t)X_{t-i}| \\ &\leq (r+1)\max\{1, |\alpha_1|, \cdots, |\alpha_r|\} \max_{l=t-r,\ldots,t} |X_l|^2 \\ &\leq d_1|Y_t|^2 \\ &\leq d_1 V(Y_t). \end{aligned} \tag{2.14}$$

Na penúltima desigualdade usamos que todas as normas são equivalentes em espaços de dimensão finita. De (2.14) e (2.2) concluímos (2.1) para $f = max\{1, |g|\}$, $b_0 = (b_1 + \beta/d_1)$ e $C = B \cup \{x \in R^r\ ;\ |V(x)| \leq 1/d_1\}$. Também $\pi(V^2(Y_t)) \leq d_2\pi(\max_{i=t-r..r} |Y_t|^4) < \infty$ (Anderson 1994).

Observando que se (2.1) vale para $\frac{\beta}{d_1}g$ então vale para g, podemos usar o Teorema 2.5 e concluir

$$\limsup_{n\to\infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n\sigma^2 E(X_1^2)\log\log n}} = 1 \quad q.c.$$

Então,

$$\limsup_{n\to\infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n\log\log n}} = \sqrt{\sigma^2 E(X_1^2)} \quad q.c.$$

Da mesma forma concluímos (2.12) e (2.13). □

Corolário 2.13. *Seja $\{X_t\}_{t\in\mathbb{Z}}$ um AR(r) satisfazendo (2.3), $k \geq r$, $\hat{\theta}_k = (\hat{\alpha}_1, \cdots, \hat{\alpha}_k)$ o estimador de máxima verossimilhança de $\theta_r = (\alpha_1, \cdots, \alpha_r, 0, \cdots, 0) \in \mathbb{R}^k$ e h_n, tal que $\lim_{n\to\infty} h_n = \infty$, então*

$$\lim_{n\to\infty} \mathrm{P}\left(\left\|\sqrt{n}(\theta_r - \hat{\theta}_k)\right\| > h_n\right) = 0.$$

Demonstração. Usando o Lema 2.12 e que para n grande A_n é positiva definida temos

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{P}(\left\|\sqrt{n}(\theta_r - \hat{\theta}_k)\right\| > h_n) &= \lim_{n\to\infty} \mathrm{P}\left(\left\|\frac{D^1_\theta \log L_{n,k}(\theta_r) A_n^{-1}}{\sqrt{n}}\right\| > h_n\right) \\
&\leq \lim_{n\to\infty} \mathrm{P}\left(\frac{1}{\lambda_k}\left\|\frac{D^1_\theta \log L_{n,k}(\theta_r)}{\sqrt{n}}\right\| > h_n\right) \\
&= \lim_{n\to\infty} \mathrm{P}\left(\sum_{i=1}^{k}\left|\frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{n}}\right|^2 > \lambda_k^2 h_n^2\right) \\
&= 0.
\end{aligned}$$

Onde $1/\lambda_k$ é o maior autovalor de A_n^{-1}. □

Corolário 2.14. *Para um AR(r), se $k > r$, então*

$$\limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r)}{\log\log n} \leq \frac{2k\sigma^2 E(X^2)}{\lambda_k} \ q.c.$$

onde λ_k é o menor autovalor de A_2.

Demonstração. A partir do Lema 2.12 temos que

$$\limsup_{n\to\infty} \left\|\frac{D^1_\theta \log L_{n,k}(\theta_r)}{\sqrt{2n\log\log n}}\right\|^2 \leq k\sigma^2 E(X_1^2).$$

Aplicando o Teorema 1.6, item (ii), e usando os Lemas 2.11 e 2.12 temos o desejado. □

Usando outra técnica, podemos obter o Teorema abaixo, que é útil se for possível determinar o limitante superior em (1.16), ao aplicar o Lema 2.12, de forma a não depender de k ou r. Se for possível, podemos encontrar o $c_n = O(\log\log n)$ preciso que gera um estimador $\hat{r}_{edc}$ fortemente consistente.

Teorema 2.15. *Se $\mathbb{X}$ é um AR(r) satisfazendo a Condição 2.3 e $k > r$, então*

$$\limsup_{n\to\infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r)}{\log\log n} \leq (k-r)c_2 \quad q.c.$$

Demonstração. Tomando a série de Taylor para $\log L_{n,k}(\hat{\theta}_r)$ em $\hat{\theta}_k$ e usando (2.9), temos

$$\log L_{n,k}(\hat{\theta}_r) = \log L_{n,k}(\hat{\theta}_k) + (\hat{\theta}_r - \hat{\theta}_k) D^1_\theta \left(\log L_{n,k}(\hat{\theta}_k)\right) + \frac{1}{2}(\hat{\theta}_r - \hat{\theta}_k) D^2_\theta \left(\log L_{n,k}(\hat{\theta}_k)\right)(\hat{\theta}_r - \hat{\theta}_k)^T$$

Como $\log L_{n,k}$ é maximizada por $\hat{\theta}_k$, temos que $D^1_\theta\left(\log L_{n,k}(\hat{\theta}_k)\right) = 0$ e

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\hat{\theta}_r)}{\log\log n} &= \limsup_{n\to\infty} \frac{-(\hat{\theta}_r - \hat{\theta}_k) D^2_\theta\left(\log L_{n,k}(\hat{\theta}_k)\right)(\hat{\theta}_r - \hat{\theta}_k)^T}{2\log\log n} \\
&= \limsup_{n\to\infty} \frac{\sqrt{n}(\hat{\theta}_r - \hat{\theta}_k)}{\sqrt{2\log\log n}} \frac{-D^2_\theta\left(\log L_{n,k}(\hat{\theta}_k)\right)}{n} \frac{\left(\sqrt{n}(\hat{\theta}_r - \hat{\theta}_k)\right)^T}{\sqrt{2\log\log n}}
\end{aligned} \tag{2.15}$$

Usando Shibata (1976), temos que (2.15) pode ser definido como a norma

$$||x||_1^2 := x \frac{-D^2_\theta\left(\log L_{n,k}(\hat{\theta}_k)\right)}{n} x^T$$

que satisfaz

$$||\hat{\theta}_k - \hat{\theta}_r||_1^2 = \hat{\sigma}_r^2 - \hat{\sigma}_k^2 \tag{2.16}$$

onde $\hat{\sigma}_l^2$ é o estimador de σ^2 se $l \geq r$. Por Hannan & Quinn (1979),

$$\hat{\sigma}_k^2 = \hat{\sigma}_r^2 \prod_{i=r+1}^{k} (1 - \hat{\alpha}_i^2(i)) \tag{2.17}$$

onde $\hat{\alpha}_i(l)$ é o estimador de máxima verossimilhança de α_i, considerando l a ordem verdadeira. Usando (2.16) e (2.17)

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\log L_{n,k}(\hat{\theta}_k) - \log L_{n,k}(\hat{\theta}_r)}{\log\log n} &= \limsup_{n\to\infty} \frac{n||\hat{\theta}_r - \hat{\theta}_k||_1^2}{2\log\log n} \\
&= \limsup_{n\to\infty} \frac{n[\hat{\sigma}_r^2 - \hat{\sigma}_r^2 \prod_{i=r+1}^{k} (1 - \hat{\alpha}_i^2(i))]}{2\log\log n} \\
&= \limsup_{n\to\infty} \frac{n\hat{\sigma}_r^2[1 - 1 + \sum_{i=r+1}^{k} \hat{\alpha}_i^2(i) + o(\log\log n/n)]}{2\log\log n} \\
&\leq \sigma^2 \sum_{i=r+1}^{k} \limsup_{n\to\infty} \left(\frac{\sqrt{n}(\hat{\alpha}_i(i) - 0)}{\sqrt{2\log\log n}}\right)^2 \\
&\leq c_2(k-r).
\end{aligned} \tag{2.18}$$

O resultado segue de (1.4). Usamos também que $\hat{\sigma}_r^2 \underset{q.c.}{\to} \sigma^2$ e a inequação (1.16), quando se aplica o Lema 2.12. O argumento utilizado tem um valor teórico interessante. Entretanto, basta verificar diretamente (2.6) e observar que

$$\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r) = n\left(\hat{\sigma}_r^2 - \hat{\sigma}_k^2\right)$$

e continuar as contas a partir de (2.18). $\square$

Teorema 2.16. *Se $\mathbb{X}$ é um AR(r) satisfazendo a Condição 2.3, $k > r$ e $h_3 \in (0, \infty)$, então*

$$\lim_{n\to\infty} \mathrm{P}\left(\log L_n(\hat{\theta}_k) - \log L_n(\hat{\theta}_r) > h_3\right) > 0.$$

Demonstração. Usando (2.6) e Hannan & Quinn (1979), temos

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{P}\left(\log L_{n,k}(\hat{\theta}_k) - \log L_{n,r}(\hat{\theta}_r) > h_3\right) &= \lim_{n\to\infty} \mathrm{P}\left(n(\hat{\sigma}_r^2 - \hat{\sigma}_k^2) > h_3\right) \\
&= \lim_{n\to\infty} \mathrm{P}\left(n[\hat{\sigma}_r^2 - \hat{\sigma}_r^2 \prod_{i=r+1}^{k} (1 - \hat{\alpha}_i^2(i))] > h_3\right) \\
&= \lim_{n\to\infty} \mathrm{P}\left(n\hat{\sigma}_r^2[1 - 1 + \sum_{i=r+1}^{k} \hat{\alpha}_i^2(i) + o(1/n)] > h_3\right) \\
&= \lim_{n\to\infty} \mathrm{P}\left(\sum_{i=r+1}^{k} (\sqrt{n}(\hat{\alpha}_i(i) - 0))^2 > \frac{h_3}{\hat{\sigma}_r^2}\right) \\
&\geq \lim_{n\to\infty} \mathrm{P}\left(\sqrt{n}(\hat{\alpha}_i(i) - 0) > \sqrt{\frac{h_3}{\hat{\sigma}_r^2}}\right) \\
&> 0.
\end{aligned}$$

Usamos a normalidade assintótica de $\sqrt{n}(\hat{\alpha}_i(i) - 0)$, que pode ser encontrada no Teorema 5.5.7 de Anderson (1994). $\square$

Proposição 2.17. *Seja $\mathbb{X}$ um AR(r) satisfazendo a Condição 2.3, $k < r$, então*

$$\lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_r) - \log L_n(\hat{\theta}_k)}{n} > 0 \quad q.c.$$

Demonstração. Tomando a série de Taylor para $\log L_{n,r}(\hat{\theta}_k)$ em $\hat{\theta}_r$ e usando (2.9), temos

$$\log L_{n,r}(\hat{\theta}_k) = \log L_{n,r}(\hat{\theta}_r) + (\hat{\theta}_k - \hat{\theta}_r) D_\theta^1\left(\log L_{n,r}(\hat{\theta}_r)\right) + \frac{1}{2}(\hat{\theta}_k - \hat{\theta}_r) D_\theta^2\left(\log L_{n,r}(\hat{\theta}_r)\right)(\hat{\theta}_k - \hat{\theta}_r)^T$$

Como $L_{n,r}$ é maximizada por $\hat{\theta}_r$, temos que $D^1_\theta\left(\log L_{n,r}(\hat{\theta}_r)\right) = 0$ e

$$\begin{aligned}
\lim_{n\to\infty} \frac{\log L_{n,r}(\hat{\theta}_r) - \log L_{n,r}(\hat{\theta}_k)}{n} &= \lim_{n\to\infty} \frac{-(\hat{\theta}_k - \hat{\theta}_r) D^2_\theta\left(\log L_{n,r}(\hat{\theta}_r)\right)(\hat{\theta}_k - \hat{\theta}_r)^T}{2n} \\
&= \frac{1}{2}\lim_{n\to\infty}(\hat{\theta}_k - \hat{\theta}_r)\frac{-D^2_\theta\left(\log L_{n,r}(\hat{\theta}_r)\right)}{n}\left(\hat{\theta}_k - \hat{\theta}_r\right)^T \\
&= \frac{1}{2}\lim_{n\to\infty}(\hat{\theta}_k - \hat{\theta}_r) A_2 \left(\hat{\theta}_k - \hat{\theta}_r\right)^T \\
&= c\lim_{n\to\infty}\left\|\hat{\theta}_k - \hat{\theta}_r\right\| \\
&> 0.
\end{aligned}$$

O resultado segue de (1.4), que A_2 é positiva definida (Lema 2.11) e que $\lim_{n\to\infty}\left\|\hat{\theta}_k - \hat{\theta}_r\right\| > 0$ para $k < r$. □

Embora a demonstração da Proposição anterior seja elegante, é utilizado diretamente o fato das derivadas de ordem 3 de $\log L_{n,k}$ serem nulas. Com isso, a técnica utilizada nessa demonstração não é aplicável no caso geral.

Teorema 2.18. *Seja $\mathbb{X}$ um $AR(r)$ satisfazendo a Condição 2.3 e $\hat{r}_{edc}$ como na Definição 2.10.*

(i) $\hat{r}_{edc}$ é fortemente consistente ($\hat{r}_{edc} \underset{q.c.}{\rightarrow} r$) se

$$\lim_{n\to\infty}\frac{c_n}{n} = 0 \quad e \quad \liminf_{n\to\infty}\frac{c_n}{\log\log n} = \infty. \tag{2.19}$$

(ii) $\hat{r}_{edc}$ é consistente ($\hat{r}_{edc} \underset{P}{\rightarrow} r$) se

$$\lim_{n\to\infty}\frac{c_n}{n} = 0 \quad e \quad \liminf_{n\to\infty} c_n = \infty.$$

(iii) $\hat{r}_{edc}$ é inconsistente se

$$\limsup_{n\to\infty} c_n < \infty. \tag{2.20}$$

Demonstração. O conjunto dos possíveis valores de θ_r, Θ_r, definido em (2.3), é aberto e $\theta_r \in \Theta_r$. De Anderson (1994), temos que o estimador de máxima verossimilhança $\hat{\theta}_k$ é fortemente consistente para $k \geq r$, isto é, vale (1.10). As derivadas de $\log L_{n,k}$, (2.7) e (2.8) são contínuas em relação a θ e a x_1^n e portanto são mensuráveis em relação a x_1^n. Temos de (2.9) que (1.11) é satisfeita.

(i) Usando o Lema 2.11 temos que as condições satisfeitas para a aplicação do item (ii) do Teorema 1.6 e com isso temos (H2) satisfeita. A hipótese (H1) segue da Proposição 2.17. O resultado segue do item (i) do Teorema 1.5.

(ii) Usando o Lema 2.11 e o Corolário 2.13, concluímos (H3) usando os Teoremas 1.6 e 1.7. Portanto temos o resultado a partir do Teorema 1.5, item (ii).

(iii) De forma semelhante, usando o Teorema 2.16, temos (H4) e concluímos o resultado pelo item (iii) do Teorema 1.5.

□

Corolário 2.19. *Seja* $\mathbb{X}$ *um* $AR(r)$ *satisfazendo a Condição 2.3,* $K \geq r$*. Então o estimador* $\hat{r}_{bic2}$*, como definido abaixo, é fortemente consistente.*

$$\hat{r}_{bic2} = \underset{k \in \{0,\dots,K\}}{\operatorname{argmin}} \left\{ -\log L_n(\hat{\theta}_k) + \frac{k}{2} \log n \right\} \tag{2.21}$$

para $\log L_n(\hat{\theta}_k)$ *como definida em* (2.6).

Demonstração. $c_n = \log n/2$ e satisfaz (2.19). □

Corolário 2.20. *Seja* $\mathbb{X}$ *um* $AR(r)$ *satisfazendo a Condição 2.3,* $K \geq r$*. Então o estimador* $\hat{r}_{aic2}$*, como definido abaixo, é inconsistente.*

$$\hat{r}_{aic2} = \underset{k \in \{0,\dots,K\}}{\operatorname{argmin}} \left\{ -\log L_n(\hat{\theta}_k) + 2k \right\}$$

para $\log L_n(\hat{\theta}_k)$ *como definida em* (2.6).

Demonstração. $c_n = 2$ e satisfaz (2.20). □

Embora o estimador $\hat{r}_{bic2}$ definido em (2.21) seja fortemente consistente, observou-se nas simulações numéricas que ele não apresenta performance satisfatória para casos em que $\sigma^2 \neq 1$. Isso ocorre porque $-2\log L_n(\hat{\theta}_k)/n$ coincide com o estimador de σ^2 supondo que a ordem seja k, e por isso as diferenças $\log L_n(\hat{\theta}_{k+1}) - \log L_n(\hat{\theta}_k)$ dependem, de certa forma, dessa grandeza. Os estimadores AIC, BIC e HQC resolvem esse problema aproximando $-\log L_n(\hat{\theta}_k)/n$ por $\log\left[-\log L_n(\hat{\theta}_k)/n\right]$. Entretanto, essa aproximação reduz a oscilação do estimador. Para solucionar isso, propomos o seguinte estimador de ordem.

