



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

PREVISÃO PARA O PREÇO FUTURO DO CACAU ATRAVÉS DE UMA SÉRIE UNIVARIADA DE TEMPO: UMA ABORDAGEM UTILIZANDO O MÉTODO ARIMA

MARCEL CASTRO DE MORAES; ALDOUS PEREIRA ALBUQUERQUE;

UFPB

JOÃO PESSOA - PB - BRASIL

marcelcmoraes@gmail.com

APRESENTAÇÃO COM PRESENÇA DE DEBATEDOR

COMERCIALIZAÇÃO, MERCADOS E PREÇOS AGRÍCOLAS

RESUMO

Neste artigo realiza-se uma previsão para o preço médio mensal do cacau (recebido pelo produtor brasileiro) a partir de uma série temporal que compreende o período janeiro/1970-agosto/2005, totalizando 428 observações. Para isto, adota-se a metodologia denominada Box-Jenkins – utilizada para análise de séries univariadas de tempo. Identifica-se para tal previsão o método autoregressivo integrado com média móvel (ARIMA). Em seguida, são apresentados cinco modelos candidatos para a realização da previsão de dados, onde são adotados os critérios *Akaike Information Criterion* (AIC), *Schwartz Bayesian Criterion* (SBC) e Erro Quadrado Médio (EQM), para a escolha do melhor modelo. Verifica-se por meio de uma previsão *ex-ante*, fundamentada nos dados amostrais, uma pequena queda no preço futuro do produto para os próximos doze meses.

Palavras-chave: Cacau, Modelagem ARIMA, Previsão de Preços, *Commodities* Agrícolas.

1. INTRODUÇÃO

De acordo com o ISAE/FGV (2003), o cacau é um produto nobre e tradicional da agricultura brasileira que vem atravessando um período recente de recuperação econômica – especialmente no sul da Bahia e em Rondônia – após um prolongado período de crise (década de 80 até meados da década de 90 do século passado). Também para o ISAE/FGV (2003), essa crise é originária de uma conjuntura externa desfavorável (redução dos preços internacionais devido a um aumento da produção mundial e dos estoques nos países produtores) e de uma conjuntura interna ainda pior (devastação causada pela doença conhecida como “vassoura de bruxa” nos cacauais brasileiros – especialmente no Sul da Bahia, até hoje a maior região produtora).

Quando processado na indústria, a amêndoa de cacau gera derivados como licor, manteiga e torta que são a matéria prima do chocolate.

Segundo estudo realizado pelo ISAE/FGV (2003), a produção nacional de cacau em amêndoas está concentrada no Sul da Bahia (aproximadamente, 83% da oferta brasileira, especialmente nos municípios de Camacã, Ilhéus e Itabuna), seguida – em uma proporção bem inferior – pelo Pará (9,9% da produção nacional, focalizada nas cidades de Medicilândia, Umará, Altamira e Tomé-açú) e por Rondônia (5,4% da oferta brasileira, centralizada nos municípios localizados na região próxima à Transamazônica). Ainda assim, segundo Toda Fruta (2005), o Brasil ainda importa 20% de toda amêndoa de cacau moída na indústria nacional de chocolate.

Tabela I: Participação dos Estados Brasileiros na Produção Nacional de Cacau (2003)

Ranking	Estado	Participação na Produção Nacional (%)
1	Bahia	83
2	Pará	9,9
3	Rondônia	5,4
4	Outros Estados	1,7

Fonte: ISAE/FGV

De acordo com dados do Ministério do Desenvolvimento, Indústria e Comércio do Governo Brasileiro (MDIC), a produção mundial de cacau vem apresentando um processo de crescimento ao longo do tempo. Entre 1993 e 2004 a produção mundial passou de 2.486 mil toneladas para 3.452 mil toneladas, representando um aumento de 38,86%. A produção que mais cresceu foi a de Costa do Marfim, saindo de 840 mil toneladas para 1.405 mil toneladas, ocupando a primeira posição no ranking mundial. Em seguida aparecem as produções de Gana com 736 mil toneladas, Indonésia com 415 mil toneladas, Nigéria com 175 mil, Brasil com 164 mil toneladas e Camarões com 145 mil, conforme a tabela II.

Tabela II: Produção Mundial de Cacau (2004)

Ranking	País	Produção Mundial (1.000 t)
1	Costa do Marfim	1.405
2	Gana	736
3	Indonésia	415
4	Nigéria	175
5	Brasil	164
6	Camarões	145

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC) e Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC)

Somente a Costa do Marfim já participa com 56% da produção de cacau do continente africano e 41% da produção de cacau no mundo, uma quantidade relativamente alta totalizando quase a metade da produção mundial. O continente africano já participa com 72%

da produção mundial de cacau. Em seguida vem o Continente Asiático junto com Oceania com uma participação de 15% e as Américas com uma participação de 13%, valor relativamente baixo quando comparado com o continente africano.

Tabela III: Participação dos Continentes na Produção Mundial de Cacau (2004)

Ranking	Continente	Participação na Produção Mundial (%)
1	África	72
2	Ásia e Oceania	15
3	Américas	13
Total		100

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC) e Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC)

A participação da produção brasileira na produção mundial em 1994 estava em 300 mil toneladas (ao ano) e a sua participação na produção mundial estava em 12,07%. A produção chegou, aproximadamente, a 123,5 mil toneladas em 2000 chegando a participar com 4,01% da produção mundial deste bem, recuperando-se a partir de 2001 e consolidando-se em 2004 quando obteve 163,8 mil toneladas – aumentando sua participação na produção mundial para 4,75%.

Tabela IV: Produção Brasileira de Cacau

Ano	Produção Nacional (1.000 t)	Participação Nacional na Produção Mundial (%)
1994	300	12,07
2000	123,5	4,01
2004	163,8	4,75

Fonte: Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC) e Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC)

Segundo o Ministério Brasileiro do Desenvolvimento, Indústria e Comércio (MDIC) – mais especificamente, a Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC) –, esse aumento na produção nacional nos últimos anos se deve principalmente às novas técnicas introduzidas pelo CEPLAC e pelo dinamismo dos produtores de cacau em acreditar e implantá-los em sua propriedade.

