



UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIA E TECNOLOGIA
CURSO DE ESTATÍSTICA

IANNE PORFIRIO DE QUEIROZ

**MODELAGEM DA AUTOCORRELAÇÃO RESIDUAL PELO
MÉTODO DOS MÍNIMOS QUADRADOS GENERALIZADOS COM
APLICAÇÃO AO MERCADO DE CARNE DE FRANGO**

Campina Grande

2014

IANNE PORFIRIO DE QUEIROZ

**MODELAGEM DA AUTOCORRELAÇÃO RESIDUAL PELO
MÉTODO DOS MÍNIMOS QUADRADOS GENERALIZADOS COM
APLICAÇÃO AO MERCADO DE CARNE DE FRANGO**

Trabalho de Conclusão de Curso apresentado à banca examinadora no curso de Bacharelado em Estatística da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Tiago Almeida de Oliveira

**Campina Grande
2014**

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

Q38m Queiroz, Ianne Porfirio de.

Modelagem da autocorrelação residual pelo método dos mínimos quadrados generalizados com aplicação ao mercado de carne de frango [manuscrito] / Ianne Porfirio de Queiroz. - 2014.
36 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

"Orientação: Prof. Dr. Tiago Almeida de Oliveira, Departamento de Estatística".

1. Regressão múltipla. 2. Autocorrelação residual. 3. Mínimos Quadrados Generalizados. 4. Produção agroindustrial.
I. Título. 21. ed. CDD 519.5

IANNE PORFIRIO DE QUEIROZ

**MODELAGEM DA AUTOCORRELAÇÃO RESIDUAL PELO
MÉTODO DOS MÍNIMOS QUADRADOS GENERALIZADOS COM
APLICAÇÃO AO MERCADO DE CARNE DE FRANGO**

Trabalho de Conclusão de Curso apresentado à banca examinadora no curso de Bacharelado em Estatística da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

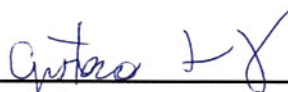
Aprovado em: 18 / 12 / 2014

BANCA EXAMINADORA

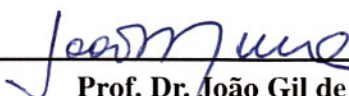


Prof. Dr. Tiago Almeida de Oliveira
(Orientador)

Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Gustavo Henrique Esteves
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. João Gil de Luna
Universidade Estadual da Paraíba (UEPB)

RESUMO

O objetivo deste trabalho consistiu no estudo da influência do mercado de carne bovina sobre o preço da carne de frango, dois importantes bens da produção agroindustrial brasileira, considerando a hipótese de que esta relação é quantitativamente relevante. Para isto, ajustou-se um Modelo de Regressão Linear Múltipla, tomando-se o preço do frango em função das variáveis que caracterizam o mercado de carne bovina, além da adição de variáveis independentes de aspectos mais gerais. Utilizou-se o procedimento *stepwise* combinado com o Critério de Akaike (AIC) para seleção do melhor ajuste, cujo poder de explicação foi avaliado pelo Coeficiente de Determinação ajustado. O modelo obtido foi submetido à validação através da análise de resíduos, com uso de técnicas gráficas, além de testes como de Shapiro-Wilk, de Durbin-Watson e de Breusch-Pagan. Verificada a autocorrelação residual como única violação aos pressupostos da análise, recorreu-se à estimação via método de Mínimos Quadrados Generalizados (GLS), considerando as variáveis explicativas selecionadas no melhor ajuste. Modelou-se a matriz de covariâncias com um modelo autorregressivo de segunda ordem, obtido a partir de procedimento iterativo. Este ajuste foi avaliado a partir do $Pseudo - R_{GLS}^2$, métrica desenvolvida especificamente para avaliação de modelos GLS. O modelo final, sob o ponto de vista teórico, ajustou-se de maneira razoável aos dados: apenas 30,25% das variações da variável respostas foram explicadas pelo modelo. Sob a perspectiva prática, no entanto, trata-se de resultado plausível, uma vez que apenas variáveis que caracterizam o mercado de carne bovina (e sua interseção com o mercado de carne de frango) foram consideradas no estudo, verificando-se, portanto, a relevância da interação entre os dois setores. Sugere-se, para melhor quantificação desta relação, a adição de outras variáveis, além de abordagens alternativas.

Palavras-chaves: Regressão Múltipla. Setor de Carnes. Autocorrelação Residual. Mínimos Quadrados Generalizados.

ABSTRACT

The objective of this work consisted of studying the influence of the beef market on the price of chicken, two important goods of the Brazilian agribusiness production, considering the hypothesis that this link is quantitatively relevant. For this, it was adjusted a multiple linear regression model, by taking the price of chicken depending on variables that characterize the beef market, plus the addition of independent variables more generally. It was used the procedure stepwise combined with the Akaike Criterion (AIC) for the best fit selection, which explanatory power was evaluated by the coefficient of determination adjusted. The obtained model was submitted to validation by residual analysis, using graphical techniques, as well as tests like Shapiro-Wilk, Durbin-Watson and Breusch-Pagan. Checked residual autocorrelation as a single violation of the analytical approach, it was used the estimation through Generalized Least Squares (GLS) method, considering the explanatory variables selected in the best fit. The covariance matrix was modeled with an autoregressive model of second order, obtained from iterative procedure. This adjustment was evaluated by *Pseudo* - R_{GLS}^2 , metric developed specifically for GLS models evaluation. The final model, from a theoretical point of view, set in a reasonable manner to the data: only 30.25% of the variations of the response variable were explained by the model. Under practical perspective, however, this is a plausible result, since only variables that characterize the beef market (and its intersection with the chicken meat market) were included in the study, verifying therefore the relevance of the interaction between the two sectors. We suggest, to better quantify this relationship, the addition of other variables, as well as alternative approaches.

Key-words: Multiple Regression. Meat industry. Residual autocorrelation. Generalized Least Squares.

SUMÁRIO

1	INTRODUÇÃO	6
2	CONSIDERAÇÕES SOBRE O MODELO MICROECONÔMICO . . .	8
2.1	As equações de oferta e demanda	8
2.2	Análise de equilíbrio geral	8
2.3	Determinação do modelo teórico	9
3	CONSIDERAÇÕES SOBRE O MODELO DE REGRESSÃO LINEAR MÚLTIPLA	11
3.1	O modelo teórico	11
3.2	Estimação dos parâmetros pelo método dos mínimos quadrados ordiná- rios	12
3.3	Testes de Hipóteses e Intervalos de Confiança	13
3.3.1	Intervalos de confiança e predição	14
3.4	O Coeficiente de Determinação	15
3.5	Seleção de variáveis	15
3.6	Validação e diagnósticos do modelo	17
3.6.1	Multicolinearidade e singularidade	17
3.6.2	Homogeneidade das variâncias	18
3.6.3	Independência dos erros	18
3.6.4	Análise da normalidade	19
3.6.5	Linearidade	21
3.6.6	Observações influentes	21
3.7	Estimação dos parâmetros pelo método dos mínimos quadrados genera- lizados	21
3.7.1	Avaliação do ajuste pelo método GLS	23
4	MATERIAL E MÉTODOS	24
4.1	Caracterização das variáveis	24
4.2	Procedimentos para ajuste e análise do modelo	25
5	RESULTADOS E DISCUSSÃO	27
5.1	Ajuste do modelo	28
5.2	Análise de Resíduos	29
5.3	Ajuste via mínimos quadrados generalizados	31
6	CONCLUSÕES	34
	Referências	35

1 INTRODUÇÃO

O agronegócio possui grande participação na economia brasileira, de modo que, em 2012, sua participação no Produto Interno Bruto do país foi de cerca de 22,15%. O Brasil demonstra grande expressividade mundial no setor, ocupando, o primeiro lugar na exportação de carne bovina e de aves.¹

Além da liderança verificada no cenário internacional, o sistema agroindustrial de carnes demonstra grande força no mercado interno brasileiro, que possui grande potencial de consumo em razão dos hábitos alimentares do país, do aumento populacional e do aumento da renda e do consumo interno após Plano Real (CARVALHO; BACHI, 2007). Em especial, a carne de frango, devido à redução de preços relativos aos outros tipos de carne e a mudanças no consumo de alimentos, com maior preocupação com aspectos de saúde. Sobre a evolução do consumo de carnes por parte dos brasileiros, Carvalho e Bachi (2007) apontam que: até a década de 1970, a carne bovina correspondia a mais da metade do total de carnes consumido pelos brasileiros, enquanto que a suína ocupava a segunda posição e a de frango, a terceira; nos anos de 1980, a maior preocupação com uma alimentação mais saudável estimulou o consumo de carnes brancas; na última virada de século, a carne de frango superou a de origem bovina no consumo das famílias brasileiras, alcançando consumo médio anual, por indivíduo, de 35 quilos, rebaixando a carne de origem suína à terceira posição.

Dessa maneira, uma vez que as carnes de frango e bovina são as mais consumidas pelo brasileiro, a quantificação da relação entre os mercados destes dois produtos torna-se importante. Com base no conhecimento desta interação, é possível desenvolver políticas públicas eficientes que valorizem estes setores, fundamentais para a matriz exportadora brasileira. Além disso, para as empresas que atuam no setor de carnes e, em geral, produzem ambos os produtos, torna-se factível uma decisão ótima para alocação de recursos de acordo com os eventos externos à produção.

Neste contexto, o objetivo deste trabalho consistiu em analisar a influência do mercado de carne bovina sobre o preço da carne de frango, considerando a hipótese de que esta é relevante. Para isto, ajustou-se um Modelo de Regressão Linear Múltipla (MRLM), tomando-se o preço do frango em função de variáveis que caracterizam o mercado de carne bovina. Modelou-se a autocorrelação residual verificada via estimação pelo método dos mínimos quadrados generalizados. Desde já, faz-se necessário reconhecer as limitações da análise proposta, visto que o setor agropecuário compreende grande variedade de sistemas agroindustriais. A vasta produção primária brasileira corrobora para que outros setores exerçam influência sobre a interação entre os mercados considerados. No entanto, em razão da operacionalização do estudo, bem como da disponibilidade dos dados, estas restrições mostraram-se inevitáveis.

Este trabalho consiste de cinco capítulos, além desta introdução: o Capítulo 2 tece breves

¹ Dados extraídos do portal da Ecoagro, disponível em: <http://www.ecoagro.agr.br>. Acesso em: 11 de dezembro de 2014

considerações sobre o modelo microeconômico utilizado; o Capítulo 3 retrata alguns aspectos relevantes do MRLM; o Capítulo 4 descreve os materiais e procedimentos computacionais utilizados; o Capítulo 5 apresenta os resultados da análise; o Capítulo 6 traz as conclusões.