Corolário 2.21. *Seja $\mathbb{X}$ um $AR(r)$ satisfazendo a Condição 2.3, $K \geq r$. Então o estimador $\hat{r}_{edc}$, como definido abaixo, é fortemente consistente.*

$$\hat{r}_{edc} = \underset{k \in \{0,\ldots,K\}}{\operatorname{argmin}} \left\{ -\log L_n(\hat{\theta}_k) + \frac{k}{2} \hat{\sigma}_K^2 \log n \right\}$$

para $\log L_n(\hat{\theta}_k)$ *como definida em* (2.6) *e* $\hat{\sigma}_K^2$ *o estimador de* σ^2.

Demonstração. $c_n = \hat{\sigma}_K^2 \log n / 2$ e satisfaz (2.19). □

Vale ressaltar que o fator $\hat{\sigma}_K^2$ é utilizado para se obter uma aproximação da grandeza de σ^2 o que não requer precisão, e portanto não limita de forma substancial o valor de K.

2.2.3 Simulações numéricas

Os estimadores considerados nas simulações são, para uma amostra x_1^n,

$$\hat{r}_{aic} = \underset{k=0,\ldots,K}{\operatorname{argmin}} \{AIC(k)\},$$

$$\hat{r}_{bic} = \underset{k=0,\ldots,K}{\operatorname{argmin}} \{BIC(k)\},$$

$$\hat{r}_{hqc} = \underset{k=0,\ldots,K}{\operatorname{argmin}} \{HQC(k)\} \quad e$$

$$\hat{r}_{edc} = \underset{k=0,\ldots,K}{\operatorname{argmin}} \{EDC(k)\}.$$

Onde,

$$AIC(k) = n \log \left(\sum_{t=k+1}^{n} \frac{[x_t - (x_{t-1}\hat{\alpha}_1 + \cdots + x_{t-k}\hat{\alpha}_k)]^2}{n} \right) + 2k,$$

$$BIC(k) = n \log \left(\sum_{t=k+1}^{n} \frac{[x_t - (x_{t-1}\hat{\alpha}_1 + \cdots + x_{t-k}\hat{\alpha}_k)]^2}{n} \right) + k \log n,$$

$$HQC(k) = n \log \left(\sum_{t=k+1}^{n} \frac{[x_t - (x_{t-1}\hat{\alpha}_1 + \cdots + x_{t-k}\hat{\alpha}_k)]^2}{n} \right) + 2k \log \log n,$$

$$EDC(k) = \sum_{t=k+1}^{n} [x_t - (x_{t-1}\hat{\alpha}_1 + \cdots + x_{t-k}\hat{\alpha}_k)]^2 + k \hat{\sigma}_K^2 \log n \quad e$$

$$\hat{\sigma}_K^2 = \frac{\sum_{t=K+1}^{n} [x_t - (x_{t-1}\hat{\alpha}_1 + \cdots + x_{t-K}\hat{\alpha}_K)]^2}{n}.$$

Os $\hat{\alpha}_i$ são os estimadores de máxima verossimilhança para os parâmetros α_i. Não há fórmula explícita para a estimação dos $\hat{\alpha}_i$. Eles são a solução do seguinte sistema linear.

$$\begin{bmatrix} \hat{R}(1,1) & \dots & \hat{R}(1,k) \\ \vdots & & \vdots \\ \hat{R}(k,1) & \dots & \hat{R}(k,k) \end{bmatrix} \begin{bmatrix} \hat{\alpha}(1) \\ \vdots \\ \hat{\alpha}(k) \end{bmatrix} = \begin{bmatrix} \hat{R}(0,1) \\ \vdots \\ \hat{R}(0,k) \end{bmatrix}$$

$$\hat{R}(i,j) = \frac{1}{n} \sum_{t=k+1}^{n} X_{t-i} X_{t-j}$$

Ressalta-se que foram utilizadas as definições propostas por Shibata (1976) e Hannan & Quinn (1979) para os estimadores baseados nos critérios AIC, BIC e HQC.

Na Tabela 2.1 estão os resultados de simulações numéricas para modelos AR com ordem $r \in \{1,3,10,50\}$ e desvio padrão $\sigma = 0,1$. Para cada caso foram realizados 100 simulações. Os valores iniciais, de 1 até K, foram gerados aleatoriamente usando a distribuição $\mathcal{N}(0,\sigma^2)$. A coluna r representa a ordem e n é o tamanho da amostra, que foi escolhida empiricamente para melhor comparação entre os estimadores. As colunas "$<$", "$=$" e "$>$" representam respectivamente as quantidades de casos que o estimador "subestimou a ordem", "acertou a ordem" ou "superestimou a ordem". Nas Tabelas 2.2, 2.3 e 2.4 temos respectivamente os casos onde $\sigma = 1$, $\sigma = 10$ e $\sigma = 1000$. Para $r = 1,3,10$ foi considerado $K = 20$ e para $r = 50$ foi considerado $K = 100$. Em todos os casos simulados foi considerado $\alpha_i = 0$ para $i < r$ e $\alpha_r = 0,5$.

Observa-se que o EDC apresentou performance superior ao HQC e AIC em todos os casos. Comparado com o BIC, o EDC apresenta performance ligeiramente inferior em casos simples, onde $r \leq 3$, e superior para casos mais complexos, $r \geq 10$. Isso sugere que em casos de maior complexidade o EDC apresente performance relativa ainda maior. Ressalta-se que simulações mais detalhadas, modificando-se as escolhas de θ e n, podem nos levar a conclusões mais precisas. Entretanto, estes aspectos não são os objetivos principais deste trabalho.

Tabela 2.1: caso $\sigma = 0,1$

r	n	EDC			BIC			AIC			HQC		
		$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$
1	50	0	1	99	1	49	50	0	0	100	0	11	89
	100	0	54	46	0	82	18	0	12	88	0	52	48
	200	0	84	16	0	92	8	0	18	82	0	68	32
	500	0	94	6	0	96	4	0	22	78	0	69	31
	500000	0	99	1	0	99	1	0	21	79	0	81	19
3	50	0	2	98	10	29	61	0	0	100	0	5	95
	100	0	50	50	4	74	22	0	7	93	0	42	58
	200	0	81	19	0	87	13	0	15	85	0	66	34
	1000	0	93	7	0	94	6	0	14	86	0	69	31
	500000	0	100	0	0	100	0	0	28	72	0	93	7
10	50	0	5	95	51	3	46	0	0	100	6	5	89
	100	17	51	32	56	31	13	0	14	86	14	33	53
	200	3	77	20	12	76	12	0	19	81	1	58	41
	500	0	95	5	0	96	4	0	34	66	0	80	20
	20000	0	100	0	0	100	0	0	29	71	0	95	5
	500000	0	100	0	0	100	0	0	23	77	0	92	8
50	200	0	0	100	0	0	100	0	0	100	0	0	100
	500	9	81	10	100	0	0	0	0	100	2	70	28
	1000	0	98	2	22	76	2	0	3	97	0	72	28
	5000	0	100	0	0	100	0	0	12	88	0	89	11
	50000	0	100	0	0	100	0	0	11	89	0	91	9

Tabela 2.2: caso $\sigma = 1$

r	n	EDC			BIC			AIC			HQC		
		$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$	$<$	$=$	$>$
1	100	0	47	53	0	87	13	0	4	96	0	43	57
	200	0	84	16	0	92	8	0	18	82	0	68	32
	500	0	93	7	0	95	5	0	13	87	0	81	19
	50000	0	100	0	0	100	0	0	19	81	0	92	8
	500000	0	100	0	0	100	0	0	17	83	0	95	5
3	100	0	50	50	0	77	23	0	7	93	0	46	54
	200	0	81	19	0	87	13	0	15	85	0	66	34
	500	0	94	6	0	94	6	0	23	77	0	74	26
	500000	0	99	1	0	99	1	0	27	73	0	82	18
10	100	16	42	42	54	35	11	0	7	93	11	35	54
	200	3	77	20	12	76	12	0	19	81	1	58	41
	500	0	89	11	0	89	11	0	16	84	0	65	35
	10000	0	99	1	0	99	1	0	26	74	0	83	17
	500000	0	100	0	0	100	0	0	28	72	0	90	10
	1000000	0	100	0	0	100	0	0	20	80	0	94	6
50	500	12	65	23	99	0	1	0	1	99	4	57	39
	1000	0	94	6	23	74	3	0	1	99	0	67	33
	5000	0	99	1	0	99	1	0	5	95	0	90	10
	20000	0	99	1	0	99	1	0	10	90	0	88	12
	50000	0	100	0	0	100	0	0	11	89	0	79	21

Tabela 2.3: caso $\sigma = 10$

r	n	EDC			BIC			AIC			HQC		
		<	=	>	<	=	>	<	=	>	<	=	>
1	50	0	2	98	4	47	49	0	0	100	0	7	93
	100	0	54	46	0	83	17	0	5	95	0	47	53
	200	0	84	16	0	92	8	0	18	82	0	68	32
	500	0	95	5	0	98	2	0	13	87	0	76	24
	10000	0	100	0	0	100	0	0	18	82	0	85	15
	500000	0	100	0	0	100	0	0	24	76	0	92	8
3	50	0	2	98	12	30	58	0	0	100	0	5	95
	100	0	49	51	3	72	25	0	9	91	0	41	59
	200	0	81	19	0	87	13	0	15	85	0	66	34
	500	0	93	7	0	94	6	0	17	83	0	75	25
	5000	0	100	0	0	100	0	0	15	85	0	83	17
	500000	0	99	1	0	99	1	0	21	79	0	90	10
10	50	1	8	91	36	6	58	1	1	98	4	6	90
	100	7	59	34	41	48	11	0	9	91	5	38	57
	200	3	77	20	12	76	12	0	19	81	1	58	41
	500	0	92	8	0	93	7	0	23	77	0	71	29
	1000	0	96	4	0	97	3	0	21	79	0	81	19
	5000	0	100	0	0	100	0	0	32	68	0	88	12
	500000	0	99	1	0	99	1	0	20	80	0	92	8
50	200	0	0	100	0	0	100	0	0	100	0	0	100
	500	4	79	17	98	2	0	0	3	97	3	62	35
	1000	0	94	6	28	70	2	0	4	96	0	70	30
	2000	0	97	3	0	99	1	0	12	88	0	78	22
	5000	0	99	1	0	99	1	0	11	89	0	85	15
	100000	0	99	1	0	99	1	0	8	92	0	94	6

Tabela 2.4: caso $\sigma = 1000$

r	n	EDC			BIC			AIC			HQC		
		<	=	>	<	=	>	<	=	>	<	=	>
1	50	0	3	97	6	52	42	0	0	100	1	10	89
	100	0	59	41	0	94	6	0	8	92	0	55	45
	200	0	84	16	0	92	8	0	18	82	0	68	32
	500	0	95	5	0	96	4	0	16	84	0	76	24
	20000	0	100	0	0	100	0	0	16	84	0	84	16
	100000	2	98	0	0	100	0	0	20	80	0	90	10
3	50	0	1	99	17	33	50	0	1	99	1	7	92
	100	0	47	53	3	80	17	0	5	95	0	41	59
	200	0	81	19	0	87	13	0	15	85	0	66	34
	1000	0	99	1	0	99	1	0	24	76	0	80	20
	10000	0	99	1	0	99	1	0	12	88	0	83	17
	100000	6	94	0	0	98	2	0	23	77	0	90	10
10	50	2	3	95	41	5	54	0	1	99	5	3	92
	100	19	43	38	51	37	12	0	9	91	11	34	55
	200	3	77	20	12	76	12	0	19	81	1	58	41
	500	0	91	9	0	94	6	0	24	76	0	68	32
	10000	0	100	0	0	100	0	0	28	72	0	78	22
	100000	1	99	0	0	100	0	0	25	75	0	86	14
50	200	0	0	100	0	0	100	0	0	100	0	0	100
	500	9	67	24	100	0	0	0	1	99	5	55	40
	1000	0	98	2	24	75	1	0	4	96	0	79	21
	2000	0	99	1	0	99	1	0	5	95	0	81	19
	5000	0	100	0	0	100	0	0	11	89	0	93	7
	50000	0	100	0	0	100	0	0	9	91	0	90	10

2.3 Modelo Autoregressivo de Heteroscedasticidade Condicional (ARCH)

Processos Autoregressivos de Heteroscedasticidade Condicional (ARCH) foram propostos originalmente por Engle (1982) como uma melhor alternativa para modelagem de séries temporais em cenários econômicos. Desde então, modelos ARCH e variações vêm sendo utilizados com sucesso em econometria na modelagem de taxas de inflação (Engle 1982), séries temporais em mercado de câmbio (Domowitz & Hakkio 1985), valores de ativos (Bollerslev, Chou & Kroner 1992), dentre outros.

A consistência forte e normalidade assintótica dos estimadores para o caso ARCH foi estabelecida por Weiss (1986). Muito embora não existam trabalhos sobre a consistência de estimadores de ordem para modelos da família ARCH, os critérios de informação AIC e BIC vêm sendo utilizados sem qualquer formalização (Hughes, King & Kwek 2004).

Nessa seção a classe de estimadores EDC é definida para processo ARCH e a consistência forte é estabelecida em função do termo de penalidade. Como consequência imediata, a consistência forte do estimador de ordem BIC é demonstrada.

Processos ARCH é um caso particular dos processos BEKK-GARCH, que é objeto da próxima seção. Todavia, as demonstrações para o caso ARCH utilizam condições diferentes das exigidas para o caso geral, e por isso são mantidas.

2.3.1 Definições

No que segue, é apresentada a definição para modelos ARCH seguida de condições de ergodicidade e regularidade que são necessárias para o estabelecimento dos próximos resultados. A propriedade básica de processos ARCH é que o desvio padrão da inovação depende dos últimos r valores do processo.

Definição 2.22. *Uma sequência de variáveis aleatórias $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ é um processo Autoregressivo de Heteroscedasticidade Condicional (ARCH) de ordem $r \in \mathbb{N}$, denotado por ARCH(r), se satisfaz as condições abaixo.*

(i) Para todo $t \in \mathbb{N}$,

$$X_t = \varepsilon_t (h_t)^{\frac{1}{2}}.$$

(ii) $h_t = h(\alpha_0^r, X_{t-1}^{t-r}) = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_r X_{t-r}^2$, $\theta_r = (\alpha_0, \ldots, \alpha_r) \in \mathbb{R}^{r+1}$ *e* $\alpha_0, \alpha_r > 0$.

(iii) $\{\varepsilon_t\}_{t\in\mathbb{N}}$ *é uma sequência de variáveis aleatórias i.i.d.* $\mathscr{N}(0,1)$.

Condição 2.5 (Ergodicidade)**.** *Para* $\mathbb{X}$ *um ARCH(r), e* $\theta_r = (\alpha_0, \ldots, \alpha_r) \in \mathbb{R}^{r+1}$ *seus respectivos parâmetros, assumimos que*

$$\sum_{i=1}^{r} \alpha_i < 1. \tag{2.22}$$

Condição 2.6 (Regularidade)**.** *Para* $\mathbb{X}$ *um ARCH(r), então*

(i)

$$E(X_t^4) < \infty.$$

(ii) $\alpha_i > 0$ *para todo* $i \in \{1, \ldots, r\}$.

(iii) Existe um intervalo conhecido, $I = [c,d]$, *tal que* $\alpha_0 \in I^\circ = (c,d)$ *e* $c > 0$.

Usando a Condição 2.5, Francq & Zakoian (2010) (Teorema 2.5, página 37) demonstra que X_t admite momentos estacionários finitos de ordem 2. O Teorema 2.9 (página 45) da mesma referência provê condições para a existência de momentos pares maiores que 2. O uso do item (ii) da Condição 2.6 dispensa a exigência de momentos de ordem superiores. Em trabalhos futuros, deve ser avaliado a possibilidade de adequar a técnica para enfraquecer essa exigência. O item (iii) é a particularização da condição exigida em Jeantheau (1998) e Comte & Lieberman (2003).

