

MODELOS ESTATÍSTICOS E ECONOMÉTRICOS PARA ESTUDO DA SAZONALIDADE DE PREÇOS: O CASO DO PREÇO DA CARNE DE BOI¹

Antônio Aguirre²
Luis A. Aguirre³

RESUMO

Nesse trabalho estuda-se, utilizando diferentes métodos de análise, o componente de variação sazonal de uma série de preços recebidos pelos produtores de gado do Estado de São Paulo, série que se estende de 1954 a 1995.

O primeiro método empregado na análise é descritivo: consiste em calcular as “médias sazonais” e os “fatores sazonais”. Na seqüência, mostra-se como se estimam as médias sazonais mediante análise de regressão com uma equação que tem onze variáveis *dummy* como regressores. Contudo, a avaliação dessa regressão mostra que seus resultados são insatisfatórios devido, principalmente, ao problema de autocorrelação dos resíduos.

Uma nova especificação da equação de regressão é proposta, sendo necessário utilizar os critérios ERR (*Error Reduction Ratio*) e o de Akaike para determinar a ordem e o número ótimo de termos autoregressivos a ser incluídos na equação. Esses resultados mostram um ajuste mais satisfatório dos dados.

1 INTRODUÇÃO

É amplamente sabido que o preço da carne de boi apresenta, na Região Centro-Sul do Brasil, uma marcada variação intra-anual, denominada “variação sazonal” ou “sazonalidade”. Tal fenômeno acontece devido à variabilidade das condições climáticas ao longo do ano. Com efeito, no período do ano em que as condições de umidade e temperatura são favoráveis à

- 1 Os autores agradecem o apoio financeiro da FAPEMIG e do CNPq. Agradecem, também, os valiosos comentários de um parecerista anônimo.
- 2 Professor do Departamento de Ciências Econômicas da FACE e pesquisador do CEDEPLAR da UFMG.
- 3 Professor do Departamento de Engenharia Eletrônica da EE/UFMG e pesquisador do CPDEE/UFMG.

existência de boas pastagens, a oferta de carne é relativamente grande.⁴ É a época da safra. Como a demanda é mais estável que a oferta, nessa época do ano os preços tendem a ser menores que a média anual. O contrário acontece no período da entressafra.

Existem pelo menos duas razões que justificam estudar as características da estrutura sazonal do preço da carne de boi. A primeira é de natureza simplesmente descritiva e tenta responder às seguintes perguntas: quais meses do ano correspondem à safra, quais à entressafra, e qual a diferença entre os preços observados em ambas as épocas, medida em moeda constante? As respostas a essas perguntas são, necessariamente, de natureza empírica, ou seja, devem basear-se em dados referentes a uma certa etapa da cadeia de comercialização (preços ao produtor, ao consumidor, atacado ou varejo) e correspondentes a uma região (ou mercado, ou cidade) determinada do País. A segunda razão é ajustar a série para eliminar a variação sazonal e obter uma série “dessazonalizada”. Tal procedimento é, às vezes, considerado um pré-requisito para se realizar outras análises. A idéia básica, nesse caso, é que, após a eliminação do fator (ou componente) sazonal da série, é possível concentrar a análise nos outros componentes, como tendência e ciclo.

Nesse artigo, tentar-se-á apenas responder às perguntas mencionadas no parágrafo acima, não nos preocupando a obtenção de séries dessazonalizadas. Para isso, será usada uma série histórica dos preços médios mensais ao produtor da carne bovina (preço do boi-gordo), no Estado de São Paulo.⁵ Apesar de tratar-se de preços recebidos pelos produtores, é provável que a sazonalidade dessa série seja muito próxima, se não exatamente igual, à apresentada pelos preços da carne de boi pagos pelos consumidores no mesmo Estado. Levando em conta o aspecto regional, ao contrário, não há bases suficientes para afirmar *a priori* que o padrão sazonal observado nessa série seja o mesmo existente em outras regiões do País. Contudo, é razoável admitir que a estrutura de variação sazonal encontrada na série de São Paulo, seja válida, também, para estados vizinhos, como é o caso do Estado de Minas Gerais.

A organização do artigo é a seguinte. Na próxima seção será feita uma breve resenha bibliográfica e na seção 3 serão discutidas várias metodologias de cálculo da sazonalidade de preços, sendo apresentados os dados e os resultados da análise estatística. Finalmente, na última seção, apresentam-se algumas conclusões.

4 Não consideramos aqui qualquer possível defasagem entre a existência de “condições ótimas de clima” e a quantidade ofertada de carne.

5 Os dados sobre preços são compilados e divulgados pela Secretaria de Estado da Agricultura do Estado de São Paulo.

2 REVISÃO DA LITERATURA

Existem dois pontos de vista diferentes sobre a natureza da sazonalidade nos dados econômicos. O primeiro considera a variação sazonal como uma parte fundamental de muitas séries temporais econômicas e recomenda que, quando está presente, o analista deve tratar de explicá-la. Assim, idealmente, um modelo econométrico para a variável dependente y_t deveria explicar qualquer sazonalidade nela presente mediante variações sazonais nas variáveis independentes, ou incluindo variáveis explicativas ligadas ao clima, ou variáveis *dummy* sazonais.

Outro ponto de vista, associado a Sims (1974), considera que a sazonalidade é mais um tipo de ruído que contamina os dados econômicos. Segundo esse autor, não se pode esperar que a teoria econômica explique esse tipo de variação que, no caso de estar presente nas variáveis independentes, é semelhante ao “problema de erros nas variáveis”. Portanto, na modelagem econômica devem-se usar dados sazonalmente ajustados, isto é, séries corrigidas de maneira tal que representem o que elas teriam realmente sido na ausência de sazonalidade.

Praticamente todas as estatísticas econômicas geradas pelo Governo dos Estados Unidos, por exemplo, são dessazonalizadas pelo método conhecido como Census X-11 (Shiskin *et al.*, 1967)⁶. Estudos de simulação realizados para analisar os efeitos desse procedimento sobre as propriedades das séries detectaram que o Census X-11 introduz forte autocorrelação espúria negativa nas defasagens sazonais e algumas autocorrelações positivas mais fracas em defasagens menores. A técnica não somente falha na tentativa de eliminar toda a sazonalidade, mas também introduz algumas regularidades espúrias (Wallis, 1974).

Hylleberg (1994) faz uma revisão dos novos métodos na área, como aqueles que existem para realizar testes de raízes unitárias sazonais e cointegração sazonal. Esse autor afirma que a principal causa que impulsionou os novos estudos foi a compreensão de que a variação sazonal explica a maior parte da variância total em muitas séries temporais econômicas, e de que essa variação sazonal não é regular. Também informa que agora a tendência é juntar considerações econômicas às técnicas estatísticas na análise da sazonalidade.

