APLICAÇÃO DO MODELO ARIMA PARA PREVISÃO DO PREÇO DO FRANGO INTEIRO RESFRIADO NO GRANDE ATACADO DO ESTADO DE SÃO PAULO

PAULO ANDRÉ CAVALCANTI CAMPOS
Ademir Clemente
AGNALDO ANTÔNIO LOPES DE CORDEIRO

Resumo:

O objetivo deste estudo foi elaborar um modelo de previsões para o preço do frango inteiro resfriado no grande atacado do estado de São Paulo, utilizando a metodologia ARIMA ou de Box-Jenkins, de previsões de séries temporais, em sua forma univariada. Utilizou-se a série histórica mensal entre os anos de 1996 a 2005, os testes de previsão ex post foram realizados para os anos de 2004 e 2005. Todos os preços da série, foram atualizados pelo IGP-DI da FGV para o mês de dezembro de 2005, como forma de eliminar o efeito da inflação. Após os testes foram determinados 4 modelos que demonstraram consistência estatística e bom desempenho de previsões. Foram realizados testes de previsões trimestrais anuais e para 2 anos, e todos os modelos apresentaram boa performance, o que não era esperado para os testes de 2 anos, já que a metodologia ARIMA univariada é reconhecidamente eficiente, apenas para previsões de curto prazo. Os bons resultados alcançados pelo modelo fornecem boas expectativas para seu uso como apoio ferramental a profissionais de diversas áreas como planejamento, orçamento, investimentos, entre outras.

Área temática: *Aplicação de Modelos Quantitativos na Gestão de Custos*

Aplicação do modelo ARIMA para previsão do preço do frango inteiro resfriado no grande atacado do estado de São Paulo

Resumo

O objetivo deste estudo foi elaborar um modelo de previsões para o preço do frango inteiro resfriado no grande atacado do estado de São Paulo, utilizando a metodologia ARIMA ou de Box-Jenkins, de previsões de séries temporais, em sua forma univariada. Utilizou-se a série histórica mensal entre os anos de 1996 a 2005, os testes de previsão *ex post* foram realizados para os anos de 2004 e 2005. Todos os preços da série, foram atualizados pelo IGP-DI da FGV para o mês de dezembro de 2005, como forma de eliminar o efeito da inflação. Após os testes foram determinados 4 modelos que demonstraram consistência estatística e bom desempenho de previsões. Foram realizados testes de previsões trimestrais anuais e para 2 anos, e todos os modelos apresentaram boa performance, o que não era esperado para os testes de 2 anos, já que a metodologia ARIMA univariada é reconhecidamente eficiente, apenas para previsões de curto prazo. Os bons resultados alcançados pelo modelo fornecem boas expectativas para seu uso como apoio ferramental a profissionais de diversas áreas como planejamento, orçamento, investimentos, entre outras.

Palavras-chave: Previsão de preços. Modelo ARIMA. Séries temporais.

Área Temática: Aplicação de Modelos Quantitativos na Gestão de Custos.

1 Introdução

O desejo de antecipar o conhecimento de fatos futuros é intrínseco ao ser humano, desde os tempos remotos, o homem vem tentando de diversas formas obter informações que ainda estão para acontecer, são tantos que procuram e outros tantos que alegam ser capazes de realizar tal desejo, que no código de processo criminal do estado de Nova York foi inserida uma pena para os que alegam ter esta capacidade, como descrito a seguir: “Pessoas que simulam prever o futuro são consideradas desordeiras nos termos da subdivisão 3, seção 901 do código criminal e estarão sujeitas a uma multa de US$ 250 e / ou seis meses de prisão”. Seção 889, Código de Processo Criminal do Estado de Nova York, apud Pindyck e Rubinfeld (2004, p. XIX). O tema principal abordado neste trabalho, Previsões de Séries Temporais, não difere muito em seus fins, ao exposto anteriormente no início desta introdução, a real diferença são os meios ou o ferramental metodológico utilizado para alcançar os desejos de antecipação de fatos ainda não concretizados.

No mundo dos negócios, altamente competitivo, um dos maiores se não for o maior diferencial estratégico, é a capacidade ou habilidade de se antecipar as tendências do mercado. O tema “previsões” se insere neste contexto estratégico competitivo no mundo da economia e negócios empresariais.

Previsões no contexto deste trabalho, refere-se a prospecções ou inferências de variáveis econômicas para períodos futuros, utilizando modelos estatísticos (econométricos) que captam o comportamento passado de uma variável ou utilizam o comportamento de outras variáveis (explicativas) para o processo de estimação. Particularmente o foco deste estudo é o modelo ARIMA (Auto-Regressivo Integrado de Médias Móveis) de previsões de séries temporais, em sua forma univariada (sem variáveis explicativas) ou também conhecido como modelo ou metodologia de Box-Jenkins.

