**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

Modelagem de séries temporais para fins de previsão

Hiron Pereira Farias

Tese apresentada para obtenção do título de Doutor em
Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

**Piracicaba
2019**

Hiron Pereira Farias
Licenciatura Plena em Matemática

Modelagem de séries temporais para fins de previsão

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:

Prof. Dr. **CARLOS TADEU DOS SANTOS DIAS**

Tese apresentada para obtenção do título de Doutor em
Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2019

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Farias, Hiron Pereira

Modelagem de séries temporais para fins de previsão / Hiron Pereira Farias.
– – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – –
Piracicaba, 2019 .

62 p.

Tese (Doutorado) – – USP / Escola Superior de Agricultura “Luiz de
Queiroz”.

1. Série Temporal 2. Variáveis climáticas 3. *Commodity* 4. Modelagem
ARMA(p, q) . I. Título.

DEDICATÓRIA

Dedico esse trabalho a minha esposa Paula, ao meu filho Pedro e ao meu filho de quatro patas Bento, pela importância que eles tem na minha vida.

AGRADECIMENTOS

A Deus, por ter me dado força nessa minha caminhada.

Ao meu pai Nilton Elias Pereira, a minha mãe Noemia Julia Pereira e aos meus 13 irmãos, pois, sem o apoio dessa grande família não seria possível alcançar esse objetivo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

A meu orientador, Prof. Dr. Carlos Tadeu dos Santos Dias, pela orientação nesses anos.

Aos professores e funcionários do Departamento de Ciências Exatas da ESALQ/USP, em especial a secretária Solange de Assis Paes Sabadin por estar sempre pronta a ajudar a resolver qualquer pendência com assuntos relacionados à Pós-graduação.

Aos colegas da Pós-graduação em Estatística e Experimentação Agronômica da ESALQ/USP, em especial a Fernando de Pol Mayer, pela elaboração do template em capítulos permitindo assim a elaboração da minha tese.

A todos os educadores que passaram por minha vida e também àqueles que direta ou indiretamente contribuíram para a conclusão deste trabalho, deixo aqui minha gratidão.

A Universidade Federal de Goiás - UFG e a Universidade do Sul e Sudeste do Pará - UNIFESSPA, pela liberação para fazer o doutorado.

SUMÁRIO

RESUMO	7
ABSTRACT	8
1 Introdução	9
1.1 Conceitos importantes em séries temporais	10
1.2 Modelagem univariada de séries temporais	11
1.2.1 Modelo ARMA	11
1.2.2 Identificação	13
1.2.3 Estimação	13
1.2.4 Diagnóstico do modelo ajustado	14
1.2.5 Previsão usando modelo ARMA	15
1.2.6 Seleção de modelos baseado nos erros de previsão	15
Referências	16
2 Modelagem de séries temporais multivariadas para fins de previsão	17
Resumo	17
Abstract	17
2.1 Introdução	17
2.2 Material e Métodos	17
2.2.1 Modelo autorregressivo vetorial (VAR)	18
2.2.1.1 Modelo VAR(1)	21
2.2.1.2 Representação multivariada de Wold	21
2.2.2 Construção de modelos VAR	22
2.2.2.1 Identificação	22
2.2.2.2 Testes sequenciais da razão de verossimilhança	22
2.2.2.3 Critério de informação	23
2.2.2.4 Estimação	24
2.2.2.5 Diagnóstico	24
2.2.2.6 Previsão	25
2.3 Resultados e Discussão	26
2.3.1 Modelagem univariada	26
2.3.1.1 Modelagem univariada: Modelo SARIMA	27
2.3.1.2 Modelagem univariada: SARIMAX	30
2.3.2 Modelo VAR(p) sazonal	32
2.4 Conclusões	34
Referências	35
3 Previsão da média e variância condicional da série de preço do milho	37
Resumo	37
Abstract	37
3.1 Introdução	37
3.2 Material e Métodos	38
3.2.1 Identificação	39

3.2.2	Estimação	39
3.2.3	Diagnósticos	40
3.2.4	Previsão	40
3.3	Resultados e Discussão	41
3.4	Conclusões	44
	Referências	44
4	Seleção dos modelos ARIMA e WAVDS-ARIMA via metodologia backtesting para fins de previsão da <i>commodity</i> milho	45
	Resumo	45
	Abstract	45
4.1	Introdução	46
4.2	Material e Métodos	46
4.2.1	Análise wavelet	46
4.2.1.1	Arquitetura do pré-processamento via wavelets	49
4.2.2	O sistema de decomposição wavelets - WavDS	49
4.2.3	Modelo autorregressivo integrado médias móveis - ARIMA	51
4.2.4	Previsão	53
4.2.5	Modelagem de séries temporais: metodologia Box-Jenkins	54
4.2.6	Previsão fora da amostra	56
4.2.7	Comparação de modelos	57
4.3	Resultados e Discussão	58
4.4	Conclusões	60
	Referências	61

RESUMO

Modelagem de séries temporais para fins de previsão

Nesse trabalho, exploramos técnicas para análise de séries temporais para fins de previsão. Para tanto, foram considerados dados observados de três séries climáticas e de uma série econômica. Para análise das séries climáticas, foi considerada a modelagem multivariada em comparação com os subseqüentes modelos univariados de cada série. Os modelos multivariados e univariados foram comparados com base em seus respectivos resultados preditivos. Para análise da série econômica, considerou-se a modelagem ARMA-GARCH, cuja média condicional e variância condicional são modeladas conjuntamente. Para essa mesma série foi realizada uma modelagem ARIMA em que considerou-se dois casos. No primeiro, a modelagem foi realizada na série original. No segundo, foi realizada na pré-modelagem uma filtragem na série, denominada de sistema de decomposição Wavelet- WavDS, com o objetivo de melhorar o poder preditivo. Na seleção dos modelos ARIMA, considerou-se a metodologia backtesting, em que as previsões são realizadas de forma sequencial, o modelo selecionado foi o que apresentou menor raiz quadrada do erro quadrático médio de previsão (REQM). Toda análise estatística realizada nesse trabalho foi com auxílio do software livre R.

Palavras-chave: Série temporal; Variáveis climáticas; *Commodity*; Modelagem ARMA(p, q).

ABSTRACT

Time-series modeling for prediction purposes

In this study, we explored techniques of time-series analysis for prediction purposes. For that, we considered data observed from three climate series and one economic series. For the analysis of the climate series, we considered the multivariate modelling in comparison with the subsequent univariate models of each series. The multivariate and univariate models were compared based on their respective predictive results. For the analysis of the economic series, the ARMA-GARCH modeling was considered, whose conditional average and conditional variance are modeled together. For this same series, the ARIMA modeling was used, considering two cases. At first, the modeling was performed in the original series. In the second, we carried out a filtering in the series during pre-modeling, called Wavelet-WavDS decomposition system, in order to improve the predictive power. In the selection of ARIMA models, we considered the backtesting methodology in which forecasts are performed in sequence. The model selected showed the lowest square root mean of the prediction square error (REQM). All statistical analyses performed in this work were carried out using the free software R.

Keywords: Time series; Climate variables; Commodity; ARMA Modeling (p, q).

1 INTRODUÇÃO

Para Box et al. (2015), uma **série temporal** é um conjunto de observações geradas sequencialmente no tempo. Na modelagem de séries temporais, a maioria dos modelos considera apenas uma classe especial de processo estocásticos, os processos estocásticos estacionários, assim definidos por possuírem uma forma de equilíbrio estável, ou seja, independentemente do período ou janela que observarmos a série, a sua média e a sua variância permanecerão sem alteração. Sob a hipótese de estacionaridade, a covariância entre os valores de uma série temporal, separados por k intervalos de tempo ou defasagem k , deve ser o mesmo para todo t . Na literatura de séries temporais, esta covariância é denominada de autocovariância de defasagem (atraso ou lag) k . Dentre os processos estocásticos estacionários, este estudo avaliou os processos estocásticos fracamente estacionários, isto é, exigindo que a média e a variância sejam constantes ao longo do tempo e que a covariância de uma série temporal separada por uma defasagem k dependa apenas de k para todo t . Segundo Box et al. (2015), estacionaridade fraca é importante porque ela fornece a estrutura básica para a previsão. Vale salientar que ao invés de dizermos que a série é um processo estocástico fracamente estacionário, diremos apenas que a série é um processo estacionário, para ficar em conformidade com os textos de séries temporais.

No sistema produtivo agrícola, as condições meteorológicas representam fatores exógenos que afetam o crescimento, o desenvolvimento e a produtividade das plantas. Durante seu ciclo, a planta responde diretamente às condições meteorológicas, que formam uma combinação de fatores cujos efeitos podem variar desde os mais favoráveis até os mais desfavoráveis ao crescimento e à produtividade vegetais (Monteiro, 2009). O conhecimento das relações entre as condições físicas do ambiente, em especial, o solo e a atmosfera, e as diversas espécies cultivadas permite a obtenção de informações mais precisas acerca da influência do tempo e do clima no crescimento, no desenvolvimento e na produtividade das culturas. De modo geral, as principais variáveis meteorológicas que afetam o crescimento, o desenvolvimento e a produtividade das culturas são a chuva, a temperatura do ar e a radiação solar, havendo ainda a influência do fotoperíodo, da umidade do ar e do solo, da velocidade e da direção do vento (Monteiro, 2009).

Valores extremos das variáveis climáticas são modificadores ambientais que limitam a produtividade, como aumento da temperatura e decréscimo da precipitação, que alteram principalmente processos como a fotossíntese e a respiração (Monteiro, 2009).

As commodities são mercadorias em estado bruto ou de simples industrialização, negociadas em escala mundial. A comercialização é estabelecida no mercado financeiro, com preços normalmente em dólar americano e que oscilam de acordo com a oferta e a demanda internacionais. Entre as commodities mais comuns estão as que servem como matéria-prima para a produção de outros bens, por exemplo, as commodities agrícolas, chamadas de soft commodities, como o algodão, a soja, o açúcar e o milho. Por se tratar de bens pouco variáveis, transações de commodities podem envolver mercadorias que sequer foram plantadas, como por exemplo as commodities agrícolas. Sacas de milho podem ser negociadas mesmo que elas não existam fisicamente nos silos. É o chamado contrato futuro, em que o produtor negocia as próximas safras na bolsa de valores. A mercadoria não necessariamente será entregue ao comprador, que por sua vez, já pode ter negociado os títulos com outro agente em função da volatilidade dos preços, na busca pelo lucro. Para mais detalhes acesse o site <https://www.dicionariofinanceiro.com/commodities/>.

1.1 Conceitos importantes em séries temporais

Seja y_t uma sequência de variáveis aleatórias indexadas por algum tempo subscrito t . Para Zivot e Wang (2007), Chame essa sequência de variáveis aleatórias uma série temporal. A série temporal, y_t , é **estacionária** se

$$E(y_t) = \mu, \text{ para todo } t.$$

$$\text{cov}(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \text{ para todo } t \text{ e qualquer } k.$$

O parâmetro γ_k é chamado de autocovariância de ordem k ou defasagem k de y_t . Para séries temporais estacionárias, γ_k depende apenas de k e mede a dependência linear entre Y_t e Y_{t-k} , ou seja, mede a dependência dinâmica de y_t em seu passado y_{t-k} . Um resultado útil é que qualquer função de uma série temporal estacionária é também uma série temporal estacionária. A análise de séries temporais lineares concentra-se em estudar a dependência dinâmica da série y_t .

Agora, consideremos, y_t , uma série temporal estacionária. O coeficiente de correlação entre y_t e y_{t-k} , denotado por ρ_k , será denominado de autocorrelação de defasagem k ou lag k . Formalmente, definido pela expressão

$$\rho_k = \frac{\text{cov}(y_t, y_{t-k})}{\sqrt{\text{var}(y_t)\text{var}(y_{t-k})}} = \frac{\gamma_k}{\gamma_0}, \quad (1.1)$$

em que $\text{var}(y_t) = \text{var}(y_{t-k}) = \gamma_0$, pois y_t é estacionária. A partir da Equação (1.1), temos que $\rho_0=1$, $\rho_k = \rho_{-k}$, $-1 < \rho_k < 1$ e que o valor de ρ_k independe da escala de medição da série temporal. A coleção de autocorrelação, $\{\rho_k\}$, é denominada de **função de autocorrelação - FAC**. Multiplicando a Equação (1.1) por σ_y^2 , tem-se $\gamma_k = \rho_k \sigma_y^2$, indicando que o conhecimento da função de autocorrelação, $\{\rho_k\}$, e a variância da série, y_t , é equivalente ao conhecimento da função de autocovariância $\{\gamma_k\}$.

Processo estocástico: Segundo Box et al. (2015), um fenômeno estatístico que evolui no tempo de acordo com as leis probabilísticas é chamado de processo estocástico. Muitas vezes nos referimos a isso simplesmente como um processo, omitindo a palavra “estocástico”.

Série ergódica: Segundo Zivot e Wang (2007), uma série temporal estacionária, y_t , é ergódica se os momentos amostrais convergem em probabilidade para os momentos populacionais, isto é, se $\bar{y} \xrightarrow{P} \mu$, $\hat{\gamma}_j \xrightarrow{P} \gamma_j$ e $\hat{\rho}_j \xrightarrow{P} \rho_j$.

Ruído branco: Para Tsay (2005), uma série temporal é denominada um ruído branco se a_t é uma sequência de variáveis aleatórias independentes e identicamente distribuídas (iid) com média zero e variância finita. Em particular, se a_t é normalmente distribuída com média 0 e variância σ_a^2 , a série é denominada ruído branco gaussiano. Vale ressaltar que na literatura de séries temporais a maioria dos textos omite a expressão gaussiano (Wei (2006) ; Chan e Cryer (2008)).

Série temporal linear: Segundo Tsay (2014), uma série temporal, Y_t , é dita ser linear se pode ser escrita como

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}, \quad (1.2)$$

em que μ é a média de Y_t , $\psi_0 = 1$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ variância finita, e $\{a_t\}$ é uma série de ruído branco. Uma vez que a_t denota a nova informação no instante t da série temporal é muitas vezes

referida como a inovação ou choque no tempo t e será denotado por $a_t \sim RB(0, \sigma_a^2)$. Segundo Zivot e Wang (2007), o teorema da decomposição de Wold, afirma que qualquer série temporal estacionária, Y_t , tem uma representação média móvel de ordem finita na forma da Equação (1.2). Não é difícil mostrar que se na Equação (1.2) a série temporal, $\{Y_t\}$, representa a soma de uma constante μ mais um processo ruído branco gaussiano, $\{a_t\}$, então:

A média é

$$E(Y_t) = \mu.$$

A variância é

$$\text{var}(Y_t) = \gamma_0 = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2.$$

A autocovariância é

$$\gamma_k = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}.$$

Consequentemente, os ψ -pesos estão relacionados com as autocorrelações de Y_t da seguinte forma:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}, \text{ para todo } k \geq 0.$$

Modelos lineares de séries temporais são modelos estatísticos empregados para descrever o padrão dos pesos de Y_t .

1.2 Modelagem univariada de séries temporais

A modelagem univariada de séries envolve o uso de modelos estocásticos para descrever e analisar a estrutura dinâmica de uma série temporal. Com esse objetivo em mente, descrevemos o modelo autorregressivo e de média móveis, ou processo ARMA, que é amplamente utilizado para modelar séries temporais. As propriedades dos modelos ARMA resultantes são discutidas em preparação para seu uso na construção de modelos nos capítulos subsequentes. Em seguida, uma breve discussão sobre métodos para identificação de modelos, estimação de parâmetros e verificação de modelos são descritos. Finalmente, apresentamos uma equação para realização de previsão fora da amostra de estimação e critérios de seleção de modelos ARMA com base nos erros de previsão.

1.2.1 Modelo ARMA

Segundo Hamilton (1994), um processo ARMA(p, q) inclui ambos os termos autorregressivos (AR) e média móveis (MA) e pode ser expresso da seguinte forma:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \quad (1.3)$$

em que $\{a_t\}$ é uma série de ruído branco, p e q são inteiros não-negativos, indicando a ordem do processo ARMA, denotado por $\text{ARMA}(p, q)$.

O processo $\text{ARMA}(p, q)$ pode ser representado na forma de operador defasagem, expresso como

$$\phi(B)Y_t = \phi_0 + \theta(B)a_t \quad (1.4)$$

em que:

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ é o operador autorregressivo de ordem p ;

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ é o operador médias móveis de ordem q .