2 CONSIDERAÇÕES SOBRE O MODELO MICROECONÔMICO

2.1 As equações de oferta e demanda

De acordo com Pyndick e Rubinfeld (2005, p.18), a curva de oferta de um bem é a relação entre a quantidade que os produtores estão dispostos a um determinado preço, sendo dada por:

$$Q_S = Q_S(P), \quad (2.1)$$

em que Q_S se refere à quantidade demandada e $Q_S(P)$ corresponde ao preço correspondente. Visto que, quanto maior o preço, maior a disposição dos produtores em ofertar o bem, a curva de oferta tem comportamento ascendente. Outros fatores também afetam a curva de oferta, como os custos dos fatores de produção - preço das matérias prima; taxa de juros (custo do capital); salários (custo de mão de obra) - deslocando-a no plano conforme efeitos de expansão ou contração (PYNDICK e RUBINFELD, 2005). Uma maneira alternativa de representar esta relação é através da função de oferta inversa, que descreve o preço em função da quantidade:

$$P_S = Q_S + W + P_{mp} + r, \quad (2.2)$$

em que: P_S é o preço de oferta; Q_S é a quantidade ofertada; W representa os salários; P_{mp} é o preço das matérias primas; r é a taxa de juros. Outro fator que altera a oferta é a tecnologia. No entanto, esta é considerada fixa no curto prazo.

A equação de demanda, por sua vez, representa a quantidade que os consumidores desejam comprar a um determinado preço. A curva possui formato descendente, sendo descrita por:

$$Q_D = Q_D(P). \quad (2.3)$$

Outro fatores que afetam a demanda são: os preços de bens substitutos; de bens complementares; hábitos; preferências; renda. Visto que hábitos e preferências podem ser considerados fixos no curto prazo, pode-se representar a função de demanda inversa por:

$$P_D = Q_D + P_{sb} + P_C + Y, \quad (2.4)$$

em que: P_D é o preço de demanda; Q_D é quantidade demanda; P_{sb} é o preço de bens substitutos; P_C , o preço de bens complementares; Y é a renda. Conforme Pindick e Rubinfeld (2005), o equilíbrio de mercado ocorre ante a igualdade de quantidades ofertadas e demandadas a um determinado preço. Assim, o preço de mercado é determinado de forma conjunta da oferta e da demanda, pelo ponto de interseção das duas curvas.

2.2 Análise de equilíbrio geral

A análise da equação de oferta e demanda para um único bem é útil para determinação de equilíbrio parcial. De acordo com Pindyck e Rubinfeld (2005):

Frequentemente, uma análise de equilíbrio parcial é suficiente para a compreensão do comportamento de mercado. Entretanto, as inter-relações entre os mercados podem ser importantes. (...) Diferentemente do que ocorre com a análise de equilíbrio parcial, a **análise de equilíbrio geral** determina os preços e as quantidades em todos os mercados simultaneamente; além disso, ela explicitamente leva em conta os efeitos de feedback. Um efeito de feedback é um ajuste de preço ou de quantidade em determinado mercado causado pelos ajustes de preços ou de quantidades em mercados correlatos. (PINDYCK; RUBINFELD, 2005, p. 499-500)

Com estruturação cada vez mais complexa das relações entre economias e entre setores específicos nas últimas décadas, houve aprofundamento do interesse pelo tema. Especialmente, no pós-guerra, com a ascendência do planejamento econômico, a interação entre mercados ganhou cada vez mais espaço como objeto de estudo de análises econômicas.

2.3 Determinação do modelo teórico

Especificamente para o mercado de carne bovina, Medeiros (2006) sugere a equação linear para a oferta inversa:

$$P_{boi} = P_{bezerro} + P_{boimagro} + P_{frango} + P_{dolar} + IPP + Q_S, \quad (2.5)$$

em que: P_{boi} é o preço da arroba de boi; $P_{bezerro}$ refere-se ao preço recebido pelo bezerro (por cabeça); $P_{boimagro}$ consiste no preço recebido pelo boi magro (por cabeça); P_{frango} é preço recebido pelo quilo de frango; P_{dolar} é o preço do dólar em reais; IPP refere-se ao Índice de Preços Pagos aos Produtores; Q_S refere-se a quantidade ofertada de carne bovina.

Os preços do bezerro e do boi magro, ambos negociados para engorda no setor, consistem em matérias prima, configurando-se, portanto, ao lado do IPP (remuneração dos produtores), como componentes de custo. O preço do frango é imputado no modelo como preço de um bem substituto, sendo, assim, uma variável da demanda. A variável preço do dólar, por sua vez, ao representar a taxa de câmbio, agrega: custos, visto que alguns elementos utilizados para produção de carnes são importados; preço relativo, já que constitui um termo de comparação entre os mercados interno e externo; renda, ou restrição orçamentária, uma vez que a taxa de câmbio impacta a capacidade de consumo.

Tendo em vista os objetivos do trabalho, considera-se plausível a obtenção do preço do frango em função das variáveis do mercado de carne bovina como modelagem para a interação entre os dois setores. Visto que, na prática, não é possível controlar os valores das variáveis independentes (apenas o período de observação), a troca da variável resposta não elimina o seu caráter estocástico. Além disso, aceitou-se o formato linear da relação entre as variáveis, sugerido por Medeiros (2006), visto que este mantém o modelo simples, do ponto de vista operacional, além de ser uma aproximação conveniente da realidade. Assim, o modelo teórico tomado para este estudo é descrito por:

$$P_{frango} = P_{boi} + P_{boimagro} + P_{bezerro} + P_{dolar} + IPA + Q_S. \quad (2.6)$$

A substituição do IPP pelo Índice de Preços ao Produtor Amplo (IPA) refere-se ao fato de este concatenar todos os custos de produção em sua definição.

3 CONSIDERAÇÕES SOBRE O MODELO DE REGRESSÃO LINEAR MÚLTIPLA

Com base na definição fornecida por Abbad e Torres (2002), pode-se descrever o modelo univariado de Regressão Linear Múltipla como um conjunto de técnicas estatísticas que respalda a análise da influência de várias variáveis independentes sobre uma variável dependente, cujo resultado é a equação da reta – de dimensão equivalente à quantidade de variáveis envolvidas no estudo – que possui o melhor ajuste, em termos de predição, ao conjunto de dados. Na prática, este tipo de modelagem pode ser aplicada para atender a diversos objetivos, dentre os quais Rencher e Schaalje (2007) destacam:

1. Predição. Estimativas dos parâmetros individuais $\beta_0, \beta_1, \dots, \beta_k$ possuem menor importância para a previsão do que a influência geral das variáveis X sobre y . Todavia, boas estimativas são necessárias para alcançar bom desempenho preditivo.
2. Descrição ou explicação de dados. Cientistas ou engenheiros usam o modelo estimado para resumir ou descrever os dados observados.
3. Estimação de parâmetros. Os valores dos parâmetros estimados podem ter implicações teóricas para um modelo postulado.
4. Seleção de variáveis ou triagem. A ênfase consiste em determinar a importância de cada variável de previsão na modelagem das variações em y . Aquelas associadas com uma importante quantidade de variação são retidas; aquelas que contribuem pouco são deletadas.
5. Controle de produção. Assume-se uma relação de causa e efeito entre y e as variáveis X . O modelo estimado pode ser utilizado para controlar a saída de um processo pela variação das entradas. Via experimentação sistemática, é possível alcançar a saída ótima. (RENCHE e SCHAALJE, 2007, p.2)

3.1 O modelo teórico

De acordo com Rencher e Schaalje (2007, p. 137), o MRLM é representado por:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon, \quad (3.1)$$

em que: y é a variável dependente; $x_j (j = 1, 2, \dots, k)$ representa a j -ésima variável independente; β_j refere-se ao coeficiente de regressão relacionado à j -ésima variável independente, ou, de outro modo, representa a derivada parcial de y em relação x_j ; ε representa as perturbações aleatórias em y que não são provenientes das variáveis explicativas consideradas.

Em estudos empíricos, conforme estes autores, consideramos uma amostra de n observações de associações entre as variáveis resposta e independentes para estimar os coeficientes da reta. Neste caso, pode-se representar o modelo como:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, i = 1, 2, \dots, n. \quad (3.2)$$

Assim, é possível adotar as seguintes notações matriciais (RENCHE e SCHAALJE,

2007, p. 139):

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (3.3)$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3.4)$$

em que y é o vetor de observações da variável estocástica de tamanho $n \times 1$; $\boldsymbol{\beta}$ é o vetor de coeficientes da regressão, parâmetros a serem estimados, de tamanho $(k + 1)$; X corresponde à matriz do modelo, de tamanho $n \times (k + 1)$; e $\boldsymbol{\varepsilon}$ refere-se ao erro teórico, vetor de tamanho $n \times 1$.

Para operacionalização do cálculo de estimadores, com base em estatística e probabilidade, algumas condições devem ser assumidas. Neste sentido, Abbad e Torres (2002, p. 20), enumeram pressuposições relativas: “(1) a multicolinearidade, (2) a singularidade, (3) a homogeneidade nas variâncias, (4) a normalidade e (5) a linearidade”.

Rencher e Schaalje (2007, p.139) afirmam que os pressupostos do modelo, em relação à notação (3.4), podem ser expressos de duas formas equivalentes: $E(\boldsymbol{\varepsilon}) = 0$ e $cov(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$ (em que cov refere-se à matriz de variâncias e covariâncias e \mathbf{I}_n é matriz identidade de ordem n); ou $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ e $cov(\mathbf{y}) = \sigma^2\mathbf{I}_n$. Além destas, destaca-se a pressuposição de normalidade multivariada, sobre a qual é elaborado o modelo teórico, necessária, portanto, para fins de inferência.