O principal país consumidor de cacau é os Estados Unidos, com aproximadamente 689 mil toneladas ao ano, seguido pela Alemanha, França e Reino Unido, com 280, 218 e 215 mil toneladas, respectivamente. O Brasil atualmente consome 98 mil toneladas de cacau segundo a Organização Internacional de Cacau (ICCO).

Tabela V: Principais Países Consumidores de Cacau (2004)

País	Consumo Nacional (1.000 t)
EUA	689
Alemanha	280
França	218
Reino Unido	215

Fonte: ICCO

Quando se analisa o consumo *per capita* de cacau por país, a classificação já se altera. O principal consumidor passa a ser a Bélgica com 5,63 kg/hab, seguido da Suíça, França e Reino Unido, com 4,09, 3,66 e 3,6 kg/hab.

Tabela VI: Principais Países Consumidores Per Capita de Cacau

País	Consumo Per Capita (Kg)
-------------	--------------------------------

Bélgica	5,63
Suíça	4,09
França	3,66
Reino Unido	2,6

Fonte: ICCO

Diante das características evidenciadas pelos cenários nacional e internacional do cacau, torna-se de grande importância realizar previsões de valores relacionados a este produto, como seu preço, o valor exportado, o valor importado, consumo interno, produção, assim como outras variáveis econômicas associadas a ele.

Sendo assim, este artigo destina-se a realizar uma previsão do preço médio do cacau recebido pelo produtor (15 Kg) para um horizonte de 12 meses a partir do mês posterior à última informação obtida a respeito dessa variável (agosto/2005). Vale ressaltar que a série temporal de preços do cacau é trabalhada do período janeiro/1970 até agosto/2005, encontrando-se deflacionada pelo índice geral de preços-disponibilidade interna (IGP-DI) e sob a mesma unidade monetária (R\$), obtendo, dessa forma, um preço atualizado tendo como base o período janeiro/2006. O critério de ter adotado o IGP-DI como deflator da série encontra-se na simples tentativa de se padronizar a coleta de dados; uma vez que a série de preços foi elaborada pela Fundação Getulio Vargas (FGV) – instituto que elabora o índice utilizado neste trabalho.

A metodologia usada é a de Box-Jenkins enfatizando suas propriedades, calculando a chamada função de autocorrelação (FAC) e função de autocorrelação parcial (FACP). A identificação do modelo que melhor realiza a previsão de preços é feita pela abordagem ARMA (ARIMA).

Portanto, este artigo encontra-se composto por cinco seções: esta introdução, metodologia adotada para a previsão, análise dos dados e previsão, conclusão e, por último, pelas referências bibliográficas utilizadas na pesquisa.

2. METODOLOGIA

Deve-se perceber que o objetivo é obter um bom modelo que explique a trajetória da série, identificando, estimando, checando o modelo que melhor represente-a e, assim, possa proporcionar uma previsão plausível e consistente. Utiliza-se a metodologia do modelo de Box-Jenkins para realizar essa previsão.

Emprega-se o termo Box-Jenkins em homenagem a George E. P. Box e Gwilym M. Jenkins que são responsáveis por formalizar o procedimento usado para análises de séries univariadas de tempo. Utiliza-se também a notação B-J para referenciar modelos univariados de tempo e muitas vezes para estabelecer o método referido como modelos ARIMA. Séries univariadas de tempo B-J/ARIMA estão baseadas somente sobre o passado da sua própria variável para fins de previsões, ou seja, não são fundamentadas sobre qualquer teoria ou quaisquer outras variáveis. Neste caso, a estratégia para a construção do modelo será arquitetada pelos próprios dados da série temporal. Portanto, a intenção aqui é realizar uma previsão de preços com os próprios dados da série.

A metodologia de Box-Jenkins divide-se nas seguintes etapas: especificação matemática do modelo; identificação do modelo (obtenção de p,d,q) e FAC e FACP; estimação de parâmetros; checagem do modelo por meios de testes estatísticos; e, por fim, previsões cabíveis. Tais etapas serão consideradas nas próximas subseções.

Os modelos de séries temporais possibilitam descrever um processo estocástico utilizando apenas valores passados da variável dependente e do termo de erro. Assim, dada uma série y_t , os modelos de séries temporais podem ser descritos em desvios como segue:

$$y_t = \phi_1 \cdot y_{t-1} + \dots + \phi_p \cdot y_{t-p} + \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \dots - \theta_q \cdot \varepsilon_{t-q}$$

Ou, escrevendo de outra forma:

$$(1 - \phi_1 \cdot B - \dots - \phi_p \cdot B^p) y_t = (1 - \theta_1 \cdot B - \dots - \theta_q \cdot B^q) \varepsilon_t$$

Onde B é o operador de defasagem ($B \cdot y_t = y_{t-1}$), ϕ e θ são os parâmetros do modelo e ε_t o termo de erro. As variáveis defasadas de y_t representam a parte autoregressiva do modelo (AR), enquanto que as defasagens de ε_t representam a parte de média móvel (MA). O modelo apresentado acima é denominado ARMA(p,q), sendo que “p” representa a ordem de defasagem do termo autoregressivo e “q” a ordem de defasagem do termo de média móvel.