Verificado que os polinômios $\phi(B)$ e $\theta(B)$ quando fatorados não possuem fatores comuns, pois, caso contrário a ordem pode ser reduzida. Se além disso, constatado que as raízes das equações características $\phi(B) = 0$ e $\theta(B) = 0$ das partes AR e MA, respectivamente, são todas maiores que um em valor absoluto, então o modelo ARMA é fracamente estacionário e invertível. Neste caso, verifica-se que a média é dada por

$$E(Y_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} = \frac{\phi_0}{\phi(1)}. \quad (1.5)$$

Ao reescrevermos a Equação (1.5) na forma

$$\phi_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$$

e substituírmos na Equação (1.3), tem-se o processo $\text{ARMA}(p, q)$ escrito em termo dos desvios em relação a média:

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}. \quad (1.6)$$

As autocovariâncias são obtidas multiplicando ambos os membros da Equação (1.6) por $Y_t - \mu$ e tomando a esperança. Para $j > q$, a equação resultante toma a forma

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \phi_3 \gamma_{j-3} + \cdots + \phi_p \gamma_{j-p}, \quad j = q+1, q+2, q+3, \cdots \quad (1.7)$$

Para Chan e Cryer (2008), o modelo geral $\text{ARMA}(p, q)$ sujeito à condição de que a_t é independente de $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$, uma solução estacionária para a Equação (1.3) existe se, e somente se, todas as raízes da equação característica AR

$$\phi(B) = 0 \quad (1.8)$$

são todas em valor absoluto maiores que um. Neste caso, ambos membros da Equação (1.4) podem ser divididos por $\phi(B)$, obtendo

$$Y_t = \mu + \psi(B)a_t \quad (1.9)$$

em que

$$\psi(B) = \frac{\theta(B)}{\phi(B)} = \frac{(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)}{(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)} = \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \dots$$

Os coeficientes $\{\psi_i\}$ são referidos como a função de resposta ao impulso do modelo ARMA. A Equação (1.9) é dita ser a forma média móvel de um processo ARMA(p, q) estacionário, então o modelo (1.3) pode ser escrito como um processo linear geral da seguinte forma

$$Y_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \dots \quad (1.10)$$

em que os coeficientes ψ_j são determinados a partir de

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= -\theta_1 + \phi_1 \\ \psi_2 &= -\theta_2 + \phi_2 + \phi_1 \psi_1 \\ &\vdots \\ \psi_j &= -\theta_j + \phi_p \psi_{j-p} + \phi_{p-1} \psi_{j-p+1} + \dots + \phi_1 \psi_{j-1}. \end{aligned} \quad (1.11)$$

em que tomamos $\psi_j = 0$ para $j < 0$ e $\theta_j = 0$ para $j > q$.

Para Tsay (2005), a representação (1.10) mostra explicitamente o impacto do choque passado a_{t-i} no valor atual da série $\{y_t\}$. Para uma série estacionária, os coeficientes ψ_i decaem exponencialmente à medida que i aumenta. Isto está sob a condição de que o efeito do choque a_{t-i} na série deve diminuir com o tempo. Assim, para um modelo ARMA estacionário, o choque a_{t-i} não tem um impacto permanente sobre a série. Finalmente, a representação MA na Equação (1.10) fornece uma prova simples de reversão à média de uma série temporal estacionária. Para um modelo geral ARMA(p, q), exigimos tanto estacionaridade quanto invertibilidade. Para um modelo ARMA, uma condição suficiente para a invertibilidade é que todas as raízes do polinômio característico $\theta(B)$ sejam maiores que 1 em valor absoluto.

1.2.2 Identificação

A identificação da ordem p e q de um modelo ARMA estacionário e invertível será realizada analisando a função de autocorrelação - FAC, a função de autocorrelação parcial - FACP e a função de autocorrelação estendida - EACF do processo estocástico gerador dos dados. A FAC define a defasagem do MA, a FACP define a defasagem do AR e no caso de um ARMA(p, q) a FAC começará a decair a partir da defasagem q , e a FACP, a partir da defasagem p . Para mais detalhes ver (Tsay, 2014) e (Box et al., 2015).

1.2.3 Estimação

Após a identificação das ordens dos parâmetros do modelo ARMA, passamos a fase de estimação dos parâmetros. No caso geral de modelos ARMA(p, q), estimação utilizando máxima verossimilhança e estimação utilizando mínimos quadrados condicionais e incondicionais (e estimativa de Yule-Walker no caso de modelos de AR) levam a estimadores ótimos. A prova

desse resultado geral pode ser encontrada em vários textos sobre análise de séries temporais teóricas (por exemplo, Brockwell e Davis, 1991, ou Hannan, 1970, para mencionar alguns). Vamos denotar os parâmetros do coeficiente ARMA por $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ (Shumway e Stoffer, 2011).

1.2.4 Diagnóstico do modelo ajustado

Após estimados os $m = p+q+1$ parâmetros do modelo proposto, tem-se a necessidade de verificar o pressuposto de que a componente do erro aleatório, a_t , é ruído branco. Para tanto, calcula-se os valores estimados pelo modelo, denotados por $\hat{y}_t$, consequentemente obtem-se os resíduos estimados, $\hat{a}_t$, que é a diferença entre os valores observados da série e os valores estimados a partir do modelo ajustado, isto é,

$$\hat{a}_t = y_t - \hat{y}_t, \quad t = 1, \dots, T.$$

A série $\{\hat{a}_t\}$ é chamada de série residual, e a partir dela, pode-se estimar a variância, denotada por σ_a^2 , da seguinte maneira

$$\hat{\sigma}_a^2 = \frac{\sum_{t=1}^T \hat{a}_t^2}{T - m}. \quad (1.12)$$

As autocorrelações dos resíduos de defasagem k , denotada por $\hat{\rho}_k$, são calculadas como

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^T \hat{a}_t^2}. \quad (1.13)$$

Sob a hipótese de ruído branco da série residual $\{\hat{a}_t\}$ e utilizando a aproximação de Bartlett, os erros padrão das autocorrelações estimadas são calculados da seguinte maneira

$$e.p.(\hat{\rho}_k) = \frac{1}{\sqrt{T}}, \text{ para } k > 0. \quad (1.14)$$

Uma forma visual de verificar a significância da autocorrelação, ρ_k , é plotar o correlograma (gráfico da FAC) das autocorrelações estimadas $\hat{\rho}_k$ versus k , e os limites $\pm \frac{1.96}{\sqrt{T}}$, caso o valor $\hat{\rho}_k$ exceda os limites tem-se um forte indicativo da significância de ρ_k . Para testar a hipótese nula de que o modelo ajusta-se bem aos dados, isto é, em outras palavras, que $\{a_t\}$ é ruído branco, em 1978 Ljung e Box propuseram uma estimativa mais precisa da variância de ρ_k de uma série de ruído branco e consequentemente sugeriu o seguinte teste mais poderoso.

Teste de Ljung-Box: Segundo Box et al. (2015) e Chan e Cryer (2008), supondo que tenhamos as K primeiras autocorrelações estimadas $\hat{\rho}_k$ ($k = 1, 2, \dots, K$) de um modelo ARMA(p, q), então a estatística de teste dada por

$$Q = T(T+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{T-k}, \quad (1.15)$$

em que T é o tamanho da amostra utilizada na estimação do modelo, segundo Ljung-Box, a estatística Q tem, uma distribuição qui-quadrado aproximadamente com $(K - p - q)$ graus de

liberdade. Como conclusão do teste, rejeita-se a hipótese nula de ajuste adequado do modelo se o valor- p for menor que o nível de significância α adotado. Uma outra forma de visualização da adequação do modelo é variar os valores de K , para $K > p + q$, na Equação (1.15) e plotar o gráfico valor- p contra K , assim, se o modelo ajustado for adequado para descrever a dependência serial dos dados, todos os valores- p deverão ser maiores que α .

1.2.5 Previsão usando modelo ARMA

Denotando a origem de previsão por T (Tamanho da sub-amostra de estimação) e a informação até o instante T por $\mathcal{F}_T$. Segundo Tsay (2005), a previsão de h -passos à frente de Y_{T+h} pode ser obtida da seguinte forma

$$\hat{y}_T(h) = E(y_{T+h}|\mathcal{F}_T) = \phi_0 + \sum_{i=1}^p \phi_i \hat{y}_T(h-i) + \sum_{i=1}^q \theta_i a_t(h-i), \quad \text{para } h = 1, 2, 3, \dots, H \quad (1.16)$$

em que:

1. $\hat{y}_T(h-i) = y_{T+h-i}$, se $h-i \leq 0$;
2. $a_T(h-i) = 0$, se $h-i > 0$
3. $a_T(h-i) = a_{T+h-i}$, se $h-i \leq 0$

Segundo DE LOSSO (2011) na modelagem de séries temporais é comum selecionar vários modelos igualmente bons. Se o objetivo é previsão, podemos utilizar critérios de seleção baseados nos erros de previsões.

1.2.6 Seleção de modelos baseado nos erros de previsão

Para calcular o erro de previsão, deve-se deixar fora da modelagem, uma quantidade H dos valores finais da série de tamanho $T + H$. Ajustado o modelo, faz-se a previsão da série h -passos a frente. O erro de previsão, denotado por $e_T(h)$ é a diferença entre o valor observado fora da amostra e o previsto, definido como

$$e_T(h) = y_{T+H} - \hat{y}_T(h), \quad (1.17)$$

em que y_{T+H} é o valor observado fora da amostra e $\hat{y}_T(h)$ o valor predito de origem T e horizonte h .

De posse do erro de previsão, pode-se calcular o Erro Quadrático Médio (MSE), o Erro Absoluto Médio (MAE) ou o Erro Percentual Absoluto Médio (MAPE), respectivamente, como

$$\text{MSE} = \frac{1}{H} \sum_{h=1}^H e^2(h), \quad (1.18)$$

$$\text{MAE} = \frac{1}{H} \sum_{h=1}^H |e(h)|, \quad (1.19)$$

$$\text{MAPE} = \left[\frac{1}{H} \sum_{h=1}^H \left| \frac{e(h)}{y_{t+h}} \right| \right] 100\%. \quad (1.20)$$

O modelo selecionado será o que apresentar menor MSE ou MAE ou MAPE. Para tomada de decisão o pesquisador tem que escolher um dos métodos de seleção, pois, é muito comum esses critérios divergirem na escolha do melhor modelo.

Referências

- Box, G., Jenkins, G., Reinsel, G., e Ljung, G. (2015). *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley.
- Chan, K. e Cryer, J. (2008). Time series analysis with applications in r.
- DE LOSSO, R. (2011). Econometria de series temporais. *São Paulo: Cengage*, 1:363–386.
- Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.
- Monteiro, J. E. B. (2009). *Agrometeorologia dos cultivos : o fator meteorológico na produção agrícola*. Instituto Nacional de Meteorologia – INMET.
- Shumway, R. H. e Stoffer, D. S. (2011). *Time series analysis and its applications*. Springer.
- Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.
- Tsay, R. S. (2014). *An introduction to analysis of financial data with R*. John Wiley & Sons.
- Wei, W. W. (2006). *Time series analysis: univariate and multivariate*. Pearson Education.
- Zivot, E. e Wang, J. (2007). *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media.

2 MODELAGEM DE SÉRIES TEMPORAIS MULTIVARIADAS PARA FINS DE PREVISÃO

Resumo

No agronegócio, ao fazer escolhas entre cursos alternativos de ação, os tomadores de decisões frequentemente precisam de previsões de variáveis climáticas. Nesse trabalho, com intuito de colaborar com os tomadores de decisões, exploramos a modelagem de séries temporais climáticas com a finalidade de fazer previsões. Para tanto, foram considerados dados observados de três séries climáticas precipitação, temperatura mínima e umidade relativa média. Na modelagem, foram utilizados modelos univariados da família ARMA e modelos multivariados da família VAR. Na teoria desenvolvida, deu-se ênfase a modelagem multivariada de séries temporais uma vez que para o caso univariado basta considerar o vetor com apenas uma série.

Palavras-chave: Agronegócio; Componente sazonal; Modelagem VAR(p); Previsão.

Abstract

In agribusiness, to make choices between alternative courses of action, decision makers often need prediction of multiple inter-related variables. In this work, with the aim to support decision makers, we explored the climate time-series modeling for predictions. We considered data on precipitation, minimum temperature and average relative humidity of three climate series. In modeling, univariate models of the ARMA family and multivariate models of VAR family were used. In the theory developed, there was emphasis on the multivariate time-series modeling, once in the case of univariate modeling, it is necessary to consider the vector with only one series.

Keywords: Agribusiness; Seasonal component; Vector autoregressive model; Prediction.

2.1 Introdução

No Brasil, o agronegócio é responsável por um terço do Produto Interno Bruto (PIB) e a soja é uma das principais commodities negociadas na bolsa de valores, contribuindo para um superávit primário na balança comercial. No entanto, variáveis climáticas, como precipitação, temperatura, fotoperíodo, e outras, são fatores que podem influenciar na produtividade da soja, milho e outras commodities que possuem um papel importante balança comercial brasileira. Os modelos de séries temporais multivariados são úteis em estudos das relações dinâmicas e modelagem para fins de previsão de variáveis climáticas. Nesse caso, podemos utilizar esses modelos para prever as variáveis climáticas e tomar decisões sobre qual cultura plantar e quanto plantar.

2.2 Material e Métodos

A análise de séries temporais multivariada considera simultaneamente várias séries temporais (Tsay, 2013). É um ramo da análise estatística multivariada, mas que lida especificamente

com dados dependentes. Estudar essa análise estatística é importante porque, na vida real, as decisões geralmente envolvem múltiplos fatores ou variáveis inter-relacionadas. Compreender as relações entre esses fatores e fornecer previsões precisas dessas variáveis é importante na tomada de decisões.

Segundo Box et al. (2015), a análise multivariada de séries temporais envolve o uso de modelos estocásticos para descrever e analisar as relações entre várias séries temporais. No estudo de processos multivariados, uma estrutura é necessária para descrever não apenas as propriedades das séries individuais, mas também as possíveis relações cruzadas entre as séries. Duas finalidades principais para analisar e modelar a série em conjunto são:

- entender as relações dinâmicas ao longo do tempo entre as séries.
- melhorar a precisão das previsões para séries individuais, utilizando as informações adicionais disponíveis nas séries relacionadas.

Com esses objetivos, este estudo apresenta alguns conceitos e ferramentas básicas necessários para modelar séries temporais multivariadas. Em seguida, descreve-se o vetor autorregressivo, ou modelos VAR, que são amplamente utilizados em trabalhos aplicados. Esses modelos são examinados e métodos para identificação de modelos, estimação de parâmetros e verificação de modelos são descritos.

Segundo Lütkepohl (2005), se as observações de séries temporais estiverem disponíveis para uma variável de interesse e os dados do passado contêm informações sobre o desenvolvimento futuro de uma variável, é plausível usar como previsão alguma função dos dados coletados no passado.

Ao lidar com um conjunto variáveis, muitas vezes o valor atual de uma variável não está relacionado apenas a seus valores passados no tempo, dependendo também de valores passados de outras variáveis. Se todas essas variáveis estão relacionadas, para um propósito digamos de previsão, então faz sentido usar seu possível conteúdo de informações adicionais na previsão. Em outras palavras, denotando as variáveis relacionadas por $y_{1,t}, y_{2,t}, \dots, y_{n,t}$, a previsão da n -ésima variável pode ser expressa como

$$\hat{y}_{k,T+h} = f_k(y_{1,T}, \dots, y_{n,T}, \dots, y_{1,T-p}, \dots, y_{n,T-p}), \quad k = 1, \dots, n \quad (2.1)$$

em que o objetivo principal da análise de séries temporais múltiplas é determinar as funções adequadas $f_1, \dots, f_n$ que podem ser utilizados para obter previsões com boas propriedades para as variáveis do sistema (Lütkepohl, 2005).

2.2.1 Modelo autorregressivo vetorial (VAR)

Seja $\mathbf{Y}_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ denotando um vetor $(n \times 1)$ de variáveis de séries temporais. Se o processo $\mathbf{Y}_t$ de ordem $(n \times 1)$ segue um modelo VAR(p), então segundo Tsay (2013), o modelo autorregressivo vetorial de defasagem p (ou lag p) tem a forma

$$\mathbf{Y}_t = \Phi_0 + \Phi_1 \mathbf{Y}_{t-1} + \Phi_2 \mathbf{Y}_{t-2} + \dots + \Phi_p \mathbf{Y}_{t-p} + \mathbf{a}_t \quad (2.2)$$

em que:

$\Phi_0 = (\phi_{10}, \phi_{20}, \dots, \phi_{n0})'$, é um vetor de constantes $(n \times 1)$;

Φ_k , são matrizes $(n \times n)$, $k = 1, 2, \dots, p$, com $\Phi_p \neq \mathbf{0}$ e;

$\mathbf{a}_t$, uma sequência de vetores aleatórios independentes e identicamente distribuídos com vetor de médias nulo e matriz de covariância, Σ_a , positiva definida. Dizemos neste caso que $\mathbf{a}_t$ é um ruído branco ou uma inovação e denotaremos por $\mathbf{a}_t \sim RB(\mathbf{0}, \Sigma_a)$.

Sendo $\mathbf{I}_n$ uma matriz identidade de ordem n , a Equação (2.2) pode ser reescrita na forma

$$\Phi(B)\mathbf{Y}_t = \Phi_0 + \mathbf{a}_t \quad (2.3)$$

em que $\Phi(B) = \mathbf{I}_n - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p = \mathbf{I}_n - \sum_{i=1}^p \Phi_i B^i$ é o operador autorregressivo vetorial de ordem p , ou ainda, um polinômio matricial $(n \times n)$ em B .