Ericsson *et al.* (1994) discutem a questão do uso de séries sazonalmente ajustadas e/ou de séries não ajustadas e descrevem testes para

6 O método denominado Census X-11 separa os componentes de uma série temporal em tendência, índice sazonal, dias úteis (*trading day*) e componente irregular. Usa-se, basicamente, para dessazonalizar as estatísticas oficiais do Governo dos E.U.A.

comparar modelos estimados com cada tipo de dados. O trabalho consiste numa detalhada descrição de uma aplicação do método “geral-a-específico”, de David Hendry, à demanda de moeda na Inglaterra, e os procedimentos necessários para realizar os testes correspondentes.

Franses (1996) faz uma revisão dos métodos mais recentes na modelagem econométrica de séries temporais com sazonalidade. A principal ênfase é colocada em modelos que incorporam, de forma explícita, uma descrição da sazonalidade em vez de eliminá-la mediante métodos de ajuste sazonal. As principais questões investigadas nesse trabalho são:

- i) que proporção da variância total das séries macroeconômicas pode ser atribuída à sazonalidade?;
- ii) é a variação sazonal constante ao longo do tempo?;
- iii) é possível separar de forma adequada a variação sazonal, a tendência, e as variações cíclicas de uma dada série temporal?

Com relação à última questão Franses conclui que, para algumas variáveis macroeconômicas, é muito difícil (se não impossível) separar as variações sazonais das não-sazonais. Essa conclusão sugere que, às vezes, é mais adequado incorporar explicitamente a variação sazonal na modelagem econométrica, ainda que se pague o preço de ter modelos menos parcimoniosos do que aqueles baseados em séries sazonalmente ajustadas. Em particular, discutem-se nesse trabalho a natureza dos modelos econométricos com raízes unitárias sazonais para séries temporais univariadas e multivariadas.

No Brasil existem alguns trabalhos específicos sobre a variação sazonal do preço da carne de boi. No primeiro deles Garcia (1982) utiliza análise harmônica para estudar a sazonalidade do preço da carne de boi magro em Campo Grande, MS, e de boi-gordo em Araçatuba, SP. O segundo estudo analisa a sazonalidade do preço da carne de boi-gordo em três estados do Nordeste: Maranhão, Piauí e Ceará (Silva, Lemos, 1986). Para isso, utiliza análise da variância e detecta diferenças significativas dos preços entre estados, meses e anos.

Em outro artigo Margarido *et al.* (1996) estudam a influência das cotações do dólar no mercado paralelo e do índice pluviométrico das principais regiões bovinocultoras paulistas sobre os preços recebidos pelos produtores de boi-gordo de São Paulo para o período 1984-1994. Os autores usam modelos ARIMA e de funções de transferência para mostrar que os efeitos das variações das precipitações pluviométricas são transferidos parcialmente ao preço da arroba de boi-gordo com defasagem temporal de cinco meses.

Finalmente, utilizando médias trimestrais dos preços recebidos pelos produtores de boi-gordo no Estado de São Paulo, (Aguirre, 1997) realiza

testes de raízes unitárias sazonais para estabelecer a ordem de integração dessa série. Os resultados desse trabalho indicam que a série tem uma raiz unitária na frequência zero mas não nas frequências sazonais. Isso implica que as variações presentes na série são uma mistura de sazonalidade determinística e estocástica estacionária que pode ser modelada com variáveis *dummy* sazonais.

3 METODOLOGIAS DE CÁLCULO DA SAZONALIDADE DE PREÇOS

No período de quase 42 anos coberto pela série sob análise, os preços nominais foram expressos nas várias unidades monetárias existentes no País e sofreram grandes aumentos devido aos processos inflacionários ocorridos nesse lapso de tempo. Por esse motivo, os primeiros ajustes realizados nos dados dizem respeito à homogeneização das unidades e à deflação dos valores, para expressá-los em moeda de poder aquisitivo constante. Para esse último fim, usou-se o IGP-DI (Índice Geral de Preços-Disponibilidade Interna) da Fundação Getúlio Vargas. Os preços médios de cada mês estão expressos em reais de dezembro de 1995 por arroba (Gráfico 1). Os valores da série variam entre um mínimo de R\$ 19,42 (em janeiro de 1957) e um máximo de R\$ 81,98 (em dezembro de 1986, época do Plano Cruzado), com média de R\$ 35,58 e desvio padrão de R\$ 11,24 (preços de dezembro de 1995 por arroba).

Dividindo a série em subperíodos, parece evidente a existência de diferentes patamares. Os valores da média, desvio padrão e coeficiente de variação de quatro subperíodos estão apresentados na Tabela 1.

Tabela 1
ESTATÍSTICAS DA SÉRIE TRIMESTRAL DE PREÇOS
DA ARROBA DE BOI-GORDO
(R\$ de Dez. 1995)

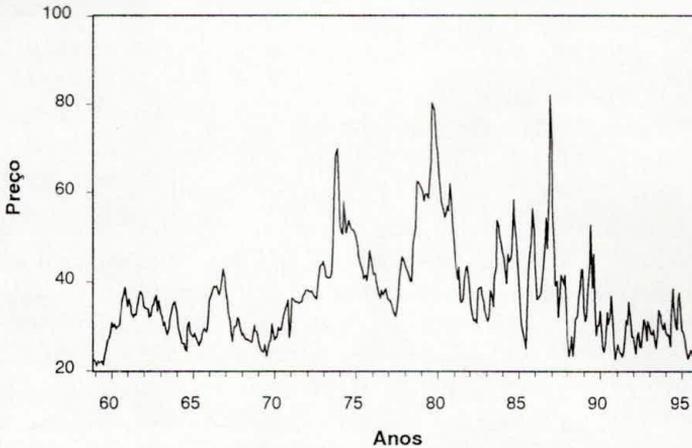
Datas	Número de observações	Média	Desvio padrão	Coefficiente de variação (%)
Mar./54-dez./59	70	23,51	2,70	11,48
Jan./60-dez./72	156	32,31	4,63	14,13
Jan./73-dez./87	180	46,13	10,96	23,76
Jan./88-dez./95	96	29,92	5,64	18,85
Mar./54-dez./95	502	35,58	11,24	31,59

Fonte: Elaboração própria, a partir de dados da Secretaria da Agricultura do Estado de São Paulo.

Gráfico 1

PREÇO TRIMESTRAL DO BOI-GORDO ESTADO DE SÃO PAULO – 1954/1995

(Em reais de Dez./95 por arroba)



Fonte: Elaboração própria, a partir de dados da Secretaria da Agricultura do Estado de São Paulo.