Principais propriedades do processo ARCH

(a) $E(X_t) = 0$.

$$E(X_t) = E(E(X_t|\mathscr{F}_{t-1})) = E(0) = 0.$$

(b) $E(X_t X_{t-k}) = 0$, se $k \geq 1$

$$E(X_t X_{t-k}) = E(E(X_t X_{t-k}|\mathscr{F}_{t-1})) = E(X_{t-k} E(X_t|\mathscr{F}_{t-1})) = 0.$$

(c) $h_t \geq \alpha_0 + \alpha_1^i X_{t-i}^2 \prod_{p=1}^{i-1} \varepsilon_{t-p}^2 \geq \alpha_1^i X_{t-i}^2 \prod_{p=1}^{i-1} \varepsilon_{t-p}^2$, se $i \geq 1$. Basta utilizar um argumento indutivo para

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_r X_{t-r}^2 \\ &\geq \alpha_0 + \alpha_1 X_{t-1}^2 \\ &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 h_{t-1}. \end{aligned}$$

(d) Um ARCH(r) $\mathbb{X}$ pode ser imerso em uma sequência de modelos aninhados $\mathbb{M} = \bigcup_{k=0}^{\infty} M_k$, tomando $\Theta_k = I \times [0,1]^k$ e, para $\theta_k = (\alpha_0, \cdots, \alpha_k) \in \Theta_k$

$$f(x_1^n, \theta_k) = C(x_1^k) \prod_{t=1+k}^{n} \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{X_t^2}{2h_t}}.$$

Funções $\log L_{n,k}$ e suas derivadas

A log-verossimilhança de $m_k = f(X_1^n, \theta_k) \in M_k$ dada por

$$\begin{aligned} \log\left(\prod_{t=1+k}^{n} f(X_t|X_{t-k-1}^{t-1})\right) + C_1(X_1^k) &= \sum_{t=1+k}^{n} \log\left(\frac{1}{\sqrt{2\pi h_t}} e^{-\frac{X_t^2}{2h_t}}\right) + C_1(X_1^k) \\ &= \sum_{t=1+k}^{n} \left\{-\frac{X_t^2}{2h_t} - \frac{1}{2}\log h_t - \log\sqrt{2\pi}\right\} + C_1(X_1^k) \\ &= \sum_{t=1+k}^{n} \left\{-\frac{X_t^2}{2h_t} - \frac{1}{2}\log h_t\right\} - (n-k)\log\sqrt{2\pi} + C_1(X_1^k) \end{aligned} \tag{2.23}$$

Definimos abaixo o logaritmo de $L_{n,k}$, que satisfaz a Definição 1.3.

$$\log L_{n,k}(\theta) = \sum_{t=1+k}^{n} \left\{-\frac{X_t^2}{2h_t} - \frac{1}{2}\log h_t\right\}. \tag{2.24}$$

Na literatura (exemplo em Bollerslev (1986) e Engle (1982)) a função (2.24) é assumida diretamente como a log-verossimilhança. Para $i, j, l \in \{1, \cdots, k\}$,

$$\begin{aligned} \frac{\partial \log L_{n,k}(\theta)}{\partial \alpha_i} &= \sum_{t=1+k}^{n} \left\{\frac{X_t^2}{2h_t^2}\frac{\partial h_t}{\partial \alpha_i} - \frac{1}{2h_t}\frac{\partial h_t}{\partial \alpha_i}\right\} \\ &= \sum_{t=1+k}^{n} \left\{\frac{X_{t-i}^2}{2h_t}\left(\frac{X_t^2}{h_t} - 1\right)\right\}. \end{aligned} \tag{2.25}$$

$$
\begin{aligned}
\frac{\partial^2 \log L_{n,k}(\theta)}{\partial \alpha_i \alpha_j} &= \sum_{t=1+k}^{n} \left\{ -\frac{1}{2h_t^2} \frac{\partial h_t}{\partial \alpha_i} \frac{\partial h_t}{\partial \alpha_j} \frac{X_t^2}{h_t} + \left(\frac{X_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \alpha_j} \left\{ \frac{1}{2h_t} \frac{\partial h_t}{\partial \alpha_i} \right\} \right\} \\
&= \sum_{t=1+k}^{n} \left\{ -\frac{X_{t-i}^2 X_{t-j}^2 X_i^2}{2h_i^3} - \frac{X_{t-i}^2 X_{t-j}^2 X_i^2}{2h_i^3} + \frac{X_{t-i}^2 X_{t-j}^2}{2h_i^2} \right\} \\
&= \sum_{t=1+k}^{n} \left\{ -\frac{2X_{t-i}^2 X_{t-j}^2 X_i^2}{2h_i^3} + \frac{X_{t-i}^2 X_{t-j}^2}{2h_i^2} \right\}. \qquad (2.26)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^3 \log L_{n,k}(\theta)}{\partial \alpha_i \alpha_j \alpha_l} &= \sum_{t=1+k}^{n} \left\{ \frac{1}{h_t^3} \frac{\partial h_t}{\partial \alpha_l} \left(\frac{\partial h_t}{\partial \alpha_i} \frac{\partial h_t}{\partial \alpha_j} \frac{X_t^2}{h_t} \right) \right. \\
&\quad - \frac{1}{2h_t^2} \left(\frac{\partial}{\partial \alpha_l} \left\{ \frac{\partial h_t}{\partial \alpha_i} \frac{\partial h_t}{\partial \alpha_j} \right\} \frac{X_t^2}{h_t} + \frac{\partial h_t}{\partial \alpha_i} \frac{\partial h_t}{\partial \alpha_j} \left(-\frac{X_t^2}{h_t^2} \frac{\partial h_t}{\partial \alpha_l} \right) \right) \\
&\quad \left. + \left(-\frac{X_t^2}{h_t^2} \frac{\partial h_t}{\partial \alpha_l} \right) \frac{\partial}{\partial \alpha_j} \left\{ \frac{1}{2h_t} \frac{\partial h_t}{\partial \alpha_i} \right\} + \left(\frac{X_t^2}{h_t} - 1 \right) \frac{\partial^2}{\partial \alpha_j \alpha_l} \left\{ \frac{1}{2h_t} \frac{\partial h_t}{\partial \alpha_k} \right\} \right\} \\
&= \sum_{t=1+k}^{n} \left\{ \frac{X_{t-i}^2 X_{t-j}^2 X_{t-l}^2}{h_t^3} \frac{X_t^2}{h_t} \right. \\
&\quad + \frac{X_{t-i}^2 X_{t-j}^2 X_{t-l}^2}{2h_t^3} \frac{X_t^2}{h_t} + \frac{X_{t-i}^2 X_{t-j}^2 X_{t-l}^2}{2h_t^3} \frac{X_t^2}{h_t} \\
&\quad \left. + \left(\frac{X_t^2}{h_t} - 1 \right) \frac{X_{t-i}^2 X_{t-j}^2 X_{t-l}^2}{h_t^3} \right\} \\
&= \sum_{t=1+k}^{n} \left\{ \frac{2X_{t-i}^2 X_{t-j}^2 X_{t-l}^2 X_t^2}{h_t^4} + \left(\frac{X_t^2}{h_t} - 1 \right) \frac{X_{t-i}^2 X_{t-j}^2 X_{t-l}^2}{h_t^3} \right\} \\
&= \sum_{t=1+k}^{n} \left\{ \frac{3X_{t-i}^2 X_{t-j}^2 X_{t-l}^2 X_t^2}{h_t^4} - \frac{X_{t-i}^2 X_{t-j}^2 X_{t-l}^2}{h_t^3} \right\}. \qquad (2.27)
\end{aligned}
$$

Convencionamos $0/0 = 0/\infty = 0$. Para $i, j, l \in \{0, \ldots, k\}$ (2.25), (2.26) e (2.27) valem se definirmos "$X_{t-0}^2 := 1$".

Definição 2.23. *Para* $\{X_t\}_{t \in \mathbb{N}}$ *um* $ARCH(r)$, $k \geq r$, $\log L_n(\hat{\theta}_k)$ *como definido em* (2.24), $\theta_r \in \Theta_k \subseteq \mathbb{R}^{k+1}$*, definimos*

$$
m(i, j) = -E\left(\frac{\partial^2 \log L_n(\theta_r)}{\partial \alpha_i \alpha_j} \right). \qquad (2.28)
$$

2.3.2 Consistência do estimador de ordem de dependência

Os resultados desenvolvidos nessa seção seguem a mesma técnica utilizada para a definição do estimador EDC em processos AR. Para fazer isso, os processos ARCH são inseridos no contexto de cadeias de Markov para obtenção da Lei Forte dos Grandes Números. Foi necessário a utilização de resultados desenvolvidos para Martingales para a aplicação da Lei do Logaritmo Iterado.

Na definição abaixo o estimador de ordem EDC é definido de forma particular para processos ARCH. O Teorema 2.47 provê condições para c_n que definem a classe de estimadores EDC fortemente consistentes.

Definição 2.24. *Para $\mathbb{X}$ um ARCH(r) e $K \geq r$, definimos o estimador* EDC *de r por*

$$\hat{r}_{edc} = \underset{k \in \{0,\dots,K\}}{\operatorname{argmin}} \left\{ -\log L_{n,k}(\hat{\theta}_k) + (k+1)c_n \right\}$$

para $\log L_{n,k}(\hat{\theta}_k)$ *como definida em* (2.24) *e* $\{c_n\}_{n\in\mathbb{N}}$ *uma sequência de números positivos.*

O processo de imersão de processos ARCH em cadeias de Markov foi proposto, em caso mais geral, por Boussama (1998). A partir disso vêm sendo utilizado por outros autores, tais como Comte & Lieberman (2003) e Hafner & Preminger (2009*a*, 2009*b*). Abaixo segue a particularização desse resultado.

Definição 2.25. *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um ARCH(r). A cadeia de Markov k-derivada de $\mathbb{X}$ é o processo $\mathbb{Y} = \{Y_t\}_{t\in\mathbb{N}}$ com valores em $\mathbb{R}^k$,*

$$Y_t = (X_t, X_{t-1}, \dots, X_{t-k+1})'. \tag{2.29}$$

Teorema 2.26 (Boussama (1998))**.** *Seja $\{X_t\}_{t\in\mathbb{N}}$ um $ARCH(r)$, $k \geq r$, satisfazendo a Condição 2.5, então $\mathbb{Y}$, a cadeia de Markov k-derivada de $\mathbb{X}$, é Harris positiva e geometricamente ergódica.*

O Teorema 2.27 estabelece certa regularidade no comportamento assintótico de alguns objetos, que são necessários para o desenvolvimento dos próximos resultados.

Teorema 2.27. *Seja* $\{X_t\}_{t\in\mathbb{N}}$ *um ARCH*(r), $k \geq r$, $\log L_{n,k}$ *como definido em* (2.24), $\hat{\theta}_k \in \Theta_k \subseteq \mathbb{R}^{k+1}$ *satisfazendo as Condições 2.5 e 2.6 e* $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$, $s \in [0,1]$, *então*

(i)

$$\lim_{n\to\infty} -\frac{D_\theta^2 \log L_{n,k}(\dot{\theta})}{n} = A_2 \quad q.c.$$

Onde

$$A_2 = \begin{bmatrix} m(0,0) & \dots & m(0,k) \\ \vdots & \ddots & \vdots \\ m(k,0) & \dots & m(k,k) \end{bmatrix}.$$

(ii) A_2 *é positiva definida.*

(iii) *Existe* $c \in (0,\infty)$, *tal que*

$$\lim_{n\to\infty} \left| \frac{\frac{\partial^3 \log L_{n,k}(\dot{\theta})}{\partial \alpha_i \alpha_j \alpha_l}}{n} \right| \leq c.$$

Demonstração. (i) Usando a definição de h_t, temos que

$$\begin{aligned}
E\left\{\left|-\frac{1}{2h_t^2}\frac{\partial h_t}{\partial \alpha_i}\frac{\partial h_t}{\partial \alpha_j}\frac{X_t^2}{h_t} + \left(\frac{X_t^2}{h_t}-1\right)\frac{\partial}{\partial \alpha_j}\left\{\frac{1}{2h_t}\frac{\partial h_t}{\partial \alpha_i}\right\}\right|\right\} &= E\left\{\left|-\frac{X_{t-i}^2X_{t-j}^2X_t^2}{2h_t^3} - \frac{X_{t-i}^2X_{t-j}^2X_t^2}{2h_t^3} + \frac{X_{t-i}^2X_{t-j}^2}{2h_t^2}\right|\right\} \\
&\leq E\left\{\frac{2X_{t-i}^2X_{t-j}^2X_t^2}{2h_t^3} + \frac{X_{t-i}^2X_{t-j}^2}{2h_t^2}\right\} \\
&= E\left\{E\left\{\frac{2X_{t-i}^2X_{t-j}^2X_t^2}{2h_t^3}\middle|\mathscr{F}_{i-1}\right\} + \frac{X_{t-i}^2X_{t-j}^2}{2h_t^2}\right\} \\
&= E\left\{\frac{2X_{t-i}^2X_{t-j}^2}{2h_t^2} + \frac{X_{t-i}^2X_{t-j}^2}{2h_t^2}\right\} \\
&= E\left\{\frac{3X_{t-i}^2X_{t-j}^2}{2h_t^2}\right\} \\
&= E\left\{\frac{3X_{t-i}^2X_{t-j}^2}{2(\alpha_0+\alpha_1X_{i-1}^2+\cdots+\alpha_rX_{i-r}^2)(\alpha_0+\alpha_1X_{i-1}^2+\cdots+\alpha_rX_{i-r}^2)}\right\} \\
&\leq E\left\{\frac{3X_{t-i}^2X_{t-j}^2}{2\alpha_1^iX_{t-i}^2\left(\prod_{p=1}^{i-1}\varepsilon_{t-p}^2\right)\alpha_1^jX_{t-j}^2\left(\prod_{p=1}^{j-1}\varepsilon_{t-p}^2\right)}\right\} \\
&\leq E\left\{\frac{3}{2\alpha_1^{i+j}\left(\prod_{p=1}^{i-1}\varepsilon_{t-p}^2\right)\left(\prod_{p=1}^{j-1}\varepsilon_{t-p}^2\right)}\right\} \\
&< \infty.
\end{aligned}$$

O caso $j=0$ ou $i=0$ é análogo. Usando (2.26), e aplicando os Teoremas 2.26 e 2.5, temos que

$$\lim_{n\to\infty} \frac{\frac{\partial^2 \log L_{n,k}(\theta_r)}{\partial \alpha_i \alpha_j}}{n} = -m(i,j) \quad q.c.$$

De Francq & Zakoian (2010) (eq. 7.89, pg. 179) temos que

$$\lim_{n\to\infty} \left| \frac{\frac{\partial^2 \log L_{n,k}(\theta_r)}{\partial \alpha_i \alpha_j}}{n} - \frac{\frac{\partial^2 \log L_{n,k}(\dot{\theta})}{\partial \alpha_i \alpha_j}}{n} \right| = 0 \quad q.c.$$

e portanto

$$\lim_{n\to\infty} \frac{\frac{\partial^2 \log L_{n,k}(\dot{\theta})}{\partial \alpha_i \alpha_j}}{n} = -m(i,j) \quad q.c.$$

(ii) Para $\theta_k = (a_0, \ldots, a_k) \in \mathbb{R}^{k+1}$ qualquer, tal que $\|\theta_k\| > 0$, temos

$$\begin{aligned}
\theta_k A_2 \theta_k^T &= \theta_k E\left\{-D^2_\theta(\log L_{n,k}(\theta_r))\right\}\theta_k^T \\
&= E\left\{ \frac{1}{2h_t^2}\begin{bmatrix} a_0 & \ldots & a_k \end{bmatrix} \begin{bmatrix} 1 & X^2_{t-1} & \ldots & X^2_{t-k} \\ X^2_{t-1} & X^2_{t-1}X^2_{t-1} & \ldots & X^2_{t-k}X^2_{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ X^2_{t-k} & X^2_{t-k}X^2_{t-1} & \ldots & X^2_{t-k}X^2_{t-k} \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_k \end{bmatrix} \right\} \\
&= E\left\{ \frac{(a_0 + a_1 X^2_{t-1} + \cdots + a_k X^2_{t-k})^2}{2h_t^2} \right\} \\
&= E\left\{ \frac{h_t^2(\theta_k)}{2h_t^2(\theta_r)} \right\} \\
&> 0.
\end{aligned}$$

(iii) Esse é um caso particular da equação B3 de Comte & Lieberman (2003).

□

Como, pelo Teorema 2.26, os modelos ARCH são inseridos no contexto de cadeias de Markov, seria natural a utilização da Lei do Logaritmo Iterado apresentada no Teorema 2.5. Entretanto, as Condições 2.1 ou 2.2 não estão estabelecidas de forma suficiente para processos ARCH. Dessa forma, optou-se nesse trabalho por utilizar o Teorema 2.6 na demonstração do próximo Lema.

Lema 2.28. *Seja* $\{X_t\}_{t\in\mathbb{N}}$ *um* $ARCH(r)$ *satisfazendo as Condições 2.5 e 2.6,* $k \geq r$, $\log L_{n,k}$ *como definido em* (2.24)*,* $i \in \{0,\cdots,k\}$*, então*

$$\limsup_{n\to\infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n\log\log n}} = E\left(\frac{X_{t-i}^4}{2h_t^2}\right)^{1/2} < \infty \quad q.c.,$$

e

$$\liminf_{n\to\infty} \frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}}{\sqrt{2n\log\log n}} = -E\left(\frac{X_{t-i}^4}{2h_t^2}\right)^{1/2} > -\infty \quad q.c.$$

Demonstração. Assumimos $\mathscr{F}_t = \sigma(X_1,\ldots,X_t)$, $Z_t = t^\delta$, $\delta > 1$,

$$U_t = \frac{X_{t-i}^2}{2h_t}\left(\frac{X_t^2}{h_t} - 1\right) \quad e \quad W_n = \left[nE\left(\frac{X_{n-i}^4}{2h_n^2}\right)\right]^{1/2}.$$

Para aplicar o Teorema 2.6, basta demonstrar as hipóteses (L1-L4) que seguem.