Considere o modelo bivariado VAR(1)

$$\mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \mathbf{a}_t, \quad (2.4)$$

O qual pode ser escrito explicitamente como

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \quad (2.5)$$

ou de forma equivalente,

$$\begin{aligned} y_{1t} &= \phi_{10} + \phi_{1,11}y_{1,t-1} + \phi_{1,12}y_{2,t-1} + a_{1t} \\ y_{2t} &= \phi_{20} + \phi_{1,21}y_{1,t-1} + \phi_{1,22}y_{2,t-1} + a_{2t} \end{aligned}$$

em que $\phi_{1,11}$ e $\phi_{1,22}$ refletem a dependência de cada componente em seu próprio passado. O parâmetro $\phi_{1,12}$ mostra a dependência de $y_{1,t}$ em $y_{2,t-1}$ na presença de $y_{1,t-1}$, enquanto $\phi_{1,21}$ mostra a dependência de $y_{2,t}$ em $y_{1,t-1}$ na presença de $y_{2,t-1}$. Assim, se $\phi_{1,12} \neq 0$ e $\phi_{1,21} \neq 0$, existe uma relação de feedback entre os dois componentes. Por outro lado, se os elementos fora da diagonal da matriz de parâmetros Φ_1 são zero, isto é, $\phi_{1,12} = \phi_{1,21} = 0$, então $y_{1,t}$ e $y_{2,t}$ não são dinamicamente correlacionados. No entanto, eles ainda são contemporaneamente correlacionados, a menos que Σ_a seja uma matriz diagonal.

O processo VAR(p) é estável se seu polinômio característico tem raízes fora do círculo de unidade complexa. Formalmente $\mathbf{Y}_t$ é estável se

$$\det(\mathbf{I}_n - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p) \neq 0 \quad \text{para } |B| \leq 1. \quad (2.6)$$

Segundo Lütkepohl (2005), esta condição é chamada de condição de estabilidade. A condição (2.6) fornece uma ferramenta para verificar a estabilidade de um processo VAR. Um conceito importante na representação de modelos em análise de séries temporais, que possibilita a obtenção de resultados de modelagem útil a partir da realização de uma amostra finita da série, é

o de estacionaridade. Segundo Lütkepohl (2005), uma condição de estacionaridade é dada pelo seguinte teorema

Teorema Um processo $\text{VAR}(p)$ estável, $t = 0, \pm 1, \pm 2, \dots$, é estacionário.

Como estabilidade implica estacionaridade, a condição de estabilidade (2.6) é frequentemente referida como condição de estacionaridade na literatura de séries temporais. Outras definições de estacionaridade são frequentemente usadas na literatura. Como por exemplo, um processo estocástico é estacionário se seu primeiro e segundo momentos são invariantes no tempo. Sobre o teorema vale ressaltar que um processo estacionário não necessariamente é estável.

Para Morettin (2011), suponha que a série na Equação (2.2) é estacionária, neste caso temos que:

$$E(\mathbf{Y}_{t-p}) = \boldsymbol{\mu}, \quad p = 0, 1, 2, \dots, p-1,$$

$$(\mathbf{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \dots - \boldsymbol{\Phi}_p)\boldsymbol{\mu} = [\boldsymbol{\Phi}(1)]\boldsymbol{\mu} = \boldsymbol{\Phi}_0$$

e conseqüentemente

$$\boldsymbol{\mu} = [\boldsymbol{\Phi}(1)]^{-1}\boldsymbol{\Phi}_0.$$

Fazendo $\mathbf{Z}_t = \mathbf{Y}_t - \boldsymbol{\mu}$ na equação (2.2), temos

$$\mathbf{Z}_t = \boldsymbol{\Phi}_1\mathbf{Z}_{t-1} + \boldsymbol{\Phi}_2\mathbf{Z}_{t-2} + \dots + \boldsymbol{\Phi}_p\mathbf{Z}_{t-p} + \mathbf{a}_t, \quad (2.7)$$

a série temporal ajustada pela média.

Pós-multiplicando a Equação (2.7) por $\mathbf{Z}_{t-l}$ e tomando a esperança, tem-se a matriz de autocovariância, expressa como se segue

$$\mathbf{Z}_t\mathbf{Z}_{t-l} = \boldsymbol{\Phi}_1\mathbf{Z}_{t-1}\mathbf{Z}_{t-l} + \boldsymbol{\Phi}_2\mathbf{Z}_{t-2}\mathbf{Z}_{t-l} + \dots + \boldsymbol{\Phi}_p\mathbf{Z}_{t-p}\mathbf{Z}_{t-l} + \mathbf{a}_t\mathbf{Z}_{t-l}$$

e

$$E(\mathbf{Z}_t\mathbf{Z}_{t-l}) = \boldsymbol{\Phi}_1E(\mathbf{Z}_{t-1}\mathbf{Z}_{t-l}) + \boldsymbol{\Phi}_2E(\mathbf{Z}_{t-2}\mathbf{Z}_{t-l}) + \dots + \boldsymbol{\Phi}_pE(\mathbf{Z}_{t-p}\mathbf{Z}_{t-l}) + E(\mathbf{a}_t\mathbf{Z}_{t-l}).$$

Logo

$$\boldsymbol{\Gamma}(l) = \boldsymbol{\Phi}_1\boldsymbol{\Gamma}(l-1) + \boldsymbol{\Phi}_2\boldsymbol{\Gamma}(l-2) + \dots + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}(l-p), \quad l > 0$$

que são as equações de Yule-Walker no caso de um modelo $\text{VAR}(p)$.

Para $l = 0$, tem-se

$$\boldsymbol{\Gamma}(0) = \boldsymbol{\Phi}_1\boldsymbol{\Gamma}(-1) + \boldsymbol{\Phi}_2\boldsymbol{\Gamma}(-2) + \dots + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}(-p) + \boldsymbol{\Sigma}_a,$$

usando o fato que $\boldsymbol{\Gamma}(l) = \boldsymbol{\Gamma}'(-l)$ concluímos que

$$\boldsymbol{\Gamma}(0) = \boldsymbol{\Phi}_1\boldsymbol{\Gamma}'(1) + \boldsymbol{\Phi}_2\boldsymbol{\Gamma}'(2) + \dots + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}'(p) + \boldsymbol{\Sigma}_a.$$

Estas equações podem ser utilizadas para calcular $\boldsymbol{\Gamma}(l)$ recursivamente, para $l \geq p$. Para $|l| < p$, temos que usar a representação $\text{VAR}(1)$ de um processo $\text{VAR}(p)$.

Para medir a dependência linear dinâmica de uma série temporal estacionária $\mathbf{Y}_t$, definimos a matriz de covariância cruzada de lag l como

$$\boldsymbol{\Gamma}_l = \text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-l}) = \text{Cov}(\mathbf{Z}_t, \mathbf{Z}_{t-l}) = E(\mathbf{Z}_t\mathbf{Z}_{t-l}), \quad (2.8)$$

para $l = 0$, tem-se a matriz de covariância de $\mathbf{Y}_t$, neste caso, tem-se $\mathbf{\Gamma}_0 = \mathbf{\Sigma}_y$. A matriz de correlação cruzada (MCC) $\boldsymbol{\rho}_l$ é definida da seguinte forma

$$\boldsymbol{\rho}_l = \mathbf{D}^{-1} \mathbf{\Gamma}_l \mathbf{D}^{-1}, \quad (2.9)$$

em que $\mathbf{D} = \text{diag}(\mathbf{\Gamma}_0)$.

2.2.1.1 Modelo VAR(1)

Seja $\mathbf{Y}_t$ um VAR(p), um VAR(1) np -dimensional correspondente é definido como

$$\mathbf{Y}_t = \boldsymbol{\nu} + \boldsymbol{\Phi} \mathbf{Y}_{t-1} + \mathbf{U}_t, \quad (2.10)$$

em que :

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix}_{(np \times 1)}, \quad \boldsymbol{\nu} = \begin{bmatrix} \phi_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}_{(np \times 1)}, \quad \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_n & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}_{(np \times np)} \quad \text{e}$$

$$\mathbf{U}_t = \begin{bmatrix} \mathbf{a}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}_{(np \times 1)}.$$

Escrevendo a Equação (2.10) na forma da Equação (2.7), temos

$$\mathbf{Y}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}(\mathbf{Y}_{t-1} - \boldsymbol{\mu}) + \mathbf{U}_t, \quad (2.11)$$

em que :

$$\boldsymbol{\Sigma}_U = \begin{bmatrix} \Sigma_a & 0_n \\ 0_n & 0_n \end{bmatrix} \quad \text{e}$$

$$\text{vec}[\Gamma(0)] = \mathbf{I}_{n^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi}^{-1} \text{vec}(\boldsymbol{\Sigma}_U) \quad (2.12)$$

em que $\text{vec}(\cdot)$ é um operador que transforma uma matriz em um vetor coluna empilhando as colunas da matriz dando assim origem a um vetor coluna, por exemplo, para uma matriz qualquer de dimensão $(m \times n)$ ao aplicarmos o operador $\text{vec}(\cdot)$ tem-se um vetor coluna $(mn \times 1)$. Enquanto $\otimes$ é o produto Kronecker ou produto direto entre matrizes. Para entendermos como funciona o produto de Kronecker, consideremos duas matrizes $A = (a_{i,j})$ e $B = (b_{i,j})$ de dimensões $(m \times n)$ e $(p \times q)$ respectivamente. Neste caso, o produto de Kronecker é calculada da seguinte forma

$$A \otimes B = (a_{i,j})B$$

é uma matriz de dimensão $(mp \times nq)$.

2.2.1.2 Representação multivariada de Wold

Segundo Zivot e Wang (2007), se $\mathbf{Y}_t$ é um vetor $(n \times 1)$ de séries temporais multivariadas estacionárias, então $\mathbf{Y}_t$ tem uma representação de Wold da seguinte forma

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{a}_t + \boldsymbol{\Psi}_1 \mathbf{a}_{t-1} + \boldsymbol{\Psi}_2 \mathbf{a}_{t-2} + \cdots = \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \mathbf{a}_{t-k}, \quad (2.13)$$

em que $\boldsymbol{\Psi}_0 = \mathbf{I}_n$ e $\mathbf{a}_t$ é um ruído branco multivariado com vetores de média $\mathbf{0}$ e variância $E(\mathbf{a}_t \mathbf{a}_t') = \boldsymbol{\Sigma}_a$. Em (2.13), $\boldsymbol{\Psi}_k$ é uma matriz $(n \times n)$ com (i, j) -ésimo elemento Ψ_{ij}^k . Na notação de operador, a forma de Wold é

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(B) \mathbf{a}_t, \quad (2.14)$$

em que $\boldsymbol{\Psi}(B) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k B^k$.

Os momentos de $\mathbf{Y}_t$ são dados por

$$E(\mathbf{Y}_t) = \boldsymbol{\mu}.$$

$$\text{var}(\mathbf{Y}_t) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\Sigma} \boldsymbol{\Psi}_k'.$$

2.2.2 Construção de modelos VAR

Segundo Morettin (2011), a construção de modelos VAR segue o mesmo ciclo de identificação, estimação e diagnóstico que a modelagem univariada. A seguir tem-se a orientação de execução das etapas do ciclo iterativo.

2.2.2.1 Identificação

Para os modelos VAR, especificação do modelo é selecionar a ordem p . Vários métodos têm sido propostos na literatura para selecionar a ordem de um VAR. Nós discutiremos duas abordagens. A primeira adota a estrutura multivariada de regressão linear múltipla e usa testes sequenciais de razão de probabilidade, enquanto que a segunda emprega critérios de informação.

2.2.2.2 Testes sequenciais da razão de verossimilhança

Para Morettin (2011), uma maneira de identificar a ordem p de um modelo VAR(p) consiste em ajustar, sequencialmente, modelos autorregressivos vetoriais de ordem $1, 2, \dots, k$ e testar a significância dos coeficientes (matrizes).

Considere, pois, os modelos

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\Phi}_0^{(1)} + \boldsymbol{\Phi}_1^{(1)} \mathbf{Y}_{t-1} + \mathbf{a}_t^{(1)} \\ \mathbf{Y}_t &= \boldsymbol{\Phi}_0^{(2)} + \boldsymbol{\Phi}_1^{(2)} \mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2^{(2)} \mathbf{Y}_{t-2} + \mathbf{a}_t^{(2)} \\ &\vdots \\ \mathbf{Y}_t &= \boldsymbol{\Phi}_0^{(k)} + \boldsymbol{\Phi}_1^{(k)} \mathbf{Y}_{t-1} + \boldsymbol{\Phi}_2^{(k)} \mathbf{Y}_{t-2} + \cdots + \boldsymbol{\Phi}_k^{(k)} \mathbf{Y}_{t-k} + \mathbf{a}_t^{(k)}. \end{aligned}$$

Os parâmetros podem ser estimados por mínimos quadrados ordinários, que fornecem estimadores consistentes e eficientes.

Testamos, então,

$$H_0 : \boldsymbol{\Phi}_k^k = \mathbf{0} \text{ contra } H_1 : \boldsymbol{\Phi}_k^k \neq \mathbf{0} \quad k = 1, 2, \dots$$

O teste da razão de verossimilhança é baseado nas estimativas das matrizes de covariâncias dos resíduos dos modelos ajustados. Para a k -ésima equação, considere

$$\hat{\mathbf{a}}_t^{(k)} = \mathbf{Y}_t - \hat{\Phi}_0^{(k)} - \hat{\Phi}_1^{(k)} \mathbf{Y}_{t-1} - \dots - \hat{\Phi}_k^{(k)} \mathbf{Y}_{t-k}.$$

A matriz de covariâncias dos resíduos, que estima Σ_a , é dada por

$$\widehat{\Sigma}_k = \frac{1}{(T-k)} \sum_{t=k+1}^T \hat{\mathbf{a}}_t^{(k)} (\hat{\mathbf{a}}_t^{(k)})', \quad k > 0 \quad (2.15)$$

em que para $k = 0$, $\hat{\mathbf{a}}_t^{(0)} = \mathbf{Y}_t - \bar{\mathbf{Y}}$. A estatística da razão de verossimilhanças para o teste é dada por

$$RV(k) = (T-k) \ln \frac{|\widehat{\Sigma}_{k-1}|}{|\widehat{\Sigma}_k|} \quad (2.16)$$

que tem distribuição qui-quadrado com n^2 graus de liberdade, $\chi_{n^2}^2$.

O procedimento apresentado não é único e uma outra maneira de identificar a ordem de um VAR é usar algum critério de informação.

2.2.2.3 Critério de informação

Para (Zivot e Wang, 2007), a ordem da defasagem para o modelo VAR(p) pode ser determinada, usando critérios de seleção do modelo. A abordagem geral é ajustar modelos VAR(p) com ordens $p = 0, 1, 2, \dots, k = p_{max}$ e escolher o valor k que minimiza alguns critérios de seleção de modelo. Os critérios de seleção de modelo para modelos VAR(p) tem a forma

$$IC(p) = \ln \left| \widehat{\Sigma}(p) \right| + C_T \varphi(n, p)$$

em que $\widehat{\Sigma}(p) = T^{-1} \sum_{t=1}^T \hat{\mathbf{a}}_t \hat{\mathbf{a}}_t'$ é a matriz de covariância residual sem uma correção do número de graus de liberdade de um modelo VAR(p), C_T é uma sequência indexada pelo tamanho da amostra T , e $\varphi(n, p)$ é uma função de penalidade a qual penaliza grandes modelos VAR(p). A seleção da penalidade é relativamente subjetiva, pois, diferentes penalidades resultam em diferentes critérios de informação. O primeiro componente está preocupado com a bondade do ajuste do modelo aos dados, enquanto o segundo componente penaliza modelos com um número grande de parâmetros. Os três critérios de informação mais comuns são os de Akaike (AIC), Schwarz-Bayesiano (BIC) e Hannan-Quinn (HQ):

$$\begin{aligned} \text{AIC}(k) &= \ln \left| \widehat{\Sigma}_k \right| + \frac{2}{T} kn^2 \quad (\text{Akaike}), \\ \text{BIC}(k) &= \ln \left| \widehat{\Sigma}_k \right| + \frac{\ln(T)}{T} kn^2 \quad (\text{Schwarz}), \\ \text{HQ}(k) &= \ln \left| \widehat{\Sigma}_k \right| + \frac{\ln[\ln(T)]}{T} kn^2 \quad (\text{Hannan-Quinn}), \end{aligned}$$

em que kn^2 é o número total de parâmetros estimados em todas as equações.

2.2.2.4 Estimação

Segundo Tsay (2013), após identificado o valor de p e supondo uma amostra $\{y_t | t = 1, 2, \dots, T\}$ observada de um modelo VAR(p), nosso objetivo reside em estimar os parâmetros $(\Phi_0, \Phi_1, \dots, \Phi_p)$ e Σ_a . Para facilitar a estimação podemos reescrever o modelo VAR(p) como

$$\mathbf{y}'_t = \mathbf{x}'_t \boldsymbol{\beta} + \mathbf{a}'_t \quad (2.17)$$

em que :

$\mathbf{x}'_t = (1, y'_{t-1}, \dots, y'_{t-p})'$ um vetor de dimensão $(np + 1)$;

$\boldsymbol{\beta}' = [\Phi_0, \Phi_1, \dots, \Phi_p]$ e Σ_a é uma matriz de dimensão $n \times (np + 1)$.

Na forma matricial a Equação 2.17 pode ser escrita como

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{A} \quad (2.18)$$

em que:

$\mathbf{Y}$: é uma matriz $(T - p) \times n$;

$\mathbf{X}$: é uma matriz de delineamento $(T - p) \times (np + 1)$;

$\mathbf{A}$: uma matriz $(T - p) \times n$.