Pode-se observar um aumento das médias e dos desvios padrão nos três primeiros subperíodos e uma diminuição no último. Como o desvio padrão aumenta mais do que proporcionalmente (com relação à média), o coeficiente de variação cresce até 1987 e no último subperíodo diminui.

Como a série apresenta três máximos locais (em abril de 1974, setembro de 1979 e dezembro de 1986), escolhendo um desses máximos é possível visualizar a existência de uma tendência crescente seguida de outra decrescente.⁷ Realizado o teste de mudança estrutural de Chow usando alternadamente os três pontos como data da modificação, em todos os casos rejeita-se a hipótese nula de ausência de mudança estrutural. O maior valor da estatística F corresponde à primeira data ($F = 18,96$).

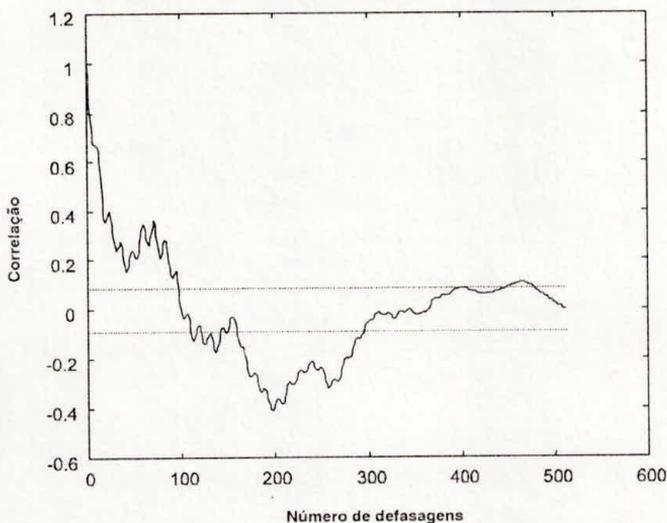
Também aparecem claramente representados os ciclos plurianuais associados com a atividade pecuária – o ciclo do gado – amplamente estudado na literatura sobre o tema. Devido à escala utilizada, as variações sazonais, principal objeto do presente trabalho, não são muito evidentes, mas sua presença pode ser intuída se observarmos as flutuações de alta frequência presentes no mencionado gráfico. Na seqüência serão apresentadas evidências adicionais sobre tais variações.

⁷ Estas tendências serão eliminadas adiante para elaborar o espectro de potência apresentado no Gráfico 3.

Duas técnicas complementares que podem ser usadas para melhor caracterizar a série em análise são o correlograma (no domínio do tempo) e o espectro de potência (no domínio da frequência). O correlograma dos preços médios mensais do boi-gordo aparece no Gráfico 2. Pelo fato dos valores de tal correlograma não diminuírem rapidamente é evidente a forte autocorrelação presente na série (na defasagem de ordem 72 o coeficiente ainda é 0,30). As flutuações observadas, com duração próxima às doze defasagens, indicam a presença das variações sazonais. Um correlograma com essas características é consequência dos fenômenos observados nos dados: tendências, variação sazonal e variação cíclica. A tendência faz com que os valores das autocorrelações não diminuam rapidamente para zero, a variação sazonal explica as flutuações de alta frequência, e o ciclo plurianual as flutuações de baixa frequência.

Gráfico 2

**CORRELOGRAMA DO PREÇO DO BOI-GORDO
ESTADO DE SÃO PAULO - 1954/1995**



Fonte: Elaboração própria, a partir de dados da Secretaria da Agricultura do Estado de São Paulo.

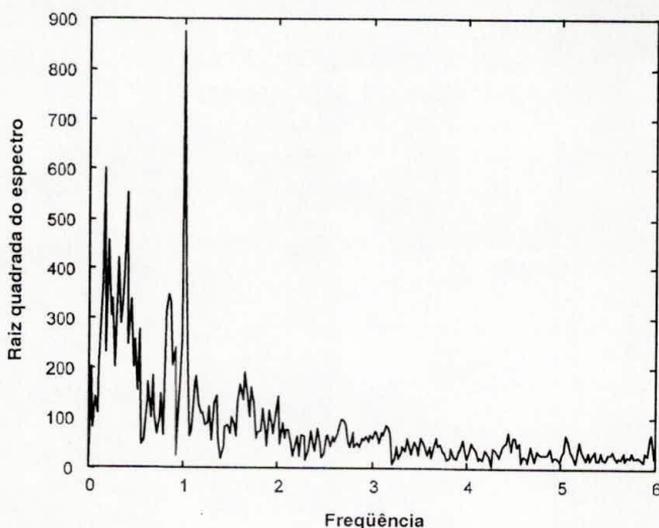
O Gráfico 3 apresenta o espectro de potência dos desvios dos preços com relação a duas tendências lineares.⁸ A função de densidade espectral ou espectro de potência contém informações sobre a série original,

8 Esse gráfico apresenta a raiz quadrada da convolução da transformada de Fourier dos desvios sem nenhum tipo de alisado.

obtidas focalizando-a de um ponto de vista diferente: o domínio da frequência. Essas informações indicam quais as frequências que explicam as maiores proporções da variância total da série. Dessa forma, essa técnica pode ser útil na identificação de ciclos escondidos em meio às variações aleatórias ou ruídos (Aguirre, 1995). No caso da série que se está analisando, o espectro de potência indica a existência de dois movimentos cíclicos importantes: o primeiro na frequência 0,15 e o segundo na frequência unitária.

Gráfico 3

ESPECTRO DE POTÊNCIA DOS DESVIOS DOS PREÇOS COM RELAÇÃO A DUAS TENDÊNCIAS LINEARES ESTADO DE SÃO PAULO – 1954/1995



Fonte: Elaboração própria, a partir de dados da Secretaria da Agricultura do Estado de São Paulo.

O pico maior, correspondente à frequência f_1 , está associado a um período de 6,22 anos (75 meses). Isso significa que nos dados existe um ciclo de baixa frequência, que se repete a cada 6,22 anos, e que explica uma proporção importante da variância total da série. Esse pico corresponde ao denominado “ciclo do gado”.

O pico relativamente alto, e estatisticamente significativo, observado na frequência 1 significa que um outro ciclo também é importante na explicação da variância total da série. Esse ciclo de maior frequência – que

se completa num período de um ano – descreve as variações intra-anuais (variações sazonais).⁹

3.1 Sazonalidade: método descritivo

A forma mais simples e tradicional de representar a estrutura de variação sazonal de uma série temporal consiste em calcular as “médias sazonais”, ou seja, a média de todas as observações correspondentes aos meses de janeiro presentes na série, a média de todas as referentes aos meses de fevereiro, de março *etc.* A diferença entre as médias sazonais de cada mês e a média global da série fornece os “fatores sazonais” (FS).