(L1) Pela desigualdade de Chebyshev temos que

$$\begin{aligned}
P(|U_t| > Z_t) &= P\left(\left|\frac{X_{t-i}^2}{2h_t}\left(\frac{X_t^2}{h_t} - 1\right)\right| > t^\delta\right) \\
&\leq P\left(\left|\frac{1}{\alpha_i}\left(\frac{X_t^2}{h_t} - 1\right)\right| > t^\delta\right) \\
&= P\left(\left|\left(\frac{X_t^2}{h_t} - 1\right)\right| > \alpha_i t^\delta\right) \\
&\leq \frac{1}{\alpha_i^2 t^{2\delta}} E\left[(\varepsilon_t^2 - 1)^2\right] \\
&= \frac{1}{\alpha_i^2 t^{2\delta}} E\left[\varepsilon_t^4 - 2\varepsilon_t^2 + 1\right] \\
&= \frac{2}{\alpha_i^2 t^{2\delta}}.
\end{aligned} \tag{2.30}$$

Então temos que

$$\sum_{t=1}^{\infty} P(|U_t| > Z_t) \leq \sum_{t=1}^{\infty} \frac{2}{\alpha_i^2 t^{2\delta}} < \infty$$

e usando o Lema de Borel-Cantelli temos que

$$P\left(\left\{\omega : \mathbb{I}(|U_t| > t^\delta) = 1 \ \ infinitas \ \ vezes\right\}\right) = P\left(\left\{\omega : |U_t| > t^\delta \ \ infinitas \ \ vezes\right\}\right) = 0$$

e portanto

$$\lim_{n\to\infty} \frac{\sum_{t=1}^{n} U_t\mathbb{I}(|U_t| > Z_t) - E[U_t\mathbb{I}(|U_t| > Z_t)|\mathscr{F}_{t-1}]}{\sqrt{2W_n^2 \log\log W_n^2}} = 0 \quad q.c.$$

(L2)

$$\begin{aligned} E(U_t|\mathscr{F}_{t-1}) &= \frac{X_{t-i}^2}{2h_t} E\left[\left(\frac{X_t^2}{h_t} - 1\right)|\mathscr{F}_{t-1}\right] \\ &= 0 \quad e \end{aligned}$$

$$\begin{aligned} E\left[E(U_t^2|\mathscr{F}_{t-1})\right] &= E\left[\frac{X_{t-i}^4}{4h_t^2}\left(\varepsilon_t^2 - 1\right)^2\right] \\ &= E\left[\frac{X_{t-i}^4}{4h_t^2}\right] E\left[\varepsilon_t^4 - 2\varepsilon_t^2 + 1\right] \\ &= E\left(\frac{X_{t-i}^4}{2h_t^2}\right) \leq E\left(\frac{1}{2\alpha_i}\right) < \infty \end{aligned}$$

e portanto, usando o item (i) do Teorema 2.5 (LFGN), temos que

$$\lim_{n\to\infty} \frac{\sum_{t=1}^{n} E[U_t^2|\mathscr{F}_{t-1}]}{nE\left(\frac{X_{1-i}^4}{2h_1^2}\right)} = 1 \quad q.c.$$

Usando o Teorema da Convergência Dominada, temos que

$$\lim_{t\to\infty} E\left[U_t\mathbb{I}(|U_t| \leq t^{\delta})|\mathscr{F}_{t-1}\right] = 0 \quad q.c.$$

Agora, considerando $\varepsilon > 0$ arbitrário, é necessário encontrar uma cota superior somável em t para

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

e aplicar Borel-Cantelli para obter

$$\lim_{t\to\infty}\left[E(U_t^2\mathbb{I}(|U_t| \leq t^{\delta})|\mathscr{F}_{t-1}) - E(U_t^2|\mathscr{F}_{t-1})\right] = 0 \quad q.c.$$

e aplicar o Teorema Médio de Cesàro (Apêndice A) para concluir

$$\begin{aligned}
\lim_{n\to\infty} \frac{\sum_{t=1}^{n} E[U_t^2 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] - E[U_t \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}]^2}{W_n^2} &= \lim_{n\to\infty} \frac{\sum_{t=1}^{n} E[U_t^2|\mathscr{F}_{t-1}]}{nE\left(\frac{X_{1-i}^4}{2h_1^2}\right)} \\
&= 1 \quad q.c.
\end{aligned}$$

Usando a desigualdade generalizada de Chebyshev (Apêndice B)

$$\begin{aligned}
P&\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta)|\mathscr{F}_{t-1})\right| > \varepsilon\right] \\
&\leq \frac{1}{\varepsilon} E\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta)|\mathscr{F}_{t-1})\right|\right]
\end{aligned}$$

e

$$\begin{aligned}
E\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta)|\mathscr{F}_{t-1})\right|\right] &= E\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta)|\mathscr{F}_{t-1}) \pm E(U_t^2 I(|U_t| > t^\delta)|\mathscr{F}_{t-1})\right|\right] \\
&= E\left[E(U_t^2 I(|U_t| > t^\delta)|\mathscr{F}_{t-1})\right] \\
&= E\left[U_t^2 I(|U_t| > t^\delta)\right] \\
&\leq E\left[U_t^4\right]^{1/2} E\left[I(|U_t| > t^\delta)\right]^{1/2} \\
&= E\left[U_t^4\right]^{1/2} \mathrm{P}(|U_t| > t^\delta)^{1/2}.
\end{aligned}$$

Usando (2.30) e (2.31) para $c > 0$ apropriado, obtemos

$$\begin{aligned}
E\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^\delta)|\mathscr{F}_{t-1})\right|\right] &\leq c\left(\frac{2}{\alpha_i^2 t^{2\delta}}\right)^{1/2} \\
&= c\frac{2^{1/2}}{\alpha_i t^\delta}
\end{aligned}$$

que é somável em t.

(L3) Temos que

$$
\begin{aligned}
0 &\leq E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] \\
&\leq E[U_t^4|\mathscr{F}_{t-1}] \\
&= E\left[\left(\frac{X_{t-i}^2}{2h_t}\right)^4 \left(\frac{X_t^2}{h_t}-1\right)^4 |\mathscr{F}_{t-1}\right] \\
&\leq \frac{1}{\alpha_i^4} E\left[\left(\varepsilon_t^2-1\right)^4 |\mathscr{F}_{t-1}\right] \\
&= c_1 E\left[\left(\varepsilon_t^2-1\right)^4 |\mathscr{F}_{t-1}\right]
\end{aligned}
\tag{2.31}
$$

para $\delta_1 \in (\frac{1}{2}, 1)$, usando as desigualdades de Chebyshev e de Jensen,

$$
\begin{aligned}
\mathrm{P}\left[c_1 E\left[\left(\varepsilon_t^2-1\right)^4 |\mathscr{F}_{t-1}\right] > t^{\delta_1}\right] &\leq \frac{c_1^2}{t^{2\delta_1}} E\left[E\left[\left(\varepsilon_t^2-1\right)^4 |\mathscr{F}_{t-1}\right]^2\right] \\
&\leq \frac{c_1^2}{t^{2\delta_1}} E\left[\left(\varepsilon_t^2-1\right)^8\right] \\
&\leq c_2 \frac{1}{t^{2\delta_1}}
\end{aligned}
$$

e portanto,

$$
\sum_{t=1}^{\infty} \mathrm{P}\left[E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] > t^{\delta_1}\right] \leq \sum_{t=1}^{\infty} \frac{c_2}{t^{2\delta_1}} < \infty.
$$

Usando o Lema de Borel-Cantelli,

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

e logo

$$
\lim_{n\to\infty} \sum_{t=1}^{n} \frac{E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}]}{W_t^4} \leq \lim_{n\to\infty} \sum_{t=1}^{n} \frac{1}{t^{2-\delta_1}} < \infty \quad q.c.
$$

(L4) Como $E(X_t^4) > 0$, caso contrário $X_t \equiv 0$ q.c., então, usando a estacionaridade, temos

$$
\lim_{n\to\infty} \frac{W_n}{W_{n+1}} = \lim_{n\to\infty} \frac{\left[nE\left(X_{1-i}^4/h_1^2\right)\right]^{1/2}}{\left[(n+1)E\left(X_{1-i}^4/h_1^2\right)\right]^{1/2}} = 1 \quad q.c. \;\; e
$$

$$\lim_{n\to\infty} W_n = E\left(X_{1-i}^4/h_1^2\right)^{1/2} \lim_{n\to\infty}\sqrt{n} = \infty.$$

□

Ressalta-se que a desigualdade generalizada de Chebyshev e o Teorema Médio de Cesàro foram utilizados por Hafner & Preminger (2009*a*), em situação semelhante, para demonstrar a normalidade assintótica do estimador do parâmetro θ_r no caso de processos FACTOR-GARCH.

Corolário 2.29. *Seja* $\{X_t\}_{t\in\mathbb{N}}$ *um ARCH*(r)*, satisfazendo as Condições 2.5 e 2.6,* $k \geq r$*,* $\log L_{n,k}$ *como definido em* (2.24)*, então existe* $c_5 \in (0,\infty)$ *tal que*

$$\limsup_{n\to\infty} \frac{\left\|D_\theta^1 \log L_{n,k}(\theta_r)\right\|}{\sqrt{2\log\log n}} \leq c_5 \quad q.c.$$

Demonstração.

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\left\|D_\theta^1 \log L_{n,k}(\theta_r)\right\|}{\sqrt{2\log\log n}} &\leq \limsup_{n\to\infty} \sum_{i=0}^{k} \frac{\left|\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}\right|}{\sqrt{2n\log\log n}} \\
&\leq \sum_{i=0}^{k} \limsup_{n\to\infty} \frac{\left|\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}\right|}{\sqrt{2n\log\log n}} \\
&\leq \sum_{i=0}^{k} E\left(\frac{X_{t-i}^4}{2h_t^2}\right)^{1/2} \\
&= c_5 \quad q.c.
\end{aligned}$$

□

O próximo Lema contém resultados técnicos que são utilizados no que segue.

Lema 2.30. *Seja* $\mathbb{X}$ *um ARCH*(r) *satisfazendo as Condições 2.5 e 2.6,* $k \geq 0$*,* $A \subseteq \Theta_k$*, então*

(i) Se $\overline{\theta} = (\alpha_0, \ldots, \alpha_k) \in A$

$$E\left|h_t(\overline{\theta})\right| < \infty,$$

$$E\left|\log(h_t(\overline{\theta}))\right| < \infty \quad e$$

$$E\left|\log\left(\frac{1}{2\pi h_t(\theta)} e^{-\frac{X_t^2}{2h_t(\theta)}}\right)\right| < \infty.$$

(ii) Para $i \in \{0, \ldots, k\}$,

$$E \sup_{\theta \in A} \left| \frac{X_{t-i}^2}{2h_t(\theta)} \left(\frac{X_t^2}{h_t(\theta)} - 1 \right) \right| < \infty.$$

Demonstração. (i) Como $E\left(X_t^2\right) < \infty$,

$$\begin{aligned} E\left|h_t(\bar{\theta})\right| &\leq E\left|\alpha_0 + \sum_{i=1}^{k} \alpha_i X_{t-i}^2\right| \\ &\leq \alpha_0 + \sum_{i=1}^{k} \alpha_i E\left(X_{t-i}^2\right) < \infty. \end{aligned}$$

Temos também que

$$E\left(\log(h_t(\bar{\theta}))\right) \leq E\left(h_t\right) < \infty.$$

Além disso, usando c como na Condição 2.6, temos

$$E\left(\log(h_t(\bar{\theta}))\right)^- \leq \max\{-\log(c), 0\} < \infty.$$

Para a última inequação temos

$$E\left|\log\left(\frac{1}{2\pi h_t(\bar{\theta})} e^{-\frac{X_t^2}{2h_t(\theta)}}\right)\right| \leq c_1 + \frac{1}{c} E\left|X_t^2\right| + E\left|\log(h_t)\right| < \infty.$$

(ii)

$$\begin{aligned} E \sup_{\theta \in A} \left| \frac{X_{t-i}^2}{2h_t(\theta)} \left(\frac{X_t^2}{h_t(\theta)} - 1 \right) \right| &\leq E\left|\frac{X_{t-i}^2}{c} \frac{X_t^2}{c}\right| \\ &\leq E\left|\frac{X_{t-i}^4}{c}\right|^{1/2} E\left|\frac{X_t^4}{c}\right|^{1/2} < \infty. \end{aligned}$$

□

Abaixo demonstramos H1 para processos ARCH. O resultado análogo demonstrado para processos AR utilizou um argumento simples, fundamentado na expansão de Taylor e considerando que as derivadas de terceira ordem de $\log L_{n,k}$ são nulas nos processos AR, o que não ocorre no geral. Assim, optou-se por desenvolver uma demonstração (abaixo) utilizando a comparação das densidades pela divergência de Kullback-Leibler em um contexto mais geral. Para isso foi necessário atentar para alguns detalhes de convergência de sequências.

Teorema 2.31. *Seja $\mathbb{X}$ um ARCH(r) satisfazendo as Condições 2.5 e 2.6, $k < r$, então*

$$\lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_r) - \log L_n(\hat{\theta}_k)}{n} > 0 \quad q.c.$$

Demonstração. Basta mostrar o resultado considerando

$$\log L_{n,k}(\theta) = \sum_{t=1+k}^{n} l_t(\theta)$$

e

$$l_t(\theta) = \log\left(f(\theta)\right) = \log\left(\frac{1}{\sqrt{2\pi h_t}} e^{-\frac{X_t^2}{2h_t}}\right)$$

em (2.23). Usando o Lema 2.30, temos que

$$E\left[|l_t(\theta)|\right] < \infty$$

e portanto, usando o Teorema 2.5,

$$\lim_{n\to\infty} \frac{\log L_n(\theta_r)}{n} = \lim_{n\to\infty} \frac{\sum_{t=1+k}^{n} l_t(\theta_r)}{n} = E\left(l_1(\theta_r)\right) = c_1 < \infty \quad q.c.$$

Usando o Teorema do valor médio, para $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_r$, $s \in (0,1)$, n suficientemente grande e $B_\delta(\theta_r)$ uma vizinhança suficientemente pequena de θ_r, temos

$$\begin{aligned}
\left| \frac{\sum_{t=1+r}^{n} l_t(\hat{\theta}_r)}{n} - \frac{\sum_{t=1+r}^{n} l_t(\theta_r)}{n} \right| &= \left| \frac{\sum_{t=1+r}^{n} D_\theta^1 l_t(\dot{\theta})}{n} (\hat{\theta}_r - \theta_r) \right| \\
&\leq \sup_{\theta \in B_\delta(\theta_r)} \left\| \frac{\sum_{t=1+r}^{n} D_\theta^1 l_t(\dot{\theta})}{n} \right\| \left\| (\hat{\theta}_r - \theta_r) \right\|
\end{aligned}$$

Aplicando o Teorema 2.5, usando o item (ii) do Lema 2.30 e a consistência forte de $\hat{\theta}_r$,

$$\left| \frac{\sum_{t=1+r}^{n} l_t(\hat{\theta}_r)}{n} - \frac{\sum_{t=1+r}^{n} l_t(\theta_r)}{n} \right| \underset{q.c.}{\rightarrow} 0. \tag{2.32}$$

Por outro lado, temos que, como o estimador é de máxima verossimilhança e $\Theta_k \subset \Theta_r$,

$$\begin{aligned} \lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_k)}{n} &\leq \lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_r)}{n} \\ &= c_1 \quad q.c. \end{aligned}$$

e portanto,

$$\lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_k)}{n} \leq \limsup_{n\to\infty} \frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_k)}{n} = c_2 \leq c_1 \quad q.c.$$

Seja n_i uma subsequência de n tal que

$$\lim_{n_i\to\infty} \frac{\sum_{t=1+k}^{n_i} l_t(\hat{\theta}_k)}{n_i} = c_2 \quad q.c.$$

Usando que $\overline{\Theta}_k$ é compacto, tome n_j uma subsequência de n_i tal que $\hat{\theta}_k(n_j) \rightarrow \overline{\theta}_k \in \overline{\Theta}_k$ q.c. Temos assim que

$$\begin{aligned} \lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_k)}{n} &\leq \limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k)}{n} \\ &= \lim_{n_j\to\infty} \frac{\log L_{n_j}(\hat{\theta}_k(n_j))}{n_j} \quad q.c. \end{aligned}$$

Aplicando o mesmo raciocínio usado em (2.32), concluímos

$$\lim_{n_j\to\infty} \frac{\log L_{n_j}(\overline{\theta}_k)}{n_j} = E\left(l_1(\overline{\theta})\right) \quad e$$

$$\left| \frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_k)}{n} - \frac{\sum_{t=1+k}^{n} l_t(\overline{\theta})}{n} \right| \underset{q.c.}{\rightarrow} 0.$$

Além disso,

$$\lim_{n\to\infty}\left[\frac{\log L_n(\hat\theta_r)}{n}-\frac{\log L_n(\hat\theta_k)}{n}\right] \quad \geq \quad E\left[\log\left(\frac{f(\bar\theta_k)}{f(\theta_r)}\right)\right]$$

Por outro lado, $E\left[\log\left(\frac{f(\bar\theta_k)}{f(\theta_r)}\right)\right]$ é a divergência de Kullback-Leibler, que é positiva se $f(\bar\theta_k)\neq f(\theta_r)$, e como $\theta_r \notin \bar\Theta_k \subseteq \mathbb{R}^{\gamma(k)}$, temos que $\theta_r \neq \bar\theta_k$ e então $f(\bar\theta_k)\neq f(\theta_r)$. Onde concluímos

$$\lim_{n\to\infty}\left[\frac{\log L_n(\hat\theta_r)}{n}-\frac{\log L_n(\hat\theta_k)}{n}\right] \quad \geq \quad E\left[\log\left(\frac{f(\bar\theta_k)}{f(\theta_r)}\right)\right] > 0.$$

□

Teorema 2.32. *Seja* $\mathbb{X}$ *um ARCH*(r) *satisfazendo as Condições 2.5 e 2.6 e* $\hat r_{edc}$ *como na Definição 2.24. Então,* $\hat r_{edc}$ *é fortemente consistente (*$\hat r_{edc} \underset{q.c.}{\to} r$*) se*

$$\lim_{n\to\infty}\frac{c_n}{n}=0 \quad e \quad \liminf_{n\to\infty}\frac{c_n}{\log\log n}=\infty. \tag{2.33}$$

Demonstração. O conjunto dos possíveis valores de θ_r, Θ_r, definido em (2.22), é aberto e $\theta_r \in \Theta_r$. De Weiss (1986), temos que o estimador de máxima verossimilhança $\hat\theta_k$ é fortemente consistente para $k \geq r$, isto é, vale (1.10). As derivadas de $\log L_{n,k}$, (2.25) e (2.26) são contínuas em relação a θ e a x_1^n e portanto são mensuráveis em relação a x_1^n. Temos de (2.27) que (1.11) é satisfeita. Usando o Teorema 2.27 e o Corolário 2.29 temos as condições satisfeitas para a aplicação do item (ii) do Teorema 1.6 e com isso temos (H2) satisfeita. A hipótese (H1) é segue do Teorema 2.31. O resultado segue do item (i) do Teorema 1.5. □

Corolário 2.33. *Seja* $\mathbb{X}$ *um ARCH*(r) *satisfazendo as Condições 2.5 e 2.6 ,* $K \geq r$*. Então o estimador* $\hat r_{bic}$*, como definido abaixo, é fortemente consistente.*

$$\hat r_{bic} = \underset{k\in\{0,\ldots,K\}}{\operatorname{argmin}}\left\{-\log L_{n,k}(\hat\theta_k)+\frac{(k+1)}{2}\log n\right\}$$

para $\log L_{n,k}(\hat\theta_k)$ *como definida em* (2.24).