Supondo que a inovação $\mathbf{a}_t$ de um modelo VAR(p) estacionário segue uma distribuição normal multivariada com média zero e matriz de covariância Σ_a positiva definida. então, os estimadores de máxima verossimilhança são:

$$\text{vec}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}\mathbf{X}')^{-1} \mathbf{X}'\mathbf{Y}$$

e

$$\hat{\Sigma}_a = \frac{1}{T - p} \sum_{t=p+1}^T \hat{\mathbf{a}}_t \hat{\mathbf{a}}'_t = \frac{1}{T - p} \hat{\mathbf{A}}' \hat{\mathbf{A}}$$

Os parâmetros de um modelo VAR(p) também podem ser estimados pelos métodos de mínimos quadrados ou Bayesiano. Para uma discussão e demonstração dos métodos de estimação mencionados, ver Lütkepohl (2005) e Tsay (2013).

2.2.2.5 Diagnóstico

Seja $\hat{\mathbf{A}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ a matriz residual de um modelo VAR(p) ajustado. Segundo Tsay (2013), a matriz de correlação de lag l da série residual é definida como

$$\hat{C}_l = \frac{1}{T - p} \hat{\mathbf{A}}' \mathbf{B}^l \hat{\mathbf{A}} \quad , \quad l \geq 0 \quad (2.19)$$

em que $\mathbf{B}$ é uma matriz quadrada de dimensão $(T - p)$ definida como

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{0}'_{T-p-1} \\ \mathbf{I}_{T-p-1} & \mathbf{0}_{T-p-1} \end{bmatrix},$$

em que $\mathbf{0}_h$ é um vetor de zeros de dimensão h . Assim, a matriz de correlação cruzada residual é definida como

$$\hat{\mathbf{R}}_l = \hat{\mathbf{D}}^{-1} \hat{\mathbf{C}}_l \hat{\mathbf{D}}^{-1}, \quad (2.20)$$

em que $\hat{\mathbf{D}}$ é a matriz diagonal dos erros padrões da série residual, isto é,

$$\hat{\mathbf{D}} = \sqrt{\text{diag}(\hat{\boldsymbol{\Sigma}}_a)}.$$

No caso particular, para $l = 0$, tem-se que $\hat{\mathbf{C}}_0 = \hat{\boldsymbol{\Sigma}}_a$ e $\hat{\mathbf{R}}_0$ são as matrizes de covariância e correlação residual, respectivamente.

Segundo Tsay (2013), a hipótese de interesse na verificação do modelo é

$$H_0 = \mathbf{R}_1 = \mathbf{R}_2 = \dots = \mathbf{R}_m = \mathbf{0} \text{ contra } H_1 = \mathbf{R}_j \neq \mathbf{0} \text{ para alguns } 1 \leq j \leq m,$$

em que m é um inteiro positivo especificado. Sob a hipótese nula, H_0 , para a série residual, a estatística de Portmanteau ajustada é definida como

$$Q(m) = T^2 \sum_{l=1}^m \frac{1}{(T-l)} \text{tr}(\hat{\mathbf{R}}_l' \hat{\mathbf{R}}_0^{-1} \hat{\mathbf{R}}_l \hat{\mathbf{R}}_0^{-1}). \quad (2.21)$$

Sob a hipótese nula, a estatística teste $Q(m)$ tem distribuição assintoticamente qui-quadrado com $[n^2(m-p)]$ graus de liberdade (g.l). No entanto, para que o número de graus de liberdade seja positivo, m deve ser maior do que p .

2.2.2.6 Previsão

Para Tsay (2013), seja T a origem de previsão, $h \geq 1$ o horizonte de previsão e $\mathcal{F}_t$ a informação disponível até o tempo T (inclusive). Estamos interessados em prever um valor Y_{T+h} de origem T e horizonte h que será denotada por $\hat{Y}_T(h)$.

Segundo Morettin (2011), considerando, o modelo VAR(p), com parâmetros supostos conhecidos, $\mathbf{a}_t$ uma sequência i.i.d. e $\mathcal{F}_t = \{\mathbf{Y}_s : s \leq t\}$, obtemos a esperança condicional de $\mathbf{Y}_{t+h}$ dado $\mathcal{F}_t$ dada por

$E(\mathbf{Y}_{t+h}|\mathcal{F}_t) = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 E(\mathbf{Y}_{t+h-1}|\mathcal{F}_t) + \dots + \boldsymbol{\Phi}_p E(\mathbf{Y}_{t+h-p}|\mathcal{F}_t)$, pois $E(\mathbf{a}_{t+h}|\mathcal{F}_t) = \mathbf{0}$, para todo $h > 0$.

Considere o modelo VAR(p) na forma padrão como na Equação (2.2) com seus parâmetros estimados pelo método de máxima verossimilhança e denotados por $\hat{\boldsymbol{\Phi}}_0, \hat{\boldsymbol{\Phi}}_1, \hat{\boldsymbol{\Phi}}_2, \dots, \hat{\boldsymbol{\Phi}}_p$, a previsão resultante de origem T e horizonte h é

$$\hat{\mathbf{Y}}_t(h) = \hat{\boldsymbol{\Phi}}_0 + \hat{\boldsymbol{\Phi}}_1 \hat{\mathbf{Y}}_t(h-1) + \hat{\boldsymbol{\Phi}}_2 \hat{\mathbf{Y}}_t(h-2) + \dots + \hat{\boldsymbol{\Phi}}_p \hat{\mathbf{Y}}_t(h-p). \quad (2.22)$$

que $\hat{\mathbf{Y}}_t(j) = \mathbf{Y}_{t-j}$, para $j \leq 0$.

Para $h = 1$, obtemos

$$\hat{\mathbf{Y}}_t(1) = \hat{\boldsymbol{\Phi}}_0 + \hat{\boldsymbol{\Phi}}_1 \mathbf{Y}_t + \hat{\boldsymbol{\Phi}}_2 \mathbf{Y}_{t-1} + \dots + \hat{\boldsymbol{\Phi}}_p \mathbf{Y}_{t-p+1}$$

e para $h = 2$ temos

$$\hat{\mathbf{Y}}_t(2) = \hat{\boldsymbol{\Phi}}_0 + \hat{\boldsymbol{\Phi}}_1 \hat{\mathbf{Y}}_t(1) + \hat{\boldsymbol{\Phi}}_2 \mathbf{Y}_t + \dots + \hat{\boldsymbol{\Phi}}_p \mathbf{Y}_{t-p+2}$$

de modo que as previsões podem ser obtidas recursivamente.

A matriz de erro quadrado médio de previsão de horizonte h é calculada por

$$\widehat{\Sigma}(h) = \sum_{j=0}^{h-1} \widehat{\Psi}_j \widehat{\Sigma}_a \widehat{\Psi}_j', \quad (2.23)$$

em que as matrizes $\widehat{\Psi}_j$ são obtidas recursivamente por

$$\widehat{\Psi}_j = \sum_{k=1}^{p-1} \widehat{\Psi}_{j-k} \widehat{\Phi}_k, \quad (2.24)$$

com $\Psi_0 = I_n$ e $\Phi_j = 0$, $j > p$.

2.3 Resultados e Discussão

Nessa aplicação foram modeladas três séries de variáveis climáticas são elas: precipitação acumulada média (precip), temperatura mínima (tmin) e umidade relativa média (urm). Os dados foram obtidos da estação convencional Araguaina, localizada no estado do Tocantins-TO. São 264 observações mensais de cada uma das variáveis no período de janeiro de 1995 a Dezembro de 2016. O objetivo será ajustar modelos para fins de previsão, os modelos ajustados serão: univariados para cada uma das séries, modelo autorregressivo vetorial - VAR(p). O período considerado na modelagem será de agosto de 2000 a agosto de 2016, constarão de 193 observações mensais ($T=193$) e quatro observações fora da amostra ($h = 4$) para fins de seleção de modelo, que compreende os meses de setembro, outubro, novembro e dezembro de 2016. Vale ressaltar que toda a análise estatística foi realizada com o software livre R e a maior parte das funções utilizadas foram dos pacotes vars e MTS.

2.3.1 Modelagem univariada

Consideremos $\mathbf{Y}_t = (y_{1t}, y_{2t}, y_{3t})'$ uma série temporal tridimensional ($n=3$), em que y_{1t} , y_{2t} e y_{3t} representam as séries temporais precip, tmin e urm, respectivamente. O caso univariado é um caso particular da modelagem multivariada, sendo neste caso $n=1$, portanto foram ajustados três modelos. As Figuras 2.1(a), 2.2(a) e 2.3(a) apresentam os gráficos das séries temporais que foram modeladas para fins de previsão.

Estamos interessados na previsão $\mathbf{Y}_{T+4}$ baseado nas observações $\{y_1, y_2, \dots, y_{193}\}$. Para este fim, o modelo proposto será

$$\widehat{\mathbf{Y}}_t = g(y_1, y_2, \dots, y_{193}),$$

em que $\widehat{\mathbf{Y}}_{T+4}$ indica a previsão de origem $T = 193$ e horizonte $h = 4$ quatro meses a frente.

A modelagem consiste em especificar a função $g(\cdot)$ com base na avaliação das séries observadas. Nas Figuras 2.1, 2.2 e 2.3, observa-se a presença da componente sazonal nas três séries a serem modeladas. Para tanto, apresentamos no contexto univariado como modelar a presença dessa componente na série. Sendo que no contexto multivariado procede-se de modo análogo.

Figura 2.1. Gráfico da série precipitação com seu correlograma e periodograma, respectivamente.

2.3.1.1 Modelagem univariada: Modelo SARIMA

Na análise gráfica das séries Y_{it} para $i = 1, 2, 3$ a componente tendência não é observada, pois, as séries desenvolvem ao redor de um valor constante ao longo do tempo. Porém considere-se a componente sazonal, devido à forma senoidal das séries ao longo do tempo. Segundo Morettin e Toloi (2011), o modelo considerado é um Sazonal Autorregressivo Integrado Média Moveis - SARIMA de ordem $(p, d, q)(P, D, Q)[s]$. Então, Y_{it} satisfaz o modelo

$$\phi(B)\Phi(B^s)(1 - B^s)^D(1 - B)^d Y_{it} = \theta(B)\Theta(B^s)a_t \text{ para } i = 1, 2, 3 \quad (2.25)$$

em que:

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o operador autorregressivo de ordem p , estacionário;

$\theta(B) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_q B^{qs}$ é o operador de médias móveis de ordem q , invertível;

$\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$ é o operador autorregressivo sazonal de ordem P estacionário e s é indicador da periodicidade da série;

$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}$ é o operador de médias móveis sazonal de ordem Q , invertível;

(a) Gráfico da série temperatura mínima.

(b) Correlograma

(c) Periodograma

Figura 2.2. Gráfico da série temperatura mínima com seu correlograma e periodograma, respectivamente.

$\Delta_s = (1 - B^s)$ é o operador diferença sazonal;

$\Delta_s^D = (1 - B^s)^D$, D indicando o número de diferenças sazonais.

Nas Figuras 2.1(b), 2.2(b) e 2.3(b) tem-se as FAC que apresentam lags significativos ao redor dos lags 1, 12, 36 e 48. Nas Figuras 2.1(c), 2.2(c) e 2.3(c) tem-se os periodogramas com picos na frequência 16, todas essas observações juntamente com o gráficos das séries apresentadas são evidências da componente sazonal, que por sua vez foi confirmada pelo teste de periodicidade de Fisher, para detalhes ver (Morettin e Toloi, 2011). O teste tem como hipótese nula não existência de periodicidade. Obteve-se valores- p menores que 0,001 para y_{1t} , y_{2t} e y_{3t} . Neste caso as séries analisadas são sazonais com periodicidade $s = \frac{T}{f} = \frac{193}{16} = 12,0625 \cong 12$ em que f é a frequência no periodograma correspondendo ao maior pico.

Na Tabela 2.1 encontram-se os modelos que melhor se ajustaram as séries, ou seja, modelos que apresentaram menor BIC e/ou MSE .

Após a seleção do modelo adequado para cada uma das séries, foi realizado o teste Box-Ljung, cuja a hipótese nula é de que os resíduos são ruídos brancos. Para os modelos que melhor ajustaram as séries y_{1t} , y_{2t} e y_{3t} , os valores- p foram 0,1914; 0,2913 e 0,9959 respectivamente, indicando assim não haver rejeição da hipótese nula. Na Tabela 2.2 observam-se os valores

(a) Gráfico da série Umidade relativa média

(b) Correlograma

(c) Periodograma

Figura 2.3. Gráfico da série umidade relativa com seu correlograma e periodograma, respectivamente.

Tabela 2.1. Modelo ajustado seguindo o ciclo iterativo Box-Jenkins.

série	modelo ajustado
y_{1t}	SARIMA(0,0,0)(2,1,2)[12]
y_{2t}	SARIMA(2,0,0)(0,1,2)[12]
y_{2t}	SARIMA(1,0,0)(1,1,1)[12]

fora da amostra e os valores previstos pelos modelos ajustados para cada uma das três séries, precip, tmin e urm. Para as variáveis y_{2t} e y_{3t} , observa-se valores previstos muito próximos dos observados, indicando um bom ajuste. No entanto, para a série y_{1t} os valores preditos e observados apresentam uma distância considerável.

As Figuras 2.4, 2.5 e 2.6 apresentam uma análise gráfica da qualidade dos ajustes das séries y_{1t} , y_{2t} e y_{3t} . Essas figuras evidenciam um boa qualidade de ajuste, pois, as FAC e FACP dos resíduos dos modelos ajustados não possuem lags significativos uma vez que todos pertencem ao intervalo pontilhado e os valores originais e preditos das séries comportam-se de maneira semelhante ao longo do tempo.

O objetivo dos próximos ajustes será de melhorar as previsões da série y_{1t} , pois, como

Tabela 2.2. Valores observados e preditos referentes aos ajustes das séries y_{1t} , y_{2t} e y_{3t}

h	mês	y_{1t}	$\hat{y}_{1t}$	y_{2t}	$\hat{y}_{2t}$	$\hat{y}_{3t}$	$\hat{y}_{3t}$
1	Setembro	102,20	51,46	20,71	20,01	64,14	61,54
2	Outubro	80,90	107,49	21,41	21,42	74,41	70,90
3	Novembro	291,00	159,77	21,88	21,84	79,54	79,51
4	Dezembro	98,10	179,33	22,12	21,76	81,84	80,95

(a) Gráficos da série precipitação acumulada média em azul com seus respectivos valores preditos em vermelho.

(b) Fac dos resíduos .

(c) Facp dos resíduos.

Figura 2.4. Gráficos da precipitação, seus valores preditos e suas Fac e Facp residuais, respectivamente.

observou-se na modelagem univariada, foi a previsão que apresentou maior distanciamento, entre os valores observados fora da amostra de previsão e preditos.

2.3.1.2 Modelagem univariada: SARIMAX

Seguindo com o objetivo de melhorar as estimativas da série y_{1t} (precip), foram ajustados modelos SARIMAX em que o X significa variáveis exógenas. Foram feitas várias combinações para o vetor X , porém o melhor ajuste para precip, foi quando considerou conjuntamente variáveis y_{2t} e y_{3t} como exógenas, isto é, $X = \{y_{2t}, y_{3t}\}$. Os modelos que melhor se ajustaram à série y_{1t} estão representados na Tabela 2.3 com seus respectivos BIC e MAE.

(a) Gráficos da série temperatura mínima em azul com seus respectivos valores preditos em vermelho.

(b) Fac dos resíduos.

(c) Fapc dos resíduos.

Figura 2.5. Gráficos da temperatura mínima, seus valores preditos e suas Fac e Fapc residuais, respectivamente.

Tabela 2.3. Modelos SARIMAX ajustados para a séries y_{1t} com seus valores de BIC e MAE, respectivamente.

SARIMAX	BIC	MAE
SARIMAX(1,0,0)(2,1,2)[12]	1968,54	47,596
SARIMAX(1,0,1)(1,1,2)[12]	1969,80	46,4365

Para ambos os modelos, a hipótese nula de ruído branco não foi rejeitada pelo teste de Box-Ljung não foi significativo ao nível de 5%, pois o valor- p foi maior que 0,05 para os resíduos do modelo 1 e para o modelo 2, indicativo de que não há evidência contra a hipótese nula de ruído branco. Neste caso, os modelos podem ser usados para fazer previsão. Na Tabela 2.4, encontram-se os valores observados fora da amostra e os valores preditos pelos dois modelos SARIMAX ajustados da série y_{1t} .

(a) Gráficos da série Umidade relativa média em azul com seus respectivos valores preditos em vermelho.

(b) Fac dos resíduos.

(c) Facp dos resíduos.

Figura 2.6. Gráficos da urm, seus valores preditos e suas Fac e Facp residuais, respectivamente.

Tabela 2.4. Valores observados fora da amostra e valores estimados da séries y_{1t}

h	y_{1t}	modelo 1	modelo 2
1	102,20	43,91	33,98
2	80,90	92,89	74,24
3	291,00	173,58	154,57
4	98,10	170,35	159,23

2.3.2 Modelo VAR(p) sazonal

Utilizando a função VARselect do pacote vars, que tem implementado os critérios de informação para seleção da ordem de um VAR(p) apresentados na seção sobre identificação, selecionou-se os modelos SVAR(1) com constante e SVAR(2) sem constante, logo foram ajustados três modelos VAR(p) sazonal, cujas as especificações acompanhadas de seus valores de BIC, MAE e MSE são apresentados na Tabela 2.5.