Simbolicamente,

$$FS = M_s - M_g$$

onde: M_s = média sazonal e M_g = média global.

Tabela 2

**MÉDIAS SAZONAIS E FATORES SAZONAIS DOS PREÇOS
DA CARNE DE BOI-GORDO NO ESTADO DE SÃO PAULO**

R\$ de Dez./95 por arroba

Mês	Média sazonal	Fator sazonal
Jan.	37,84	2,26
Fev.	35,96	0,38
Mar.	34,71	-0,87
Abr.	34,64	-0,94
Mai	34,49	-1,09
Jun.	34,27	-1,31
Jul.	35,38	-0,20
Ago.	37,39	1,81
Set.	39,61	4,03
Out.	40,75	5,17
Nov.	41,20	5,62
Dez.	39,26	3,68

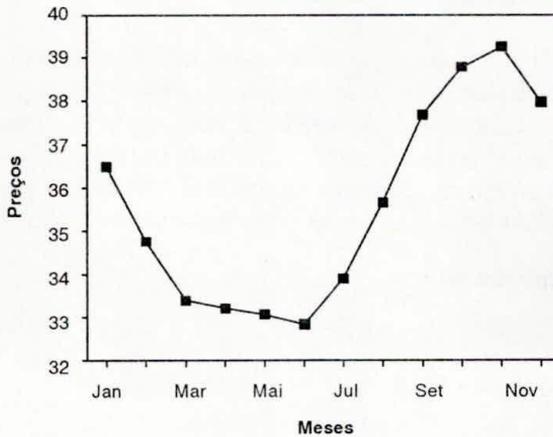
Fonte: Elaboração própria, a partir de dados da Secretaria da Agricultura do Estado de São Paulo.

⁹ A significância estatística dos picos é estabelecida num outro gráfico (periodograma) onde o logaritmo decimal do espectro de potência é devidamente alisado.

Gráfico 4

ESTRUTURA SAZONAL

Preços do boi-gordo em São Paulo



Fonte: Tabela 2.

O resultado desses cálculos está apresentado na Tabela 2 e no Gráfico 4, onde pode observar-se que os preços médios do período de fevereiro a julho são menores que a média global da série, enquanto que os restantes são superiores à mencionada média¹⁰. Essas informações definem a extensão das duas épocas sazonais conhecidas como safra e entressafra, respectivamente.

3.2 Sazonalidade: uso de variáveis *dummy*

Existem várias formas de implementar a análise da variação sazonal utilizando variáveis *dummy*. Um desses procedimentos¹¹ consiste em criar, para cada mês, uma variável binária com valores unitários para as observações correspondentes a esse mês, e valores zero para todos os outros. Assim, cada mês tem sua própria *dummy*. Contudo, se a regressão inclui um intercepto, apenas onze *dummies* formam o conjunto de variáveis explicativas. No caso que nos ocupa o modelo é o seguinte:

10 Essas médias e fatores sazonais correspondem ao período dezembro de 1958 a dezembro de 1995, num total de 445 observações. A redução da amostra tem por finalidade produzir resultados comparáveis com aqueles das próximas seções.

11 Ver, por exemplo, Gujarati (1995).

$$P = \alpha + X\beta + u \quad (1)$$

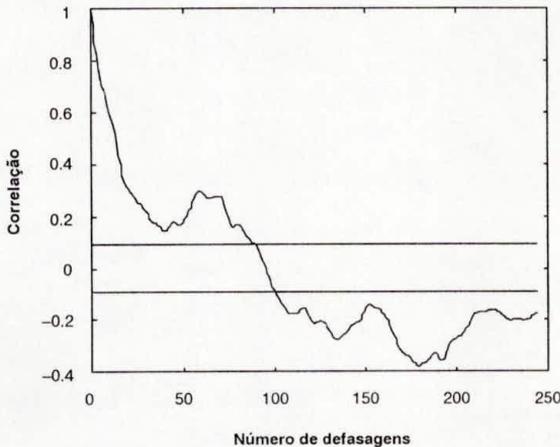
onde P é um vetor de N observações da variável explicada, α é o intercepto, X é uma matriz ($N \times 11$), β é um vetor de parâmetros sazonais, e u é um vetor de variáveis aleatórias idênticas e independentemente distribuídas com média zero e variância σ^2 .

Ao se usar apenas onze *dummies*, o mês não representado passa a ser um ponto de referência ou “base de comparação”, com relação ao qual são comparados os outros meses. Os meses cujas *dummies* têm coeficientes de regressão estatisticamente iguais a zero, são sazonalmente semelhantes ao mês base. Aqueles meses cujos coeficientes de regressão são positivos e significativos correspondem, por sua vez, a períodos nos quais os preços são sazonalmente maiores que o preço do mês base.

Na Tabela 3 pode-se observar que os coeficientes estimados pela regressão coincidem com as médias sazonais da Tabela 2. Com efeito, a constante da regressão estima a média do mês base (junho) e os outros coeficientes estimam a quantia que deve ser adicionada para se obter, a partir do mês base, a média sazonal do mês em questão.¹²

Gráfico 5

CORRELOGRAMA DOS RESÍDUOS DA EQUAÇÃO (1)



Fonte: Elaboração própria.

¹² Por esse motivo, Gujarati (1995) chama os coeficientes de regressão, nesse caso, de “interceptos diferenciais”.

Apesar da regressão reproduzir exatamente os valores das médias sazonais, a avaliação global desse ajuste não é satisfatória, já que o coeficiente de determinação é baixo e, principalmente, os resíduos apresentam autocorrelação (ver correlograma no Gráfico 5). Em consequência, os erros padrão dos coeficientes estimados (e os testes deles decorrentes) não são confiáveis. Esse problema será tratado na seqüência.

Tabela 3
RESULTADOS DA ANÁLISE DE REGRESSÃO
(Mês base de comparação: junho)

Variável	Coefficiente	Erro padrão	t de Student	Probabilidade
C	34.27281	1.780663	19.24721	0.0000
JAN	3.562537	2.518238	1.414694	0.1579
FEV	1.685470	2.518238	0.669305	0.5037
MAR	0.435054	2.518238	0.172761	0.8629
ABR	0.367405	2.518238	0.145898	0.8841
MAI	0.215571	2.518238	0.085604	0.9318
JUL	1.106201	2.518238	0.439276	0.6607
AGO	3.119721	2.518238	1.238851	0.2161
SET	5.335549	2.518238	2.118762	0.0347
OUT	6.476287	2.518238	2.571753	0.0105
NOV	6.929286	2.518238	2.751640	0.0062
DEZ	4.991159	2.501616	1.995174	0.0467
R^2	0.050297	Média var dep		37.12963
R^2 ajustado	0.026170	D. P. var dep		10.97593
E. P. da regressão	10.83135	C. I. Akaike		4.791486
Σ resíduos ao quadrado	50798.78	Estatística F		2.084723
Estatística D-W	0.099958	Probab. de F		0.020272

Fonte: Elaboração própria.