Demonstração. $c_n = \log n/2$ e satisfaz (2.33). □

2.4 Modelo ARCH multivariado generalizado (BEKK-GARCH)

Desde sua criação por Engle (1982), o modelo ARCH teve diversas generalizações e adaptações em que se destacam os modelos GARCH (Bollerslev 1986), NGARCH (Engle & Ng 1993), EGARCH (Nelson 1991) e generalizações multivariadas como BEKK-GARCH (Engle & Kroner 1995), CCC-GARCH (Bollerslev 1990), VEC-GARCH (Bollerslev, Engle & Wooldridge 1988), dentre outros.

Os modelos multivariados tem especial aplicação na seleção de portfólios e precificação de ativos (Hafner & Preminger 2009*b*). Na classe dos modelos multivariados, o BEKK-GARCH se destaca por ser geral e por haver avanços significativos disponíveis na literatura. Dos citados, apenas o VEC-GARCH é mais geral que o modelo BEKK-GARCH e, mesmo assim, os casos VEC-GARCH que não são representáveis na modelagem BEKK-GARCH são de certa forma degenerados (Stelzer 2008).

Boussama (1998), utilizando técnicas de geometria algébrica, inseriu os modelos BEKK-GARCH no contexto de cadeias de Markov e demonstrou a ergodicidade geométrica desses modelos a partir de determinadas condições. Com algumas alterações, esses resultados também foram publicados por Boussama, Fuchs & Stelzer (2011).

Comte & Lieberman (2003) utilizou os resultados de Boussama (1998) para demonstrar as condições propostas por Jeantheau (1998) para obter a consistência forte de estimadores de máxima verossimilhança para processos BEKK-GARCH. Usando as condições de Basawa & Heyde (1976), a normalidade assintótica do estimador de máxima verossimilhança também é estabelecida por Comte & Lieberman.

Assim como em processos ARCH, não há a formalização de estimadores de ordem para modelos BEKK-GARCH, embora os critérios de informação AIC e BIC vêm sendo utilizados (Francq & Zakoian 2010).

Nessa Seção, a classe de estimadores EDC é definida para modelos BEKK-GARCH e são demonstradas a consistência forte para uma subclasse, que inclui o estimador de ordem baseado no critério de informação BIC. A técnica utilizada é semelhante à utilizada para o caso de modelos ARCH, a maior diferença está no uso de cálculo matricial, que é necessário para a manipulação dos objetos existentes na definição multivariada.

2.4.1 Definições

No que segue é apresentada a definição de modelo BEKK-GARCH e da notação utilizada nos próximos resultados.

Definição 2.34. *Uma sequência de variáveis aleatórias* $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ *com valores em* $\mathbb{R}^m$ *é um processo Autoregressivo de Heteroscedasticidade Condicional Generalizado na modelagem BEKK (BEKK-GARCH) de ordem* $(p,q) \in \mathbb{N}^2$*, denotado por BEKK-GARCH(p,q), se satisfaz as condições abaixo.*

(i) *Para todo* $t \in \mathbb{N}$*,*

$$X_t = (H_t)^{\frac{1}{2}} \varepsilon_t.$$

(ii) *Para* C*,* A_{ls} *e* B_{ls} *matrizes* $(m \times m)$*,* C *positiva definida e* $N \in \mathbb{N}$*,*

$$H_t = C + \sum_{l=1}^{q} \left(\sum_{s=1}^{N} A_{ls} X_{t-l} X_{t-l}' A_{ls}' \right) + \sum_{l=1}^{p} \left(\sum_{s=1}^{N} B_{ls} H_{t-l} B_{ls}' \right).$$

(iii) $\{\varepsilon_t\}_{t\in\mathbb{N}}$ *é uma sequência de variáveis aleatórias i.i.d em* $\mathbb{R}^m$ *com distribuição* $\mathcal{N}(0, I_m)$*, para* I_m *a matriz identidade* $(m \times m)$*.*

Definição 2.35. *Seja* $A = (a_{ij}) \in \mathcal{M}$ *uma matriz e* $\mathcal{M}$ *o espaço das matrizes* $(m \times m)$*.*

(i) *O operador* $vec : \mathcal{M} \to \mathbb{R}^{m\times m}$ *é definido por*

$$vec(A) = (a_{11}, a_{21}, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1m}, \ldots, a_{mm})'$$

(ie., empilha as colunas de A).

(ii) *O operador* $vech : \mathcal{M} \to \mathbb{R}^{m(m+1)/2}$ *é definido por*

$$vech(A) = (a_{11}, a_{21}, \ldots, a_{m1}, a_{22}, \ldots, a_{m2}, \ldots, a_{mm})'$$

(ie., empilha as colunas da parte triangular inferior de A).

(iii) $D_m = \left(d_{ij}\right)$ *é a matriz duplicação, em que*

$$d_{ij} = \mathbb{I}[(i,j) \in A_d] \quad e$$

$$\begin{aligned} A_d \quad = \quad & \left\{(a,b) \in \mathbb{N}^2 : (a,b) = ((j-1)m+i,(j-1)(m-j/2)+i) \;\; ou \right. \\ & (a,b) = ((i-1)m+j,(j-1)(m-j/2)+i) \\ & \left. i,j \in \mathbb{N} \quad e \quad i \geq j \right\}. \end{aligned}$$

(iv) $D_m^+ = (D_m' D_m)^{-1} D_m'$.

(v) Se B é matriz $(m \times n)$ *e C é matriz* $(m' \times n')$*, então o Produto de Kronecker* $B \otimes C$ *é a matriz* $(mm' \times nn')$ *com os blocos* $a_{ij}B$*.*

(vi) O raio espectral de A é definido por $\rho(A) = \max\{|\lambda_i| : \lambda_i \;\; é \;\; autovalor \;\; de \;\; A\}$*.*

(vii) $\|A\|$ *é a norma de Frobenius, isto é,*

$$\|A\|^2 = \sum_{i,j} a_{ij}^2.$$

Usando a notação definida, temos o seguinte resultado que pode ser encontrado em Francq & Zakoian (2010).

Proposição 2.36 (Francq & Zakoian (2010))**.** *Seja A matriz* $(m \times m)$*, B e C matrizes tais que o produto ABC esteja bem definido. Então*

$$vec(A) = D_m vech(A),$$

$$vech(A) = D_m^+ vec(A) \;\; e$$

$$vec(ABC) = (C' \otimes A) vec(B).$$

A condição a seguir, necessária para a obtenção de ergodicidade de processos BEKK-GARCH, foi proposta por Boussama (1998) e é utilizada por diversos trabalhos, dentre eles Comte & Lieberman (2003).

Condição 2.7 (Ergodicidade)**.** *Para $\mathbb{X}$ um BEKK-GARCH(p,q),*

$$\rho\left(\sum_{l=1}^{q}\tilde{A}_l+\sum_{l=1}^{p}\tilde{B}_l\right)<1 \tag{2.34}$$

para

$$\tilde{A}_l=D_m^{+}\sum_{s=1}^{N}(A_{ls}\otimes A_{ls})D_m \quad e \quad \tilde{B}_l=D_m^{+}\sum_{s=1}^{N}(B_{ls}\otimes B_{ls})D_m.$$

Principais propriedades do processo BEKK-GARCH

(a) $E(X_t)=\vec{0}$.

$$E(X_t)=E(E(X_t|\mathscr{F}_{t-1}))=E(0)=0.$$

(b) $E(X_tX_{t-k})=\vec{0}$, se $k\geq 1$.

$$E(X_tX_{t-k})=E(E(X_tX_{t-k}|\mathscr{F}_{i-1}))=E(X_{t-k}E(X_t|\mathscr{F}_{i-1}))=0.$$

(c) $E(X_tX_t'|\mathscr{F}_{t-1})=H_t \quad q.c.$ (Comte & Lieberman 2003).

(d) Supondo $N=1$, um BEKK-GARCH(k_1,k_2) $\mathbb{X}$ pode ser imerso em uma classe de modelos parcialmente aninhados $\mathbb{M}=\{M_k\}_{k\in\mathbb{N}^2}$ tomando, $\bar{k}=\max\{k_1,k_2\}$,

$$\Theta_k=\Omega_0\times\Omega_1\times\cdots\times\Omega_{2\bar{k}}.$$

Onde $\Omega_i=\{0\}^{m^2}$ se $i/2>k_2$ e i é impar ou $i/2>k_1$ e i é par, nos demais casos $\Omega_i\subseteq\mathbb{R}^{m^2}$ com interior não vazio.

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

assumindo $A_i=0$ se $i>k_2$, $B_i=0$ se $i>k_2$,

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

e

$$f(x_1^n,\theta_k)=C_1(x_1^{\bar{k}})\prod_{t=1+\bar{k}}^{n}\frac{1}{\sqrt{(2\pi)^{m/2}\det(H_t)}}exp\left(-\frac{1}{2}x_t'H_t^{-1}x_t\right)$$

para

$$H_t=C+\sum_{l=1}^{k_2}A_lX_{t-l}X_{t-s}'A_l'+\sum_{l=1}^{k_1}B_lH_{t-l}B_l'.$$

Funções $\log L_{n,k}$ **e suas derivadas**

Para uma amostra x_1^n, temos que a log-verossimilhança de $m_k = f(x_1^n, \theta_k)$, $k \in \mathbb{N}^2$, é dada por

$$\begin{aligned}
&\log\left(\prod_{t=1+\bar{k}}^{n} \frac{1}{\sqrt{(2\pi)^{m/2}\det(H_t)}} exp\left(-\frac{1}{2}x_t' H_t^{-1} x_t\right)\right) + C_1(x_1^{\bar{k}}) \\
&= \sum_{t=1+\bar{k}}^{n} \log\left(\frac{1}{\sqrt{(2\pi)^{m/2}\det(H_t)}} exp\left(-\frac{1}{2}x_t' H_t^{-1} x_t\right)\right) + C_1(x_1^{\bar{k}}) \\
&= \sum_{t=1+\bar{k}}^{n} \left\{-\frac{1}{2}x_t' H_t^{-1} x_t - \frac{1}{2}\log\det(H_t) - \log\sqrt{(2\pi)^{m/2}}\right\} + C_1(x_1^{\bar{k}}) \\
&= \sum_{t=1+\bar{k}}^{n} \left\{-\frac{1}{2}x_t' H_t^{-1} x_t - \frac{1}{2}\log\det(H_t)\right\} - (n-\bar{k})\log\sqrt{(2\pi)^{m/2}} + C_1(x_1^{\bar{k}})
\end{aligned} \tag{2.35}$$

Definimos abaixo o logaritmo de $L_{n,k}$, que satisfaz a Definição 1.3.

$$\log L_{n,k}(\theta) = \sum_{t=1+\bar{k}}^{n} l_t(\theta_k) \tag{2.36}$$

para

$$l_t(\theta_k) = \left\{-\frac{1}{2}x_t' H_t^{-1} x_t - \frac{1}{2}\log\det(H_t)\right\}.$$

Na literatura (exemplo em Comte & Lieberman (2003)) a função (2.36) é assumida diretamente como a log-verossimilhança. Para $i, j \in \{1, \cdots, k\}$,

$$\begin{aligned}
\frac{\partial \log L_{n,k}(\theta)}{\partial \alpha_i} &= \sum_{t=1+\bar{k}}^{n} \left\{-\frac{1}{2}x_t' \frac{\partial H_t^{-1}}{\partial \alpha_i} x_t - \frac{1}{2det(H_t)}\left(det(H_t)\mathrm{Tr}\left[H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}\right]\right)\right\} \\
&= \sum_{t=1+\bar{k}}^{n} \left\{\frac{1}{2}x_t' H_t^{-1}\frac{\partial H_t}{\partial \alpha_i} H_t^{-1} x_t - \frac{1}{2}\mathrm{Tr}\left[H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}\right]\right\} \\
&= \sum_{t=1+\bar{k}}^{n} \frac{1}{2}\mathrm{Tr}\left[x_t' H_t^{-1}\frac{\partial H_t}{\partial \alpha_i} H_t^{-1} x_t - H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}\right] \\
&= \sum_{t=1+\bar{k}}^{n} \frac{1}{2}\mathrm{Tr}\left[x_t x_t' H_t^{-1}\frac{\partial H_t}{\partial \alpha_i} H_t^{-1} - H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}\right] \quad e
\end{aligned} \tag{2.37}$$

$$
\begin{aligned}
\frac{\partial^2 \log L_{n,k}(\theta)}{\partial\alpha_i\alpha_j} &= \sum_{t=1+\bar{k}}^{n} \frac{1}{2}\mathrm{Tr}\left\{x_t x_t' \left[\frac{\partial H_t^{-1}}{\partial\alpha_j}\frac{\partial H_t}{\partial\alpha_i}H_t^{-1} + H_t^{-1}\left(\frac{\partial^2 H_t}{\partial\alpha_i\alpha_j}H_t^{-1} + \frac{\partial H_t}{\partial\alpha_i}\frac{\partial H_t^{-1}}{\partial\alpha_j}\right)\right]\right. \\
&\qquad \left. - \left(\frac{\partial H_t^{-1}}{\partial\alpha_j}\frac{\partial H_t}{\partial\alpha_i} + H_t^{-1}\frac{\partial^2 H_t}{\partial\alpha_i\alpha_j}\right)\right\} \\
&= \sum_{t=1+\bar{k}}^{n} \frac{1}{2}\mathrm{Tr}\left\{x_t x_t' \left[-H_t^{-1}\frac{\partial H_t}{\partial\alpha_j}H_t^{-1}\frac{\partial H_t}{\partial\alpha_i}H_t^{-1}\right.\right. \\
&\qquad \left. + H_t^{-1}\left(\frac{\partial^2 H_t}{\partial\alpha_i\alpha_j}H_t^{-1} - \frac{\partial H_t}{\partial\alpha_i}H_t^{-1}\frac{\partial H_t}{\partial\alpha_j}H_t^{-1}\right)\right] \\
&\qquad \left. - \left(-H_t^{-1}\frac{\partial H_t}{\partial\alpha_j}H_t^{-1}\frac{\partial H_t}{\partial\alpha_i} + H_t^{-1}\frac{\partial^2 H_t}{\partial\alpha_i\alpha_j}\right)\right\} \\
&= \sum_{t=1+\bar{k}}^{n} \frac{1}{2}\mathrm{Tr}\left\{x_t x_t' \left[-H_t^{-1}\frac{\partial H_t}{\partial\alpha_j}H_t^{-1}\frac{\partial H_t}{\partial\alpha_i}H_t^{-1}\right.\right. \\
&\qquad \left. + H_t^{-1}\frac{\partial^2 H_t}{\partial\alpha_i\alpha_j}H_t^{-1} - H_t^{-1}\frac{\partial H_t}{\partial\alpha_i}H_t^{-1}\frac{\partial H_t}{\partial\alpha_j}H_t^{-1}\right] \\
&\qquad \left. + H_t^{-1}\frac{\partial H_t}{\partial\alpha_j}H_t^{-1}\frac{\partial H_t}{\partial\alpha_i} - H_t^{-1}\frac{\partial^2 H_t}{\partial\alpha_i\alpha_j}\right\}. \qquad (2.38)
\end{aligned}
$$

2.4.2 Consistência do estimador de ordem

Abaixo definimos a classe de estimadores EDC para processos BEKK-GARCH considerando $N = 1$ e θ como as colunas concatenadas das matrizes $\{A_l\}_{l=1\cdots q}$ e $\{B_l\}_{l=1\cdots p}$. Esse resultado pode ser generalizado para qualquer N, observando que devem ser consideradas condições para garantir que o processo seja identificável, que por Jeantheau (1998) H_t deve ser injetiva quase certamente. De forma semelhante, é possível generalizar para matrizes $A_l(\theta)$ e $B_l(\theta)$ com condições sobre as derivadas dessas matrizes em relação a θ. Esses dois casos de generalização são objetos de trabalhos futuros.