3.2 Estimação dos parâmetros pelo método dos mínimos quadrados ordinários

Considerando que todos os pressupostos, supracitados, são atendidos, o estimador dos coeficientes de regressão (representado por $\hat{\boldsymbol{\beta}}$), conforme Rencher e Schaalje (2007, p.143), é aquele que minimiza a soma de quadrados do erro (que representa os desvios das observações da variável resposta em relação aos correspondentes valores preditos pelo modelo), a qual é obtida por:

$$\sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik})^2. \quad (3.5)$$

Assim, o vetor $\hat{\boldsymbol{\beta}}$, obtido pelo Método de Mínimos Quadrados Ordinários (MMQ), é dado por (FOX, WEISBERG, 2010, p.1):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (3.6)$$

cuja matriz de covariâncias é:

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \quad (3.7)$$

O estimador não viciado da variância, s^2 , é obtido com base no estimador de MMQ de $\hat{\boldsymbol{\beta}}$, sendo sua fórmula (RENCHER e SCHAALJE, 2007):

$$s^2 = \frac{1}{n-k-1} \sum_{i=1}^n (y_i - \mathbf{X}'\hat{\boldsymbol{\beta}})^2. \quad (3.8)$$

Considerando a Soma de Quadrados do Erro (SQE), uma vez que

$$\sum_{i=1}^n (y_i - \mathbf{X}'\hat{\boldsymbol{\beta}})^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y} = SQE,$$

é possível expressar s^2 através do Quadrado Médio do Erro (QME) (*idem*):

$$s^2 = \frac{SQE}{n-k-1} = QME. \quad (3.9)$$

3.3 Testes de Hipóteses e Intervalos de Confiança

Sejam Soma de Quadrados Total (SQT), Soma de Quadrados de Regressão (SQR) e a Soma de Quadrados do Erro (SQE), definidos, conforme Renche e Schaalje (2007), por:

$$SQT = SQR + SQE = \mathbf{y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{y}; \quad (3.10)$$

$$SQR = (\mathbf{X}\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}) - n\bar{y}^2; \quad (3.11)$$

$$SQE = \mathbf{y}'\mathbf{y} - (\mathbf{X}\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}). \quad (3.12)$$

Se $\mathbf{y} \sim \mathbf{N}_m(\mathbf{X}\hat{\boldsymbol{\beta}}, \sigma^2\mathbf{I})$, é válido que a relação:

$$F = \frac{\frac{SQR}{k}}{\frac{SQE}{n-k-1}} \quad (3.13)$$

tem distribuição F de Snedcor com parâmetros (k) e ($n - k - 1$). Assim, é possível, por meio da construção de uma Tabela de Análise de Variância (ANOVA), obter um teste geral para o modelo (*idem*):

$$H_0 : \beta_j = 0, \forall j = 0, 1, 2, \dots, k$$

$$H_1 : \beta_j \neq 0 \text{ para qualquer } j = 0, 1, \dots, k$$

Seja o Quadrado Médio (QM) definido como o quociente entre Soma de Quadrado (SQ) de uma Fonte de Variação (FV) e seus respectivos Graus de Liberdade (GL), a estatística F pode ser expressa como a razão entre o Quadrado Médio de Regressão (QMR) e o Quadrado Médio do Erro (QME). Dessa forma, obtém-se a Tabela da ANOVA (Tabela 1).

A regra de decisão para este teste consiste em rejeitar a hipótese nula (H_0) se a estatística F supere o valor do quantil teórico $F_{(k, n-k-1)}$ com significância $100\alpha\%$, o que implica que, sob H_0 , a ocorrência do valor observado de F é pouco provável. Neste caso, é plausível considerar que pelo menos um dos coeficientes de regressão é significativamente diferente de zero.

Tabela 1 – Modelo de Tabela de Análise Variância para o Modelo de Regressão Linear

<i>FV</i>	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F</i>
Regressão	<i>k</i>	<i>SQR</i>	$QMR = SQR/k$	QMR/QME
Erro	$n - k - 1$	<i>SQE</i>	$QME = SQE/(n - k - 1)$	
Total	<i>n-1</i>	<i>SQT</i>		

Fonte: RENCHER e SCHAALJE, 2007, p. 188 (adaptação própria)

Para testar a significância individual do *j*-ésimo coeficiente ($H_0 : \beta_j = 0$), é possível utilizar, alternativamente, os testes t e F, cujas estatísticas são definidas, respectivamente, como (RENCHER e SCHAALJE, 2007, p.205):

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{g_{jj}}}, \tag{3.14}$$

$$F = \frac{\hat{\beta}_j^2}{s^2 g_{jj}}, \tag{3.15}$$

em que g_{jj} é o *j*-ésimo elemento da diagonal da matriz $(\mathbf{X}'\mathbf{X})^{-1}$. As respectivas regras de rejeição da hipótese nula dos testes t e F são: $|t| > t_{(\alpha/2, n-k-1)}$; $F > F_{(\alpha, 1, n-k-1)}$.

3.3.1 Intervalos de confiança e predição

Conforme Rencher e Schaalje (2007, p. 213), uma vez que $t_j = \frac{\hat{\beta}_j - \beta_j}{s\sqrt{g_{jj}}}$ tem distribuição central t de Student, obtém-se um intervalo de confiança para β_j ($IC(\beta_j, \gamma)$) dado por :

$$IC(\beta_j, \gamma) = \hat{\beta}_j \pm t_{(\alpha/2, n-k-1)} s\sqrt{g_{jj}}. \tag{3.16}$$

A interpretação deste resultado é de que β_j está contido no referido intervalo com $\gamma = 1 - \alpha$

Além disso, é possível construir intervalos de predição para estimação de valores correspondentes a períodos que extrapolam a abrangência de observação dos dados, seja esta temporal ou de outro tipo (espacial, por exemplo). Uma vez que o valor futuro y_0 pode ser estimado recorrendo-se a $E(y_0) = \hat{y}_0 = \mathbf{x}_0' \hat{\beta}_0$ e sabendo que y_0 e as observações (\hat{y}_0) são independentes, obtém-se o Intervalo de Predição ($IP(y_0, \gamma)$):

$$IP(y_0, \gamma) = \mathbf{x}_0'(\hat{\beta}_0) \pm t_{(\alpha/2, n-k-1)} s\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}. \tag{3.17}$$

Entende-se este resultado como o intervalo no qual uma observação futura (observada no período $n + q, q > 0$) estará contida com probabilidade γ . A abrangência do intervalo de predição é maior do que o intervalo de confiança em razão da maior variância do valor futuro y_0 em relação ao y estimado (\hat{y}_0).

3.4 O Coeficiente de Determinação

Uma maneira de avaliar o poder de explicabilidade do modelo é através do Coeficiente de Determinação, dado por (REENCHER; SCHAALJE, 2007):

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT}. \quad (3.18)$$

Pela própria definição de sua fórmula, o R^2 reflete a proporção das variações na variável resposta explicadas pelo modelo, assumindo valores entre zero e um, de modo que valores próximos a um indicam boa qualidade do ajuste do modelo aos dados. Conforme Reencher e Schaalje (2007), o coeficiente de determinação aumenta com a inclusão de variáveis regressoras, visto que, proporcionalmente, a soma de quadrados do erro diminui. Assim, seu uso indiscriminado poderia conduzir a um modelo que não fosse parcimonioso. Para modelos de regressão múltipla, um ajuste na fórmula que penaliza o acréscimo indiscriminado de variáveis explicativas fornece o R^2 ajustado:

$$R_a^2 = \frac{(n-1)R^2 - k}{n - k - 1}. \quad (3.19)$$

O R_a^2 consiste numa importante indicação preliminar para a qualidade do ajuste e para a decisão de buscar novas variáveis explicativas. Todavia, seus resultados devem ser relativizados para cada caso e analisados de maneira conjunta com as técnicas de seleção de variáveis e de diagnóstico do modelo.

3.5 Seleção de variáveis

Uma importante questão relacionada à análise de regressão múltipla se refere a obtenção de um modelo que proporcione maior eficiência na explicabilidade da variável dependente sem adição indevida de variáveis independentes. Dentre os métodos de seleção de variáveis, destacam-se o de maior R_P^2 , o de menor s_P^2 , *backward*, *forward*, *stepwise* e o Critério de Akaike (AIC). Conforme Paula (2004), nenhum dos vários métodos disponíveis é consistente, de modo que, mesmo para grandes amostras, não selecionam uma variável explicativa com probabilidade 1.

O método *forward* inicia com o modelo mais simples, em que y é uma constante, isto é: $y = \beta_0$. Supondo que sejam consideradas q variáveis explicativas, para cada uma, ajusta-se o modelo:

$$y = \beta_0 + \beta_j x_j, \quad j = 1, \dots, q \quad (3.20)$$

Seja p o menor nível descritivo dos q modelos ajustados para o teste de hipóteses:

$$\begin{aligned} H_0 &: \beta_j = 0, \\ H_1 &: \beta_j \neq 0, \end{aligned}$$

a variável correspondente entra no modelo se $p < p_e$, em que p_e é um nível descritivo crítico, escolhido como critério de entrada. O processo ocorre de maneira iterativa, levando em consideração o ajuste com variáveis já selecionadas em passos anteriores (isto é, uma vez que uma variável seja selecionada para o modelo, esta não será mais descartada), até que ocorra $p > p_e$ (PAULA, 2004).

O método *backward*, por outro lado, parte do modelo mais completo, que inclui todas as q possíveis variáveis explicativas consideradas:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_q x_q. \quad (3.21)$$

Realiza-se o teste de significância dos coeficientes e, sendo p o maior p-valor dos q testes, retira-se a variável correspondente se $p > p_s$, em que p_s é o nível descritivo crítico escolhido como critério de descarte. Iterativamente, repete-se o processo, tendo em conta os descartes já realizados (de modo que nenhuma variável descartada pode ser reconsiderada), até que não haja descarte, quando verificar-se $p < p_s$ (PAULA, 2004).

O método *stepwise*, conforme Paula (2004):

É uma mistura dos dois procedimentos anteriores. Iniciamos o processo com o modelo $[y = \beta_0]$. Após duas variáveis terem sido incluídas no modelo, verificamos se a primeira não sai do modelo. O processo continua até que nenhuma variável seja incluída ou seja retirada do modelo. Geralmente adotamos $0,15 \leq P_E, P_S \leq 0,25$. Uma sugestão seria usar $P_E = P_S = 0,20$. (PAULA, 2004, p. 78)

Ainda sobre o método, Alves *et al.* (2013) destaca a possibilidade de combiná-lo com outros critérios:

(...) é feito de forma iterativa, adicionando (passo *forward*) e removendo variáveis (passo *backward*), a partir de um critério de seleção, um dos critérios de seleção mais usados é o teste F, mas também pode ser feito com o coeficiente correlação linear múltipla, erro quadrático total, critério de informação de Akaike Hoking (1976). (ALVES *et al.*, 2013, p.2)

O método AIC (*Akaike's Information Criterion*) foi introduzido em 1973 e se popularizou no meio acadêmico rapidamente pelas possibilidades de aplicações práticas. Como uma extensão do princípio da máxima verossimilhança, propondo uma combinação entre estimação pontual e teste de adequação do modelo em um princípio único de comparação entre os valores da log-verossimilhança média (DELEEUW, 1992). Conforme Akaike (1973), a esperança da log-verossimilhança é dada por:

$$E(\log f(X|\hat{\theta})) = E\left(\int f(x|\theta) \log f(x|\hat{\theta}) dx\right), \quad (3.22)$$

em que $\hat{\theta}$ são estimadores do vetor de parâmetros θ da distribuição de probabilidade cuja densidade é $f(x|\theta)$; X é uma variável aleatória que segue esta distribuição.