Na Tabela 2.6, encontram-se os valores observados e os valores preditos pelos modelos VAR sazonal. Para ambos os modelos VAR ajustados foi realizado o teste de Portmanteau para testar a hipótese de que os erros são não correlacionados e verificou-se que em nenhum dos testes

Tabela 2.5. Modelos VAR ajustados com seus respectivos valores de BIC, MAE e MSE.

modelo	SVAR	BIC	MAE	MSE
1	SVAR(1) com constante	3.590,563	75,679	6.549,618
2	SVAR(2) sem constante	3.615,163	78,952	6.993,905
3	SVAR(2) com constante	3.590,480	78,200	6.852,555

a hipótese nula de ruído branco foi rejeitada, uma vez que os valores- p foram 0,08446 ; 0,09269 e 0,3372 , respectivamente.

Tabela 2.6. Valores observados y_{1t} e preditos $\hat{y}_{1t}$ referentes aos modelos 1, 2 e 3

mês	h	y_{1t}	modelo1	modelo 2	modelo 3
setembro	1	102,20	46,79	42,61	41,82
outubro	2	80,90	124,74	131,11	128,72
novembro	3	291,00	205,17	206,51	205,25
dezembro	4	98,10	215,75	219,60	216,94

A Tabela 2.7 e a Figura 2.7 apresentam uma explicação para a discrepância entre os valores estimados pelos modelos que buscaram modelar a precipitação e os valores deixados fora da amostra para avaliação das previsões. Percebe-se que, em qualquer um dos modelos ajustados os valores preditos e observados dentro da amostra foram em geral destoantes.

Na Tabela 2.7 tem-se o resumo estatístico da série y_{1t} , em que estão representadas as medidas: O primeiro quartil ($q1$), a mediana (md), a média ($media$) o terceiro quartil ($q3$), a distância interquartil ($dq = q3 - q1$), o limite inferior ($LI = q1 - 1,5 \times dq$) e o limite superior ($LS = q3 + 1,5 \times dq$). As medidas LI e LS serão utilizadas para definir valores atípicos na série em estudo.

Tabela 2.7. Resumo estatístico dos valores observados de y_{1t}

mês	$q1$	mediana	media	$q3$	dq	LI	LS
setembro	22,48	32,00	51,44	65,58	43,10	-42,17	130,23
outubro	94,50	133,70	127,60	162,00	67,50	-6,75	263,25
novembro	131,80	198,60	207,00	257,70	125,90	-57,05	446,55
dezembro	142,60	230,70	216,90	273,60	131,00	-53,90	470,10

A Figura 2.7 mostra como os valores da série precipitação acumulada média ($precip$) são distribuídos em cada mês ao longo dos anos em estudo. O ponto azul representa a média dentro de cada mês, o ponto vermelho a mediana e o intervalo em vermelho destaca o intervalo que vai de $q1$ a $q3$. Ao observar a Tabela 2.7, percebe-se que todos os valores fora da amostra na modelagem da série y_{1t} encontram-se abaixo de $q1$ ou acima de $q3$ e conseqüentemente distantes da média e mediana. A Figura 2.7 apresenta uma visão gráfica da distribuição da precipitação dentro de cada mês. Percebe-se que, de setembro a março, temos uma precipitação acumulada média crescente com pico máximo nos meses de fevereiro e março. Com isso, podemos concluir que, devido as instabilidades apresentadas para os meses de setembro a dezembro de 2016, quando comparado com o mesmo período dos anos anteriores. O modelo univariado ajustado com a finalidade de prever a precipitação não é um bom modelo no entanto, quando comparamos os valores preditos com a média e mediana do mês em que foi feita a previsão, percebe-se ser um

bom modelo. Porém, o modelo não foi capaz de prever valores que distam muito da média ou mediana.

Figura 2.7. Valores da série y_{1t} dentro dos meses ao longo dos anos, média, mediana, primeiro quartil e terceiro quartil plotados

2.4 Conclusões

O presente estudo mostrou que:

1. As séries $tmin$ e urm apresentam bom ajuste com a modelagem univariada;
2. As séries $tmin$ e urm não contribuíram para melhor previsão da série $precip$, pois, as previsões que utilizaram os modelos univariados SARIMAX, bem como os modelos multivariados VARX e sazonal VAR, não foram capazes de diminuir a diferença constatada na modelagem univariada entre os valores fora da amostra com os das previsões;
3. A precipitação para os meses de setembro, outubro, novembro e dezembro do ano de 2016, foram valores distantes da média e mediana dos anos anteriores para o mesmo período de previsão, ora abaixo do primeiro quartil, ora acima do terceiro quartil, com isso, as previsões não foram satisfatórias;
4. A modelagem de séries temporais tem limitação nas previsões mensais para anos com precipitação atípicas em relação aos anos anteriores.

Referências

- Box, G., Jenkins, G., Reinsel, G., e Ljung, G. (2015). *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Morettin, P. A. (2011). *Econometria financeira: um curso em séries temporais financeiras*.
- Morettin, P. A. e Tolo, C. (2011). *Análise de séries temporais*. Editora:Blucher.
- Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.
- Zivot, E. e Wang, J. (2007). *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media.

3 PREVISÃO DA MÉDIA E VARIÂNCIA CONDICIONAL DA SÉRIE DE PREÇO DO MILHO

Resumo

Nesse trabalho, exploramos técnicas para análise de séries temporais para fins de previsão. Para tanto, o objetivo foi modelar simultaneamente a estrutura média condicional e variância condicional de uma série temporal, combinando um modelo ARMA com um modelo de heterocedasticidade condicional GARCH, respectivamente. A série utilizada foi a série de preço do milho por apresentar resíduos do modelo ARMA(p, q) não correlacionados, mas no entanto, a série do quadrado dos resíduos apresentou uma estrutura de correlação. O modelo geral proposto foi um ARMA-GARCH com a componente do erro aleatório na modelagem da variância condicional suposta ter distribuição normal, t -Student, distribuição erro generalizado (GED), Skew normal, skew t -Student ou uma Skew GED. Após ajustes de modelos combinados ARMA-GARCH, o modelo selecionado foi o que apresentou menor AIC.

Palavras-chave: Volatilidade; Modelagem GARCH; Modelagem ARMA-GARCH; Previsão.

Abstract

In this work, we explored techniques for the analysis of time series for prediction purposes. To this end, the aim was to model simultaneously the structure of conditional average and conditional variance of a time series, combining a model ARMA with a conditional heteroscedasticity model GARCH, respectively. The series used was the corn price for presenting the ARMA model ($p; q$) not correlated; however, the square series of residues presented a correlation structure. The general model proposed was ARMA-GARCH with random error component in the modeling of the conditional variance expected to have normal distribution, the t -Student, generalized error distribution (GED), Skew normal, skew t -Student or skew GED. After adjustments of models combined ARMA-GARCH, the model with the lowest AIC was selected.

Keywords: Volatility; GARCH Modeling; ARMA-GARCH Modeling; Prediction.

3.1 Introdução

Segundo Box et al. (2015), a suposição feita na modelagem ARMA é que os erros de predição são variáveis aleatórias independentes com variância constante e que é independente do passado, no entanto, esta suposição parece inconsistente com a heteroscedasticidade, muitas vezes vista para séries temporais em negócios e economia, em particular.

Segundo Tsay (2013), modelos heteroscedásticos condicionais podem ser classificados em duas categorias gerais. Aqueles na primeira categoria usam uma função exata para governar a evolução de σ_t^2 , enquanto aqueles na segunda categoria usam uma equação estocástica para descrever σ_t^2 . O modelo GARCH pertence à primeira categoria, enquanto o modelo de volatilidade estocástica está na segunda categoria.

Segundo Zivot e Wang (2007), o comportamento volátil nos mercados financeiros é geralmente chamado de “volatilidade”. A volatilidade tornou-se um conceito muito importante

em diferentes áreas da teoria e prática financeira, como gerenciamento de risco, seleção de portfólio, precificação de derivativos, etc. Em termos estatísticos, a volatilidade é geralmente medida pela variância ou desvio padrão.

3.2 Material e Métodos

Para Chan e Cryer (2008), no caso mais geral, a estrutura média condicional pode ser modelada por um modelo ARMA(p, q), com o termo ruído branco, a_t , do modelo ARMA modelado por algum modelo GARCH(m, n), denominado modelo ARMA(p, q)-GARCH(m, n). Consideremos a série temporal estacionária $\{y_t \mid t = 1, \dots, T\}$ modelada de acordo com o modelo ARMA(p, q)-GARCH(m, n) expresso da seguinte forma

$$Y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + a_t \quad (3.1)$$

$$a_t = \sigma_t \varepsilon_t \quad (3.2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^n \beta_j \sigma_{t-j}^2 \quad (3.3)$$

em que, $\alpha_0 > 0$, $\alpha_m > 0$, $\alpha_i \geq 0$, $i = 1, \dots, m-1$, $\beta_n > 0$, $\beta_j \geq 0$, $j = 1, \dots, n-1$ e $\sum_{i=1}^{\max(m,n)} (\alpha_i + \beta_i) < 1$, segundo Chan e Cryer (2008), condição necessária e suficiente para a

estacionaridade de um modelo GARCH (m, n).

A Equação (3.1) é considerada a forma geral para a equação média condicional, com termos ARMA e componente do erro aleatório, a_t , suposto ser ruído branco, denominada de choque ou inovação. Na Equação (3.2), $\{\varepsilon_t\}$ é uma sequência de variáveis aleatórias distribuídas de forma independente e identicamente distribuída, cada uma com média zero e variância unitária (também conhecidas como as inovações), e ε_t é independente de a_{t-j} , $j = 1, 2, \dots$. A inovação ε_t é presumida ter variância unitária de modo que a variância condicional de a_t é igual a $\sigma_{t|t-1}^2$, isto é, $E(a_t^2 | a_{t-j}, j = 1, 2, \dots) = \sigma_{t|t-1}^2$. Neste trabalho, ε_t é suposta seguir uma distribuição normal padrão, ou uma distribuição t -Student padronizada, ou uma distribuição de erro generalizada (GED), ou uma distribuição Skew normal, Skew t -Student ou Skew GED. A Equação (3.3) exibe o modelo GARCH(m, n) considerado neste artigo o modelo geral na modelagem da variância condicional, σ_t^2 , que depende dos resíduos quadrados nos p períodos anteriores e da variância condicional nos q períodos anteriores. Aqui vale ressaltar que, não é fácil especificar as ordens (m, n) na Equação (3.3), no entanto, a prática mostrada que o modelo GARCH(1,1) ajusta-se bem na maioria das aplicações.

Para identificação das ordens do modelo GARCH, é vantajoso expressar o modelo para a variância condicional em termos de resíduos ao quadrado, para tanto, faça $\sigma_t^2 = a_t^2 - \eta_t$ e substitua na Equação (3.3), obtendo assim a expressão

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,n)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{j=1}^n \beta_j \eta_{t-j} \quad (3.4)$$

em que a Equação (3.4) tem a forma de um modelo ARMA[$\max(m, n), n$] para a_t^2 , no entanto, $\{\eta_t\}$ é uma série não-correlacionada com média zero, mas não é iid e a variância incondicional de a_t é

$$\text{var}(a_t) = E(a_t^2) = \alpha_0 \left\{ 1 - \left[\sum_{i=1}^{\max(m, n)} (\alpha_i + \beta_i) \right] \right\}^{-1} \quad (3.5)$$

que é finita se

$$\sum_{i=1}^{\max(m, n)} (\alpha_i + \beta_i) < 1. \quad (3.6)$$

3.2.1 Identificação

Na modelagem ARMA-GARCH, as ordens ARMA podem ser identificadas com base na série temporal, y_t , enquanto as ordens GARCH podem ser identificadas com base nos resíduos quadrados do modelo ARMA ajustados. Para Tsay (2013), construir uma equação média condicional equivale a remover a média da amostra dos dados, nesse caso, seja $a_t = y_t - \mu_t$ os resíduos da equação média. A série $\{a_t^2\}$ é então utilizada para verificar a heteroscedasticidade condicional, também conhecida como efeito ARCH. Segundo Zivot e Wang (2007), antes de estimar um modelo ARCH completo para uma série temporal, é geralmente uma boa prática testar a presença de efeitos ARCH nos resíduos. Se não houver efeitos ARCH nos resíduos, o modelo ARCH é desnecessário e não especificado.

3.2.2 Estimação

Para Chan e Cryer (2008), o preço de não especificar corretamente a forma distributiva da inovação é uma perda na eficiência da estimativa, embora, com grandes conjuntos de dados, a conveniência computacional da abordagem de probabilidade gaussiana possa superar a perda de eficiência da estimativa.

Supondo na Equação (3.3), $n = 0$ e que a_t tenha distribuição t -Student padronizada com ν graus de liberdade, denotado por $a_t \sim t_\nu$. Nesse caso, para Box et al. (2015) e Tsay (2013), a estimação conjunta dos parâmetros ν e dos parâmetros nas equações média e variância condicional, apresentados nas Equações (3.1) e (3.3), respectivamente, podem ser estimados pelo método de máxima verossimilhança condicional, sendo o $\ln$ da função de verossimilhança expresso como se segue

$$\begin{aligned} \ln(L) = n \left\{ \ln \left[\Gamma \left(\frac{\nu + 1}{2} \right) \right] - \ln \left[\Gamma \left(\frac{\nu}{2} \right) \right] - \frac{1}{2} \ln [\pi(\nu - 2)] \right. \\ \left. - \frac{1}{2} \sum_{t=1}^n \left[\ln(\sigma_t^2) + (1 + \nu) \ln \left(1 + \frac{a_t^2}{(\nu - 2)\sigma_t^2} \right) \right] \right\} \end{aligned} \quad (3.7)$$

em que a_t^2 e σ_t^2 , podem serem obtidos recursivamente utilizando as equações:

$$a_t = y_t - \phi_0 - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j a_{t-j} \quad e$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \dots + \alpha_n \sigma_{t-n}^2, \text{ respectivamente.}$$

Aqui vale ressaltar que utilizou-se o pacote fGarch do R para ajustar o modelo ARMA(p, q)-GARCH(m, n) aos dados da série temporal.

3.2.3 Diagnósticos

Para Chan e Cryer (2008), antes de aceitar um modelo ajustado e interpretar as suas descobertas, é essencial verificar se o modelo está corretamente especificado, ou seja, se os pressupostos do modelo são suportados pelos dados. Se alguns pressupostos do modelo principal parecem ser violados, então um novo modelo deve ser especificado; instalado e verificado novamente até encontrar um modelo que forneça um ajuste adequado aos dados.

Denotemos o desvio padrão condicional estimado por, $\hat{\sigma}_{t|t-1}$. Segundo Chan e Cryer (2008) os resíduos padronizados definidos como

$$\varepsilon_t = \frac{a_t}{\hat{\sigma}_{t|t-1}}. \quad (3.8)$$

Os resíduos padronizados são muito úteis para verificar a especificação do modelo. A suposição de normalidade das inovações pode ser explorada plotando o gráfico de escores normais quantil-quantil.

3.2.4 Previsão

Para Zivot e Wang (2007), uma importante tarefa de modelagem da volatilidade condicional é gerar previsões precisas tanto para o valor futuro de uma série temporal quanto da volatilidade condicional.

Denotando a origem de previsão por t e a informação por $\mathcal{F}_t$, segundo Tsay (2005), a previsão de h -passos à frente de y_{t+h} pode ser obtida do modelo

$$\hat{y}_t(h) = E(y_{t+h}|\mathcal{F}_t) = \phi_0 + \sum_{i=1}^p \phi_i \hat{y}_t(h-i) + \sum_{i=1}^q \theta_i a_t(h-i) \quad (3.9)$$

em que: $\hat{y}_t(h-j) = y_{t+h-j}$ para $h-j \leq 0$, $a_t(h-j) = 0$ se $h-j > 0$ e $a_t(h-j) = a_{t+h-j}$ se $h-j \leq 0$. Considerando um modelo GARCH(1,1), dados os parâmetros α_0 , α_1 e β_1 , a variância condicional pode ser computada recursivamente pela expressão

$$\sigma_{t|t-1}^2 = \alpha_0 + \alpha_1 a_{t-1|t-2}^2 + \sigma_{t-1|t-2}^2, \text{ para } t \geq 2. \quad (3.10)$$

com valor inicial, $\sigma_{1|0}^2$, definido sob a hipótese de estacionaridade como variância incondicional estacionária $\sigma_t^2 = \alpha_0(1 - \alpha_1 - \beta_1)^{-1}$. Para uma expressão geral para a previsão da variância condicional, veja Chan e Cryer (2008) página 297.

A modelagem para fins de previsão consistirá em 4 passos: Primeiro, ajusta-se um modelo ARMA(p, q) para modelar a equação média condicional, em seguida, obtém-se os resíduos e teste a hipótese de serem ruído branco. Segundo, sendo os resíduos do modelo ARMA(p, q) ajustados, ruído branco, considere a série $\{a_t^2\}$ para averiguar a existência de heteroscedasticidade condicional, conhecido como feito ARCH, caso detecte o efeito ARCH prossiga. Terceiro, especificar um modelo para volatilidade condicional e em seguida estimar conjuntamente as

equações da média e variância condicional. Quarto, obter os resíduos padronizados para testar o modelo especificado e refinar caso necessário.