Definição 2.37. *Para $K \in \mathbb{N}^2$, $r = (p,q) \in \mathbb{N}^2$, $k = (k_1,k_2) \in \mathbb{N}^2$, $\mathbb{X}$ um BEKK-GARCH(r) e $K \geq r$, definimos o estimador* EDC *de r por*

$$
\hat{r}_{edc} = \underset{k\leq K}{\operatorname{argmin}} \left\{-\log L_{n,k}(\hat{\theta}_k) + [m^2(1+k_1+k_2)]c_n\right\}
$$

para $\log L_{n,k}(\hat{\theta}_k)$ como definida em (2.36) *e $\{c_n\}_{n\in\mathbb{N}}$ uma sequência de números positivos.*

Os dois próximos resultados tratam da imersão dos modelos BEKK-GARCH em cadeias de Markov, que podem ser encontrados em Boussama (1998) ou Boussama, Fuchs & Stelzer (2011).

Definição 2.38. *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um BEKK-GARCH(r) e $k \geq r$. A cadeia de Markov k-derivada de $\mathbb{X}$ é o processo $\mathbb{Y} = \{Y_t\}_{t\in\mathbb{N}}$ com valores em $\mathbb{R}^{\zeta(k)}$,*

$$Y_t = (vech(H_{t+1})', vech(H_t)', \ldots, vech(H_{t-k_1+2})', X_t', X_{t-1}', \ldots, X_{t-k_2+1}')' \tag{2.39}$$

onde

$$\zeta(k) = \frac{m(m+1)(k_1-2)}{2} + m(k_2-1).$$

Teorema 2.39 (Boussama (1998))**.** *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um BEKK-GARCH(r), $k \geq r$, satisfazendo a Condição 2.7, então $\mathbb{Y}$, a cadeia de Markov k-derivada de $\mathbb{X}$, é Harris positiva e geometricamente ergódica.*

As condições abaixo são propostas por Comte & Lieberman (2003) e foram baseadas nas condições propostas por Jeantheau (1998) para o estabelecimento da consistência forte dos estimadores. A única exceção é o item (v), que em Comte & Lieberman são exigidos momentos finitos de ordem 8 e nessa tese é necessário a finitude de momentos de ordem 16 para a demonstração do Lema 2.43.

Condição 2.8. *Para $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um BEKK-GARCH(r), θ_r seu respectivo parâmetro e Θ_r seu conjunto de parâmetros.*

(i) Θ_r *é compacto.*

(ii) θ_r *é ponto interior de* Θ_r.

(iii) Existe $c > 0$ *tal que*

$$\inf_{\theta\in\Theta_r} \det(C(\theta)) \geq c.$$

(iv) $H_t(\theta) = H_t(\theta')$ *q.c. se e somente se* $\theta = \theta'$.

(v) X_t *admite momentos finitos de ordem 16.*

O próximo Teorema sumariza os principais resultados de Comte & Lieberman (2003) que são utilizados a seguir.

Teorema 2.40 (Comte & Lieberman (2003))**.** *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um BEKK-GARCH(r), $k \geq r$, $\theta_r = (\alpha_1, \ldots, \alpha_{\gamma(k)})$ seu parâmetro verdadeiro, $\hat{\theta}_k = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{\gamma(k)})$ o estimador de máxima verossimilhança de θ_r, em que as Condições 2.7 e 2.8 são satisfeitas. Então é verdade que*

(i)

$$\lim_{n\to\infty} -\frac{D^2_\theta \log L_{n,k}(\theta_r)}{n} = A_2 \quad q.c.,$$

onde

$$A_2 = -E\left(\frac{\partial^2 l_t(\theta_r)}{\partial\theta\partial\theta'}\right). \tag{2.40}$$

(ii) A_2 é positiva definida.

(iii)

$$\frac{D^1_\theta \log L_n(\theta_r, k)}{\sqrt{n}} \underset{D}{\to} \mathcal{N}(0, A_1),$$

onde

$$A_1 = E\left(\frac{\partial l_t(\theta_r)}{\partial\theta}\frac{\partial l_t(\theta_r)}{\partial\theta'}\right),$$

é finita e não depende de t.

(iv) Para todo $i, j, l \in \{1, \ldots, \gamma(k)\}$,

$$E\left(\sup_{\|\theta-\theta_r\|\leq\delta}\left|\frac{\partial^3 l_t(\theta)}{\partial\alpha_i\alpha_j\alpha_l}\right|\right) < c(\delta).$$

(v) Para todo $i \in \{1, \ldots, \gamma(k)\}$, $\frac{\partial \log L_{n,k}(\theta_r)}{\partial\alpha_i}$ é martingale com segundo momento finito.

(vi)

$$\sqrt{n}(\hat{\theta}_k - \theta_r) \underset{D}{\to} \mathcal{N}(0, 2A_2^{-1}).$$

(vii) O estimador $\hat{\theta}_k$ é fortemente consistente.

(viii) Existe $c_1 \in (0, \infty)$, que não depende de t ou θ, tal que

$$\left\|H_t^{-1}\right\| \leq c_1.$$

(ix)

$$E\left(\left|\log\left(det(H_t(\theta_r))\right)\right|\right) < \infty.$$

No Teorema 2.41 e no Lema 2.42, alguns resultados de Comte & Lieberman (2003) são ajustados para serem utilizados no que segue.

Teorema 2.41. *Seja* $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ *um BEKK-GARCH(r),* $k \geq r$*,* $\theta_r = (\alpha_1, \ldots, \alpha_{\gamma(k)})$ *seu parâmetro verdadeiro,* $\hat{\theta}_k = (\hat{\alpha}_1, \ldots, \hat{\alpha}_{\gamma(k)})$ *o estimador de máxima verossimilhança de* θ_r*,* $\log L_{n,k}$ *como definido em* (2.36)*,* $\hat{\theta}_k \in \Theta_k$ *satisfazendo as Condições 2.7 e 2.8,* $\dot{\theta} = s\theta_r + (1-s)\hat{\theta}_k$ *e* $s \in [0,1]$ *e* $B_\delta(\theta_r) \subset \Theta_k$ *uma vizinhança de* θ_r*, então*

(i) Existe $c \in (0,\infty)$*, tal que, para todo* $i, j, l \in \{1, \ldots, \gamma(k)\}$*,*

$$\limsup_{n\to\infty} \left\| \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta\in B_\delta(\theta_r)} \frac{\partial^3 l_t(\theta)}{\partial\alpha_i\alpha_j\alpha_l} \right\| \leq c.$$

(ii)

$$\lim_{n\to\infty} -\frac{D^2_\theta \log L_n(\dot{\theta})}{n} = A_2 \quad q.c.$$

para A_2 *como definido em* (2.40).

(iii)

$$E\left(\left|\log\left[det(H_t(\theta_r))\right] + X_t'H_t^{-1}X_t\right|\right) < \infty.$$

Demonstração. (i) Usando o item (iv) do Teorema 2.40 e o Teorema 2.39, basta aplicar o Teorema 2.5.

(ii) De forma análoga à utilizada no Lema 5 de Hafner & Preminger (2009*a*), como $D^2_\theta l_t(\theta)$ e $D^3_\theta l_t(\theta)$ são contínuas em relação a θ e $\hat{\theta}_k$ é fortemente consistente, então, pelo Teorema do valor médio

$$\left\| \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_t(\dot{\theta})}{\partial\alpha_i\alpha_j} - \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 l_t(\theta_r)}{\partial\alpha_i\alpha_j} \right\| \leq \sup_{\theta\in B_\delta(\theta_r)} \left\{ \left\| \frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial\theta'}\left(\frac{\partial^2 l_t(\theta)}{\partial\alpha_i\alpha_j}\right) \right\| \cdot \left\|\dot{\theta} - \theta_r\right\| \right\}.$$

Usando o item (i) e a consistência forte de $\hat{\theta}$ obtemos o resultado.

(iii)

$$\begin{aligned}
E\left(\left|\log\left[det(H_t(\theta_r))\right] + X_t'H_t^{-1}X_t\right|\right) &\leq E\left(\left|\log\left[det(H_t(\theta_r))\right]\right|\right) + E\left(\left|X_t'H_t^{-1}X_t\right|\right) \\
&\leq E\left(\left|\log\left[det(H_t(\theta_r))\right]\right|\right) + E\left(\|X_t\|^2\right) E\left(\left\|H_t^{-1}\right\|\right)
\end{aligned}$$

que é finito usando o Teorema 2.40, itens (viii) e (ix), e a Condição 2.8.

□

Lema 2.42. *Seja* $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ *um BEKK-GARCH(k) em que as Condições 2.7 e 2.8 são satisfeitas. Então existe* $c_2 \in (0,\infty)$*, que não depende de* t *ou* i*, tal que*

$$E\left(\sup_{\theta\in\Theta_k}\left[\left\|\frac{\partial H_t}{\partial\alpha_i}(\theta)\right\|^8\right]\right) < c_2.$$

Demonstração. Pela demonstração do Lema A.2 de Comte & Lieberman (2003), temos que $D^1_\theta H_t$ admite momentos finitos de ordem p se X_t admite momentos finitos de ordem $2p$. Portanto, usando a Condição 2.8, temos o desejado. □

Como ocorre para o caso de modelos ARCH, não foi possível satisfazer as Condições 2.1 ou 2.2 para modelos BEKK-GARCH de forma suficiente para a aplicação da Lei do Logaritmo Iterado (LLI) apresentado no Teorema 2.5. Dessa forma, foi necessário a utilização da LLI para Martingales apresentada no Teorema 2.6. A demonstração segue de forma análoga à do Lema 2.28. As diferenças significativas estão na utilização de cálculo matricial e no uso do Lema 2.42, basicamente, em substituição ao item (ii) da Condição 2.6.

Lema 2.43. *Seja* $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ *um BEKK-GARCH(r),* $k \geq r$*,* $\theta_r = (\alpha_1,\ldots,\alpha_{\gamma(k)})$ *seu parâmetro verdadeiro, em que as Condições 2.7 e 2.8 são satisfeitas, então, para todo* $i \in \{1,\ldots,\gamma(k)\}$*,*

$$\limsup_{n\to\infty}\frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial\alpha_i}}{\sqrt{2n\log\log n}} = \left[E\left(\frac{\partial l_1(\theta_r)}{\partial\alpha_i}^2\right)\right]^{1/2} \quad q.c.,$$

$$\liminf_{n\to\infty}\frac{\frac{\partial \log L_{n,k}(\theta_r)}{\partial\alpha_i}}{\sqrt{2n\log\log n}} = -\left[E\left(\frac{\partial l_1(\theta_r)}{\partial\alpha_i}^2\right)\right]^{1/2} \quad q.c.$$

e $E\left(\frac{\partial l_1(\theta_r)}{\partial\alpha_i}^2\right)$ *é finito.*

Demonstração. Usando o item (v) do Teorema 2.40 e assumindo $\mathscr{F}_{t-1} = \sigma(X_1,\ldots,X_t)$, $Z_t = t^\delta$, $\delta > 1$,

$$U_t = \frac{\partial l_t(\theta_r)}{\partial\alpha_i} \quad e \quad W_n = \left[nE\left(\frac{\partial l_t(\theta_r)}{\partial\alpha_i}^2\right)\right]^{1/2}.$$

Onde, por (2.37),

$$\frac{\partial l_t(\theta_r)}{\partial \alpha_i} = \frac{1}{2}\mathrm{Tr}\left(X_t X_t' H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}H_t^{-1} - H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}\right).$$

Para aplicar o Teorema 2.6, basta demonstrar as hipóteses (L1-L4) que seguem.

(L1) Pela desigualdade de Chebyshev, temos que

$$\begin{aligned} P(|U_t| > Z_t) &= P\left(\left|\frac{\partial l_t(\theta_r)}{\partial \alpha_i}\right| > t^{\delta}\right) \\ &\leq \frac{1}{t^{2\delta}}E\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right) \end{aligned} \tag{2.41}$$

Usando o item (iii) do Teorema 2.40,

$$\sum_{t=1}^{\infty} P(|U_t| > Z_t) \leq E\left(\frac{\partial l_1(\theta_r)}{\partial \alpha_i}^2\right)\sum_{t=1}^{\infty}\frac{1}{t^{2\delta}} < \infty.$$

Pelo Lema de Borel-Cantelli, temos que

$$P\left(\left\{\omega : \mathbb{I}(|U_t| > t^{\delta}) = 1 \ \ infinitas \ \ vezes\right\}\right) = P\left(\left\{\omega : |U_t| > t^{\delta} \ \ infinitas \ \ vezes\right\}\right) = 0$$

e portanto

$$\lim_{n\to\infty}\frac{\sum_{t=1}^{n} U_t\mathbb{I}(|U_t| > Z_t) - E[U_t\mathbb{I}(|U_t| > Z_t)|\mathscr{F}_{t-1}]}{\sqrt{2W_n^2\log\log W_n^2}} = 0 \quad q.c.$$

(L2)

$$\begin{aligned} E(U_t|\mathscr{F}_{t-1}) &= \frac{1}{2}\mathrm{Tr}\left[E(X_t X_t'|\mathscr{F}_{t-1})H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}H_t^{-1} - H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}\right] \\ &= \frac{1}{2}\mathrm{Tr}\left[H_t H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}H_t^{-1} - H_t^{-1}\frac{\partial H_t}{\partial \alpha_i}\right] \\ &= 0. \end{aligned}$$

Além disso, pelo item (v) do Teorema 2.40,

$$E(E(U_t^2|\mathscr{F}_{t-1})) = E(U_t^2) < \infty$$

e portanto, usando o item (i) do Teorema 2.5 (LFGN), temos que

$$\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} E[U_t^2|\mathscr{F}_{t-1}]}{nE\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right)} = 1 \quad q.c.$$

Usando o Teorema da Convergência Dominada, temos que

$$\lim_{t\to\infty} E(U_t \mathbb{I}(|U_t| \leq n)|\mathscr{F}_{t-1}) = 0 \quad q.c.$$

Agora, considerando $\varepsilon > 0$ arbitrário, é necessário encontrar uma cota superior somável em t para

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

e aplicar Borel-Cantelli para obter

$$\lim_{t\to\infty} \left[E(U_t^2 \mathbb{I}(|U_t| \leq t^{\delta})|\mathscr{F}_{t-1}) - E(U_t^2|\mathscr{F}_{t-1})\right] = 0 \quad q.c.$$

e aplicar o Teorema Médio de Cesàro (Apêndice A) para concluir

$$\begin{aligned}
\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} E[U_t^2 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] - E[U_t \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}]^2}{W_n^2} &= \lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} E[U_t^2|\mathscr{F}_{t-1}]}{nE\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right)} \\
&= 1 \quad q.c.
\end{aligned}$$

Usando a desigualdade generalizada de Chebyshev

$$\begin{aligned}
P\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^{\delta})|\mathscr{F}_{t-1})\right| > \varepsilon\right] \\
\leq \frac{1}{\varepsilon} E\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^{\delta})|\mathscr{F}_{t-1})\right|\right]
\end{aligned}$$

e

$$\begin{aligned}
E\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^{\delta})|\mathscr{F}_{t-1})\right|\right] &= E\left[\left|E(U_t^2|\mathscr{F}_{t-1}) - E(U_t^2 I(|U_t| \leq t^{\delta})|\mathscr{F}_{t-1}) \pm E(U_t^2 I(|U_t| > t^{\delta})|\mathscr{F}_{t-1})\right|\right] \\
&= E\left[E(U_t^2 I(|U_t| > t^{\delta})|\mathscr{F}_{t-1})\right] \\
&= E\left[U_t^2 I(|U_t| > t^{\delta})\right] \\
&\leq E\left[U_t^4\right]^{1/2} E\left[I(|U_t| > t^{\delta})\right]^{1/2} \\
&= E\left[U_t^4\right]^{1/2} \mathrm{P}(|U_t| > t^{\delta})^{1/2}.
\end{aligned}$$

Usando (2.41), (2.45) e o Lema 2.42 para $c > 0$ apropriado, obtemos

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

que é somável em t.