Nota: MQO // Variável dependente: PREÇO

Amostra (ajustada): Dezembro 1958 - Dezembro 1995

Observações incluídas: 445

3.3 Regressão dinâmica

A regressão estimada na subseção anterior representa um ajuste insatisfatório dos dados, com um coeficiente de determinação muito baixo. Na realidade, se o ajuste da função (1) estimasse uma única média, o coeficiente de determinação seria zero, já que não haveria variação explicada. No caso das variáveis *dummy*, contudo, a porção de variação explicada surge do fato dos desvios das observações de cada mês serem calculados com relação a diferentes médias sazonais.

Na regressão da Tabela 3 a Estatística de Durbin-Watson indica que os resíduos estão forte e positivamente autocorrelacionados. Isso implica na quebra de uma das hipóteses sobre a natureza dos erros aleatórios u da relação (1). A partir do valor amostral da estatística D-W é possível estimar o valor aproximado do coeficiente do processo autoregressivo de primeira ordem dos resíduos, da seguinte forma: $1 - DW/2$. Esse valor, superior a 0,95, mede a autocorrelação dos resíduos que, por sua vez, indica a existência de erros de especificação na estrutura do modelo estimado. Se a estrutura do modelo está incorretamente especificada, os estimadores dos parâmetros são inconsistentes e o modelo torna-se inútil como marco para fazer testes estatísticos de hipóteses. No nosso caso, sabe-se perfeitamente que não foram incluídas nem as tendências nem a flutuação cíclica da série. Portanto, é necessário mudar a estrutura do modelo para levar em conta esses fatores e cabe perguntar se os coeficientes sazonais estimados continuarão sendo os mesmos caso obtidos com uma especificação diferente que não apresente as deficiências da anterior.

A nova especificação proposta incluirá valores defasados da variável dependente como variáveis explicativas. Em tal caso a estrutura dinâmica em termos do máximo número de defasagens da variável dependente deve ser especificada.

É importante notar que a inclusão dos termos autoregressivos tem como principal objetivo eliminar a autocorrelação nos resíduos. Este procedimento é fundamental uma vez que o objetivo é apenas testar a estabilidade dos coeficientes das variáveis *dummy* sazonais, o que só poderá ser bem feito se os estimadores desses coeficientes forem consistentes. Nenhuma interpretação econômica será dada aos coeficientes dos termos autoregressivos.

Considere-se um modelo dinâmico muito geral que inclui vários tipos de modelos de defasagens distribuídas. A forma mais geral desse modelo é conhecida como ARDL (*Autoregressive Distributed Lag Model*). A partir dessa forma geral diferentes pressupostos definem diversos modelos. Assim,

por exemplo, a seguinte especificação é denominada modelo ARMAX (*Auto-regressive Moving Average with Exogenous Variables*):

$$A(L)y_t = \alpha + \beta(L)x_t + C(L)\varepsilon_t \quad (2)$$

onde $A(L)$, $B(L)$ e $C(L)$ são polinômios no operador de defasagens. Nessa forma o modelo tem variáveis dependentes e independentes defasadas, e um termo de erro que segue um processo de média móvel.

Um modelo ARMA pode ser obtido eliminando os termos referentes às variáveis independentes (exógenas) da equação (2), ou seja, eliminando os termos $B(L)x_t$. Os modelos ARMA e ARMAX estão amplamente difundidos na análise de séries temporais no contexto de regressões dinâmicas, havendo usos e interpretações alternativas (Greene, 1993).

3.3.1 Especificação de modelos

O modelo ARMAX mostrado na equação (2) é uma representação matemática de um sistema qualquer do qual se observa a seqüência temporal de uma de suas variáveis. No caso da equação (2) a seqüência observada é $y_1, y_2, \dots, y_{t-1}, y_t$. Lembrando que $A(L)$, $B(L)$ e $C(L)$ são polinômios no operador de defasagens e supondo que os primeiros elementos de $A(L)$ e $C(L)$ sejam iguais a um, a equação (2) pode ser reescrita da seguinte maneira:

$$y_t = \alpha + a_1 y_{t-1} + \dots + a_{ny} y_{t-ny} + b_1 x_{t-1} + \dots + b_{nu} x_{t-nu} + c_1 \xi_{t-1} + \dots + c_{n\xi} \xi_{t-n\xi} + \xi_t \quad (3)$$

Ou na forma vetorial:

$$\hat{y}_t = \varphi^T \hat{\Theta} + \xi_t \quad (4)$$

onde $\varphi^T = [1 \ y_{t-1} \ \dots \ y_{t-ny} \ x_{t-1} \ \dots \ x_{t-nu} \ \xi_{t-1} \ \dots \ \xi_{t-n\xi}]$ é o vetor de regressores, $\hat{\Theta} = [\alpha \ a_1 \ \dots \ a_{ny} \ b_1 \ \dots \ b_{nu} \ c_1 \ \dots \ c_{n\xi}]^T$ é o vetor de parâmetros a determinar, ξ_t é uma variável aleatória e o chapéu indica predição no caso de variáveis e valores estimados no caso de parâmetros. Assim, a equação explica a variável observada como uma combinação linear de: observações passadas da mesma variável, variáveis exógenas e variáveis aleatórias. Estas variáveis são as que compõem o vetor φ . Na equação (4) a variável aleatória ξ_t é o erro estocástico, que se supõe tenha distribuição $N(0, \sigma)$.¹³

¹³ Note-se que essa especificação também pode ser escrita sem incluir erros defasados no vetor de regressores. Nesse caso, o termo estocástico da relação, digamos u_t , seria um processo autoregressivo em ξ_t .

Um dos problemas mais difíceis de resolver na modelagem de um sistema ou uma série temporal é o de especificar a estrutura do modelo a ser utilizado, ou seja, escolher quais regressores devem ser utilizados para explicar a série de observações. Este problema é denominado, por autores de outras áreas do conhecimento, como o **problema de seleção de estrutura** (Billings *et al.*, 1989; Aguirre, 1994).

Apesar da solução do problema de especificação não ser necessariamente trivial, o requisito básico é simples. Ao determinar a estrutura do modelo é fundamental que ela seja capaz de “absorver” a informação relevante da seqüência de observações. Para que isto aconteça é necessário que haja uma certa correspondência entre os regressores e o tipo de comportamento presente nos dados. Por exemplo, se houver uma relação quadrática entre uma determinada variável e a observação, é importante que o quadrado de tal variável seja um dos regressores. Ou ainda, se a série temporal for de um sistema dinâmico, é importante que haja termos autoregressivos entre os regressores. Esta última observação será útil na especificação do modelo para a estimação dos coeficientes de sazonalidade no problema que nos ocupa.