De acordo com Greene (2003), se a observação y_t é gerada pela média ponderada de somente as “p” primeiras observações próximas anteriores da variável acrescida de um erro aleatório ε_t , então ela pode ser modelada por um processo AR(p), dado por:

$$y_t = \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \dots + \phi_p \cdot y_{t-p} + \varepsilon_t$$

Onde:

$$\varepsilon_t = y_t - \hat{y}_t$$

É importante colocar que este modelo será considerado autoregressivo de ordem “p” – isto é, um AR(p) – se a seguinte condição for satisfeita:

$$E(y_t / y_{t-1}) = \mu + \phi \cdot y_{t-1}$$

Para o processo ser estacionário, a média (μ) deve ser constante e a condição necessária (mas não suficiente) de estacionariedade é dada por:

$$\phi_1 + \phi_2 + \dots + \phi_p < 1$$

Utilizando-se do mesmo raciocínio do processo AR(p), se a observação y_t é gerada pela média ponderada dos q primeiros valores passados de um processo de “ruído branco”, então ela pode ser modelada por um processo MA(q), dado por:

$$y_t = \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \theta_2 \cdot \varepsilon_{t-2} - \dots - \theta_q \cdot \varepsilon_{t-q}$$

Neste caso, a condição de estacionariedade é verificada se o valor de “q” for finito.

Um processo ARMA(p,q) é verificado quando a observação y_t é gerada tanto pela média ponderada das “p” primeiras observações próximas anteriores da variável, quanto pela média ponderada dos “q” primeiros valores passados de um processo de “ruído branco”. Esse processo é descrito por:

$$y_t = \phi_1 \cdot y_{t-1} + \dots + \phi_p \cdot y_{t-p} + \varepsilon_t - \theta_1 \cdot \varepsilon_{t-1} - \dots - \theta_q \cdot \varepsilon_{t-q}$$

Neste caso, a média e as condições de estacionariedade são dadas pela parcela do processo AR.

Se as observações y_t forem geradas tanto pela média ponderada das “p” primeiras observações próximas anteriores da variável, quanto pela média ponderada dos “q” primeiros valores passados de um processo de “ruído branco”, e ainda pertencerem a um processo não estacionário, será necessário diferenciar a série original dos dados “d” vezes até se obter uma série estacionária. Dessa maneira a série será considerada um processo ARIMA(p,d,q), dado por:

$$\phi(B) \Delta^d y_t = \theta(B) \varepsilon_t$$

Onde:

$$\Delta^d = (1 - B)^d$$

Segundo Moreira Júnior [s.d.], em geral, o número de “d” é no máximo igual a 2. Necessita-se de $d = 1$ quando a série é não estacionária quanto ao nível do processo, isto é, quando a média durante um período é uma e, depois desse período, ocorre uma mudança na média. Torna-se necessário que $d = 2$ quando a série é não estacionária também quanto à

inclinação, ou seja, quando a série oscila em uma direção em um certo período e depois muda de direção.

Um ponto de grande importância para a modelagem ARIMA é saber determinar qual o processo (modelo) que melhor explica a dinâmica dessa série em estudo. Portanto, pretende-se saber se a dinâmica temporal de uma dada variável é mais bem explicada por um processo autoregressivo de ordem “p” [AR(p)]; por um processo de média móvel de ordem “q” [MA(q)]; por um processo autoregressivo com média móvel de ordem “p,q” [ARMA(p,q)]; ou ainda, por um processo autoregressivo integrado com média móvel de ordem “p,d,q” [ARIMA(p,d,q)].

As etapas de análise para determinação do modelo que melhor explica a dinâmica da série temporal em análise, através da metodologia Box-Jenkins, são caracterizadas a seguir.

3.1. Identificação:

A identificação é a fase mais crítica da metodologia Box-Jenkins; sendo possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal utilizando critérios distintos de escolha (FAC, FACP, Akaike, Hannan, Schwartz etc.). Em geral, os modelos devem ser parcimoniosos, isto é, contendo um número pequeno de parâmetros.

Esta etapa consiste em escolher (p,d,q) do modelo ARIMA. Neste caso, analisam-se a FAC e FACP, e tenta-se identificar o modelo. O processo busca determinar a ordem de (p,d,q) tomando por base o comportamento das funções de autocorrelação (FAC) e autocorrelação parcial (FACP), assim como também seus respectivos correlogramas.

Segundo Greene (2003), a autocorrelação é a correlação entre membros de uma série de observações ordenadas no tempo. Assim, a Função de Autocorrelação (FAC) que calcula a correlação entre y_t e y_{t-p} é definida por:

$$\rho_p = \frac{\gamma_p}{\gamma_0} = \frac{Cov(y_t, y_{t-p})}{Var(y_t)}$$

Mais especificamente:

$$r_p = \frac{\hat{\gamma}_p}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-p} (y_t - \bar{y})(y_{t-p} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

A representação gráfica de r_p contra “p” (quantidade de defasagens ou *lag*) é conhecida como correlograma amostral da FAC. Os valores de r_p também podem ser obtidos pelos coeficientes das regressões. Matematicamente, tem-se:

$$\begin{aligned} y_t &= \rho_1 \cdot y_{t-1} + \varepsilon_t \\ y_t &= \rho_2 \cdot y_{t-2} + \varepsilon_t \\ &\vdots \\ y_t &= \rho_p \cdot y_{t-p} + \varepsilon_t \end{aligned}$$

A Função de Autocorrelação Parcial (FACP) mede a correlação entre y_t e y_{t-p} depois de eliminada a influência dos y 's intermediários, sendo definida por:

$$\phi_{pp} = \rho_p, \text{ se } p = 1 \quad \text{ou} \quad \phi_{pp} = \frac{P_p^*}{P_p}, \text{ se } p > 1$$

Onde P_p é a matriz de autocorrelação e P_p^* é a matriz P_p com a última coluna substituída pelo vetor de autocorrelações ρ_p . Matricialmente, tem-se:

$$\rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_p \end{bmatrix} \quad P_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \quad \phi_p = \begin{bmatrix} \phi_{p1} \\ \phi_{p2} \\ \dots \\ \phi_{pp} \end{bmatrix}$$

Onde:

$$\rho_p = P_p \cdot \phi_p$$

Isto é:

$$\rho_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} \phi_{p1} \\ \phi_{p2} \\ \dots \\ \phi_{pp} \end{bmatrix}$$