3.3 Resultados e Discussão

A série utilizada neste trabalho foi do preço diário do milho em reais, denotada por $\{y_t | t = 1, \dots, T\}$, observada no período de 04/01/2010 a 30/04/2013, constando de $T = 828$ valores, foi extraída do site do CEPEA/ESALQ. A série preço do milho amostrada foi dividida em duas subamostras. A primeira, constando das primeiras 800 observações, compreendendo o período de 04/01/2010 à 20/03/2013, denominada de subamostra de estimação (ou modelagem). A segunda, constituída das últimas 28 observações, compreende o período de 21/03/2013 à 30/04/2013, denominada de subamostra de previsão. O nível de significância adotado nesse trabalho foi $\alpha = 0,05 = 5\%$, valor esse que será comparado com o valor-p para a rejeição, ou não, da hipótese nula. Toda análise estatística foi realizada utilizando o software estatístico livre R com uso de pacotes e funções específicas para modelagem de séries temporais (R CORE TEAM, 2017).

A Figura 3.1(a) apresenta a série do preço do milho na primeira diferença e as Figuras 3.1(b) e 3.1(c) são a FAC e FACP da referida série. A FAC e a FACP mostram correlações significativas nas primeiras defasagens, sendo assim, necessária ajustar um modelo ARMA(p, q) para modelar a estrutura média condicional.

A Tabela 3.1 apresenta os modelos ajustados, os valores- p do teste de Ljung-Box, com hipótese nula de ruído branco para a série de resíduos e a estatística do teste Ljung-Box aplicado a série dos resíduos ao quadrado, denotada por $Q(30)$ em que foi considerada as 30 primeiras correlações.

Tabela 3.1. Valores- p e estatísticas de Ljung-Box das séries dos resíduos e quadrado dos resíduos no ajuste da a média condicional, respectivamente

modelo	valor-p	Q(30)
ARMA(6,0)	0,7259	257,79*
ARMA(1,2)	0,4875	317,37*
ARMA(1,4)	0,3709	318,84*
ARMA(3,3)	0,5903	268,04*

* teste significativo

A estatística de Ljung-Box para os resíduos mostra que os modelos ajustados para modelar a equação da média condicional apresentam bons ajustes, um vez que não há evidência contra a hipótese nula de ruído branco para a série $\{a_t\}$ dos modelos ajustados, visto que os valores- p são maiores que o nível de significância do teste que é de 5% ($\alpha = 0,05$). A Figura 3.2 apresenta a FAC do quadrado dos resíduos para cada um dos modelos ajustados e mostra a existência de uma correlação significativa. O teste de Ljung-Box da série $\{a_t^2\}$, cuja hipótese é que não há efeito ARCH, constatou a presença do efeito ARCH, um vez que os valores- p são menores que $\alpha = 0,05$, rejeitando assim a hipótese nula. Sendo assim foram estimados conjuntamente os parâmetros das equações da média e variância condicional. Vale ressaltar que no ajuste do

(a) Gráfico da série preço do milho na primeira diferença

(b) FAC (ou correlograma) da série milho diferenciada

(c) FACP da série milho na primeira diferença

Figura 3.1. As linhas tracejadas são os limites de confiança $\pm \frac{1.96}{\sqrt{T}}$ das correlações e correlações parciais da série diferenciada.

Tabela 3.2. Modelo ARMA-GARCH ajustados e seus respectivos AIC

modelo	AIC
ARMA(6,0)+(1,1)	-0,35902
ARMA(1,2)+(1,1)	-0,34960
ARMA(1,4)+(1,1)	-0,35372
ARMA(3,3)+(1,1)	-0,36021*

* modelo selecionado

modelo ARMA(p, q)-GARCH(m, n) foram consideradas todas as possíveis distribuições para a componente do erro aleatório da Equação (3.2), isto é, a série $\{\epsilon_t\}$ foi considerada ter uma distribuição normal, t -Student, GED, skew normal, skew t -Student e skew GED, sendo que a distribuição t -Student foi a que proporcionou um menor AIC para os quatro modelos ajustados.

A Tabela 3.2, apresenta os modelos ARMA-GARCH ajustados e seus respectivos valores de AIC. O modelo selecionado foi ARMA(3,3)-(1,1) por apresentar menor valor AIC.

A Figura 3.3 apresenta a FAC e FACP dos resíduos do modelo ARMA(3,3)-GARCH(1,1).

(a) Modelo ARMA(6,0).

(b) Modelo ARMA(1,2).

(c) Modelo ARMA(1,4).

(d) Modelo ARMA(3,3).

Figura 3.2. Gráficos das funções de autocorrelações do quadrado dos resíduos referente aos modelos ARMA ajustados .

A FAC e FACP mostrar que o modelo selecionado ajusta-se bem aos dados, o que foi confirmado pelo teste de Ljung-Box aplicado nas séries dos resíduos padronizados e no quadrado dos resíduos padronizados, apresentando valores- p 0,60298 e 0,52840, respectivamente.

(a) Modelo ARMA(3,3)-GARCH(1,1).

(b) Modelo ARMA(3,3)-GARCH(1,1).

Figura 3.3. FAC e FACP dos resíduos padronizados do modelo ARMA(3,3)-GARCH(1,1).

A Figura 3.4, apresenta o histograma e o gráfico QQ-plot dos resíduos padronizados do modelo selecionado ARMA(3,3)-GARCH(1,1). O histograma revela uma assimetria positiva, enquanto que o QQ-plot evidência um excesso de curtose positiva. Uma distribuição com excesso

Figura 3.4. Histograma e QQ-plot dos resíduos padronizados do modelo ARMA(3,3)-GARCH(1,1).

de curtose positiva é dito ter caudas pesadas, implicando que a distribuição coloca mais massa na cauda do seu suporte do que uma distribuição normal. O revelado pelo histograma e QQ-plot dos resíduos é confirmado pelos valores obtidos da assimetria e curtose que são 1,048773 e 11,41639, respectivamente.

3.4 Conclusões

Ao final da modelagem da média condicional e da variância condicional da série do preço diário do milho, pode-se concluir que:

1. A modelagem ARMA-GARCH foi necessária e o modelo ARMA(3,3)-GARCH(1,1) com componente do erro aleatório seguindo uma distribuição *t*-Student ajustou-se bem aos dados;
2. A análise gráfica do histograma e QQ-plot dos resíduos padronizados revelaram uma assimetria positiva e um excesso de curtose positiva significativa;

Referências

- Box, G., Jenkins, G., Reinsel, G., e Ljung, G. (2015). *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley.
- Chan, K. e Cryer, J. (2008). *Time series analysis with applications in r*.
- Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.
- Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.
- Zivot, E. e Wang, J. (2007). *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media.

4 SELEÇÃO DOS MODELOS ARIMA E WAVDS-ARIMA VIA METODOLOGIA BACKTESTING PARA FINS DE PREVISÃO DA *COMMODITY* MILHO

Resumo

A modelagem de série temporal para fins de previsão serve como subsídio para o tomador de decisões, e devido a importância, vários são os trabalhos publicados com a finalidade de aprimorar a precisão das previsões. Neste artigo, temos como objetivo, apresentar as vantagens de um pré-processamento (ou filtragem) na série antes da modelagem, desenvolvida por (de Melo, 2018) e divulgar uma metodologia não muito utilizada para seleção de modelos para fins de previsão, denominada *backtesting*. A *commodity* milho pela sua importância no cenário brasileiro foi escolhida como base de dados para a modelagem. Antes da modelagem foi utilizado uma nova metodologia de pré-processamento via *Wavelets* da série preço diário do milho, denominado Sistema de Decomposição Wavelet-WavDS, originando uma nova série. Na modelagem tanto da série filtrada quanto da série não-filtrada, foram utilizados modelos $ARIMA(p, d, q)$. Na fase de seleção de modelos, em ambas modelagens para fins de previsão, utilizou-se a metodologia *backtesting*. De posse dos dois modelos selecionados pode verificar-se que a filtragem WavDS aplicada na série de preço do milho proporcionou uma redução significativa no Erro Quadrático Médio-EQM, redução essa em torno de 86% e consequentemente um aumento na acurácia.

Palavras-chave: Série temporal; Pré-processamento; Transformada Wavelet; Modelagem ARIMA; Backtesting; Previsão.

Abstract

Time series modelling for prediction purposes serves as an aid for the decision-maker. Thus, many works have been published aiming to improve accuracy of forecasts. In this study, we present the advantage of a preprocessing (or filtering) in the series before the modeling, developed by (de Melo, 2018) and spread a methodology not much used for selection of models for forecasting purposes, named Backtesting. Corn was chosen as the basis for data modeling for its importance in the Brazilian scenario. Before the modeling, a new methodology for preprocessing via Wavelets daily corn price series was used, called the Wavelet-WavDS Decomposition System, giving rise to a new series. In the modeling of the filtered and unfiltered series, ARIMA models (p, d, q) were used. In the selection of models, in both modeling systems for forecasting purposes, the Backtesting methodology was used. Both models show that the WavDS filtering applied on corn price series provided a significant reduction in Mean Square Error-NDE by roughly 86%, increasing thus its accuracy.

Keywords: Time series; Preprocessing; Raw Wavelet; ARIMA Modeling; Backtesting; Forecasting.

4.1 Introdução

Segundo Box et al. (2015), uma série temporal é um conjunto de observações tomadas sequencialmente no tempo. Os modelos de séries temporais buscam modelar a estrutura de correlação temporal e são excelentes para fins de previsão de variáveis climáticas. Para Chan e Cryer (2008), um dos principais objetivos da construção de um modelo para uma série temporal é prever os valores para essa série em tempos futuros. De igual importância é a avaliação da precisão dessas previsões. Para Tsay (2014), não existe um modelo verdadeiro para uma determinada série temporal. Todos os modelos estatísticos são aproximações utilizadas para descrever a dependência dinâmica dos dados. É então comum ver que vários modelos se encaixam bem em um dado conjunto de dados, o que requer a comparação de modelos.

O Brasil possui uma posição de destaque no cenário mundial na produção de milho, por se tratar de um produto de origem primária é classificado como uma *commodity*, sendo assim, seu preço é determinado no mercado financeiro, isto é, cotado na bolsa de valores. O preço do milho, em grande parte, influencia o comportamento de determinados setores econômicos. Isso significa que as oscilações em seu preço influenciam outras atividades, como por exemplo, a suinocultura, a avicultura, a pecuária e, em geral, todos os produtos que contém milho na sua composição. Por isso, ter modelos que nos permitem prever o preço futuro do milho é de suma importância na tomada de decisão.

A modelagem da série de preços diários de milho possui três objetivos. Primeiro, apresentar a metodologia *backtesting*, útil para seleção de modelos para fins de previsão de variáveis climáticas. Segundo, utilizando a metodologia *Backtesting* com previsão um passo a frente, encontrar um modelo apropriado que seja útil para previsão do preço diário do milho. Terceiro, verificar se uma filtragem na série antes de realizar a modelagem aumenta a acurácia de previsão como conclui (de Melo, 2018).

4.2 Material e Métodos

4.2.1 Análise wavelet

Uma *Wavelet* pode ser definida como uma função de oscilação no tempo e está definida no espaço de Hilbert. A Transformada *Wavelet* (TW) pode decompor séries temporais em combinações lineares de funções *Wavelets*. Existem várias famílias de funções *Wavelets*, cada uma satisfazendo uma série de propriedades, a partir das quais suas formulações são obtidas. A TW é uma importante ferramenta matemática na análise de séries temporais (Sidney, 1998). Sua representação envolve a decomposição da série temporal $f(\cdot)$ no tempo $t \in \mathbb{R}$ em funções bases *Wavelets* $\psi(t)$ dadas por:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (4.1)$$

em que $a, b \in \mathbb{R}$ com $a > 0$.

Ao representar uma série temporal pelos coeficientes *Wavelets*, através das transformadas, os parâmetros utilizados para representar os coeficientes são escala, a , relativo a frequência

e translação, b , relativo ao tempo de amostragem considerado. A alteração dos parâmetros a e b geram várias *Wavelets*, $\psi(t)$, que são comumente chamadas *Wavelet*-mãe (Morettin, 2014).

A descoberta de bases ortonormais na forma da Equação (4.1) de suporte compacto, sintetizou a família de *Wavelets* e possibilitou uma análise mais eficiente do que a obtida com outros sistemas de funções (Sidney, 1998). Para algumas escolhas especiais de ψ , as funções $\psi_{a,b}$ que estão apresentadas na Equação (4.1) constituem uma base ortonormal para o espaço de funções do $L^2(R)$. Uma maneira de gerar as *Wavelets* com estas características é a utilização da função escala $\phi(t)$ que é a solução da seguinte equação:

$$\phi(t) = \sqrt{2} \sum_k l_k \phi(2t - k). \quad (4.2)$$

Segundo Sidney (1998) e Morettin (2014), a função ϕ dada na Equação (4.2) gera uma família ortonormal em $L^2(R)$,

$$\phi_{j,k}(t) = 2^{\frac{j}{2}} \phi(2^j(t - k)), \text{ com } j, k \in \mathbb{Z}. \quad (4.3)$$

Nestas condições, a *Wavelet*-mãe ψ pode ser obtida a partir da função escala ϕ por meio da seguinte equação:

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k), \quad (4.4)$$

de modo que:

$$h_k = (-1)^k l_{k-1}. \quad (4.5)$$

Os coeficientes l_k e h_k são coeficientes de filtros digitais FIR - filtros de resposta a impulso de duração finita - que também são conhecidos como filtros passa-altas e filtros passa-baixas, respectivamente. Estes filtros são utilizados para calcular a Transformada *Wavelet* Discreta (TWD) (Verma e Verma, 2012) e (Morettin, 2014).

Os coeficientes são calculados por meio das seguintes expressões:

$$l_k = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \phi(2t - k) dt, \quad (4.6)$$

$$h_k = \sqrt{2} \int_{-\infty}^{\infty} \psi(t) \psi(2t - k) dt. \quad (4.7)$$

Contudo, considerando o sistema ortonormal $\{\phi_{j,k}(t), \psi_{j,k}(t)\}$, qualquer função $f(t) \in L^2(R)$ pode ser escrita de acordo com a seguinte expressão (Morettin, 2014)

$$f(t) = \sum_k cA_{j_0,k} \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_k cD_{j,k} \psi_{j,k}(t), \quad (4.8)$$

em que cA e cD representam conjuntos de coeficientes *Wavelets* e j_0 representa a escala mais suave.

Os coeficientes de altas frequências são chamados de coeficientes de detalhes (cD) e os de baixas frequências são os coeficientes de aproximações (cA). Estes coeficientes possibilitam separar a série temporal em componentes que podem ser considerados detalhes e ruídos (altas frequências), assim como componentes que representam o comportamento fundamental da série

(componentes de baixas frequências) em vários níveis de resolução. Em virtude disto, tem-se o nome Análise de Múltipla Resolução (AMR) (Sidney, 1998) e (de Souza et al., 2007).

A teoria de Multirresolução dada por Mallat (1989) prova que filtros conjugados que caracterizam uma *Wavelet* são gerados a partir de bases ortonormais do $L^2(R)$, e que a TW é uma implementação do tipo cascata destes filtros conjugados. Uma das diversas maneiras de se introduzir a teoria sobre funções *Wavelets* ortonormais é através do conceito de AMR que tem como principal objetivo oferecer uma infraestrutura para que as funções em $L^2(R)$ possam ser representadas dependendo de diferentes escalas (ou níveis de resolução) (Mallat, 1989).

A decomposição *Wavelets* de uma série temporal $f(t)$ obtida por meio da teoria de multirresolução pode ser realizada por meio de filtros FIR (Verma e Verma, 2012). O esquema da decomposição *Wavelets* baseados em bancos de filtro FIR é apresentado na Figura 4.1. Nela, são utilizados dois filtros de decomposição *Wavelets*, um filtro passa alta e outro passa baixa, seguidos de um *down sampling* por dois que produz metade dos dados de entrada em altas e baixas frequências (Morettin, 2014).

Figura 4.1. Algoritmo de decomposição de Mallat.

Após a realização da decomposição o sinal pode ser reconstruído com a utilização da Transformada *Wavelets* Inversa (TWI). A TWI essencialmente realiza as mesmas operações em uma direção contrária. Aos mesmos coeficientes são aplicados a TWI em ordem reversa utilizando o algoritmo piramidal de reconstrução. Desta forma, ao invés de ocorrer um procedimento de decimação, as séries temporais são interpoladas. Desta forma caracteriza-se um *upsampling* por dois (notação: $\uparrow 2$), ou seja, zeros são colocados entre os coeficientes da transformada (Mallat, 1989). A representação da reconstrução da série de dados da função $f(t)$ por meio da estrutura de banco de filtros FIR no 4º nível de resolução é ilustrado na Figura 4.2.

A relação entre os níveis de resolução pode ser descrita sob a forma de um aninhamento dos espaços de funções, em que cada espaço de maior resolução contém espaços de menor resolução (Morettin, 2014).

Para de Melo (2018), a escolha da *Wavelets*-mãe é fundamental para que o procedimento de filtragem seja bem sucedido. As *Wavelets* de *Haar* e *Daubechies* frequentemente produzem bons resultados e em virtude disto são as mais utilizadas na literatura, contudo existem diversas outras.

Figura 4.2. Algoritmo de reconstrução de Mallat.

4.2.1.1 Arquitetura do pré-processamento via wavelets

O principal trabalho que envolve filtragem que é baseado na limiarização dos coeficientes *Wavelets* de uma série temporal foi apresentado por Donoho (1995). O método é dependente da forma como o ruído é comumente presente nas estruturas mais finas da série temporal e o fato da TW fornecer uma decomposição em várias escalas é o fator primordial para a adequabilidade do procedimento.