Uma maneira de escolher a estrutura mais adequada para um modelo é utilizar o critério ERR (*Error Reduction Ratio*) pelo qual os termos candidatos são ordenados segundo a proporção explicada da variância dos dados (Billings *et al.*, 1989). Dado um modelo na forma da equação (4), o ERR de cada termo pode ser determinado da seguinte maneira. Quando é aplicada a um conjunto de observações a regressão da equação (4) resulta numa equação matricial como segue:¹⁴

$$y = \Psi \Theta + \Xi.$$

A matriz de regressores Ψ pode ser decomposta utilizando técnicas de ortogonalização do tipo Gram-Schmidt, Householder *etc.* (Chen *et al.*, 1989). Nesses casos a última equação toma a seguinte forma, chamada de modelo auxiliar:

$$y = Wg + \Xi.$$

onde W é uma matriz de regressores ortogonais entre si, e g representa os seus respectivos parâmetros. A preferência pela última equação é justificada com base em dois fatos: primeiro, a última equação é numericamente mais estável e, segundo, um subproduto da decomposição ortogonal é a determinação de um índice útil para escolher os regressores do modelo.

A variância do erro de modelagem ξ , para o caso de média nula ($E\{\xi(t)\} = 0$) é igual ao erro quadrático médio e é dado por:

14 Maiores detalhes deste procedimento serão discutidos na próxima seção.

$$\text{Var}\{\xi(t)\} = \lim_{N \rightarrow \infty} \frac{1}{N} \Xi^T \Xi = \lim_{N \rightarrow \infty} \frac{1}{N} \left[Y^T Y - \sum_{i=1}^{n_0} g_i^2 w_i^T w_i \right],$$

onde g_i são os elementos do vetor de parâmetros g do modelo auxiliar e w_i são os regressores ortogonais (colunas da matriz ortogonal W).

Quando nenhum termo é incluído no modelo, o valor quadrático médio (VQM) da variável explicada $y(t)$ é igual ao valor quadrático médio dos resíduos ξ . A cada novo termo incluído no modelo, o valor quadrático médio de ξ é diminuído de um fator igual a $1/N (g^2 w_i^T w_i)$, onde w_i corresponde ao novo termo e g_i é o seu respectivo parâmetro. A redução no valor quadrático médio pode ser normalizada com relação ao VQM da série temporal. Assim, o ERR de cada termo é definido como (Billings *et al.*, 1989):

$$[ERR]_i = \frac{g_i^2 w_i^T w_i}{Y^T Y}, \quad 1 \leq i \leq n_0$$

onde n_0 é o número de termos candidatos.

Portanto, o ERR indica a porção do VQM da série temporal explicada pela inclusão de um novo termo (regressor) no modelo, e pode ser utilizado na especificação de modelos. Na prática, parte-se de um conjunto de termos candidatos a comporem o modelo e ordenam-se tais termos por ordem decrescente de importância, de acordo com o critério ERR. Uma vez ordenados todos os termos em ordem decrescente de importância, é necessário decidir onde “truncar” o modelo. Para este fim os critérios de informação, como o AIC (Akaike, 1974), são bastante úteis (Aguirre, 1994). Finalmente, os regressores e seus respectivos parâmetros podem ser facilmente obtidos a partir dos regressores e parâmetros do modelo auxiliar (Billings *et al.*, 1989).

3.3.2 Um modelo para estudo de sazonalidade

A próxima tarefa é investigar o problema de seleção de estrutura para um modelo, quando o objetivo é utilizá-lo para estimar a partir dos dados os coeficientes de sazonalidade. Ou seja, procura-se uma estrutura que capacite o modelo a absorver a informação de sazonalidade presente nas observações. O tipo de modelo a ser investigado a seguir tem a seguinte estrutura:

$$A(L)y_t = \alpha + \chi_i^T + \xi_t \quad (5)$$

que é do tipo ARX, onde $\beta \in \mathfrak{R}^{nx}$ é um vetor de coeficientes desconhecidos, $\chi \in \mathfrak{R}^{nx}$ é um vetor de variáveis independentes (*dummies*) não estocásticas correspondentes à observação feita no instante t , e o termo ξ_t é uma variável aleatória com variância constante, serialmente independente, e não-correlacionada com os regressores α e χ_t .

O modelo da equação 5 pode ser representado em forma vetorial como:

$$y_t = \begin{bmatrix} 1 & y_{t-1} & \dots & y_{t-ny} & \chi_t^T \end{bmatrix} \begin{bmatrix} \alpha \\ a_1 \\ \dots \\ a_{ny} \\ \beta \end{bmatrix} + \xi_t \quad (6)$$

que, quando aplicada a uma janela de observações, pode ser escrita em forma matricial da seguinte maneira:

$$y = \Psi \Theta + \Xi \quad (7)$$

onde $y = [y_1 \ y_2 \ \dots \ y_N]^T \in \mathfrak{R}^N$ é um vetor com N observações, $\Psi \in \mathfrak{R}^{N \times 1+ny+nx}$ é a matriz de regressores, $\Theta \in \mathfrak{R}^{1+ny+nx}$ é o vetor de parâmetros a determinar e, finalmente, $\Xi = [\xi_1 \ \xi_2 \ \dots \ \xi_N]^T \in \mathfrak{R}^N$ é o vetor de resíduos. De acordo com Draper, Smith (1981), a solução que minimiza o somatório do quadrados dos erros, $\Xi^T \Xi$, é:

$$\hat{\Theta} = [\Psi^T \Psi]^{-1} \Psi^T y \quad (8)$$

Tal estimador é consistente se as raízes do polinômio $A(L)$ ficarem fora do círculo unitário.

Se a série temporal em estudo for dinâmica, isto é, se houver correlação entre observações sucessivas, é necessário que o modelo contenha termos autoregressivos a fim de explicar tal correlação. Se num determinado modelo tais regressores estiverem faltando, os outros regressores tentarão explicar a dinâmica nos dados. Na prática, isto resultará em estimadores inconsistentes. No presente estudo, essa inconsistência significará estimativas pouco confiáveis dos coeficientes de sazonalidade, que são os parâmetros dos regressores *dummy*. Portanto, com a inclusão de termos autoregressivos em (5), é possível, em princípio, estimar corretamente o vetor de parâmetros β , uma vez que a informação dinâmica contida nas observações será explicada pela parte autoregressiva do modelo deixando assim as variáveis *dummy* absorverem apenas a informação sobre sazonalidade.