A representação gráfica de $\hat{\phi}_{pp}$ contra “p” é conhecida como correlograma amostral da FACP. Os valores de $\hat{\phi}_{pp}$ também podem ser obtidos pelos coeficientes de índice pp das regressões.

$$\begin{aligned} y_t &= \phi_{11} \cdot y_{t-1} + \varepsilon_t \\ y_t &= \phi_{21} \cdot y_{t-1} + \phi_{22} \cdot y_{t-2} + \varepsilon_t \\ &\vdots \\ y_t &= \phi_{p1} \cdot y_{t-1} + \phi_{p2} \cdot y_{t-2} + \dots + \phi_{pp} \cdot y_{t-p} + \varepsilon_t \end{aligned}$$

Procuram-se semelhanças entre as funções de autocorrelação teóricas e amostrais que sejam boas sugestões do processo (modelos AR, MA, ARMA etc.) que possam melhor explicar a dinâmica da série em estudo. Na verdade, a escolha do modelo envolve uma grande habilidade, ou seja, a estratégia para a construção do modelo é baseada nos próprios dados.

3.2. Estimação:

É com base nas melhores sugestões de modelos que se elabora o segundo passo da metodologia de Box-Jenkins: a estimação do modelo. Aqui, os modelos (processos AR, MA, ou ARMA, etc.) sugeridos na fase de identificação são ajustados e examinados. Os modelos ajustados são comparados utilizando alguns critérios.

Um dos critérios é o da parcimônia, no qual se verifica que a incorporação de coeficientes adicionais melhora o grau de ajustamento (aumenta o R^2 e reduz a soma dos quadrados dos resíduos) do modelo, mas reduz os graus de liberdade do mesmo. Uma das formas de melhorar o grau de ajustamento desse modelo aos dados da série temporal é incluir defasagens adicionais nos processos AR(p), MA(q), ARMA(p,q) e ARIMA. A inclusão de defasagens adicionais implica em aumento do número de “regressores”, o que leva a uma redução na soma do quadrado dos resíduos estimados.

Atualmente, existem vários critérios de seleção de modelos que ocasionam um *trade-off* entre uma redução na soma do quadrado dos resíduos estimados e um modelo mais parcimonioso. Os critérios mais usados são o *Akaike Information Criterion* (AIC) e o *Schwartz Bayesian Criterion* (SBC), cujas fórmulas são dadas por:

$$AIC = T \ln(\text{soma do quadrado dos resíduos}) + 2n$$

$SBC = T \ln(\text{soma do quadrado dos resíduos}) + n \ln(T)$

Onde:

n = número de parâmetros estimados;

T = número de observações utilizadas.

Geralmente, quando se trabalha com variáveis defasadas perdem-se informações sobre a série temporal em estudo. Logo, para se comparar modelos alternativos (ou concorrentes) deve-se manter fixo o número de informações utilizadas (T) para todos os modelos em comparação. O ideal é quanto menor o AIC e o SBC melhor o ajustamento do modelo. Porém, faz-se necessário comparar os AIC's e os SBC's de modelos alternativos, para se saber qual modelo melhor explica a dinâmica da série de tempo analisada.

3.3. Checagem:

Depois de identificado e estimado, procura-se saber se o modelo obtido descreve os dados adequadamente.

A análise dos resíduos de modelos alternativos (candidatos) ajustados é de extrema importância na escolha final do(s) modelo(s) que melhor explica(m) a dinâmica da série temporal em estudo. Se os resíduos são autocorrelacionados, então, a dinâmica da série não é completamente explicada pelos coeficientes do modelo ajustado. Deve-se excluir do processo de escolha modelo(s) com esta característica. Uma análise da existência (ou não) da autocorrelação serial de resíduos é feita com base nas funções de autocorrelação e autocorrelação parcial dos resíduos e seus respectivos correlogramas. Enders (1995) afirma que, na prática, o número de FAC e FACP a serem analisados é de $\frac{T}{4}$ onde “ T ” é o número de informações da série.

A estatística Q – calculada por meio do teste de Box-Pierce ou pelo teste de Ljung-Box – é verificada para testar se um determinado conjunto de autocorrelações de resíduos é (ou não) estatisticamente diferente de zero. Essa estatística possui distribuição χ_m^2 (qui-quadrado) e o teste é realizado comparando-se os valores de $Q_{calculado}$ e $Q_{tabelado}$. Se $Q_{calculado} > Q_{tabelado}$ (ou Probabilidade de $Q_{calculado} > \alpha$), dado certo nível de significância, então se deve rejeitar a hipótese nula (H_0) de que os resíduos não são autocorrelacionados. Porém, rejeitar H_0 significa aceitar a hipótese alternativa (H_a) de que pelo menos uma autocorrelação r_k é estatisticamente diferente de zero (ou seja, o erro não é ruído branco, rejeitando-se o modelo).

A interpretação dos testes para as estatísticas Q_{bp} (Box-Pierce) e Q_{lb} (Ljung-Box) seguem a mesma interpretação e tais estatísticas podem ser calculadas das seguintes formas:

$$Q_{bp} = n \sum_{k=1}^m r_k^2 \quad \text{e} \quad Q_{lb} = n(n+2) \sum_{k=1}^m \left(\frac{r_k^2}{n-k} \right)$$

Onde:

n = número de observações

k = número de parâmetros estimados

É importante ressaltar que ao estimar-se um modelo, deseja-se que o erro produzido por ele possua característica de “ruído branco”, isto é, este venha ser independente e identicamente distribuído (condição i.i.d.).

3.4. Previsão :

A previsão é uma das principais razões da popularidade da metodologia de Box-Jenkins. Em muitos casos as previsões – principalmente de curto prazo – obtidas com base em Box-Jenkins são melhores que as obtidas com base nos modelos econométricos tradicionais.

As previsões podem ser de dois tipos: *ex-ante* e *ex-post*. A previsão *ex-ante* é feita para calcular valores futuros, de curto prazo, da variável em estudo. Por outro lado, a previsão *ex-post* é realizada para gerar valores dentro do período amostral. Quanto melhor forem essas últimas, mais eficiente será o modelo estimado.