(L3) Adotando a notação $\dot{H}_t := \frac{\partial H_t}{\partial \alpha_i}$, temos que

$$\begin{aligned} 0 &\leq E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] \\ &\leq E[U_t^4|\mathscr{F}_{t-1}] \\ &\leq E\left\{\mathrm{Tr}\left[X_t X_t' H_t^{-1}\dot{H}_t H_t^{-1} - H_t^{-1}\dot{H}_t\right]^4 |\mathscr{F}_{t-1}\right\} \\ &= E\left\{\left[\mathrm{Tr}\left(X_t X_t' H_t^{-1}\dot{H}_t H_t^{-1}\right) - \mathrm{Tr}\left(H_t^{-1}\dot{H}_t\right)\right]^4 |\mathscr{F}_{t-1}\right\} \\ &= E\left\{\mathrm{Tr}\left(X_t X_t' H_t^{-1}\dot{H}_t H_t^{-1}\right)^4\right. \\ &\quad -4\mathrm{Tr}\left(X_t X_t' H_t^{-1}\dot{H}_t H_t^{-1}\right)^3 \mathrm{Tr}\left(H_t^{-1}\dot{H}_t\right) \\ &\quad +6\mathrm{Tr}\left(X_t X_t' H_t^{-1}\dot{H}_t H_t^{-1}\right)^2 \mathrm{Tr}\left(H_t^{-1}\dot{H}_t\right)^2 \\ &\quad -4\mathrm{Tr}\left(X_t X_t' H_t^{-1}\dot{H}_t H_t^{-1}\right) \mathrm{Tr}\left(H_t^{-1}\dot{H}_t\right)^3 \\ &\quad \left. +\mathrm{Tr}\left(H_t^{-1}\dot{H}_t\right)^4 |\mathscr{F}_{t-1}\right\} \end{aligned} \tag{2.42}$$

Adicionalmente,

$$\begin{aligned} \left|\mathrm{Tr}\left(X_t X_t' H_t^{-1}\dot{H}_t H_t^{-1}\right)\right| &= \left|\mathrm{Tr}\left(H_t^{1/2}\varepsilon_t \left(H_t^{1/2}\varepsilon_t\right)' H_t^{-1}\dot{H}_t H_t^{-1}\right)\right| \\ &= \left|\mathrm{Tr}\left(H_t^{1/2}\varepsilon_t\varepsilon_t' H_t^{1/2} H_t^{-1}\dot{H}_t H_t^{-1}\right)\right| \\ &= \left|\mathrm{Tr}\left(\varepsilon_t\varepsilon_t' H_t^{-1/2}\dot{H}_t H_t^{-1/2}\right)\right| \\ &\leq \left\|\varepsilon_t\varepsilon_t'\right\| \left\|H_t^{-1/2}\right\|^2 \left\|\dot{H}_t\right\| \end{aligned} \tag{2.43}$$

e, usando o Teorema 2.40 (viii) para $c \in (0,\infty)$ apropriado, temos

$$\left|\mathrm{Tr}\left(H_t^{-1/2}\dot{H}_t H_t^{-1/2}\right)\right| \leq c\left\|\dot{H}_t\right\|. \tag{2.44}$$

Onde usamos a relação $|\mathrm{Tr}(ABC)| \leq \|A\| \, \|B\| \, \|C\|$. Portanto,

$$\begin{aligned} E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] &\leq E\left(\left\|\varepsilon_t \varepsilon_t'\right\|^4 |\mathscr{F}_{t-1}\right) \left\|\dot{H}_t\right\|^4 \\ &\quad +4cE\left(\left\|\varepsilon_t \varepsilon_t'\right\|^3 |\mathscr{F}_{t-1}\right) \left\|\dot{H}_t\right\|^4 \\ &\quad +6c^2 E\left(\left\|\varepsilon_t \varepsilon_t'\right\|^2 |\mathscr{F}_{t-1}\right) \left\|\dot{H}_t\right\|^4 \\ &\quad +3c^4 \left\|\dot{H}_t\right\|^4 \\ &\leq c_1 E\left[\left\|\varepsilon_t \varepsilon_t'\right\|^2 + \left\|\varepsilon_t \varepsilon_t'\right\|^3 + \left\|\varepsilon_t \varepsilon_t'\right\|^4 + 1|\mathscr{F}_{t-1}\right] \left\|\dot{H}_t\right\|^4 \end{aligned} \tag{2.45}$$

para $\delta_1 \in (\frac{1}{2}, 1)$, usando as desigualdades de Chebyshev e Jensen,

$$\begin{aligned} \mathrm{P}\left[E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] > t^{\delta_1}\right] &\leq \mathrm{P}\left[E\left[\left\|\varepsilon_t \varepsilon_t'\right\|^2 + \left\|\varepsilon_t \varepsilon_t'\right\|^3 + \left\|\varepsilon_t \varepsilon_t'\right\|^4 + 1|\mathscr{F}_{t-1}\right] \left\|\dot{H}_t\right\|^4 > \frac{t^{\delta_1}}{c_1}\right] \\ &\leq \frac{c_2^2}{t^{2\delta_1}} E\left\{E\left[\left(\left\|\varepsilon_t \varepsilon_t'\right\|^2 + \left\|\varepsilon_t \varepsilon_t'\right\|^3 + \left\|\varepsilon_t \varepsilon_t'\right\|^4 + 1\right) \left\|\dot{H}_t\right\|^4 |\mathscr{F}_{t-1}\right]^2\right\} \\ &\leq \frac{c_2^2}{t^{2\delta_1}} E\left\{E\left[\left(\left\|\varepsilon_t \varepsilon_t'\right\|^2 + \left\|\varepsilon_t \varepsilon_t'\right\|^3 + \left\|\varepsilon_t \varepsilon_t'\right\|^4 + 1\right)^2 \left\|\dot{H}_t\right\|^8 |\mathscr{F}_{t-1}\right]\right\} \\ &\leq \frac{c_2^2}{t^{2\delta_1}} E\left\{\left(\left\|\varepsilon_t \varepsilon_t'\right\|^2 + \left\|\varepsilon_t \varepsilon_t'\right\|^3 + \left\|\varepsilon_t \varepsilon_t'\right\|^4 + 1\right)^2\right\} E\left\{\left\|\dot{H}_t\right\|^8\right\} \\ &\leq \frac{c_3}{t^{2\delta_1}} E\left[\left\|\dot{H}_t\right\|^8\right] \end{aligned}$$

e portanto, usando o Lema 2.42,

$$\sum_{t=1}^{\infty} \mathrm{P}\left[E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}] > t^{\delta_1}\right] \leq \sum_{t=1}^{\infty} \frac{c_3}{t^{2\delta_1}} E\left[\left\|\dot{H}_t\right\|^8\right] < \infty.$$

Usando o Lema de Borel-Cantelli,

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

e logo

$$\lim_{n\to\infty} \sum_{t=1}^{n} \frac{E[U_t^4 \mathbb{I}(|U_t| \leq Z_t)|\mathscr{F}_{t-1}]}{W_t^4} \leq \lim_{n\to\infty} \sum_{t=1}^{n} \frac{1}{t^{2-\delta_1}} < \infty \quad q.c.$$

(L4) Como $E\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right) > 0$, caso contrário $\frac{\partial l_t(\theta_r)}{\partial \alpha_i} \equiv 0$ q.c. e A_2 seria 0, então, usando a estacionaridade, temos

$$\lim_{n\to\infty}\frac{W_n}{W_{n+1}} = \lim_{n\to\infty}\frac{\left[nE\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right)\right]^{1/2}}{\left[(n+1)E\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right)\right]^{1/2}} = 1 \quad q.c. \;\; e$$

$$\lim_{n\to\infty} W_n = E\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right)^{1/2}\lim_{n\to\infty}\sqrt{n} = \infty.$$

□

Corolário 2.44. *Seja* $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ *um BEKK-GARCH(r) em que as Condições 2.7 e 2.8 são satisfeitas,* $k \geq r$*,* $\log L_n(\theta)$ *como definido em* (2.36)*, então existe* $c_5 \in (0,\infty)$ *tal que*

$$\limsup_{n\to\infty}\frac{\left\|D_\theta^1 \log L_{n,k}(\theta_r)\right\|}{\sqrt{2\log\log n}} \leq c_5 \quad q.c.$$

Demonstração.

$$\begin{aligned}
\limsup_{n\to\infty}\frac{\left\|D_\theta^1 \log L_{n,k}(\theta_r)\right\|}{\sqrt{2\log\log n}} &\leq \limsup_{n\to\infty}\sum_{i=0}^{\gamma(k)}\frac{\left|\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}\right|}{\sqrt{2n\log\log n}} \\
&\leq \sum_{i=0}^{\gamma(k)}\limsup_{n\to\infty}\frac{\left|\frac{\partial \log L_{n,k}(\theta_r)}{\partial \alpha_i}\right|}{\sqrt{2n\log\log n}} \\
&\leq \sum_{i=0}^{\gamma(k)} E\left(\frac{\partial l_t(\theta_r)}{\partial \alpha_i}^2\right)^{1/2} \\
&= c_5 \quad q.c.
\end{aligned}$$

□

O Lema abaixo possui resultados técnicos que são utilizados na demonstração do Teorema 2.46, que conclui a hipótese H1.

Lema 2.45. *Seja* $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ *um BEKK-GARCH(k),* $i \in \{0,\ldots,\gamma(k)\}$*, em que as Condições 2.7 e 2.8 são satisfeitas, então, adotando a notação* $\dot{H}_t := D_\theta H_t$*, temos que*

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

Demonstração. Usando o Teorema 2.40 e o Lema 2.42 temos, para uma constante c,

$$\begin{aligned}
E\sup_{\theta\in\Theta_k}\left[\left|\mathrm{Tr}(\dot{H}_tH_t^{-1}-X_tX_t'H_t^{-1}\dot{H}_tH_t^{-1})\right|\right] &\leq E\sup_{\theta\in\Theta_k}\left[\left\|\dot{H}_t\right\|\left\|H_t^{-1}\right\|+\left\|X_tX_t'\right\|\left\|H_t^{-1}\right\|^2\left\|\dot{H}_t\right\|\right]\\
&\leq E\sup_{\theta\in\Theta_k}\left[c\left\|\dot{H}_t\right\|+c^2\left\|X_tX_t'\right\|\left\|\dot{H}_t\right\|\right]<\infty.
\end{aligned}$$

□

Teorema 2.46. *Seja $\mathbb{X}=\{X_t\}_{t\in\mathbb{N}}$ um BEKK-GARCH(r), $k\ngeq r$, θ_r seu parâmetro verdadeiro, $\hat{\theta}_k$ o estimador de máxima verossimilhança de θ_r, em que as Condições 2.7 e 2.8 são satisfeitas, então*

$$\lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_r)-\log L_n(\hat{\theta}_k)}{n}>0\ q.c.$$

Demonstração. Tomando $\bar{k}\geq k,r$, temos que

$$\lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_r)-\log L_n(\hat{\theta}_k)}{n} = \lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_r)-\log L_n(\hat{\theta}_{\bar{k}})+\log L_n(\hat{\theta}_{\bar{k}})-\log L_n(\hat{\theta}_k)}{n}.$$

Aplicando o Teorema 1.11, usando os resultados acima, temos que

$$\lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_r)-\log L_n(\hat{\theta}_{\bar{k}})}{n}=0\quad q.c.$$

Observando (2.35), vemos que basta mostrar que

$$\lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_{\bar{k}})-\log L_n(\hat{\theta}_k)}{n}>0\quad q.c.$$

considerando

$$\log L_n(\theta)=\sum_{t=1+\bar{k}}^{n}l_t(\theta)$$

e

$$l_t(\theta)=\log(f(\theta))=\log\left(\frac{1}{\sqrt{(2\pi)^{m/2}\det(H_t)}}exp\left(-\frac{1}{2}x_t'H_t^{-1}x_t\right)\right)$$

em (2.35). Assim, temos que pelo Teorema 2.41, item (iii), que

$$E\left[\left|l_t(\theta_r)\right|\right]<\infty$$

e portanto usando o Teorema 2.5,

$$\lim_{n\to\infty}\frac{\log L_n(\theta_r)}{n}=\lim_{n\to\infty}\frac{\sum_{t=1+k}^{n} l_t(\theta_r)}{n}=E\left(l_1(\theta_r)\right)=c_1<\infty \quad q.c.$$

Usando o Teorema do valor médio, para $\dot{\theta}=s\theta_r+(1-s)\hat{\theta}_{\bar{k}}$, $s\in(0,1)$, n suficientemente grande e $B_\delta(\theta_r)$ uma vizinhança suficientemente pequena de θ_r, temos

$$\begin{aligned}\left|\frac{\sum_{t=1+\bar{k}}^{n} l_t(\hat{\theta}_{\bar{k}})}{n}-\frac{\sum_{t=1+\bar{k}}^{n} l_t(\theta_r)}{n}\right| &= \left|\frac{\sum_{t=1+\bar{k}}^{n} D_\theta^1 l_t(\dot{\theta})}{n}(\hat{\theta}_{\bar{k}}-\theta_r)\right|\\ &\leq \sup_{\theta\in B_\delta(\theta_r)}\left\|\frac{\sum_{t=1+\bar{k}}^{n} D_\theta^1 l_t(\dot{\theta})}{n}\right\|\left\|(\hat{\theta}_{\bar{k}}-\theta_r)\right\|\end{aligned}$$

Aplicando o Teorema 2.5 e usando o Lema 2.45 e a consistência forte de $\hat{\theta}_r$,

$$\left|\frac{\sum_{t=1+\bar{k}}^{n} l_t(\hat{\theta}_{\bar{k}})}{n}-\frac{\sum_{t=1+\bar{k}}^{n} l_t(\theta_r)}{n}\right| \underset{q.c.}{\rightarrow} 0. \tag{2.46}$$

Por outro lado, temos que, como o estimador é de máxima verossimilhança e $\Theta_k\subset\Theta_{\bar{k}}$,

$$\begin{aligned}\lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_k)}{n} &\leq \lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_{\bar{k}})}{n}\\ &= c_1 \quad q.c.\end{aligned}$$

e portanto,

$$\lim_{n\to\infty}\frac{\log L_n(\hat{\theta}_k)}{n} \leq \limsup_{n\to\infty}\frac{\sum_{t=1+k}^{n} l_t(\hat{\theta}_k)}{n}=c_2\leq c_1 \quad q.c.$$

Seja n_i uma subsequência de n tal que

$$\lim_{n\to\infty}\frac{\sum_{t=1+k}^{n_i} l_t(\hat{\theta}_k)}{n_i} = c_2 \quad q.c.$$

Usando que Θ_k é compacto, tome n_j uma subsequência de n_i tal que

$$\hat{\theta}_k(n_j) \to \bar{\theta}_k \in \Theta_k \quad q.c.$$

Temos assim que

$$\begin{aligned}\lim_{n\to\infty} \frac{\log L_n(\hat{\theta}_k)}{n} &\leq \limsup_{n\to\infty} \frac{\log L_n(\hat{\theta}_k)}{n} \\ &= \lim_{n_j\to\infty} \frac{\log L_{n_j}(\hat{\theta}_k(n_j))}{n_j} \quad q.c.\end{aligned}$$

Aplicando o mesmo raciocínio usado em (2.46), concluímos

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

$$\left| \frac{\sum\limits_{t=1+\bar{k}}^{n} l_t(\hat{\theta}_k)}{n} - \frac{\sum\limits_{t=1+\bar{k}}^{n} l_t(\bar{\theta})}{n} \right| \underset{q.c.}{\to} 0.$$

Além disso,

$$\lim_{n\to\infty} \left[\frac{\log L_n(\hat{\theta}_{\bar{k}})}{n} - \frac{\log L_n(\hat{\theta}_k)}{n} \right] \geq E\left[\log\left(\frac{f(\bar{\theta}_k)}{f(\theta_r)}\right)\right]$$

Por outro lado

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

é a divergência de Kullback-Leibler, que é positiva se $f(\bar{\theta}_k) \neq f(\theta_r)$, e como $\theta_r \notin \bar{\Theta}_k \subseteq \mathbb{R}^{\gamma(k)}$, temos que $\theta_r \neq \bar{\theta}_k$ e então $f(\bar{\theta}_k) \neq f(\theta_r)$. Onde concluímos

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

$\square$

No Teorema abaixo aplicamos os resultados desenvolvidos nessa Seção para o estabelecimento da consistência forte dos estimadores de ordem EDC para modelos BEKK-

GARCH. Como caso particular, temos a consistência forte do estimador baseado no critério de informação BIC. Nota-se que, caso seja possível encontrar cotas superiores menores para H2, essa classe pode ser estendida para termos de penalidade menores, da ordem de $O(\log\log n)$.

Teorema 2.47. *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um BEKK-GARCH(r) em que as Condições 2.7 e 2.8 são satisfeitas e $\hat{r}_{edc}$ como na Definição 2.37. Então $\hat{r}_{edc}$ é fortemente consistente ($\hat{r}_{edc} \underset{q.c.}{\to} r$) se*

$$\lim_{n\to\infty} \frac{c_n}{n} = 0 \quad e \quad \liminf_{n\to\infty} \frac{c_n}{\log\log n} = \infty. \tag{2.47}$$

Demonstração. O conjunto dos possíveis valores de θ_r, Θ_r, definido em (2.34), é aberto e $\theta_r \in \Theta_r$. De Boussama (1998), temos que o estimador de máxima verossimilhança $\hat{\theta}_k$ é fortemente consistente para $k \geq r$, isto é, vale (1.10). As derivadas de $\log L_{n,k}$, (2.37) e (2.38) são contínuas em relação a θ e a x_1^n e portanto são mensuráveis em relação a x_1^n. Usando o Teoremas 2.41 e o Corolário 2.44 temos as condições satisfeitas para a aplicação do item (ii) do Teorema 1.11 e com isso temos (H2) satisfeita. A hipótese (H1) segue do Teorema 2.46. O resultado segue do item (i) do Teorema 1.10. □

Corolário 2.48. *Seja $\mathbb{X} = \{X_t\}_{t\in\mathbb{N}}$ um BEKK-GARCH(r) em que as Condições 2.7 e 2.8 são satisfeitas e $\hat{r}_{edc}$ como na Definição 2.37. Então o estimador $\hat{r}_{bic}$, como definido abaixo, é fortemente consistente.*

$$\hat{r}_{bic} = \underset{k\leq K}{\operatorname{argmin}} \left\{ -\log L_n(\hat{\theta}_k) + \frac{[m^2(1+k_1+k_2)]}{2} \log n \right\}$$

para $\log L_n(\hat{\theta}_k)$ *como definida em* (2.35).