Partindo da ideia de maximizar uma razão média de log-verossimilhanças para maximizar a entropia do modelo escolhido, chega-se a uma fórmula prática de cálculo do AIC, dada por:

$$AIC_P = -2 \sum_{i=1}^n \log(L_p) + 2(p + 1), \quad (3.23)$$

em que: L_p é a função de verossimilhança do modelo; p , o número de parâmetros; $k = 2$, para aplicações usuais do AIC ou $k = \log(n)$ (n representa o tamanho amostral), para abordagens bayesianas. O modelo a ser selecionado será aquele com menor AIC.

3.6 Validação e diagnósticos do modelo

Conforme Paula (2004), a análise de diagnóstico, utilizada para avaliar o cumprimento dos pressupostos do modelo (relativos aos componentes aleatório e sistemático) e para detectar observações discrepantes com grande influência nos resultados do ajuste, é uma etapa importante da análise de regressão. Acerca deste assunto, Abbad e Torres (2002, p.20), que indicam quatro suposições para o modelo - multicolinearidade e singularidade; homogeneidade nas variâncias; normalidade e linearidade - afirmam que:

Embora seja imprescindível que o pesquisador examine esses pressupostos antes de iniciar suas análises, nota-se que a RM [Regressão Múltipla] é um modelo eficaz contra a violação de grande parte dos pressupostos. Por exemplo, no caso da inclusão de variáveis multicolineares ou singulares nas análises, o pesquisador estará perdendo graus de liberdade, o que consequentemente reduz o poder estatístico de suas conclusões. (ABBAD e TORRES, 2002, p.20)

3.6.1 Multicolinearidade e singularidade

Miloca e Conejo (2008, p. 2) definem multicolinearidade como a correlação entre, pelo menos, três variáveis preditivas. A singularidade, conforme HAIR *et al.* (2007, p.153), representa um caso extremo de colinearidade, configurando-se pela correlação perfeita (positiva ou negativa, ou, de ± 1) entre variáveis explicativas. Uma vez que a ocorrência desta pode ter efeitos no grau de explicação do modelo e, de forma adicional, nas estimativas dos coeficientes regressores (e, portanto, na aplicabilidade geral do ajuste), é desejável selecionar variáveis independentes que apresentem baixa multicolinearidade entre si e, em concomitância, elevadas correlações com a variável resposta (MILOCA e CONEJO, 2008). Estes autores apontam, ainda, indícios do problema em questão:

Algumas indicações da presença de multicolinearidade são:

1. valores altos do coeficiente de correlação;
2. grandes alterações nas estimativas dos coeficientes de regressão, quando uma variável independente for adicionada ou retirada do modelo, ou quando uma observação for alterada ou eliminada;
3. a rejeição da hipótese $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$, mas nenhuma rejeição das hipóteses $H_0 : \beta_i = 0, i = 1, 2, \dots, k$, sobre os coeficientes individuais de regressão;
4. obtenção de estimativas para os coeficientes de regressão com sinais algébricos contrários àqueles que seriam esperados a partir de conhecimentos teóricos disponíveis ou de experiências anteriores sobre o fenômeno estudado e
5. obtenção de intervalos de confiança com elevadas amplitudes para os coeficientes de regressão, associados a variáveis independentes importantes. (MILOCA e CONEJO, 2008, p. 2-3)

Entre as medidas para contornar este tipo de ocorrência, a literatura disponível sugere (MILOCA e CONEJO, 2008, p.3): exclusão de uma ou mais variáveis explicativas com forte correlação entre si e adição de outras variáveis independentes; utilizar o modelo apenas para previsão, evitando-se, assim, interpretar os coeficientes estimados; uso de Análise de Componentes Principais, para obter variáveis latentes, ortogonais entre si, a partir de combinações lineares das variáveis originais.

3.6.2 Homogeneidade das variâncias

Hair *et al.* (2007, p.152) define homocedasticidade como a ocorrência de variâncias constantes nos valores de uma variável independente. Conforme Rencher e Schaalje (2007, p. 228), a matriz *hat*, \mathbf{H} :

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (3.24)$$

é a matriz de projeção ortogonal, que transforma \mathbf{y} em $\hat{\mathbf{y}}$, possuindo as propriedades de simetria e idempotência. A partir de \mathbf{H} , o vetor estimado de resíduos é dado por (HAIR *et al.*, 2007):

$$\hat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (3.25)$$

Algumas comparações importantes entre as propriedades dos resíduos estimados $\hat{\boldsymbol{\varepsilon}}$ e àquelas do termo de erro do modelo são:

(...) o vetor de resíduos $\hat{\boldsymbol{\varepsilon}}$ possui a mesma média que o termo de erro $\boldsymbol{\varepsilon}$, mas, (...), $cov(\hat{\boldsymbol{\varepsilon}}) = \sigma^2(\mathbf{I} - \mathbf{H})$ difere do pressuposto $cov(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$. Assim, os resíduos $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n$ não são independentes. Todavia, em muitos casos, especialmente quando n é grande, os h'_{ij} s tendem a ser pequenos (para $i \neq j$), e a dependência mostrada em $\sigma^2(\mathbf{I} - \mathbf{H})$ não afeta indevidamente gráficos e outras técnicas de validação do modelo. Cada $\hat{\varepsilon}_i$ possui correlação perceptível com cada \mathbf{y}_j (...), mas, (...) $\hat{\varepsilon}_i$'s não são correlacionados com os $\hat{\mathbf{y}}_j$'s. (RENCHEER e SCHAALJE, 2007, p. 229 - tradução e adaptação própria).

Para testar a suposição de homocedasticidade (hipótese nula) contra a hipótese de heterocedasticidade, o teste de Breusch-Pagan (1979) mostra-se apropriado. O teste, baseado no multiplicador de Lagrange e no teste de Aitchison e Silvey, tem como estatística:

$$u_i = \frac{e_i^2}{SQE/n}, i = 1, 2, \dots, n. \quad (3.26)$$

Conforme Breusch e Pagan (1979), sob H_0 , a estatística do teste segue uma distribuição qui-quadrado com $p - 1$ (o número de parâmetros do modelo, sem contar o intercepto) graus de liberdade. Assim, rejeita-se a hipótese nula se $u_i \geq \chi^2_{[p-1;\alpha]}$.

3.6.3 Independência dos erros

Barroso *et al.* (2012) indica que testar a independência dos resíduos, a partir da verificação de existência da autocorrelação residual, é de extrema importância para validação do uso do

modelo de regressão. De acordo com estes autores, para analisar a independência dos resíduos, a literatura sugere análise gráfica por meio dos *plots* de $\hat{\varepsilon}_i$'s versus \hat{y}_j 's: se a suposição é plausível, o comportamento dos pontos no gráfico não indica tendência funcional destes. Neste contexto, a aplicação do teste de Durbin-Watson constitui ferramenta mais formal de diagnóstico. Não obstante, sua restrição à identificação de correlações de primeira ordem constitui séria limitação. Para resolver este problema, "Vinod (...) generalizou o teste proposto por Durbin e Watson (1950) possibilitando detectar a presença de autocorrelação de qualquer ordem." (BARROSO *et al.*, 2012, p. 433). Assim, por um lado, o teste original baseia-se na suposição de que um modelo autorregressivo de primeira ordem gera os erros do modelo, tendo como estatística de teste:

$$dw = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}. \quad (3.27)$$

A rejeição da hipótese nula (de que os erros são não correlacionados) pode ser avaliada por meio da comparação com valores críticos da tabela de Durbin-Watson. Não obstante, as análises por meio de algoritmos computacionais utilizam-se de cálculo de níveis descritivos. Sendo o nível descritivo a probabilidade de ocorrência da estatística observada sob H_0 , sua comparação com o nível de significância do teste conduz a uma regra de decisão: rejeita-se a hipótese nula se $p - \text{valor} < \alpha$. Por outro lado, a generalização do teste, conforme Barroso *et al.*, avalia-se as várias ordens de maneira sequencial, de modo que o j -ésimo passo do teste é descrito por:

$$\begin{aligned} H_0 : \phi_j &= 0, \text{ dado } \phi_1 = \dots = \phi_{j-1} = 0 \\ H_1 : \phi_j &\neq 0, \end{aligned}$$

em que ϕ_j representa a autocorrelação de ordem j . A estatística do teste, no j -ésimo passo, é dada por:

$$dj = \frac{\sum_{t=j+1}^n (e_t - e_{t-j})^2}{\sum_{t=1}^n e_t^2}. \quad (3.28)$$

De acordo com Barroso *et al.*, o nível descritivo do teste é calculado por meio de um procedimento *bootstrap* recursivo ¹.

3.6.4 Análise da normalidade

Para verificação do pressuposto de normalidade dos resíduos, dentre as possíveis análises gráficas, destaca-se quantil-quantil para normalidade. De maneira geral, um gráfico quantil-quantil demonstra se dois conjuntos de dados pertencem à mesma distribuição. Os pontos são formados pelos pares de quantis e, dessa maneira, a aproximação ao comportamento de uma função identidade (uma reta de inclinação 1) indica que os dois conjuntos possuem mesma distribuição. Em particular, para avaliação da normalidade, trata-se do gráfico dos resíduos

¹ Ver BARROSO *et al.*, 2012, p.435-436

estudentizados contra os valores esperados das estatísticas de ordem da normal padrão (PAULA, 2004). Os resíduos estudentizados são dados por:

$$t_{i*} = \frac{\widehat{\varepsilon}_i}{s_{(i)}(1 - h_{ii})^{1/2}}, \quad (3.29)$$

em que $s_{(i)}$ é a raiz do erro quadrático médio do modelo sem a i – sima observação e h_{ii} é o i – simo elemento da diagonal principal da matriz de projeção. As estatísticas de ordem são obtidos por (PAULA, 2004, p. 64):

$$E(Z_{(i)}) \cong \varphi^{-1}\left(\frac{i - 3/8}{n + 1/4}\right), \quad (3.30)$$

em que φ denota a função de distribuição acumulada da distribuição normal padrão. Uma vez que o distanciamento das observações amostrais podem ser fontes de dificuldade para aceitação/rejeição do pressuposto de normalidade, uma maior formalização torna-se conveniente. Há opções gráficas, como a construção de bandas de confiança para os resíduos, como o gráfico envelope (*idem*, p. 65). Por outro lado, a realização de testes de hipóteses mostra-se uma solução factível.

O teste de Shapiro-Wilk foi proposto em 1965, por S. S. Shapiro e M.B. Wilk. Conforme os autores:

A estatística de teste é obtida dividindo-se o quadrado de uma combinação linear adequada das estatísticas de ordem da amostra pela estimativa simétrica usual de variância. Esse índice é invariante em relação à escala e à origem e, portanto, a estatística é apropriada para composição de um teste da hipótese de normalidade.