Por fim, tem-se o erro quadrado médio (EQM) da previsão. Este é igual à média do quadrado da diferença entre cada valor previsto *ex-post* e o valor real observado na amostra. Ele é uma medida formal da qualidade das previsões *ex-post*. Portanto, quanto menor o EQM, melhor será o grau de ajustamento do modelo aos dados da série temporal.

3. ANÁLISE DOS DADOS E PREVISÕES

A pesquisa aqui mostrada – que teve seus dados coletados na plataforma eletrônica de informações econômicas do Instituto de Pesquisa Econômica Aplicada (IPEA), pertencente ao governo Brasileiro – faz uma breve análise sobre o comportamento do preço (em R\$) mensal do cacau no mercado brasileiro durante os anos de 1970 a 2005, totalizando 428 observações. Vale salientar que a metodologia para o cálculo desse preço mensal é entendida como uma média entre os valores pagos aos diversos produtores nacionais deste bem. Para que não se torne repetitivo, é importante ressaltar que em todos os momentos que se referir ao preço do cacau neste artigo, estará fazendo-se referência ao valor atualizado (tendo como base o período de janeiro/2006) da arroba (15 kg) do mesmo.

Fundamentada na série deflacionada do preço do cacau entre 1970-2005 pretende-se realizar uma previsão para os 12 (doze) meses seguintes para o preço médio recebido pelos produtores, através do modelo ARMA (ARIMA) realizando o diagnóstico do(s) modelo(s) estimado(s) decidindo, portanto, a melhor especificação para a previsão.

Com relação às estatísticas da série, algumas observações podem ser colocadas. O preço médio da arroba do cacau durante o período pesquisado é de R\$113,03, chegando a atingir um valor máximo de R\$428,56 (fevereiro/1977) e um valor mínimo de R\$33,24 (setembro/1990)¹. O desvio-padrão desse preço é de R\$79,78 e o coeficiente de variação de Pearson é de aproximadamente 70,58%. Segundo Oliveira (1999), esse índice revela um alto grau de dispersão dos dados, pois é maior que 30%.

Figura 1



Fonte: Dados utilizados na pesquisa
* Valores atualizados para janeiro/2006

Na figura 1 visualiza-se o comportamento dos preços do cacau (R\$/15 kg) e verifica-se que, ao longo do tempo, o mesmo tem uma tendência de queda durante o período analisado. De acordo com Lima Neto e Loureiro (2003), a tendência compreende um

¹ Ressalte-se que são valores reais com base em janeiro/2006.

movimento dominante de uma série temporal, segundo o qual a mesma se desenvolve em um longo intervalo de tempo.

Isso leva a crer que, pelo menos no mercado brasileiro, este produto e seus derivados foram ficando cada vez mais acessível às camadas com menor poder de compra no país.

A tendência (figura 2) é calculada por um processo denominado 2x12 MA, ou seja, para se calcular a tendência de preço do cacau em um determinado mês, são – num primeiro momento – observados os 6 valores anteriores e os 5 valores posteriores, além, é claro, do valor relativo ao próprio mês. Num segundo momento, esse valor calculado é somado ao valor calculado do período exatamente seguinte e dividido por 2 (dois). Matematicamente, tem-se:

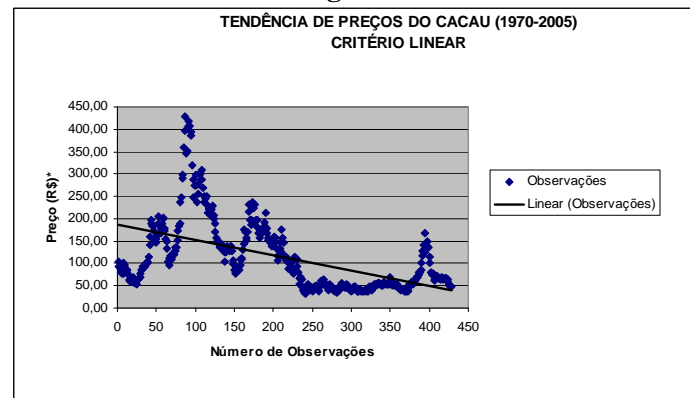
$$t_n = \frac{1}{2} \left(\frac{\sum_{i=-6}^5 p_{n+i-1}}{12} + \frac{\sum_{i=-6}^5 p_{n+i}}{12} \right)$$

Onde:

t_n → tendência calculada pelo método 2x12 MA do período n .

p_n → preço do cacau no período n .

Figura 2:



Fonte: Dados utilizados na pesquisa
* Valores atualizados para janeiro/2006

Para se compreender as variações sazonais no preço do cacau, calcula-se o índice mensal de sazonalidade. Este índice é obtido, num primeiro momento, pela divisão entre o preço mensal do produto em um determinado ano e preço médio anual daquele mesmo ano. Num segundo momento, calcula-se uma média para cada mês durante todo o período amostral (1970-2005) e obtém-se o índice mensal de sazonalidade. Matematicamente, tem-se:

$$i_n = \frac{\sum_{n=1}^k \frac{p_n}{\bar{p}_a}}{k}$$

Onde:

i_n = índice de sazonalidade do mês n

p_n = preço do cacau no mês n

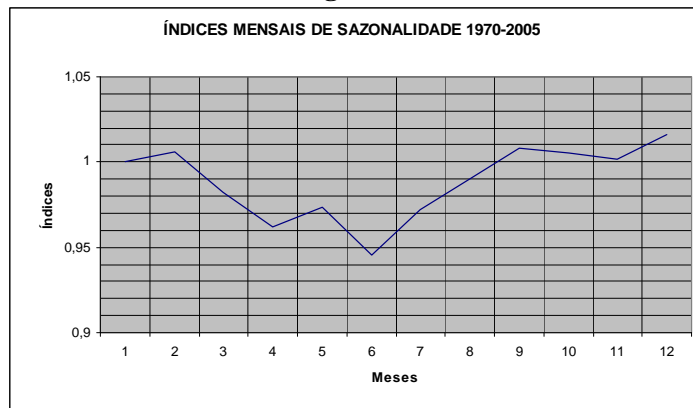
\bar{p}_a = preço médio anual do cacau

k = número de observações registradas na amostra do respectivo mês.