3.4 Teste empírico e análise de resultados

O objetivo desta seção é especificar de maneira precisa o modelo (5) para realizar a estimativa dos coeficientes de sazonalidade no contexto da regressão dinâmica. Para isso, os critérios ERR e de Akaike serão utilizados para escolher a ordem e o número exato dos termos autoregressivos a ser incluídos na relação. Resumindo o procedimento seguido, temos:

- i) utiliza-se o modelo (1) para um estudo preliminar dos coeficientes de sazonalidade. Os resíduos deste modelo são altamente correlacionados (conforme já foi mostrado);
- ii) a fim de explicar o comportamento dinâmico de tais resíduos propõe-se incluir no modelo termos autoregressivos. O resultado é uma equação do tipo (5);
- iii) a fim de estabelecer quais termos são mais relevantes para explicar a variância de tais resíduos, utiliza-se o critério ERR;
- iv) uma vez listados os termos autoregressivos por ordem de importância mediante o método ERR, a escolha do número ótimo de regressores é feita usando-se o critério de informação de Akaike, que penaliza a complexidade do modelo.¹⁵

A regressão mostrada na Tabela 4 é aquela que apresenta o valor mínimo do parâmetro representativo do critério de Akaike. A inclusão dos termos autoregressivos resulta num aumento significativo da precisão do ajuste. O R^2 ajustado é igual a 0,91 e a estatística F é fortemente significativa. Além do mais, a regressão cumpre a condição de estabilidade, já que nenhuma das inversas das raízes do polinômio autoregressivo é maior que a unidade em módulo.

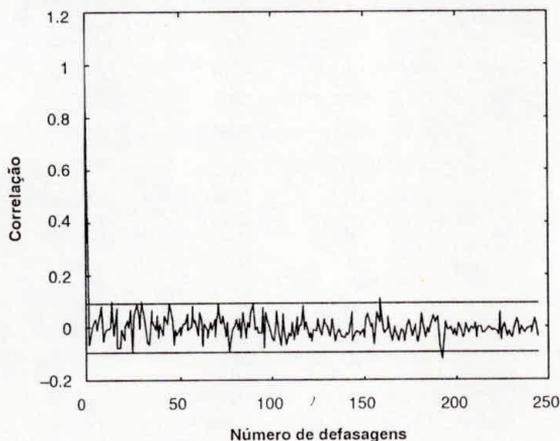
Observa-se a estabilidade dos coeficientes estimados correspondentes aos regressores *dummy* sazonais. Os coeficientes das *dummies* dos meses de fevereiro a julho não são estatisticamente diferentes da menor média sazonal – correspondente ao mês de junho – ao nível de significância de 5%. Isso confirma que o período da safra estende-se de fevereiro a julho de cada ano. Note-se que os coeficientes correspondentes aos meses de janeiro e agosto, que na Tabela 3 aparecem como não-significativos, agora são estatisticamente diferentes do coeficiente base de comparação e, portanto, pertencem à entressafra tal como indicado no Gráfico 5.

15 Esse critério foi proposto na teoria da informação, na área de modelos *state space* da engenharia (Akaike, 1974).

O teste da hipótese nula de ausência de autocorrelação nos resíduos mediante a estatística Q de Ljung-Box, com 12, 24 e 36 defasagens, não rejeita a hipótese. A versão F do teste LM de Breusch-Godfrey para a presença de autocorrelação de primeira a décima segunda ordem nos resíduos estimados também não dá resultado significativo. Essas evidências apontam para o fato dos resíduos do ajuste da equação (5) serem ruído branco (ver correlograma dos resíduos no Gráfico 6). Contudo, as estatísticas de Ljung-Box calculadas para os quadrados dos resíduos são estatisticamente significativas, indicando a existência de alguma autocorrelação não-linear nos resíduos da regressão estimada. Para sanar esse problema seria necessário incluir regressores não-lineares no modelo. Apesar de o critério ERR ser apto para selecionar termos não-lineares candidatos a entrar na equação, tal tarefa não foi realizada nessa oportunidade.

Gráfico 6

CORRELOGRAMA DOS RESÍDUOS DA EQUAÇÃO (5)



Fonte: Elaboração própria.

Com relação à hipótese de que os erros do modelo têm todos a mesma variância, por um lado, o teste ARCH rejeita a hipótese nula de igual variância, produzindo coeficientes significativos para os resíduos ao quadrado com até doze defasagens. Por outra parte, o teste de heterocedasticidade de White não rejeita a hipótese nula, indicando que os resíduos estimados podem ser uma amostra de uma única distribuição de erros.

Por não serem totalmente conclusivos, os resultados do conjunto anterior de testes alertam sobre a possível existência de heterocedasticidade nos erros do modelo estimado, o que é contrário às hipóteses da teoria de regressão que sustentam as propriedades desejáveis dos estimadores. Como

regra geral, a heterocedasticidade não afeta as propriedades de não-tendenciosidade e de consistência dos estimadores, mas alteram sua eficiência. Na presença desse problema os desvios padrão estimados tenderão a ser menores do que os verdadeiros. O uso desses resultados leva a concluir que as estimativas dos parâmetros são mais precisas do que realmente são. Haverá uma tendência no sentido de rejeitar a hipótese nula quando, de fato, não deveria ser rejeitada.

Tabela 4
RESULTADOS DA ANÁLISE DE REGRESSÃO
 (Mês base de comparação: junho)

Variável	Coefficiente	Erro padrão	t de Student	Probabilidade
C	34.78623	4.967555	7.002688	0.0000
JAN	3.450807	1.494249	2.309392	0.0214
FEV	1.562175	1.375187	1.135972	0.2566
MAR	0.333797	1.187648	0.281057	0.7788
ABR	0.300214	0.931986	0.322123	0.7475
MAI	0.185267	0.598842	0.309375	0.7572
JUL	1.133432	0.598901	1.892519	0.0591
AGO	3.151227	0.932370	3.379802	0.0008
SET	5.353283	1.188659	4.503634	0.0000
OUT	6.469701	1.377131	4.697957	0.0000
NOV	6.897360	1.497090	4.607179	0.0000
DEZ	5.393950	1.533501	3.517409	0.0000
AR(1)	1.049344	0.047222	2.222165	0.0005
AR(57)	0.041106	0.014615	2.812562	0.0000
AR(2)	-0.155779	0.048555	-3.208289	0.0051
AR(8)	0.098243	0.025237	3.892857	0.0014
AR(15)	-0.064787	0.020677	-3.133311	0.0001
R ²	0.914273	Média var dep		37.12963
R ² ajustado	0.911068	D. P. var dep		10.97593
E. P. da regressão	3.273179	C. I. Akaike		2.408977
Resíduos ao quadrado	4585.463	Estatística F		285.2869
Estatística D-W	1.958979	Probab. de F		0.000000

Fonte: Elaboração própria.