(SHAPIRO; WILK, 1965, p. 591 - Tradução própria)

A estatística de teste é dada por:

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2} \quad (3.31)$$

Em que b é uma constante determinada por:

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1}(x_{(n-i+1)} - x_{(i)}), & \text{se } n \text{ é par} \\ \sum_{i=1}^{(n+1)/2} a_{n-i+1}(x_{(n-i+1)} - x_{(i)}), & \text{se } n \text{ é ímpar} \end{cases}, \quad (3.32)$$

em que a_{n-i+1} são constantes geradas pelas médias, variâncias e covariâncias das estatísticas de ordem de uma amostra de tamanho n da distribuição normal. Shapiro e Wilk tabelaram valores para estes coeficientes, bem como os valores críticos para avaliação de W , no artigo em que formalizaram o teste. Propriedades da estatística W que devem ser destacadas são: W possui uma distribuição que depende apenas do tamanho amostral, sendo, portanto, independente de S^2 e de \bar{y} , para amostras provenientes de distribuições normais; o valor máximo de W é 1 e o valor mínimo, $\frac{na_1^2}{n-1}$ (SHAPIRO; WILK, 1965).

3.6.5 Linearidade

A linearidade refere-se às propriedades de aditividade e homogeneidade do modelo, ou, de maneira mais simples, que o modelo populacional é, de fato, uma reta (HAIR *et al.*, 2007, p.152). Portanto, a aceitação deste pressuposto relaciona-se com a verificação de adequação do modelo aos dados, por meio do teste F, procedido através da análise de variância. Além disso, o gráfico de dispersão de resíduos padronizados contra valores preditos consiste em importante ferramenta auxiliar.

3.6.6 Observações influentes

Observações aberrantes podem influenciar os diagnósticos e o a adequação do modelo aos dados. Assim, é necessário identificar esses pontos, entender os motivos de sua ocorrência, tratá-los e, em última instância, excluí-los do ajuste. Paula (2004) mostra que pode-se identificar valores aberrantes através de técnicas gráficas, utilizando o resíduo estudentizado:

Geralmente para detectarmos pontos suspeitos de serem aberrantes e/ou influentes recorreremos aos seguintes gráficos: (i) pontos aberrantes, o gráfico de t_i^* contra a ordem das observações; (ii) heteroscedasticidade, o gráfico de t_i^* contra \hat{y}_i (valores ajustados); (iii) pontos influentes, gráfico de D_i , C_i , $DFFITs_i$, h_{ii} (...) contra a ordem das observações; (iv) falta de algum termo extra, gráfico de t_i^* contra \hat{y}_i ou contra as covariáveis que estão ou não foram incluídas no modelo, (v) correlação entre as observações, gráfico de t_i^* contra o tempo ou contra a ordem em que há suspeita de correlação, e (vi) afastamentos da normalidade, gráfico normal de probabilidades. (PAULA, 2004, p.64)

Conforme o mesmo autor, Cook desenvolveu uma das técnicas mais poderosas para identificar pontos influentes. A Distância de Cook é dada por:

$$D_i = \frac{\hat{\varepsilon}_i^2}{k QME} \frac{h_{ii}}{(1 - h_{ii})^2}. \quad (3.33)$$

Se $D_i > 1$, as observações são destacadas e devem ser investigadas.

3.7 Estimação dos parâmetros pelo método dos mínimos quadrados generalizados

Conforme Fox e Weisberg (2010), considerando que a matriz de covariância dos resíduos Σ é simétrica e positiva definida, de modo que o erros possam ser autocorrelacionados e/ou apresentar heterocedasticidade, a distribuição destes é dada por $\varepsilon \sim N_m(0, \Sigma)$. Considerando, ainda, que a matriz Σ é conhecida, o estimador de Mínimos Quadrados Generalizados (GLS - *Generalized Least Squares*) que maximiza a função de verossimilhança de β , e sua respectiva matriz de covariância são dados, nesta ordem, por:

$$b_{GLS} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}, \quad (3.34)$$

$$\text{Var}(b_{GLS}) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}. \quad (3.35)$$

No entanto, em aplicações reais, Σ não é conhecida e, para ausência de restrições, possui $n(n+1)/2$ distintos elementos a serem estimados. Assim, é conveniente estabelecer uma parametrização adequadamente restritiva para Σ , de maneira que seja possível estimar o modelo via métodos como a máxima verossimilhança (*idem*, p.2). Uma maneira de parametrizar a matriz Σ , ante a presença de autocorrelação residual no modelo de regressão linear múltipla, é assumir que os erros representam observações de diferentes momentos equidistantes no tempo, de modo que possam ser descritos por uma série temporal. Assumindo como verdadeiros os pressupostos do modelo acerca dos resíduos (de média igual a zero e de variâncias homogêneas) e assumindo, ainda, que a covariância de dois erros depende apenas da distância destes no tempo, pode-se considerar que o processo que gera os resíduos é estacionário. Conforme Fox e Weisberg (2010), as covariâncias são dadas por:

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+s}) = \text{Cov}(\varepsilon_{t+s}, \varepsilon_t) = \sigma^2 \rho_s, \quad (3.36)$$

em que ρ_s é a correlação correspondente a defasagem s . Assim, a matriz de covariância tem como estrutura:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & & \vdots & \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{pmatrix} = \sigma^2 \mathbf{P}. \quad (3.37)$$

Dessa forma, através da estimação de um modelo misto, que combina um modelo autorregressivo, $AR(p)$, a um modelo de médias móveis, $MA(p)$, é possível estimar os parâmetros de Σ e, portanto, ajustar um modelo via método dos mínimos quadrados generalizados para os dados. Conforme Ehlers (2007), um processo autoregressivo médias móveis, $ARMA(p, q)$, é dado por:

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}, \quad (3.38)$$

em que $\{\varepsilon_t\}$ é um processo puramente aleatório com média zero e variância σ_ε^2 . Usando o operador diferença, o modelo pode ser representado por:

$$\phi(B)X_t = \theta(B)\varepsilon_t. \quad (3.39)$$

Os valores de $\alpha_1, \dots, \alpha_p$ e β_1, \dots, β_q são tais que, respectivamente, as raízes de $\phi(B) = 0$ e de $\theta(B) = 0$ estão fora do círculo unitário (EHLERS, 2007, p. 24). Em relação ao comportamento das funções de autocorrelação e autocorrelação parcial para os modelos ARMA, este autor explica:

Vale notar que as funções de autocorrelação e autocorrelação parcial ficam consideravelmente mais complicadas em processos ARMA. De um modo geral, para um processo $ARMA(p, q)$ estacionário a função de autocorrelação tem um decaimento exponencial ou oscilatório após a defasagem q enquanto que a $facp$ tem o mesmo comportamento após a defasagem p (Box and Jenkins 1970, p. 79). Em princípio este resultado pode ser utilizado para auxiliar na determinação da ordem (p, q) do processo mas na prática pode ser bastante difícil distinguir entre decaimentos exponenciais e oscilatórios através das estimativas destas funções. (EHLERS, 2007, p. 24)

No que concerne ao ajuste do modelo, métodos computacionais iterativos devem ser utilizados, visto que os $\{\varepsilon_t\}$ são funções não lineares dos parâmetros $\alpha_1, \dots, \alpha_p$ e β_1, \dots, β_q , de modo que os estimadores não podem ser obtidos analiticamente (EHLERS, 2007).

3.7.1 Avaliação do ajuste pelo método GLS

O coeficiente de determinação, R^2 , tradicionalmente aplicado aos modelos estimados pelo método dos mínimos quadrados ordinários não pode ser estendido ao modelo ajustado pelo método GLS, visto que, neste último, não é possível decompor a variância total entre variância amostral (proveniente do ajuste) e variância dos erros (GRUBER *et al.*, 2007). Neste sentido, torna-se necessário a utilização de uma estatística de avaliação capaz de realizar esta separação. Gruber *et al.* (2007) indica uma métrica baseada na comparação entre as variâncias residuais de um modelo com k parâmetros e do modelo nulo (variável resposta em função da média geral): o Pseudo- R_{GLS}^2 , dado por:

$$Pseudo - R_{GLS}^2 = 1 - \frac{\hat{\sigma}_\delta^2(k)}{\hat{\sigma}_\delta^2(0)}, \quad (3.40)$$

em que: $\hat{\sigma}_\delta^2(k)$ é a variância residual do modelo com k variáveis explicativas; $\hat{\sigma}_\delta^2(0)$, a variância residual para o modelo nulo. Esta métrica está pautada na redução da variância residual com a adição de variáveis explicativas: assim, a razão entre as variâncias residuais indica a porção proveniente do componente aleatório do modelo, a qual, deduzida da unidade, fornece a proporção da variância amostral. Assim como o R^2 , este coeficiente varia de 0 a 1, o que possibilita uma interpretação análoga sobre o modelo ajustado. Conforme os mesmos autores (2007), trata-se de uma extensão direta do coeficiente de determinação ajustado, visto que baseia-se na utilização do quociente de estimadores não tendenciosos das variâncias do erro e da variável resposta.

4 MATERIAL E MÉTODOS

4.1 Caracterização das variáveis

Os dados analisados, obtidos de fontes secundárias, referem-se a uma amostra composta por 195 observações (abrangendo o período de janeiro de 1998 a março de 2014) de sete variáveis: preço do quilo de frango para corte; preço da arroba (15 kg) do boi gordo rastreado; preço da unidade de boi magro; preço da unidade de bezerro; abate bovino; Índice de Preços ao Produtor Amplo (IPA); taxa de câmbio. Os preços agropecuários referem-se às médias mensais recebidas pelos produtores e foram obtidas do Instituto de Economia Agrícola ¹. O abate bovino, expresso em unidades, foi obtido do portal Ipea Data ². O IPA, foi obtido da Fundação Getúlio Vargas³. A taxa de câmbio expressa a quantidade de reais necessárias à aquisição de uma unidade de dólar e também foi obtido do Ipea Data.

Tomando por base o modelo teórico obtido no capítulo 2, em razão da indisponibilidade de dados específicos para os setores de interesse: considerou-se o abate bovino como aproximação plausível para a oferta de carne bovina, assumindo que as perdas entre produção e distribuição são negligenciáveis; utilizou-se o IPA para incluir os custos de produção no ajuste.

À exemplo do procedimento adotado por Medeiros (2006), as variáveis de preços agropecuários foram deflacionados pelo Índice Geral de Preços - Disponibilidade Interna (IGP-DI) mensal de março de 2014. A taxa de câmbio, que também configura-se como um preço, foi deflacionada pelo Índice de Preços ao Consumidor dos Estados Unidos Dessazonalizado (originalmente, *Consumer Price Index for All Urban Consumers seasonally adjusted*). Estes *benchmarks* foram obtidos do Ipea Data, abrangendo o mesmo período do que as demais variáveis. Dessa forma, "eliminando-se o efeito da inflação, os dados podem ser comparados a qualquer tempo, evitando que a previsão seja afetada por uma fonte adicional de variação"(MEDEIROS, 2006, p.63). Além disso, tornou-se possível isolar os efeitos inflacionários sobre a variável resposta no IPA, tornando factível sua identificação.