Demonstração. $c_n = \log n/2$ e satisfaz (2.47). □

Conclusão

O método de estimação de ordem baseado no critério de informação EDC, concebido originalmente para cadeias de Markov, se mostra bastante promissor no contexto de seleção de modelos.

Além dos resultados obtidos com a definição do estimador de ordem EDC para processos AR, ARCH e BEKK-GARCH, observa-se a possibilidade de aplicação do ferramental desenvolvido para estimação de tamanho de espaço de estados e ordem de dependência ocultos em cadeias de Markov Ocultas e seleção de modelos aninhados em modelos Lineares de Espaço de Estados.

Como casos particulares, foi demonstrado nesse trabalho a consistência forte dos estimadores de ordem baseados no critério de informação BIC para os modelos ARCH e BEKK-GARCH. Como ainda não existiam estudos sobre estimadores consistentes para esses casos, esses resultados se mostram altamente relevantes.

A generalização da função verossimilhança para classe de funções $\log L_{n,k}$, além de apresentar resultados práticos para a seleção de ordem utilizando o critério EDC, evidencia a possibilidade de utilização desse conceito no tratamento de outros problemas.

Para processos AR, o desenvolvimento de estimadores iniciado por Akaike (1969), com o procedimento de minimização do erro final de predição (FPE), seguido dos métodos baseados em critérios de informação (AIC, BIC e HQC), utilizavam essencialmente o mesmo radical, alterando apenas o termo de penalidade. Nessa tese, usando o ferramental desenvolvido para generalização do EDC, é proposto um novo estimador, que apresentou no geral melhor performance nas simulações numéricas realizadas.

Como trabalhos futuros, observou-se a possibilidade de existência de cotas superiores menores na hipótese H2, o que permitiria a extensão da classe de estimadores EDC fortemente consistentes para termos de penalidade assintoticamente menores.

Referências Bibliográficas

Akaike, H. 1969. “Fitting autogressive models for prediction.” *Annals of the Institute of Statistical Mathematics* 21(1):243–247.

Akaike, H. 1974. “A new look at the statistical model identification.” *Automatic Control, IEEE Transactions on* 19(6):716–723.

Akaike, H. 1979. “A Bayesian extension of the minimum AIC procedure of autoregressive model fitting.” *Biometrika* 66(2):237–242.

Anderson, T. W. 1994. *The Statistical Analysis of Time Series*. New York: Wiley.

Anderson, T. W. & L. A. Goodman. 1957. “Statistical Inference about Markov Chains.” *The Annals of Mathematical Statistics* 28(1):89–110.

Athreya, K. B. & P. Ney. 1978. “A new approach to the limit theory of recurrent Markov chains.” *Transactions of the American Mathematical Society* 245:493–501.

Athreya, K. B. & P. Ney. 1980. “Some aspects of ergodic theory and laws of large numbers for Harris recurrent Markov chains.” *Colloquia Mathematica Societatis János Bolyai. Nonparametric Statistical Inference* 32:41–56. Budapest, Hungary.

Baigorri, A. R., C. R. Gonçalves & P. A. A. Resende. 2014. “Markov chain order estimation based on the chi-square divergence.” *The Canadian Journal of Statistics* 42(4):563–578.

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

Bartlett, M. S. & D. V. Rajalakshman. 1953. “Goodness of fit tests for simultaneous autoregressive series.” *Journal of the Royal Statistical Society* 15(1):107–124.

Basawa, I. V. & C. C. Heyde. 1976. “Asymptotic properties of maximum likelihood estimators for stochastic processes.” *The Indian Journal of Statistics* 38(3):259–270.

Billingsley, P. 1961. “Statistical methods in Markov chains.” *The Annals of Mathematical Statistics* 32(1):12–40.

Bollerslev, T. 1986. “Generalized autoregressive conditional heteroskedasticity.” *Journal of Econometrics* 31(3):307–327.

Bollerslev, T. 1990. "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model." *Review of Economics and Statistics* 72(3):498–505.

Bollerslev, T., R. F. Engle & J. M. Wooldridge. 1988. "A capital asset pricing model with time-varying covariances." *Journal of Political Economy* 96(1):116–131.

Bollerslev, T., R. Y. Chou & K. F. Kroner. 1992. "ARCH modeling in finance: A review of the theory and empirical evidence." *Journal of Econometrics* 52(1-2):5–59.

Boussama, F. 1998. Ergodicité, mélange et estimation dans les modèles GARCH PhD thesis Université 7 Paris.

Boussama, Farid, Florian Fuchs & Robert Stelzer. 2011. "Stationarity and geometric ergodicity of BEKK multivariate GARCH models." *Stochastic Processes and their Applications* 121(10):2331 – 2360.

Carvalheira, J.G.V., R.W. Blake, E.J. Pollak, R.L. Quaas & C.V. Duran-Castro. 1998. "Application of an Autoregressive Process to Estimate Genetic Parameters and Breeding Values for Daily Milk Yield in a Tropical Herd of Lucerna Cattle and in United States Holstein Herds." *Journal of Dairy Science* 81(10):2738–2751.

Choi, B. 1992. *ARMA Model Identification*. New York: Springer.

Cogburn, R. 1972. The Central Limit Theorem for Markov processes. In *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press pp. 485–512.

Comte, F. & O. Lieberman. 2003. "Asymptotic theory for multivariate GARCH processes." *Journal of Multivariate Analysis* 84:61–84.

Csiszár, I. & P. C. Shields. 2000. "The Consistency of the BIC Markov Order Estimator." *The Annals of Statistics* 28(6):1601–1619.

Doeblin, W. 1937. "Sur les propriétés asymptotiques de mouvement régis par certain types de chaînes simples." *Bull. Math. Soc. Roum. Sci.* 39(1):57–115; (2), 3–61.

Doeblin, W. 1940. "Eléments d'une théorie générale des chaînes simples constantes de Markov." *Annales Scientifiques de l'Ecole Normale Supérieure* 57(III):61–111.

Domowitz, I. & C. S. Hakkio. 1985. "Conditional variance and the risk premium in the foreign exchange market." *Journal of International Economics* 19(1-2):47–66.

Doob, J. L. 1966. *Stochastic Processes*. John Wiley & Sons Inc.

Dorea, C. C. Y. 2008. "Optimal penalty term for EDC Markov chain order estimator." *Annales de l'Institut de Statistique de l'Universite de Paris (l'ISUP)* 52:15–26.

Engle, R. F. 1982. “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.” *Econometrica* 50(4):987–1007.

Engle, R. F. & K. F. Kroner. 1995. “Multivariate simultaneous generalized ARCH.” *Econometric Theory* 11(1):122–150.

Engle, R. F. & V. K. Ng. 1993. “Measuring and testing the impact of news on volatility.” *The Journal of Finance* 48(5):1749–1778.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications, Volume 1*. Wiley.

Finesso, L. 1990. Consistent Estimation of the Order for Markov and Hidden Markov Chains PhD thesis University of Maryland.

Foster, F. G. 1953. “On the stochastic matrices associated with certain queuing processes.” *The Annals of Mathematical Statistics* 24(3):355–360.

Francq, C. & J. M. Zakoian. 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*. New York: Wiley.

Good, I. J. 1955. “The likelihood ratio test for Markoff chains.” *Biometrika* 42(3/4):531–533.

Hafner, C. M. & A. Preminger. 2009*a*. “Asymptotic theory for a factor garch model.” *Econometric Theory* 25(2):336–363.

Hafner, C. M. & A. Preminger. 2009*b*. “On asymptotic theory for multivariate GARCH models.” *Journal of Multivariate Analysis* 100(9):2044–2054.

Hall, P. & C. C Heyde. 1980. *Martingale Limit Theory and its Application*. New York: Academic Press.

Hannan, E. J. 1980. “The estimation of the order of an ARMA process.” *The Annals of Statistics* 8(5):1071–1081.

Hannan, E. J. & B. G. Quinn. 1979. “The determination of the order of an autoregression.” *Journal of the Royal Statistical Society. Series B (Methodological)* 41(2):190–195.

Harris, T. E. 1956. The existence of stationary measures for certain Markov processes. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 2 University of California Press pp. 113–124.

Hoel, P. G. 1954. “A Test for Markoff Chains.” *Biometrika* 41(3/4):430–433.

Hughes, A. W., M. L. King & K. T. Kwek. 2004. “Selecting the order of an ARCH model.” *Economics Letters* 83(2):269–275.

Jeantheau, T. 1998. “Strong consistency of estimators for multivariate ARCH models.” *Econometric Theory* null(01):70–86.

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

Kolmogorov, A. N. 1936. "Anfangsgründe der Theorie der Markoffschen Ketten mit unendlichen vielen möglichen Zuständen." *Mat. Sbornik N.S. Ser* pp. 607–610.

Kullback, S. 1959. *Information theory and statistics*. New York: John Wiley and Sons.

Logan, J. A. 1981. "A structural model of the higher-order Markov process incorporating reversion effects." *Journal of Mathematical Sociology* 8:75–89.

Lopes, J. S. 2005. Determinação da Ordem de uma Cadeia de Markov Usando o Critério EDC PhD thesis Universidade de Brasília, UNB, Brasil.

Maddala, G. S. & K. Lahiri. 2009. *Introduction to Econometrics*. Chichester: Wiley.

Meyn, S. P. & R. L. Tweedie. 1993. *Markov Chains and Stochastic Stability*. London: Springer-Verlag.

Nelson, D. B. 1991. "Conditional heteroskedasticity in asset returns: a new approach." *Econometrica* 59(2):349–370.

Niemi, S. & E. Nummelin. 1982. "Central limit theorems for Markov random walks." *Commentationes Physico-Mathematicae* 54.

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

Nummelin, E. 1978. "A splitting technique for Harris recurrent chains." 43:309–318.

Nummelin, E. 1984. *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge: Cambridge University Press.

Ogata, Y. 1980. "Maximum likelihood estimates of incorrect Markov models for time series and the derivation of AIC." *Journal of Applied Probability* 17(1):59–72.

Orey, S. 1959. "Recurrent Markov chains." *Pacific Journal of Mathematics* 9(3):805–827.

Orey, S. 1971. *Limit theorems for Markov chain transition probabilities*. London: Van Nostrand Reinhold.

Ozaki, T. 1977. "On the order determination of ARIMA models." *Journal of Applied Statistics* 26(3):290–301.

Pegram, G. G. S. 1980. "An autoregressive model for multilag Markov chains." *Journal of Applied Probability* 17:350–362.

Polansky, A. M. 2007. "Detecting change-points in Markov chains." *Computational Statistics & Data Analysis* 51(12):6013–6026.

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

Raftery, A. E. 1985. "A model for high-order Markov chains." *J. R. Statist. Soc. B.* .

Resende, P. A. A. 2009. Análise comparativa de estimadores da ordem de cadeias de Markov. Master's thesis Universidade de Brasília.

Rosenblatt, M. 1964. "Equicontinuous Markov operators." *Teor. Verojatnost. i Primenen* 9:205–222.

Rosenblatt, M. 1974. "Recurrent points and transition functions acting on continuous functions." 30:173–183.

Schneider, H. 1988. Application of an autoregressive reflection model for the signal analysis of radar echoes from rotating objects. In *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*. pp. 1236–1239 vol.2.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6(2):461–464.

Shao, J. 2007. *Mathematical Statistics*. New York: Springer Verlag.

Shibata, R. 1976. "Selection of the Order of an Autoregressive model by Akaike's Information Criterion." *Biometrika* 63:117–126.

Stelzer, R. 2008. "On the relation between the vec and bekk multivariate garch models." *Econometric Theory* 24(04):1131–1136.

Tong, H. 1975. "Determination of the Order of a Markov Chain by Akaike's Information Criterion." *Journal of Applied Probability* 12(3):488–497.

Tweedie, R. L. 1974. "*R*-theory for Markov chains on a general state space I: solidarity properties and *R*-recurrent chains." 2:840–864.

Tweedie, R. L. 1975. "Sufficient conditions for regularity, recurrence and ergodicity of Markov processes." *Math. Proc. Camb. Phil. Soc.* 78:125–136.

Tweedie, R. L. 1976. "Criteria for classifying general Markov chains." 8:737–771.

van der Vaart, A. W. 2000. *Asymptotic Statistics*. New York: Cambridge University Press.

Weiss, A. A. 1986. "Asymptotic theory for ARCH models: estimation and testing." *Econometric Theory* 2(1):107–131.

Whittle, P. 1951. *Hypothesis testing in time-series analysis*. Almqvist and Wiksells.

Whittle, P. 1954. *Some recent contributions to the theory of stationary processes: A Study in the analysis of stationary time series.* Almqvist Wiksells.

Yule, G. U. 1921. "On the time-correlation problem, with especial reference to the variate-difference correlation method." *Journal of the Royal Statistical Society* 84(4):497–537.

Zhang, F. 2011. *Matrix Theory: Basic Results and Techniques*. New York: Springer.

Zhao, L., C. Dorea & C. Gonçalves. 2001. "On determination of the order of a Markov chain." *Statistical Inference for Stochastic Processes* 4(3):273–282.

APÊNDICE A – Teorema Médio de Cesàro

Teorema A.1 (Cesàro)**.** *Seja* $\{a_t\}_{t\in\mathbb{N}}$ *sequência de números reais tal que* $\lim_{t\to\infty} a_t = a$ *então*

$$\lim_{n\to\infty} \bar{a}_n = a$$

para $\bar{a}_n = \frac{1}{n}\sum_{t=1}^{n} a_t$.

Demonstração. Dado $\varepsilon > 0$, seja j tal que $t > j \Rightarrow |a_t - a| < \varepsilon$. Então

$$\begin{aligned}
\lim_{n\to\infty} \frac{\sum_{t=1}^{n} a_t}{n} &= \lim_{n\to\infty} \frac{\sum_{t=1}^{j} a_t}{n} + \frac{\sum_{t=j+1}^{n} (a_t + a - a)}{n} \\
&\leq a \lim_{n\to\infty} \frac{\sum_{t=j+1}^{n} 1}{n} + \frac{\sum_{t=j+1}^{n} |a_t - a|}{n} \\
&= a \lim_{n\to\infty} \frac{n-j}{n} + \varepsilon \\
&= a + \varepsilon.
\end{aligned}$$

Usamos que

$$\begin{aligned}
\left| \lim_{n\to\infty} \frac{\sum_{t=j+1}^{n} (a_t - a)}{n} \right| &\leq \lim_{n\to\infty} \frac{\sum_{t=j+1}^{n} |a_t - a|}{n} \\
&\leq \lim_{n\to\infty} \frac{\sum_{t=j+1}^{n} \varepsilon}{n} \\
&\leq \varepsilon \lim_{n\to\infty} \frac{n-j}{n} \\
&= \varepsilon.
\end{aligned}$$

Analogamente, demonstra-se que $\lim\limits_{n\to\infty} \frac{\sum\limits_{t=1}^{n} a_t}{n} \geq a - \varepsilon$. $\square$

Proposição A.2. *Seja* $\{a_t\}_{t\in\mathbb{N}}$ *e* $\{b_t\}_{t\in\mathbb{N}}$ *sequências de números reais tais que* $\lim\limits_{t\to\infty} |b_t - a_t| = 0$ *e* $(\sum_1^n a_t)/n \to a$*, então*

$$\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} b_t}{n} = a.$$

Demonstração. Como $\lim\limits_{t\to\infty} |a_t - b_t| = 0$, usando o Teorema A.1, temos que

$$\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} |a_t - b_t|}{n} = 0$$

então

$$\begin{aligned}
\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} b_t}{n} &= \lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} a_t + b_t - a_t}{n} \\
&= \lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} a_t}{n} + \frac{\sum\limits_{t=1}^{n} (a_t - b_t)}{n} \\
&= a.
\end{aligned}$$

$\square$

APÊNDICE B – Desigualdade Generalizada de Chebyshev

Teorema B.1 (Desigualdade generalizada de Chebyshev)**.** *Se* $\varepsilon > 0$ *e* $r > 0$, *então*

$$\mathrm{P}\left(|X| > \varepsilon\right) \leq \frac{E\left(|X|^r\right)}{\varepsilon^r}.$$

Demonstração.

$$\begin{aligned} E(|X|^r) &= \int |X|^r dP \\ &\geq \int_{|x|>\varepsilon} |X|^r dP \\ &\geq \int_{|x|>\varepsilon} \varepsilon^r dP \\ &= \varepsilon^r P(|X| > \varepsilon) \end{aligned}$$

Então,

$$P(|X| > \varepsilon) \leq \frac{E\left(|X|^r\right)}{\varepsilon^r}.$$

□