Nota: MQO // Variável dependente: PREÇO

Amostra (ajustada): Dezembro 1958 - Dezembro 1995

Observações incluídas: 445

Convergência conseguida após 3 iterações.

Por todo o anterior, as interpretações sobre a significância estatística dos coeficientes das *dummies* sazonais devem ser tomadas com alguma cautela.

4 CONCLUSÕES

No presente trabalho focaliza-se a análise no componente sazonal da série de preços recebidos pelos produtores de carne bovina do Estado de São Paulo no período de março de 1954 a dezembro de 1995. Com preços, expressos em reais de dezembro de 1995 por arroba, essa série apresenta uma tendência crescente até 1979 e, a partir daí, a tendência muda de sinal. O comportamento dos preços no primeiro subperíodo é atribuído ao fato da demanda ter apresentado, nessa época, taxas de crescimento superiores às da oferta. Já a tendência declinante coincide com um período de arrocho salarial e perda do poder aquisitivo da classe média – principal consumidora de carne bovina – e, também, com a introdução de melhoras tecnológicas na produção de carne. Além do mais, são apresentadas evidências de que a série possui um componente cíclico e outro sazonal, ambos estatisticamente significativos.

O estudo da sazonalidade é iniciado com a descrição da estrutura sazonal média dos dados, ou seja, com o cálculo das médias sazonais. Essas médias permitem calcular um conjunto de fatores sazonais que, por sua vez, podem ser utilizados para dessazonalizar a série original. Esses resultados descritivos são utilizados como marco de referência para mostrar o produto gerado pela aplicação do método de regressão com variáveis *dummy* sazonais. A regressão que tem onze variáveis *dummy* sazonais como variáveis explicativas consegue reproduzir as médias sazonais e as diferenças entre essas médias, mas o ajuste é deficiente do ponto de vista econométrico. Além do mais, o período de preços sazonalmente baixos (safra) é relativamente extenso, incluindo os meses de janeiro a agosto.

Ao se mudar a especificação para melhorar o ajuste a safra inclui os meses de fevereiro a julho, e o período de entressafra – de igual duração – estende-se do mês de agosto de um ano até janeiro do ano seguinte.

Para especificar a regressão dinâmica utiliza-se o critério ERR, combinado com o de Akaike. Na nova especificação a qualidade do ajuste melhora significativamente mas ainda persistem indícios de problemas com relação à presença de heterocedasticidade nos resíduos. Contudo, como temos o marco de referência proporcionado pelos cálculos aritméticos do procedimento mais simples, resultados esses que são reproduzidos com razoável precisão pelo modelo proposto, julgamos que o mencionado problema não é crítico nesse caso.

Na opinião de alguns econométricos não é recomendável eliminar a sazonalidade – por um método qualquer – antes de se realizar uma análise com modelos Box-Jenkins, por exemplo, porque esse ajuste poderia distorcer substancialmente a estrutura estocástica do fenômeno em estudo. Para esses autores o procedimento correto é modelar a sazonalidade de forma explícita como parte do modelo. No caso dos preços do boi-gordo no Estado de São Paulo a utilização de variáveis *dummy* permite fazer isso nos modelos de regressão já que outros estudos estabeleceram que não existem raízes unitárias sazonais nesta série.

5 REFERÊNCIAS BIBLIOGRÁFICAS

- AGUIRRE, L. A. Some remarks on structure selection for nonlinear models. *Int. J. Bifurcation and Chaos*, v. 4, n. 6, p. 1707–1714, 1994.
- . Uma introdução à análise espectral de séries temporais econômicas. *Nova Economia*, Belo Horizonte, v. 5, n. 1, p. 41–60, 1995.
- . Testing for seasonal unit roots in a quarterly series of beef cattle prices in the State of São Paulo (Brazil). *Revista de Economia e Sociologia Rural*, v. 35, n. 4, p. 151–173, 1997.
- AKAIKE, H. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, v. 19, n. 6, p. 716–723, 1974.
- BILLINGS, S. A., CHEN, S., KORENBERG, M. J. Identification of MIMO nonlinear systems using forward-regression orthogonal estimator. *International Journal of Control*, v. 49, n. 6, p. 2157–2189, 1989.
- CHEN, S., BILLINGS, S. A., LUO, W. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, v. 50, n. 5, p. 1873–1896, 1989.
- DRAPER, N. R., SMITH, H. *Applied Regression Analysis*. New York: John Wiley & Sons, 1981.
- ERICSSON, N. R., HENDRY, D. F., HONG-ANH TRAN. *Nonstationary time series analysis and cointegration*. Oxford: Oxford University Press, 1994. (HARGREAVES, C. P. [Ed.]).
- FRANSES, P. H. Recent advances in modelling seasonality. *Journal of Economic Surveys*, v. 10, n. 3, p. 299–345, 1996.
- GARCIA, E. A. C. Análise harmônica aplicada às variações do preço do boi no Pantanal Mato-Grossense. *Revista de Economia Rural*, v. 20, n. 4, p. 557–574, 1982.

- GREENE, W. H. *Econometric Analysis*. Prentice Hall: Englewood Cliffs, 1993.
- GUJARATI, D. N. *Basic Econometrics*. New York: McGraw-Hill Book Co., 1995.
- HYLLEBERG, S. Modelling seasonal variation. In: HARGREAVES, C. P. (Ed.) *Nonstationary time series analysis and cointegration*. Oxford: Oxford University Press, 1994.
- MARGARIDO, M. A., KATO, H. T., BUENO, C. R. F., CAMBOM Jr., E. Análise dos impactos das cotações do dólar paralelo e do índice pluviométrico sobre os preços do boi-gordo no Estado de São Paulo. *Revista Brasileira de Economia*, v. 50, n. 2, p. 255–278, 1996.
- SHISHKIN, J., YOUNG, A., MUSGRAVE, J. *The X-11 variant of the census method II seasonal adjustment program*. Bureau of the Census, USA, 1967. (Technical paper, 15).
- SILVA, L. M. R., LEMOS, J. J. S. Variação estacional nos preços do boi-gordo no Nordeste: um modelo de análise estática comparativa. *Revista de Economia Rural*, v. 24, n. 3, p. 331–340, 1986.
- SIMS, C. A. Seasonality in regression. *Journal of the American Statistical Association*, v. 69, p. 618–626, 1974.
- WALLIS, K. F. Seasonal adjustment and relations between variables. *Journal of the American Statistical Association*, v. 69, p. 18–31, 1974.