Em seguida, todas as variáveis consideradas para o modelo sofreram transformação logaritmica. Conforme Medeiros (2006),

Esse procedimento torna o efeito sazonal da série aditivo e estabiliza a variância do erro. A justificativa para essa transformação é a de que as previsões realizadas sem tal transformação podem considerar uma elevação na tendência e um forte padrão sazonal, o que aumenta a variação da série ao longo do tempo (MEDEIROS, 2006, p.63).

¹ Disponível em: <http://ciagri.iea.sp.gov.br/> Acesso em: 16 de agosto de 2014

² Disponível em: <http://www.ipeadata.gov.br/> Acesso em: 16 de agosto de 2014

³ Disponível em: <http://www.antigofgvdados.fgv.br/> Acesso em: 16 de agosto de 2014

4.2 Procedimentos para ajuste e análise do modelo

A partir da utilização do *software* R 3.0.1, ajustou-se o MRLM para obter a equação do preço do frango em função das variáveis do mercado de carne bovina, a qual é descrita esquematicamente por:

$$y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5 + \hat{\beta}_6 X_6, \quad (4.1)$$

em que: y é o preço do frango; X_1 refere-se ao preço do boi gordo rastreado; X_2 é o preço do boi magro; X_3 , o preço do bezerro; X_4 , o abate bovino; X_5 , o IPA; X_6 , a taxa de câmbio; $\hat{\beta}_j$ (com $j = 0, 2, \dots, 6$) são os estimadores dos coeficientes regressores. O modelo foi ajustado por meio da função *lm* do pacote *stats*, sendo, em seguida, submetido à análise de variância para testar se os coeficientes estimados eram significativamente diferentes de zero.

Procedeu-se à seleção de variáveis por meio da função *stepAIC* do pacote *MASS*, que ajusta o melhor modelo através de um algoritmo *stepwise*, utilizando o AIC como critério de identificação. O novo modelo adotado foi submetido à ANOVA, no sentido de verificar os resultados do teste F para os coeficientes das variáveis que permaneceram no modelo. A partir dos resultados obtidos, realizou-se análise de resíduo para a validação do modelo.

A verificação do pressuposto de normalidade dos resíduos foi realizada a partir da análise gráfica do gráfico quantil-quantil para normalidade, obtido através da função *qqnorm* do pacote *stats*, além do teste de Shapiro-Wilk, fornecido pelo comando *shapiro.test* do mesmo pacote. Visto que o nível descritivo para o teste é calculado de forma aproximada, optou-se pela comparação da estatística W com os valores críticos tabelados disponibilizados em Shapiro e Wilk (1965).

Para verificar a suposição de homocedasticidade, utilizou o teste de Breusch-Pagan, realizado através do comando *bptest* do pacote *lmtest*. Buscou-se detectar a existência de autocorrelação nos erros através do teste de Durbin-Watson, em suas versões originais e generalizadas, as quais são obtidas, respectivamente, pelas funções *dwtest* (do pacote *lmtest*) e *dwt* (do pacote *car*). No sentido de obter melhor suporte para esta análise, também foi gerado o gráfico dos resíduos *versus* valores ajustados, na tentativa de identificar graficamente uma possível tendência funcional dos resíduos.

A identificação de possíveis pontos influentes consistiu de análise do gráfico de Resíduos *versus* Leverage (diagonal da matriz H), além do cálculo da distância de Cook, via função *cooks.distance*, para os resíduos observados.

A modelagem da matriz de erros fundamentou-se no ajuste do modelo com uso dos estimadores de mínimos quadrados generalizados, considerando como variáveis explicativas aquelas selecionadas pelo *stepAIC*, através do uso da função *gls* do pacote *nlme*. No entanto, para proceder ao ajuste, foi necessário estimar a matriz de variâncias, a qual foi parametrizada como uma série temporal gerada por um modelo *ARMA*(p, q).

Para determinar a ordem do modelo misto, recorreu-se à análise gráfica das funções de autocorrelação e de autocorrelação parcial, obtidas, respectivamente, pelos comandos *acf* e *pacf*

(ambas do pacote *stats*). Com o intuito de conferir maior formalização às decisões acerca dos valores de p e q , utilizou-se a função *auto.arima*, implementada no pacote *forecast*, a qual retorna o melhor modelo ajustado de acordo com os critérios AIC e BIC, determinando, inclusive, se o modelo pode ser considerado estacionário ou não.

A avaliação do modelo ajustado via GLS, através do *Pseudo* – R_{GLS}^2 , foi realizada a partir do erro padrão residual deste e do erro padrão residual do modelo sem variáveis explicativas, os quais foram elevados ao quadrado para obtenção das variâncias residuais.

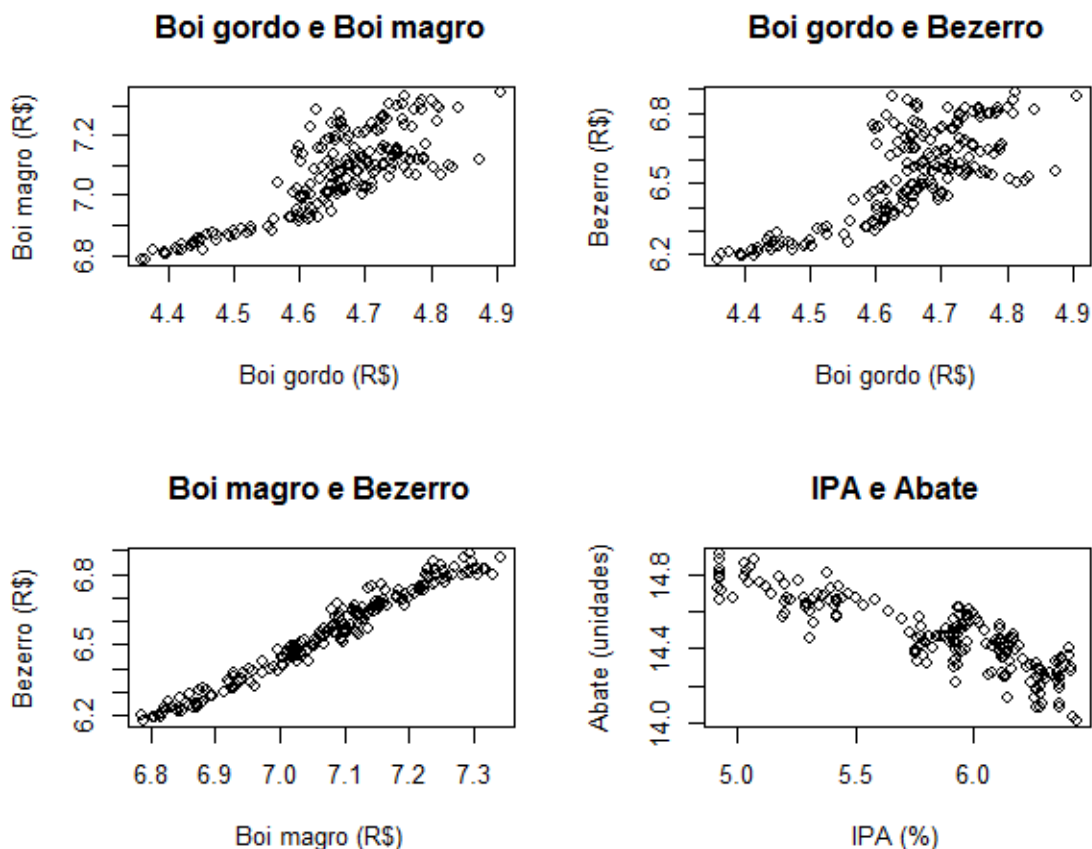
5 RESULTADOS E DISCUSSÃO

Uma análise preliminar das correlações entre as variáveis (Figura 1), indicou correlação positiva e forte entre: os preços do boi gordo e do boi magro (0,8129); os preços do boi gordo e do bezerro (0,7651). Os preços do boi magro e do bezerro obtiveram correlação positiva quase perfeita (0,9828). Essas relações decorrem das características do setor.

Uma vez que produtores compram boi magro ou bezerro para engorda, espera-se que o aumento do preço do boi gordo impulse a demanda por estes tipos, o que aumenta seus preços. Por outro lado, um aumento no preço do bezerro ou do boi magro deverá ser repassado para o preço de venda do boi gordo, que, por ser um produto alimentício de primeira necessidade, possui baixa elasticidade-preço (isto é, uma variação positiva no preço produz redução menos que proporcional no consumo).

A correlação entre os preços do bezerro e boi magro mostra-se quase perfeita pela função muito similar que estes desempenham no setor. Assim, se o preço de um destes fatores aumenta muito, os produtores deverão optar pelo outro, ocasionando queda no preço do primeiro e aumento do preço do segundo, o que indica tendência ao equilíbrio estável dos mercados destes dois produtos. Outra relação a ser destacada é a correlação forte e negativa (-87,56) entre o IPA

Figura 1 – Gráficos de dispersão selecionados



e o abate bovino. Tendo em vista que o IPA é uma variável que representa os custos de produção,

seu aumento reflete um desestímulo ao investimento e, assim, poderá ocasionar redução da oferta de carne, aqui aproximada pelo número de abates. Analogamente, se uma redução do índice for verificada, a produção e o aumento da oferta de carne bovina deverão ser estimulados. Uma vez que o IPA é um indicador macroeconômico (que concatena os diversos setores da economia brasileira), a relação de causalidade, em termos teóricos, parece estabelecida como: variações no IPA causam variações na oferta bovina. Desta forma, variações dos números de abate deverão ter pouca influência direta nas variações do IPA.