Neste caso, o principal objetivo para se calcular o índice de sazonalidade (i_n) é evidenciar em qual mês do ano o preço do cacau produzido no Brasil atinge um nível máximo

e um patamar mínimo. Pela figura 3, visualiza-se que o preço atinge um máximo no mês de dezembro e um mínimo no mês de junho.

Figura 3:



Fonte: Dados utilizados na pesquisa

Também segundo Lima Neto e Loureiro (2003), o movimento cíclico compreende as oscilações de longo prazo que podem ocorrer em torno de uma linha de tendência. Tais movimentos podem ser ou não periódicos e somente são considerados quando ocorrem depois de intervalos de tempo superiores a um ano (como observados na figura 4).

Para se calcular o movimento cíclico, atende-se ao processo metodológico seguinte:

$$mc_n = \frac{1}{12} \left(\sum_{t=-11}^0 \frac{p_{n+t}}{t_{n+t} \cdot i_{n+t}} \right)$$

Onde:

mc_n = movimento cíclico do período n

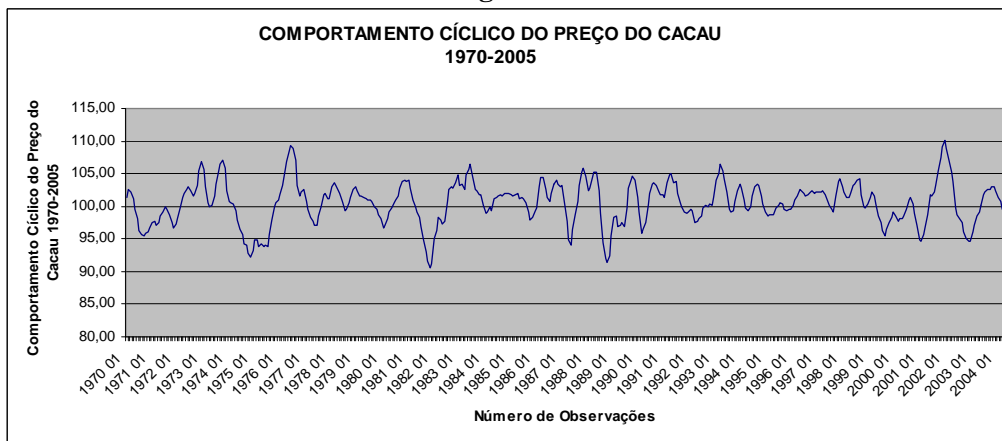
p_n → preço do cacau no período n .

t_n → tendência calculada pelo método 2x12 MA do período n .

i_n = índice de sazonalidade do mês n

A figura 4 evidencia o comportamento cíclico do preço do cacau durante o período janeiro/1970 – agosto/2005.

Figura 4:



Fonte: Dados utilizados na pesquisa

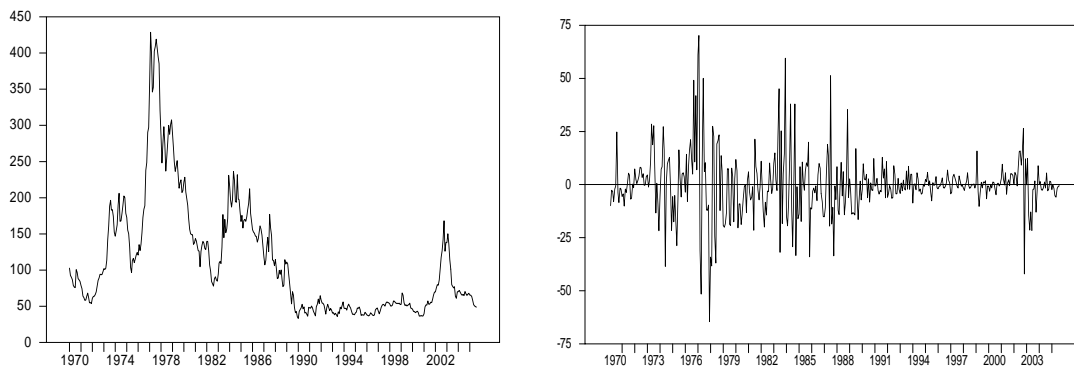
Como o objetivo deste artigo é a realização de uma previsão para a série de preços do cacau, verificam-se os critérios pertencentes à metodologia de Box-Jenkins – como parcimoniosidade nos parâmetros; invertibilidade e estacionariedade dos coeficientes dos modelos estimados; ajustamento do modelo aos dados da série temporal com base nos critérios AIC e SBC; análise da estatística Q de Box-Ljung-Pierce para identificar se os resíduos dos modelos alternativos estimados possuem características de ruído branco; e observação do erro quadrado médio da previsão (EQM) – para se desenvolver um processo de previsão consistente e plausível para a série em questão.

A partir de então, alguns desses critérios serão demonstrados, como estatísticas, gráficos, modelos, testes e previsão.

Observando o gráfico à esquerda na figura 5, pode-se observar que a série de preço estudada apresenta sua maior elevação absoluta no período 1973-1975. Outros aumentos podem ser notados durante 1984-1985, mantendo-se relativamente estável ao longo da década de 90. Em 2003 pode-se perceber um novo aumento acompanhado de uma pequena queda até início de 2005.

Já no gráfico à direita na mesma figura, pode ser observado que o comportamento dos preços em primeira diferença possui características visualmente estacionárias, pois os valores variam, agora, em torno de uma linha ao longo do período tratado. Esta série apresenta uma oscilação maior nos anos de 1976, 1977, 1982, 1987 e 2003.