A maior parte do ruído é representada pelos coeficientes *Wavelets* nas escalas mais finas e, descartando estes coeficientes, o resultado é uma filtragem natural do ruído (Joy et al., 2013). Deste modo, de acordo com Verma e Verma (2012), o esquema que envolve a filtragem *Wavelet* segue três principais etapas:

1. A decomposição da série temporal por meio da TW;
2. A estimação do limiar e a limiarização dos coeficientes *Wavelets*;
3. A aplicação da TWI no mesmo nível para a reconstrução da série temporal filtrada.

Segundo Donoho (1995), existem dois procedimentos de filtragem wavelet que promovem a limiarização de cada coeficiente *Wavelet* $d_{j,k}$. O *Hard Thresholding* (thr_{λ}^H) limiariza os coeficientes que estão abaixo de um valor específico de limiar λ ajustando-os para zero, enquanto que o *Soft Thresholding* (thr_{λ}^S) ajusta para zero os coeficientes abaixo do limiar λ bem como simplesmente escolhe outros coeficientes que estão acima do valor da escala ou do limiar. Toda a formulação matemática, pode ser encontrada em Joy et al. (2013). Apesar de atualmente existirem outros métodos de truncamento de coeficientes *Wavelets*, os métodos de truncamento *Hard* e *Soft Thresholdings* foram utilizados nesse artigo, para detalhes ver Joy et al. (2013).

Outro aspecto imprescindível é a estimação do limiar λ , esta é outra questão fundamental neste processo. Existem quatro métodos padrões de estimação deste limiar: *minimaxi*, *sqtwlog*, *rigsure* e *heursure*. Em Joy et al. (2013) pode ser encontrada suas respectivas formulações matemáticas assim como informações complementares.

4.2.2 O sistema de decomposição wavelets - WavDS

Aqui utiliza-se a metodologia de pré-processamento de séries temporais intitulado WavDS, proposta por Melo (2018), sendo realizado em duas etapas.

Na primeira etapa, é realizada a tradicional filtragem wavelet unidimensional proposta por Donoho (1995). Com a proposta, o ruído branco, aqui considerado é gaussiano, deve ser eliminado por meio da filtragem *Wavelet*. Nesta etapa, a série temporal, y_t , é dada de entrada no procedimento tradicional de filtragem via DW. Inicialmente, faz-se uma seleção dos parâmetros a e b , necessários para o procedimento de filtragem, enumerados à seguir:

- i. Regra de seleção de limiar;
- ii. Critério de estimação de limiar;
- iii. Escolha da *Wavelet*-mãe;
- iv. Nível de decomposição *Wavelet*.

Após a definição da regra de seleção do limiar λ e do critério de estimação do limiar λ , a função *Wavelet* que será utilizada como *Wavelet*-mãe, e o nível da DW possível (1 nível), devem ser heurísticamente definidos para dar início ao procedimento de filtragem. Neste momento da primeira etapa deve-se escolher a função *Wavelet* que será utilizada como *Wavelet*-mãe. Após a primeira etapa do procedimento ser realizado, o ruído da filtragem, e_t , é gerado e analisado como critério de parada para a primeira etapa. Caso o ruído tenha as características pertinentes a um ruído branco gaussiano a primeira etapa do método WavDS é concluída a série pré-processada é a *Wavelets*-mãe selecionada. Este procedimento está ilustrado na Figura 4.3.

Figura 4.3. Arquitetura do Sistema de Decomposição *Wavelet* - WavDS.

Caso o ruído da primeira etapa não tenha as características pertinentes a um ruído branco gaussiano, uma nova seleção de parâmetros é realizada a fim de que o critério de parada estabelecido para conclusão da primeira etapa seja encontrado. A modificação dos parâmetros para satisfazer o critério de parada segue a mesma ordem dada anteriormente para a seleção dos parâmetros. Analisam-se as combinações possíveis envolvendo os parâmetros i , ii e iii antes de modificar o nível de decomposição (parâmetro iv).

Por meio do método proposto, a *Wavelet*-mãe, $\psi(t)$, que proporcionar esta eliminação do ruído branco no menor nível de decomposição (na primeira etapa do WavDS) é a *Wavelet* selecionada como mãe para a segunda etapa do procedimento.

A segunda etapa do WavDS consiste na utilização da decomposição *Wavelet* no primeiro nível nos dados obtidos a partir da realização da primeira etapa do procedimento. Utiliza-se a TW e a respectiva TWI nos dados a fim de separar as componentes da série temporal em componentes de altas e baixas frequências (Ababneh et al., 2013).

Na segunda etapa do procedimento WavDS, o conjunto de dados é subdividido em dois subconjuntos por meio da TW, um formado pelos coeficientes de aproximações (tendências - componentes de baixas frequências) que é responsável pelo principal comportamento da série e outro pelos coeficientes de detalhes (variações - componentes de altas frequências), responsáveis pelos pequenos ajustes (Ababneh et al., 2013). Assim, de maneira similar aos diferentes trabalhos (Ababneh et al., 2013), descartam-se os coeficientes de detalhes (cD) que são considerados outros tipos de ruído e utiliza-se a série dos coeficientes de aproximação (cA) na TWI para produzir os dados pré-processados (y'_t) a serem utilizados no modelo de previsão.

4.2.3 Modelo autorregressivo integrado médias móveis - ARIMA

Um pré-requisito para modelar um série temporal é que ela seja estacionária, isto é, a série se desenvolva ao longo do tempo em torno de uma média constante e uma variação estável, porém, é muito comum depararmos com séries não-estacionárias. Uma maneira de lidar com esse problema é usar a diferenciação, nesse caso, transformando assim uma série não-estacionária em estacionária. Se uma série temporal, Y_t , após ser diferenciada d vezes segue um processo ARMA(p, q) estacionário e invertível, dizemos nesse caso que Y_t segue um processo ARIMA(p, d, q) e pode ser representado como

$$\phi(B)\Delta^d Y_t = \theta(B)a_t, \quad (4.9)$$

em que y_t e a_t são o valor atual e o erro aleatório no período de tempo t , respectivamente.

$\Delta = (1 - B)$ é o operador diferença e d um número natural;

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o operador autorregressivo de ordem p ;

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ é o operador médias móveis de ordem q ;

a_t é uma série temporal de ruído branco, isto é, uma sequência de variáveis aleatórias independentes e identicamente distribuídas-(iid) com média zero e variância constante, σ_a^2 , denotado por $a_t \sim \text{RB}(0; \sigma_a^2)$.

O operador defasagem, denotado por B , opera no índice de tempo de uma série temporal defasando uma unidade de tempo para formar uma nova série. De modo geral,

$$B^k Y_t = Y_{t-k}, \text{ para } k = 1, \dots, p.$$

Vale ressaltar que: Primeiro, o modelo na Equação (4.9) emprega $p + q + 2$ parâmetros desconhecidos $\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2$, que são estimados a partir dos dados. Segundo, a_t que denota a nova informação no tempo t da série temporal e é muitas vezes referida como a inovação ou choque no tempo t . Terceiro, por se tratar de uma série de ruído branco, a função de autocorrelação da série a_t , denotada por $\rho_k(a)$, é igual a zero para todo k maior que zero. Quarto, a cada diferença a série perde uma observação, sendo assim, após d diferenças, tem-se d observações a menos. Em outras palavras, se a série Y_t tem T valores observados, então a série transformada, $\Delta^d Y_t$, terá tamanho $n = T - d$.

Verificado que os polinômios $\phi(B)$ e $\theta(B)$ quando fatorados não possuem fatores comuns, pois, caso contrário a ordem pode ser reduzida, e constatado que as raízes das equações características $\phi(B) = 0$ e $\theta(B) = 0$, são todas maiores que um em valor absoluto, então o modelo ARIMA é fracamente estacionário e invertível, respectivamente.

A Equação (4.9) pode ser representada da seguinte forma

$$\varphi(B)Y_t = \theta(B)a_t \quad (4.10)$$

em que

$$\varphi(B) = \phi(B)(1 - B)^d = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d},$$

é um operador autorregressivo não-estacionário, de tal forma que d das raízes da equação característica $\varphi(B) = 0$ são a unidade e as p restantes ficam fora do círculo unitário. A Equação (4.10) será denominada, forma de equação de diferença. O uso direto da equação de diferença nos permite expressar o valor atual, Y_t , do processo em termos de seus valores passados e atuais de a_t . O modelo geral dado na Equação (4.10), com $\phi_0 = 0$, pode ser escrito na forma de equação de diferença como

$$Y_t = \varphi_1 Y_{t-1} + \dots + \varphi_{p+d} Y_{t-p-d} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (4.11)$$

Vale ressaltar que o objetivo reside na previsão e para representar a equação de previsão, a equação de diferença, Equação (4.11), é a forma mais simples de ser utilizada nas previsões de valores fora da amostra e será a equação utilizada nesse trabalho para tal finalidade. Uma forma conveniente de escrever a Equação (4.11), utilizada para calcular a variância dos erros de previsão, na forma da decomposição de Wold, uma vez que toda série temporal pode ser escrita na forma do filtro linear

$$\begin{aligned} Y_t &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \dots \\ &= a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \\ &= \psi(B)a_t, \end{aligned} \quad (4.12)$$

cuja entrada é uma sequência de choques não-correlacionados com média zero e variância comum σ_a^2 , em outras palavras, um ruído branco. Substituindo na Equação (4.10) o valor de Y_t encontrado na Equação (4.12), temos a seguinte equação

$$\varphi(B)\psi(B) = \theta(B). \quad (4.13)$$

Assim, podemos escrever a Equação (4.13) na forma

$$(1 - \varphi_1 B - \dots - \varphi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 - \theta_1 B - \dots - \theta_q B^q). \quad (4.14)$$

Segundo Box et al. (2015), os pesos ψ_j do processo ARIMA podem ser obtidos recursivamente através das equações

$$\psi_j = \varphi_1\psi_{j-1} + \varphi_2\psi_{j-2} + \dots + \varphi_{p+d}\psi_{j-p-d} - \theta_j, \text{ para } j > 0 \quad (4.15)$$

com $\psi_0 = 1$, $\psi_j = 0$ para $j < 0$, e $\theta_j = 0$ para $j > q$.

Uma outra forma de reescrever a Equação (4.9) é a seguinte

$$\phi(B)Z_t = \theta(B)a_t \quad (4.16)$$

em que $Z_t = \Delta^d Y_t$. Assim, observa-se que o modelo ARIMA(p, d, q) pode ser representado por um processo ARMA(p, q) estacionário e invertível.

4.2.4 Previsão

Para Chan e Cryer (2008), com base no histórico disponível da série temporal até o instante t , isto é, observa-se $\{Y_t, t = 1, 2, \dots, T\}$. Deseja-se prever o valor de Y_{t+h} que ocorrerão h unidades de tempo no futuro. Denotamos o tempo T a origem da previsão, h o intervalo de tempo de previsão e a previsão por $\hat{Y}_t(h)$. A previsão do erro quadrático médio mínimo é dado por

$$\hat{Y}_t(h) = E(Y_{t+h}|Y_1, Y_2, \dots, Y_T) = E(Y_{t+h}|\mathcal{F}_t), \quad (4.17)$$

em que $\mathcal{F}_t$ denota a informação disponível até o instante t . Segundo Box et al. (2015), uma observação Y_{t+h} gerada pela processo ARIMA(p, d, q) descrito pela Equação (4.11) pode ser expressa da seguinte forma

$$Y_{t+h} = \varphi_1 Y_{t+h-1} + \varphi_2 Y_{t+h-2} + \dots + \varphi_{p+d} Y_{t+h-p-d} + a_{t+h} - \theta_1 a_{t+h-1} - \theta_2 a_{t+h-2} - \dots - \theta_q a_{t+h-q}. \quad (4.18)$$

Tomando a esperança condicional, dada a informação até o instante t , na Equação (4.18), temos

$$\hat{Y}_t(h) = \sum_{j=1}^{p+d} \varphi_j \hat{Y}_t(h-j) - \sum_{j=1}^q \theta_j a_{t+h-j} \quad (4.19)$$

em que: $\hat{Y}_t(h-j) = Y_{t+h-j}$ para $h-j \leq 0$, $a_t(h-j) = 0$ se $h-j > 0$ e $a_t(h-j) = a_{t+h-j}$ se $h-j \leq 0$.

Nesse trabalho, para atingir o propósito desejado, utilizou-se dois ciclos dinâmicos. O primeiro, na fase de modelagem da série temporal amostrada utilizou a metodologia Box-Jenkins. O segundo ciclo iterativo, que compõe a parte de previsão fora da amostra, utilizou-se uma metodologia denominada na literatura de séries temporais de *backtesting*. Após a segunda fase, de posse dos erros de previsão, que serão úteis para calcular as estatísticas que irão subsidiar o pesquisador na tomada de decisão sobre qual modelo escolher para fins de previsão, calculou-se os valores dos critérios de seleção de modelos.

Na seção (4.2.5), será apresentado o ciclo iterativo de Box-Jenkins para modelagem das observações da série constituída da subamostra de estimação. Na seção (4.2.6) será apresentada

a metodologia de previsão um passo à frente, denominada de *backtesting* e na seção (4.2.7) as estatísticas de seleção de modelos, o Erro Absoluto Médio - EAM, o Erro Percentual Absoluto Médio - EPAM e o Erro Quadrático Médio de Previsão - EQMP.

4.2.5 Modelagem de séries temporais: metodologia Box-Jenkins

Consideremos a série estacionária observada, y_t , a modelagem $ARIMA(p, d, q)$ segue um ciclo iterativo de identificação, estimação e diagnóstico.

A primeira etapa do ciclo, constitui-se basicamente em identificar a ordem do modelo, em outras palavras, os valores dos inteiros não-negativos p e q . Na prática com auxílio das Funções amostrais de Autocorrelação - FAC, Autocorrelação Parcial - FACP e Autocorrelação Estendida - FACE, vários valores de p e q podem ser propostos, cuja combinação dos mesmos constitui um conjunto de modelos propostos nessa fase. Vale enfatizar que a dificuldade de encontrar um único valor para p e q , reside na subjetividade das técnicas gráficas e na dificuldade inerente da modelagem.

Para construir o gráfico da FAC amostral, o correlograma, calculamos as correlações entre y_t e y_{t-k} , denotada por $\hat{\rho}_k$, denominadas autocorrelações estimadas, dadas por

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad 0 \leq k \leq n-1 \quad (4.20)$$

em que $\bar{y}$ é média da série estacionária y_t , $n = T - d$, T o tamanho da subamostra utilizada na estimação do modelo e d o número de diferença. O gráfico das correlações amostrais $\hat{\rho}_k$ versus a defasagem k é denominado de correlograma. Para Tsay (2014), sob condições gerais $\hat{\rho}_k$ é um estimador consistente de ρ_k .

Segundo Wei (2006), a FACP amostral, denotada por $\hat{\phi}_{k,k}$, mede a correlação entre y_t e y_{t-k} ajustada para os valores intermediários $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$, sendo estimado por um método recursivo proposto por Durbin, iniciando com $\hat{\phi}_{11} = \hat{\rho}_1$ e para $k > 0$, $\hat{\phi}_{kk}$ será calculada iterativamente utilizando a expressão

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j} \quad (4.21)$$

em que

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{k,j} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j}, \quad j = 1, \dots, k.$$

A FAC é útil na determinação da ordem p do modelo autorregressivo puro $AR(p)$ e a FACP é útil na determinação da ordem q do modelo média móveis puro $MA(q)$, porém, não são informativas na determinação da ordem de um modelo $ARMA$. No entanto, Box e Tiao propõem uma nova abordagem que usa a FACE para especificar a ordem de um processo $ARMA$.

A versão teórica da FACE para um modelo ARMA(p, q) é uma tabela de dupla entrada, em que as linhas correspondem a parte AR de ordem p e as colunas a parte MA de ordem q . A principal característica da tabela é que ela contém um triângulo de “0” zeros com vértice superior esquerdo com coordenada (p, q) , para mais detalhes veja Tsay e Tiao (1984) e Tsay (2014).

Na segunda etapa, estima-se os parâmetros dos modelos propostos, sendo que a estimação pode ser realizada pelos métodos de Mínimos Quadrados - MQ ou Máxima verossimilhança - MV. Nesse artigo, o método utilizado foi o método de MV. Segundo Chan e Cryer (2008), para qualquer conjunto de observações, $\{Y_t \mid t = 1, 2, \dots, n\}$, de séries temporais ou não, a função de verossimilhança L é definida como sendo a densidade de probabilidade conjunta de obtenção dos dados efetivamente observados. Contudo, é considerado como uma função dos parâmetros desconhecidos no modelo com os dados observados mantidos fixos. Para modelos ARIMA, L será função dos parâmetros $\{\phi_i \mid i = 1, \dots, p\}$, $\{\theta_j \mid j = 1, \dots, q\}$, μ e σ_a^2 dado os valores observados da série $\{Y_t \mid t = 1, \dots, n\}$. Neste caso, os estimadores de MV são definidos como os valores dos parâmetros para os quais os dados realmente observados são mais prováveis, isto é, os valores que maximizam a função de verossimilhança. Para mais detalhes, ver Box et al. (2015) e Hamilton (1994).