5.1 Ajuste do modelo

O ajuste inicial para o modelo completo forneceu como estimativa:

$$\hat{y} = -3,219 + 0,833X_1 + 0,331X_2 - 0,527X_3 + 0,098X_4 - 0,027X_5 + 0,087X_6. \quad (5.1)$$

Uma vez que $\hat{\beta}_j$ representa a primeira derivada de y em relação a j -ésima variável, os coeficientes regressores podem ser interpretados em termos de elasticidade do preço do frango às variações individuais das demais variáveis. Assim, espera-se que a elevação em um real do preço do boi gordo rastreado (X_1) eleve em cerca de 83 centavos o preço do frango. Analogamente, uma elevação de uma unidade monetária do preço do boi magro (X_2) provocará, em média, uma elevação de R\$ 0,33 em y . A relação da variável resposta em relação ao preço do bezerro (X_3) ocorre em sentido inverso: ante uma elevação em uma unidade monetária de X_3 , espera-se uma redução de aproximadamente 053 centavos do preço do frango. O aumento do abate bovino (X_4) em uma unidade deverá provocar, em média, aumento de R\$ 0,0982 em y . Espera-se observar redução no preço do frango na ordem de R\$ 0,027 para cada elevação de um ponto percentual no IPA. Em relação à taxa de câmbio, o aumento de um real no preço do dólar (que representa uma desvalorização cambial), provocará, em média, aumento de R\$0,87 em y . O intercepto $\hat{\beta}_0$, por sua vez, pode ser interpretado como o valor esperado para Y caso todas as variáveis independentes assumissem valor nulo.

De acordo com a análise da significância dos coeficientes, apenas os coeficientes de X_1 , X_3 e X_6 foram considerados significativos (Tabela 2). O coeficiente de determinação ajustado, que quantifica a proporção das variações em y explicadas pelo modelo, foi de 0,3465.

Tabela 2 – Resumo das estimativas dos coeficientes do modelo completo

Coeficientes	Estimativa	Desvio Padrão	Estatística t	Valor p
Intercepto	-3,219	2,114	-1,523	0,129
X_1	0,833	0,158	5,265	< 0,0001
X_2	0,331	0,412	0,804	0,422
X_3	-0,527	0,267	-1,972	0,050
X_4	0,098	0,093	1,046	0,297
X_5	-0,027	0,043	-0,626	0,532
X_6	0,087	0,035	2,488	0,014

Apesar dos resultados obtidos para o modelo completo, o procedimento para seleção de variáveis resultou no modelo composto por IPA (X_5), preço do boi gordo (X_1), preço do bezerro (X_3) e taxa de câmbio (X_6), cujas estimativas estão resumidas na Tabela 3. Para este modelo, obteve-se $R_a^2 = 0,3486$.

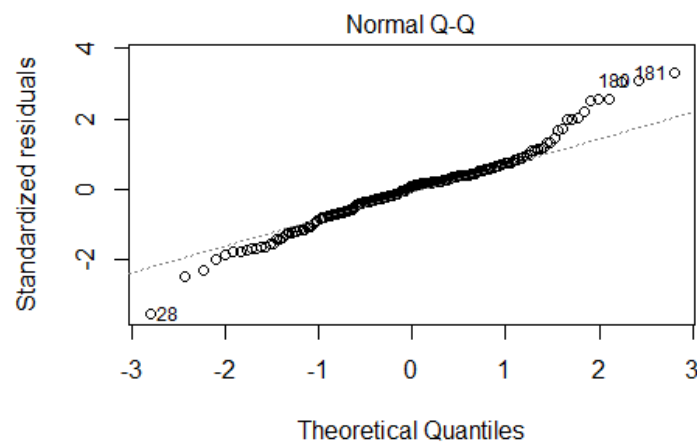
Tabela 3 – Resumo das estimativas dos coeficientes do modelo selecionado pelo stepAIC

Coeficientes	Estimativa	Desvio Padrão	Estatística t	Valor p
Intercepto	-0,773	0,385	-2,010	0,046
X_5	-0,067	0,022	-3,100	0,002
X_1	0,871	0,13548	6,427	<0,0001
X_3	-0,317	0,08346	-3,797	<0,001
X_6	0,078	0,033	2,357	0,019

5.2 Análise de Resíduos

Os resultados obtidos do modelo ajustado pelo algoritmo *stepwise* foram submetidos à análise de resíduos. Em relação à verificação do pressuposto de normalidade dos resíduos, com a aplicação do teste de Shapiro-Wilk obteve-se $W = 0,9673$. Considerando uma amostra de tamanho 50, os valores críticos tabelados da distribuição acumulada de W para 0,01 e 0,99 de probabilidade são, respectivamente, 0,930 e 0,991. Assim, o teste levou à aceitação da hipótese nula e, portanto, de que os resíduos parecem ser normalmente distribuídos. Graficamente, essa decisão é corroborada pelo gráfico Quantil-Quantil para normalidade (Figura 2), embora haja alguns pontos que se distanciem da reta que representa a função identidade.

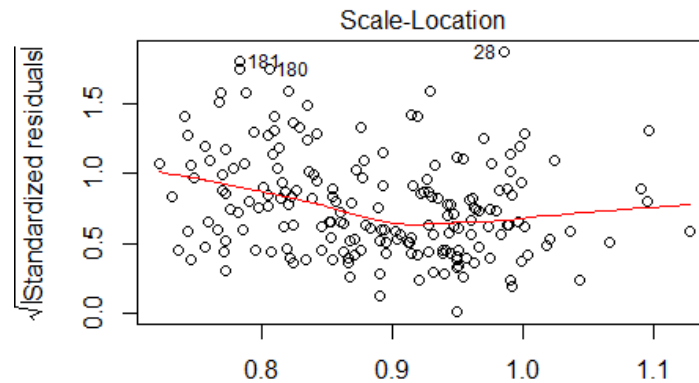
Figura 2 – Gráfico quantil-quantil para normalidade



Em relação à média, o gráfico de resíduos contra valores ajustados (Figura 3), utilizado para verificar o parâmetro de escala da distribuição dos dados, indica que as observações do vetor de erros parece distribuir-se em torno do zero. A aplicação do teste T para a média dos

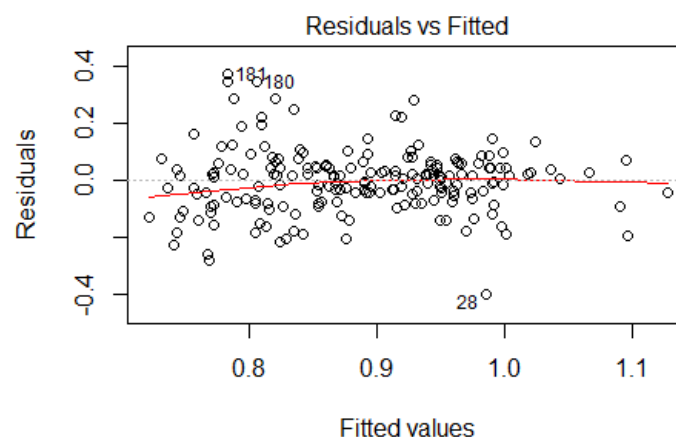
resíduos confirmou tal indício, levando à aceitação da hipótese nula (de que os erros possuem média igual a zero) com nível descritivo de 0,9978.

Figura 3 – Gráfico de Resíduos Padronizados *versus* Valores Ajustados

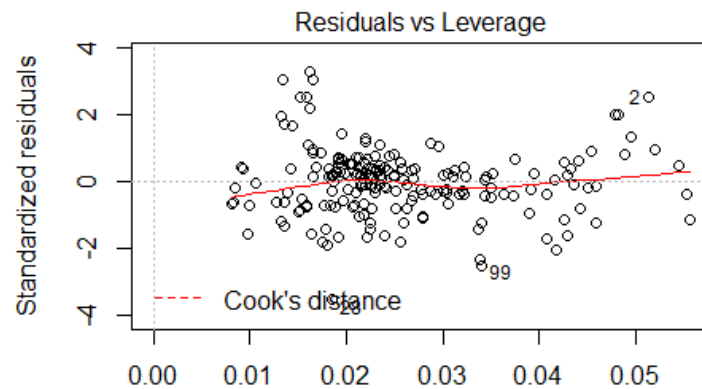


Para avaliação de homocedasticidade, o teste de Breusch-Pagan se mostrou não significativo ao nível de significância $\alpha = 0,01$ (o p-valor obtido foi de 0,01687), conduzindo à aceitação do pressuposto de homogeneidade das variâncias dos resíduos. Em relação à verificação da existência de autocorrelação residual, a análise gráfica dos resíduos contra os valores ajustados (Figura 4) mostrou um agrupamento cônico dos dados, o que se configura como indicativo de tendência funcional dos resíduos. A aplicação do teste de Durbin-Watson original confirmou essa conclusão, levando à rejeição de H_0 (de que não há autocorrelação) com um valor-p menor que 0,0001. A versão generalizada do teste corroborou a autocorrelação de primeira ordem, com nível descritivo próximo a zero.

Figura 4 – Gráfico de Resíduos *versus* Valores ajustados



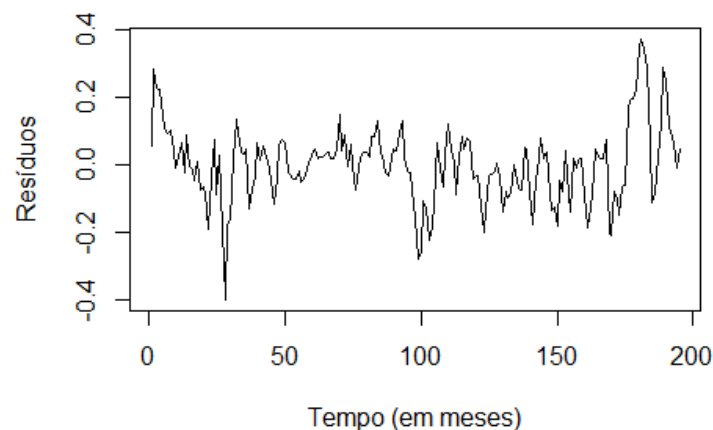
O cálculo da Distância de Cook para os resíduos não revelou nenhum ponto influente (isto é: $D_i < 1 \forall i = 1, 2, \dots, 195$). A análise do gráfico de resíduos *versus* Leverage (Figura 5), apesar de indicar três pontos suspeitos, mostrou os dados razoavelmente agrupados, de modo que seu estudo isolado foi inconclusivo.

Figura 5 – Gráfico de Resíduos *versus* Leverage

5.3 Ajuste via mínimos quadrados generalizados

A modelagem dos resíduos como uma série temporal está representada no Figura 6, o qual parece não indicar componentes como tendência ou sazonalidade na série. A confirmação dos pressupostos de homogeneidade das variâncias e de normalidade dos resíduos tornou plausível a aceitação da suposição de estacionariedade da série e, portanto, o ajuste de um modelo ARMA.