Figura 5: Série de Preços do Cacau e Série de Preços do Cacau em Primeira Diferença



Fonte: Dados utilizados na pesquisa

Uma outra análise importante é aquela vinculada às autocorrelações. Quando a autocorrelação e a autocorrelação parcial da série de dados são verificadas “em nível” (gráfico à esquerda na figura 6), percebe-se que a série de dados relativa ao preço mensal do cacau tem um comportamento não-estacionário, pois essas duas estatísticas reduzem-se lentamente por fora do intervalo de confiança.

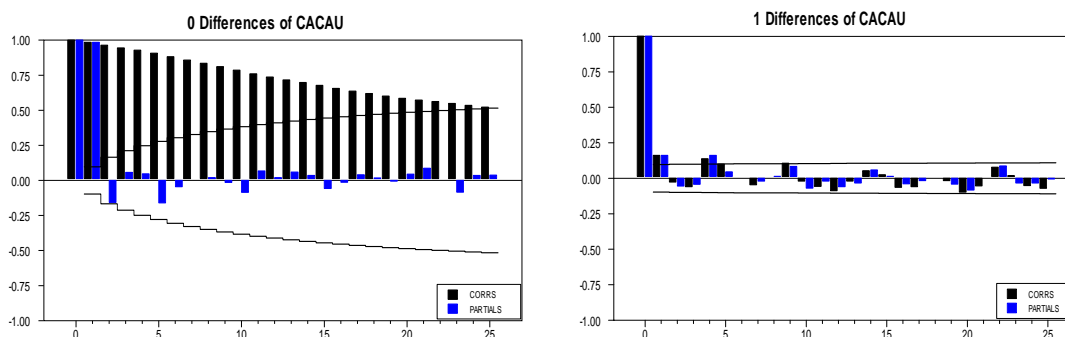
Mas quando a autocorrelação e autocorrelação parcial são analisadas em primeira diferença (gráfico à direita na figura 6), verifica-se que a série passa a assumir características de estacionariedade, haja vista o comportamento das estatísticas mencionadas não se desviarem do intervalo de confiança (exceto em regiões pontuais do mesmo). Os pontos onde as autocorrelações desviam-se do intervalo serão os indicadores de quais níveis de “p” e “q” (até porque já se sabe que $d = 1$) irão mensurar o processo ARIMA. Pelo gráfico da primeira diferença, pode-se constatar que:

$p = AR = 1, 4; e$

$q = MA = 1, 4.$

Portanto, o processo será formado por um ARIMA (2, 1, 2).

Figura 6:



Fonte: Dados utilizados na pesquisa

Analisando a tabela VII, 5 modelos são indicados como candidatos para poder realizar a previsão *ex-post* e, posteriormente, efetua-se a seleção dos modelos a partir dos critérios SQR, AIC e SBC. Verifica-se que pelo primeiro critério (menor SQR) foram selecionados os modelos 2, 3, 4 e 5. Pelo segundo critério (menores AIC e SBC) os modelos 2 e 4 foram escolhidos para teste e os modelos 3 e 5 foram eliminados. Portanto, ficaram para serem testados os modelos 2 e 4.

Tabela VII: Resultados Empíricos

Modelos	SQR	AIC	SBC
Modelo 1 ($(\ 4\ , 1, \ 1\)$)	79.676,146	4780,3447	4.792,4868
Modelo 2 ($(\ 4\ , 1, \ 1, 4\)$)	79.635,7087	4.781,6471	4.797,8366
Modelo 3 ($(\ 4\ , 1, \ 1\)$)	79.665,9968	4.781,808	4.797,997
Modelo 4 ($(\ 1, 4\ , 1, \ 4\)$)	79.557,0511	4.781,232	4.797,418
Modelo 5 ($(\ 1, 4\ , 1, \ 1, 4\)$)	79.531,8874	4.783,096	4.803,332

Fonte: Dados utilizados na pesquisa

A previsão *ex-post* é realizada para gerar valores dentro do próprio período. A idéia é de que, quanto melhor forem essas previsões, melhor será o modelo estimado. Portanto, o EQM da previsão (que é igual à média do quadrado da diferença entre cada valor previsto *ex-post* e o valor real observado na amostra) é uma medida formal da qualidade das previsões *ex-post*. Sendo assim, para efeitos de comparabilidade, quanto menor o EQM melhor será o grau de ajustamento do modelo aos dados da série temporal.

De acordo com a tabela VIII, evidencia-se que o modelo 4 apresenta um menor EQM. Dessa forma, este é selecionado para realizar a previsão *ex-ante* dos preços do cacau no mercado nacional para um horizonte de 12 meses a partir do período setembro/2005.

Tabela VIII: Erro Quadrado Médio (EQM) das Previsões *Ex-Post*

Modelos	EQM	Erro Mínimo	Erro Máximo
Modelo 2 ($(\ 4\ , 1, \ 1, 4\)$)	85,478194	0,227396	238,608936
Modelo 4 ($(\ 1, 4\ , 1, \ 4\)$)	81,25901	0,420573	226,200526

Os dados estatísticos relacionados ao modelo 4 são apresentados na tabela IX.

Primeiramente, é importante colocar que na metodologia Box-Jenkins o ideal é que todos os coeficientes estatísticos sejam aprovados pelo teste “t” (o qual evidencia que aquele parâmetro não será igual a zero). Pelo termo aqui empregado, a aprovação dos coeficientes no respectivo teste não é condição fundamental para que um modelo não possa ser caracterizado

como o representativo de uma série de dados. Isto posto, observa-se na tabela VI apenas o coeficiente de $AR\{1\}$ passa no teste comentado.

O teste “ R^2 Centrado” revela uma ótima aderência da variável dependente às variáveis independentes do modelo, isto é, o preço do cacau no período t é 97,06% explicado (em média) pela variáveis: o próprio preço defasado em 1 e 4 períodos, assim como também o termo de erro defasado em 4 períodos.