Na terceira etapa, especificados os valores p e q na fase de identificação e estimados os parâmetros do modelo via método de MV na fase de estimação, procede-se o diagnóstico do modelo ajustado. Considere que um modelo ARIMA(p, d, q) escrito na forma da Equação (4.16) foi ajustado, denominaremos às quantidades

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)Z_t \quad (4.22)$$

como os resíduos. Segundo Box et al. (2015), os resíduos são computados recursivamente a partir de $\hat{\phi}(B)Z_t = \hat{\theta}(B)\hat{a}_t$ como

$$\hat{a}_t = Z_t - \sum_{i=1}^p \hat{\phi}_i Z_{t-i} + \sum_{j=1}^q \hat{\theta}_j \hat{a}_{t-j}, \text{ para } t = 1, \dots, n. \quad (4.23)$$

A série temporal $\{\hat{a}_t \mid t = 1, \dots, n\}$ é denominada de série residual, de onde obtemos a estimativa da variância, σ_a^2 , da seguinte forma

$$\hat{\sigma}_a^2 = \frac{1}{n - p - q - 1} \sum_{t=1}^n \hat{a}_t^2, \quad (4.24)$$

e utilizando as Equações (4.20) e (4.21) pode-se calcular as funções de autocorrelação e autocorrelação parcial amostrais da série residual, denotadas por $\hat{\rho}_k(\hat{a})$ e $\hat{\rho}_{kk}(\hat{a})$, respectivamente. Sabe-se que, se a série residual $\{\hat{a}_t \mid t = 1, \dots, n\}$ comportar-se como um ruído branco, então $\hat{\rho}_k(\hat{a})$ é assintoticamente normal com média 0 e variância $\frac{1}{n}$ para qualquer inteiro positivo fixo k , isto é,

$$\hat{\rho}_k \overset{A}{\sim} N\left(0; \frac{1}{n}\right), \text{ para } k > 0.$$

Segundo Wei (2006), em 1949 Quenouille demonstrou que, sob a hipótese de que o processo subjacente é uma sequência de ruído branco, a variância de $\hat{\rho}_{kk}(\hat{a})$ pode ser aproximada por

$$\text{var}[\hat{\rho}_{kk}(\hat{a})] \simeq \frac{1}{n}. \quad (4.25)$$

Segundo Chan e Cryer (2008), um modelo ajustado deve ser examinado cuidadosamente para verificar a possível inadequação do mesmo. Se ele for adequado, a série residual deve se comportar como um ruído branco.

A primeira verificação de diagnóstico é inspecionar um gráfico dos resíduos ao longo do tempo. Se o modelo for adequado, espera-se que o gráfico tenha dispersão retangular em torno de um nível horizontal zero sem nenhuma tendência. Em seguida, uma forma visual para verificar a significância da autocorrelação residual, $\rho_k(a)$, é plotar o correlograma das autocorrelações residuais estimadas $\hat{\rho}_k(\hat{a})$ versus k , com os limites $\pm \frac{1.96}{\sqrt{n}}$. Caso o valor $\hat{\rho}_k(\hat{a})$ exceda o limite tem-se um forte indicativo da significância de $\rho_k(a)$. Finalmente, o pressuposto será verificado pela aplicação do teste modificado proposto por Ljung-Box. Segundo Box et al. (2015), supondo que tenhamos as primeiras autocorrelações $\hat{\rho}_k(\hat{a})$ ($k = 1, \dots, K$) de um modelo ARIMA(p, d, q), com estatística de teste expressa da seguinte forma

$$Q = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2(\hat{a})}{n-k}, \quad (4.26)$$

em que $n = T - d$, sendo T o tamanho da subamostra utilizada na estimação do modelo, d o número de diferenças e $\hat{\rho}_k(\hat{a})$ as autocorrelações residuais estimadas. A estatística Q tem, aproximadamente distribuição qui-quadrado com $(K - p - q)$ graus de liberdade. Rejeita-se a hipótese nula de ajuste adequado do modelo se o valor-p for menor que o nível de significância adotado pelo pesquisador.

Nesse momento vale ressaltar que, se houver a necessidade de que a série seja diferenciada para satisfazer o pré-requisito para ser estacionária, geralmente uma diferença, isto é $d = 1$, é suficiente nas maiorias das aplicações para conseguir tal propósito. A verificação desse pressuposto será realizada, calculando o valor absoluto das raízes da equação característica $\phi(B) = 0$, com auxílio da função `polyroot` do programa R.

Ao final da terceira etapa, os modelos que seguirão para a fase de previsão e seleção são os modelos cujos resíduos são ruído branco.

4.2.6 Previsão fora da amostra

A metodologia *backtesting*, consiste em comparar o desempenho dos modelos concorrentes, utilizando as previsões um passo à frente fora do conjunto de dados utilizados na modelagem, denominadas de previsões fora da amostra.

Segundo Tsay (2014), considerando o conjunto de dados da série temporal observada ou amostrada, $\{y_t \mid t = 1, \dots, T-1, T\}$, em que T denota o tamanho da série observada, o procedimento *backtesting* é executado da seguinte maneira:

1. Divida o conjunto de dados em duas subamostras de estimação (modelagem) e previsão, de tamanhos N e $H = T - N$, respectivamente. Aqui vale enfatizar, que cada subamostra deve conter dados suficientes para que a modelagem e as estatísticas utilizadas na seleção de modelo, possam ser precisas.

2. A partir da subamostra de modelagem, $\{y_t | t = 1, \dots, N\}$, realize o ajuste do modelo, utilize-o para fazer previsão um passo à frente e obtenha o erro de previsão de um passo à frente, $e_{N+1}(1) = Y_{N+1} - \hat{Y}_N(1)$. Observe que, o ponto dos dados Y_{N+1} pertence a subamostra de previsão, ou seja, é um ponto fora da subamostra utilizada na modelagem.
3. Avance a subamostra de estimativa em um ponto de dados, ou seja, $\{y_t | t = 1, \dots, N+1\}$, ajuste novamente o modelo usando $N+1$ pontos de dados, faça a previsão um passo à frente, $\hat{Y}_{N+1}(1)$, e calcule o erro de previsão, $e_{N+2}(1) = Y_{N+2} - \hat{Y}_{N+1}(1)$. Sendo Y_{N+2} o segundo ponto de dados na subamostra de previsão.
4. Repita o processo de modelagem e previsão um passo à frente até o penúltimo ponto da subamostra de previsão, obtendo assim, o erro de previsão no instante $t = N+H$, calculado a partir da expressão

$$e_{N+H}(1) = Y_{N+H} - \hat{Y}_{N+H-1}(1).$$

Observe que, ao final do processo foram geradas duas séries, ou vetores de dados. O primeiro, constituído das previsões fora da amostra, denotado por $\{\hat{Y}_t | t = N+1, \dots, N+H\} = \{\hat{Y}_{N+1}, \dots, \hat{Y}_{N+H}\}$. O segundo, o vetor de erros de previsões um passo à frente, constituído da diferença entre o vetor de dados da subamostra de previsão e o vetor constituído das previsões um passo à frente fora da amostra, denotado por $\{e_t | t = N+1, \dots, N+H\}$.

4.2.7 Comparação de modelos

Segundo Zivot e Wang (2007), o *backtesting* é frequentemente utilizado para comparar a precisão da previsão de dois ou mais modelos concorrentes. De posse da série de erros de previsão, $\{e_{N+1}, \dots, e_{N+H}\}$, as estatísticas de avaliação das previsões são calculadas para um conjunto de m modelos concorrentes.

O Erro Absoluto Médio - EAM

$$\text{EAM}(m) = \frac{1}{T-N} \sum_{t=N+1}^{N+H} |e_t(1)|. \quad (4.27)$$

O Erro Percentual Absoluto Médio - EPAM (%)

$$\text{EPAM}(m) = \left[\frac{1}{T-N} \sum_{t=N+1}^{N+H} \left| \frac{e_t(1)}{Y_t} \right| \right] \times 100. \quad (4.28)$$

O Erro Quadrático Médio - EQM

$$\text{EQM}(m) = \frac{1}{T-N} \sum_{t=N+1}^{N+H} [e_t(1)]^2. \quad (4.29)$$

Em que $Y_t = \{y_t | t = N+1, \dots, N+H\}$ é a subamostra de previsão de tamanho H . Geralmente, as estatísticas de avaliação (4.27), (4.28) e (4.29) são calculadas para cada modelo, e o modelo que produz o menor conjunto de estatísticas é considerado o melhor modelo. Neste trabalho, selecionou-se o modelo com o menor EQM como o melhor modelo para fins de previsão.

4.3 Resultados e Discussão

A série utilizada neste trabalho foi do preço diário do milho em reais, denotada por $\{y_t | t = 1, \dots, T\}$, observada no período de 04/01/2010 à 30/04/2013, constando de $T = 828$ valores, obtidos do site do CEPEA/ESALQ. A série preço do milho amostrada foi dividida em duas subamostras. A primeira, constando das primeiras 800 observações, compreendendo o período de 04/01/2010 à 20/03/2013, denominada de subamostra de estimação (ou modelagem). A segunda, constituída das últimas 28 observações, compreende o período de 21/03/2013 à 30/04/2013, denominada de subamostra de previsão. O nível de significância adotado aqui foi $\alpha = 0,05$, valor que será comparado com o valor-p para a rejeição ou não da hipótese nula. Toda análise estatística foi realizada utilizando o software estatístico livre R com uso de pacotes e funções específicas para modelagem de séries temporais (Team, 2014).

A Figura 4.4(a) apresenta o gráfico das séries preço do milho filtrada e não-filtrada e as Figuras 4.4(b) e 4.4(c) seus respectivos correlogramas, gráficos das autocorrelações amostrais contra as defasagens. Vale destacar na Figura 4.4(a) a aproximação dos valores das duas séries, pois, raramente conseguimos diferenciar uma da outra devido a sobreposição de seus gráficos. A inspeção gráfica das Figuras 4.4(a), 4.4(b) e 4.4(c) permite-nos concluir que ambas as séries são não-estacionárias, pois, o gráfico das séries não apresentam um desenvolvimento em torno de um valor constante e os correlogramas não apresentam um decaimento exponencial ou senoidal amortecida, sendo essas características gráficas de séries estacionárias.

Para garantir o pressuposto de estacionariedade, foi realizada a transformação nas séries não-filtrada e filtrada do preço diário do milho tomando a primeira diferença, $(1 - B)y_t = y_t - y_{t-1}$. De posse da série diferenciada, $(1 - B)y_t$, consideremos agora um modelo geral ARIMA $(p, 1, q)$ na modelagem da série temporal, seguiremos os passos do ciclo iterativo de identificação, estimação e diagnóstico: Os modelos selecionados da série filtrada e não-filtrada apresentados nas Tabelas 4.1 e 4.2, respectivamente, são de valores p e q propostos na análise das FAC, FACP e FACE de cada uma das séries.

Nas Tabelas 4.1 e 4.2, todos os modelos foram estimados via MV e são estacionários e invertíveis, pois, para cada um deles foram calculadas as raízes das equações características, $\phi(B) = 0$ e $\theta(B) = 0$ e todas as raízes apresentaram valor em módulo maior que 1. Além do mais, apresentam as estatísticas de seleção de modelos, o EAM, o EPAM e a raiz quadrada do EQM.

Na fase de diagnóstico dos resíduos dos modelos estimados, todos os modelos apresentados nas Tabelas 4.1 e 4.2, tem a componente do erro aleatório, a_t , ruído branco, pelo teste de Ljung-Box todos apresentaram valor-p maior que o nível de significância de 5%, com isso não há evidência contra a hipótese nula de ruído branco da componente a_t . Em particular para os modelos selecionados, apresentados nas Equações (4.30) e (4.31), os valores-p foram 0,3303 e 0,4254, respectivamente. A Figura (4.5), apresenta os correlogramas dos resíduos dos modelos selecionados na modelagem das séries de preços diários filtrada e não-filtrada, respectivamente. Indicando ser um bom ajuste, pois, todos as autocorrelações residuais estimadas até a defasagem 30 mantem se dentro do intervalo $[-\frac{1,96}{\sqrt{n}} ; \frac{1,96}{\sqrt{n}}]$ para ambas as séries.

Na fase de seleção de modelos, utilizou-se a metodologia *backtesting* para fazer previsões um passo a frente e de posse dos erros de previsões calculou-se as estatísticas de seleção, o EAM,

o EPAM e o EQM, que estão apresentadas nas Tabelas 4.1 e 4.2. O modelo selecionado foi o que apresentou menor raiz do EQM.

(a) gráfico da série preço do milho em azul e o gráfico da série filtrada do preço diário do milho em vermelho

(b) Correlograma da série não-filtrada

(c) Correlograma da série filtrada

Figura 4.4. As linhas azuis tracejadas nas Figuras 4(b) e 4(c) são os limites $\pm \frac{1,96}{\sqrt{n}}$.

Tabela 4.1. Modelos selecionados a partir da subamostra de estimação referente a série filtrada

Modelo	EAM	EPAM	REQM
ARIMA(5,1,3)	0,0236	0,0887	0,0299
ARIMA(5,1,4)	0,0239	0,0900	0,0294*
ARIMA(5,1,5)	0,0242	0,0911	0,0300

* modelo selecionado utilizando a raiz quadrada do EQM

O modelo ajustado ARIMA(5,1,4), selecionado na modelagem da série filtrada é expresso na Equação (4.30), enquanto, a Equação (4.31) apresenta o modelo ajustado ARIMA(1,1,2) selecionado na modelagem da série não-filtrada e suas respectivas variâncias residuais estimadas.

O modelo ajustado é

Tabela 4.2. Modelos selecionados a partir da subamostra de estimação referente a série não-filtrada

Modelo	EAM	EPAM	REQM
ARIMA(2,1,1)	0,1641	0,6106	0,2049
ARIMA(6,1,0)	0,1668	0,6220	0,2044
ARIMA(1,1,2)	0,1628	0,6056	0,2036*

* modelo selecionado utilizando a raiz quadrada do EQM

(a) Correlograma da série residual do modelo ajustado ARIMA(5,1,4) referente a série filtrada (b) Correlograma da série residual do modelo ajustado ARIMA(1,1,2) referente a série não-filtrada

Figura 4.5. As linhas azuis tracejadas nas Figuras 5(a) e 5(b) são os limites $\pm \frac{1,96}{\sqrt{n}}$ das autocorrelações residuais estimadas.

$$(1 - 0,8123 B - 1,0037 B^2 + 0,9526 B^3 - 0,3814 B^4 + 0,2620 B^5)(1 - B)y_t = (1 - 1,5211 B + 1,8836 B^2 + 1,0406 B^3 + 0,3710 B^4) a_t. \quad (4.30)$$

com variância amostral igual a $\hat{\sigma}_a^2 = 1,349 \times 10^{-3}$.

$$(1 - 0,8721 B)(1 - B)y_t = (1 - 0,8037 B + 0,1351 B^2) a_t \quad (4.31)$$

com variância amostral igual a $\hat{\sigma}_a^2 = 5,663 \times 10^{-2}$.

4.4 Conclusões

Após a filtragem da série preço diário do milho via WavDS, modelar tanto a série filtrada quanto a série não-filtrada, utilizando modelos ARIMA(p, d, q), selecionar o melhor para fins de previsão pela metodologia *Backtesting* e calcular a raiz quadrada do EQM, REQM, pode-se concluir que:

A filtragem via WavDS permitiu uma diminuição expressiva no REQM de previsão, pois, quando comparados os REQM das modelagens da série preço do milho (não-filtrada) com a série filtrada a redução foi de 86%.

Mesmo trabalhando com modelos de previsões diferentes o resultado obtido com o aumento na precisão da previsão com uso da filtragem WavDS nesse artigo corrobora com os obtidos por (de Melo, 2018).

Referências

- Ababneh, F., Al Wadi, S., e Ismail, M. T. (2013). Haar and daubechies wavelet methods in modeling banking sector. In *International Mathematical Forum*, volume 8, pages 551–566. HIKARI.
- Box, G., Jenkins, G., Reinsel, G., e Ljung, G. (2015). *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics. Wiley.
- Chan, K. e Cryer, J. (2008). Time series analysis with applications in r.
- de Melo, M. P. S. (2018). Modelo híbrido inteligente baseado em filtragem wavelet e redes neurais com aplicações à previsão de séries temporais. *Universidade Rural de Pernambuco*, 1(105):1–105.
- de Souza, F., Pagamisse, A., Monico, J. F. G., e Polezel, W. G. C. (2007). Comparação das bases de wavelets ortonormais e biortogonais: implementação, vantagens e desvantagens no posicionamento com gps. *Trends in Applied and Computational Mathematics*, 8(1):149–158.
- Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627.
- Hamilton, J. D. (1994). *Time series analysis*, volume 2. Princeton university press Princeton.
- Joy, J., Peter, S., e John, N. (2013). Denoising using soft thresholding. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 2(3):1027–1032.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693.
- Morettin, P. A. (2014). *Ondas e ondaletas: da análise de Fourier à análise de ondas e ondaletas de séries temporais*. São Paulo: EDUSP.
- Sidney, B. C. (1998). Introduction to wavelets and wavelet transforms: a primer.
- Team, R. C. (2014). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2013.
- Tsay, R. S. (2014). *An introduction to analysis of financial data with R*. John Wiley & Sons.
- Tsay, R. S. e Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models. *Journal of the American Statistical Association*, 79(385):84–96.

- Verma, N. e Verma, A. (2012). Performance analysis of wavelet thresholding methods in denoising of audio signals of some indian musical instruments. *International Journal of Engineering Science and Technology*, 4(5):2040–2045.
- Wei, W. W. (2006). *Time series analysis: univariate and multivariate*. Pearson Education.
- Zivot, E. e Wang, J. (2007). *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media.