Figura 6 – Gráfico de Resíduos como série de tempo

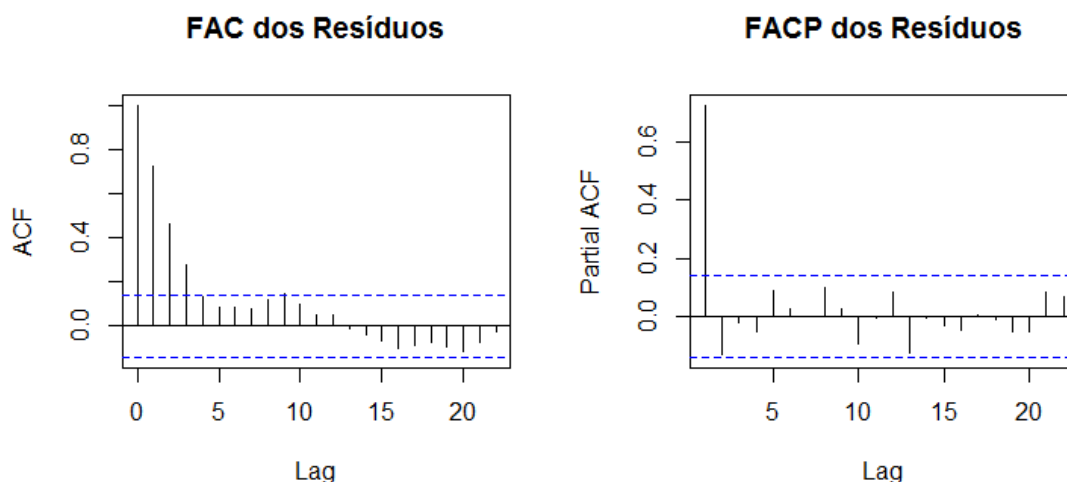


Apesar da aplicação do teste generalizado de Durbin-Watson ter indicado a primeira defasagem como ordem maior da autocorrelação, a função *auto.arima* ajustou um modelo $ARMA(2, 0)$, com parâmetros $\phi_1 = 0,8551754$ e $\phi_2 = -0,1352388$. A ordem do modelo ajustado foi corroborada pelas funções de autocorrelação e de autocorrelação parcial.

Considerando esta parametrização de Σ , o modelo estimado foi:

$$Y = -0,343 - 0,071X_5 + 0,703X_1 - 0,258X_3 + 0,074X_6. \quad (5.2)$$

Figura 7 – Correlogramas para FAC e FACP dos resíduos



Neste ajuste, foram considerados significativamente diferentes de zero os coeficientes das variáveis X_1 (preço do boi gordo) e X_3 (preço do bezerro), conforme a Tabela 4. A ANOVA (Tabela 5), no entanto, indicou todos os coeficientes como significativos, à exceção de β_6 . Esta diferença de resultados ocorreu em razão da mudança da parametrização e modelagem da matriz de resíduos, de modo que já não se verifica a congruência entre os testes T e F para os coeficientes. Conforme Paula (2004, p.79), para Modelos Lineares Generalizados de regressão normal, o teste F é mais indicado, de modo que os resultados da Tabela 5 são considerados plausíveis.

Tabela 4 – Resumo das estimativas para o modelo ajustado via GLS

Coeficientes	Estimativa	Desvio Padrão	Estatística t	Valor-p
Intercepto	-0,343	0,741	-0,463	0,644
X_5	-0,0714	0,047	-1,520	0,130
X_1	0,703	0,193	3,641	<0,001
X_3	-0,258	0,136	-1,906	0,058
X_6	0,074	0,069	1,068	0,287

Tabela 5 – ANOVA para o modelo ajustado pelo método GLS

Coeficientes	Graus de Liberdade	Estatística F	Valor-p
Intercepto	1	2021,779	<0,0001
X_5	1	7,335	0,0074
X_1	1	9,63	0,0022
X_3	1	4,9619	0,0271
X_6	1	1,1401	0,287

A interpretação das estimativas do último ajuste é de que o preço do frango para corte varia: negativamente com o aumento do IPA e do preço do bezerro; positivamente, ante aumento do preço do boi gordo e do dólar.

Uma vez que o IPA representa os custos de produção, seu impacto negativo sobre Y não parece plausível, a menos que esta relação ocorra em razão de maiores ganhos de produtividade na produção de frango de corte em relação ao resultado conjunto dos demais setores da economia brasileira.

Em relação ao preço do bezerro, o impacto negativo sobre o preço do frango pode estar relacionado à mobilidade de capitais entre os setores de carne de frango e de carne bovina: o aumento dos preços do bezerro, fator de produção do mercado bovino, pode desestimular investimento no setor, canalizando os recursos para o mercado de frango, de modo que com o aumento da oferta, haja queda de preços.

O impacto positivo do aumento do preço do boi gordo pode ser explicado a partir da ideia de equilíbrio geral: tendo em conta que as carnes bovina e de frango são bens substitutos (conforme é possível inferir levando em consideração os hábitos alimentares da população brasileira), um aumento do preço do boi gordo, que encarece a carne deste originada, estimula o consumo de frango, cujo aumento ocasiona elevação de preço até o ponto em que o quociente inicial dos preços dos dois produtos seja novamente alcançado.

Em relação à taxa de câmbio, pode-se afirmar que: por se tratar de um produto importante na matriz exportadora brasileira, desvalorizações cambiais (aumentos do preço do dólar) tornam o mercado externo mais atrativo, canalizando boa parte da produção para o exterior, de modo que a redução de oferta interna eleve o preço do frango; por outro lado, valorizações cambiais (redução do preço do dólar) tornam o mercado doméstico mais atrativo, de modo que há aumento da oferta interna e, portanto, redução de preços. O fato de que o teste F considerou o coeficiente desta variável não significativo pode indicar que as flutuações cambiais não alteram significativamente o preço do produto, visto que: os produtores líderes de mercado, que definem os preços praticados, podem escolher entre produzir para o exterior ou para o consumo doméstico com baixo custo de adequação; por se tratar de produto primário, a elasticidade preço do frango para corte é baixa, de modo que as alterações cambiais, que impactam diretamente na inflação, não modifiquem de maneira considerável seu valor no mercado.

O cálculo do $Pseudo - R_{GLS}^2$ forneceu valor de 0,302491, indicando, portanto, sob a perspectiva teórica, uma baixa qualidade do ajuste. No entanto, considerando-se os objetos deste estudo, verifica-se que cerca de 30,25% das variações do preço do frango de corte são provenientes de variáveis que caracterizam o mercado de carne bovina e a interseção deste com o mercado de frango.

6 CONCLUSÕES

O ajuste do MRLM, pelo método dos mínimos quadrados ordinários, aos dados, sob o ponto de vista teórico, retornou um ajuste com coeficientes com baixos erros padrão, sendo considerados significativos aqueles relativos às variáveis preço do boi gordo, preço do bezerro e preço do dólar. O poder de explicabilidade para o modelo completo, medido pelo R_a^2 , foi de apenas 34,65%. O procedimento de seleção de variáveis (que incluiu as variáveis IPA, preço do boi gordo, preço do bezerro e preço do dólar) ajustou um modelo com coeficiente de determinação um pouco maior (0,3486).

A análise de resíduos não confirmou o pressuposto de independência dos erros, de modo que foi ajustado um modelo pelo método dos GLS, considerando a parametrização da matriz de covariância dos resíduos pelo modelo ARMA(2,0). Este último modelo considerou significativos os coeficientes relacionados às variáveis IPA, preço do boi gordo e preço do bezerro. A avaliação deste ajuste, por meio do pseudo R^2 retornou, aproximadamente, 0,3025, o que indica baixa qualidade do ajuste.

Do ponto de vista prático, conclui-se que as variações dos preços de frango sofrem impactos consideráveis das variáveis que caracterizam o mercado bovino em relação aos componentes de custos e de preço. Especificamente, em referência à estimação via GLS, o ajuste indicou que cerca de 30,25% das variações do preço do frango são explicadas pela caracterização do mercado de carne bovina - proporção alta, visto que componentes particulares do próprio mercado de carne de frango não foram incluídos no modelo.

Não obstante, para melhor caracterização desta relação, sugere-se a adição de variáveis específicas aos setores considerados (como preço de rações, por exemplo), além de modelagem através de equações simultâneas que, concomitantemente, descrevam os dois mercados, levando em consideração a interação entre estes.

REFERÊNCIAS

ABBAD, G.; TORRES, C. V.. **Regressão múltipla stepwise e hierárquica em Psicologia Organizacional: aplicações, problemas e soluções**. Universidad Federal do Rio Grande do Norte, 2002.

AKAIKE, H. Information theory as an extension of the maximum likelihood principle. In: INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 2., Budapest, 1973. Proceedings. Budapest, Akadémia Kiadó, 1973. p.267-281.

BARROSO, L. M. A. et al. Avaliação do teste generalizado de Durbin-Watson. **Revista Brasileira de Biometria**, v. 30, p. 432-441, 2012.

BREUSCH, T.; PAGAN, A. Teste simples para heterocedasticidade e coeficiente de variação aleatória econométrica. Sociedade Econométrica, Rio de Janeiro, v. 47, p. 1287-1294, 1979.

DE CARVALHO, T.B; BACCHI, M. R. P.. Estudo da elasticidade-renda da demanda de carne bovina, suína e de frango no Brasil. In: **Anais do XXXV Encontro Nacional de Economia [Proceedings of the 35th Brazilian Economics Meeting]**. ANPEC-Associação Nacional dos Centros de Pós-graduação em Economia [Brazilian Association of Graduate Programs in Economics], 2007.

DE LEEUW, Jan. Information theory and an extension of the maximum likelihood principle by Hirotogu Akaike. **Department of Statistics, UCLA**, 1992.

EHLERS, R. S. Análise de séries temporais. **Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná**, 2007.

FOX, J.; WEISBERG, H. S. **An R companion to applied regression**. Los Angeles: Sage Publications, 2010.

GRUBER, A. M.; REIS JR, D. S.; STEDINGER, J. R. DESENVOLVIMENTO DE MÉTRICAS PARA DIAGNÓSTICO DE REGRESSÃO HIDROLÓGICA BAYESIANA COM MODELO GLS, 2007.

HAIR, J. F. et al. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2007. 593 p. il.

MEDEIROS, A. L. **REGRESSÃO MÚLTIPLA EO MODELO ARIMA NA PREVISÃO DO PREÇO DA ARROBA DO BOI GORDO**. 2006. Tese de Doutorado. UNIVERSIDADE FEDERAL DE ITAJUBÁ.

MILOCA, S. A.; CONEJO, P. D. Multicolinearidade em modelos de regressão. **XXII Semana Acadêmica de Matemática da Unioeste**, 2008.

PAULA, Gilberto Alvarenga. **Modelos de Regressão: com apoio computacional**. São Paulo: IME-USP, 2004.

PYNDICK, R. S., RUBINFELD, D. L. **Microeconomia**. Tradução: Eleutério Prado, Thelma Guimarães. 6 ed. São Paulo: Pearson Prentice Hall, 2005.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

RENCHER, A. C.; SCHAALJE, G. B. **Linear models in statistics**. John Wiley Sons, 2007.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, p. 591-611, 1965.