Por último, o nível de significância em “Q” (maior que 5%) revela que o processo é gerador de ruído branco, podendo, portanto, ser utilizado como modelo de previsão.

Tabela IX: Modelo Estimado

	AR{1}	AR{4}	MA{4}
Coefficientes	0,17437	0,300791	-0,166043
Erros-Padrão	0,0468	0,19851	0,20903
Estatísticas “t”	3,72558	1,51521	-0,79432
Significâncias	0,00022	0,13047	0,42745
R^2 Centrado	0,97069		
Significância de Q	0,0527		

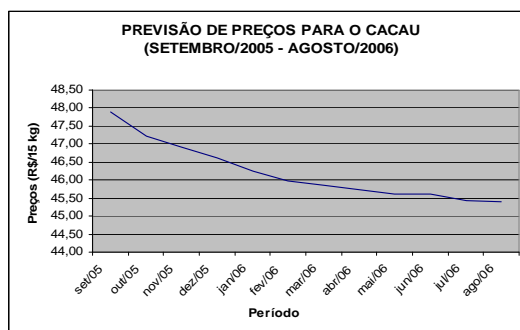
Observando a tabela X e sendo levados em consideração os critérios de previsão e checagem anteriormente comentados, a previsão de preços para o cacau produzido no Brasil segue a tendência histórica (1970-2005) de queda. Vale salientar que essas previsões são realizadas sabendo-se que as informações disponibilizadas para domínio público por parte do IPEA/FGV têm uma defasagem aproximada de 6 meses; isto é, quando esses valores foram estimados ainda não se possuía qualquer informação sobre os mesmos.

Tabela X: Previsões *Ex-Ante*

Período	Previsões
Setembro/2005	47,8793
Outubro/2005	47,2377
Novembro/2005	46,9084
Dezembro/2005	46,6287
Janeiro/2006	46,2488
Fevereiro/2006	45,9896
Março/2006	45,8454
Abril/2006	45,7361
Maio/2006	45,6028
Junho/2006	45,6028
Julho/2006	45,4405
Agosto/2006	45,3970

A figura 7 evidencia a análise gráfica do comportamento dos preços previstos para o período setembro/2005 a agosto/2006. Como comentado anteriormente, observa-se que ao longo do período da previsão, o preço do cacau recebido pelo produtor apresentou um comportamento decrescente sem manifestar qualquer oscilação brusca, iniciando em setembro/2005 em R\$47,87 e fechando a previsão em agosto/2006 em R\$45,39. Essa queda gradativa no preço estimado do produto passa a ser significativa, dado que a produção nacional do cacau tenda a aumentar, uma vez encontrando-se o mercado nacional – como citado anteriormente – num período de recuperação econômica.

Figura 7



4. CONCLUSÃO

O cacau é um produto tradicional da agricultura brasileira e que vem apresentando uma alavancagem na sua produção nos últimos anos, após um período prolongado de crise.

Neste artigo, efetuou-se uma previsão *ex-ante* para o preço do cacau empregando uma abordagem ao método Box-Jenkins. O modelo utilizado para tal questão foi o ARIMA(2,1,2). Dentre os candidatos a melhor modelo, o modelo 4 foi selecionado por meio dos critérios AIC, SBC e Erro Quadrado Médio. Como mencionado anteriormente, quanto menor AIC e SBC melhor o ajustamento do modelo. É importante ressaltar que na metodologia Box-Jenkins o ideal é que todos os coeficientes sejam aprovados pelo teste estatístico “t”, mas isso não é condição fundamental para que o modelo se torne representativo. Para o modelo escolhido, verificou-se que somente o termo AR{1} foi aprovado pelo mencionado teste. Através do teste “Q” revelou-se a presença de ruído branco para a série do preço do cacau.

Os resultados encontrados pela previsão seguiram à tendência histórica de queda. No período setembro/2005 obteve-se um preço de R\$47,88 fechando em agosto/2006 em, aproximadamente, 45,40, isto é, uma queda de 5,47% em doze meses.

Portanto, a análise e previsão – neste caso, para o preço do cacau – são de extrema importância para que tanto investidores quanto produtores possam formar suas expectativas a respeito do comportamento do preço futuro deste produto, minimizando possíveis externalidades negativas a respeito da série em estudo; exceto no caso de choques econômicos (políticas de tributação, crises internacionais etc.) no setor cacauzeiro.

5. REFERÊNCIAS

- ENDERS, Walter. Applied econometric time series. [New York]: Wiley and Sons, 1995.
- GUJARATI, Damodar N. Econometria básica. 3 ed. São Paulo: Pearson, 2000.
- GREENE, Willian H. Econometric Analysis. 5.ed. New Jersey: Prentice Hall, 2003.
- HILL, R. Carter et al. Econometria. São Paulo: Saraiva, 1999.
- ISAE/FGV. Cacau: Pontencialidades Regionais e Estudo de Viabilidade Econômica. Manaus: Suframa, 2003.
- LIMA NETO, Eufrásio de Andrade e LOUREIRO, Juliana. Mineração de séries temporais e dados sequenciais. Recife: Ufpe, 2003.
- MAKRIDAKIS, Spyros et al. Forecasting methods and applications. [New York]: Wiley e Sons, [2000].
- MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA E COMÉRCIO. Resumo sobre exportações. Disponível em: www.desenvolvimento.gov.br/arquivo/rdp/proAcao/forcompetitividade/impZonLivComercio/13citricosResumo.pdf. Acesso em 06 mar 2006.
- MOREIRA JÚNIOR, Fernando de J. Modelos Box-Jenkins e forecasting qualitativo. [S. l.: s.n.], [2002].

- OLIVEIRA, Francisco Martins de Oliveira. Estatística e probabilidade. 2. ed. São Paulo: Atlas, 1999.
- TODA FRUTA. Chat técnico sobre tecnologia para produção de cacau. [S. l.: s.n.], 2005.
- VARIAN, Hal R. Microeconomia: princípios básicos. 6.ed. Rio de Janeiro: Campus, 2003.