Como meio empírico de abordagem e exploração do modelo ARIMA, utiliza-se aqui, a série histórica mensal para os anos 1996 a 2005 do preço médio do frango inteiro resfriado para o grande atacado do estado de São Paulo. A série do preço do frango inteiro foi escolhida pela grande significância econômica do produto para o segmento empresarial, pois atualmente o Brasil figura como o maior produtor e exportador de carne de frango do mundo, sendo o estado de São Paulo um dos maiores produtores e o maior consumidor nacional de frango. Esta série em particular, é termômetro nacional e alvo de interesse de todos os “players” deste mercado. Com o exposto, surge o seguinte problema de pesquisa: Com a metodologia ARIMA de previsões de séries temporais, pode-se elaborar um modelo consistente e eficiente de previsões para a série histórica do preço do frango inteiro resfriado no grande atacado do estado de São Paulo?

Diante deste problema, objetiva-se com este estudo, elaborar um modelo ARIMA de previsões para o preço do frango inteiro resfriado no grande atacado do estado São Paulo.

A estrutura deste trabalho contempla além, desta introdução, um referencial teórico sobre o modelo ARIMA de previsões, exposto na seção 2; na seção 3 é apresentada a metodologia utilizada para a análise e elaboração dos modelos determinados; na seção 4 é feita a identificação e aplicação do modelo à série de preços; e na seção 5 conclui-se o estudo com considerações e propostas para futuras pesquisas sobre o tema.

2 Referencial teórico

Esta seção aborda o referencial teórico necessário ao desenvolvimento do modelo ARIMA, a subseção 2.1 introduz os conceitos sobre séries temporais, em seguida na 2.2 descreve-se as propriedades das séries temporais estacionárias e não-estacionárias, na subseção 2.3 aborda-se os aspectos conceituais do modelo ARIMA e na 2.4 trata-se de métodos de avaliação de performance de previsões.

2.1 Séries temporais

Estudos com dados econômicos podem ser apresentados de duas formas, dados em corte transversal (*cross section*) ou séries de tempo (*time series*). Para Fava (2000a), dados de uma variável em corte transversal são observados em um instante específico de tempo, o que varia são os grupos de unidades observadas que podem ser regiões, classes, empresas, consumidores, etc.. Para séries de tempo, observa-se a trajetória temporal de uma variável econômica ordenada seqüencialmente no tempo.

Uma série temporal pode ser definida, como um conjunto de dados observados ao longo de um período no tempo, para Morettin e Toloi (2004, p. 1), “Uma série temporal é um conjunto de observações ordenadas no tempo”. As vendas mensais de uma empresa, o consumo diário de energia elétrica de uma cidade, a cotação do preço de uma ação na bolsa de valores, são exemplos comuns de séries temporais. Uma série temporal pode ser apresentada de diversas formas, como por exemplo: diária, mensal, semestral ou anual, no entanto, obrigatoriamente toda a série deve estar representada com a mesma periodicidade. Segundo Pindyck e Rubinfeld (2004, p. 3), “Dados que descrevem o movimento de uma variável ao longo do tempo são chamados *séries temporais*, as quais podem ser diárias, semanais, mensais, trimestrais ou anuais”.

Pindyck e Rubinfeld (2004), classificam modelos de previsões em três grandes grupos: série temporal, regressão de uma única equação e equações múltiplas.

Diante das dificuldades na utilização de modelos econométricos estruturados que utilizam variáveis explicativas, os modelos de séries temporais são muito utilizados para previsão de variáveis econômicas. Pindyck e Rubinfeld (2004) classificam em dois tipos os modelos de previsões de séries temporais, modelos determinísticos e modelos estocásticos. Os dois modelos utilizam o comportamento passado da série para prever seus componentes futuros, porém os modelos determinísticos não fazem referência às fontes ou a natureza aleatória (estocástica) subjacente à série.

2.2 Séries estacionárias

Os modelos estocásticos de séries temporais são válidos apenas na aplicação em séries ditas estacionárias, para Hill, Griffiths e Judge (2003 p. 389), "Um processo estocástico (série temporal) y_t é estacionário se sua média e sua variância são constantes ao longo do tempo, e a covariância entre dois valores da série depende apenas da distância no tempo que separa os dois valores, e não dos tempos reais em que as variáveis são observadas".

Uma série temporal pode ser interpretada como um processo estocástico, Hill, Griffiths e Judge (2003) observam que uma variável econômica é aleatória porque não se pode prevê-la perfeitamente e o modelo econômico que gera uma variável de série temporal é chamado de processo estocástico ou aleatório. Uma amostra particular da série é normalmente chamada *uma realização particular do processo estocástico*. Sobre o tema, Gujarati (2000, p. 719) relata que: "A distinção entre o processo estocástico e sua realização é parecida com a distinção entre população e amostra em dados de corte [Cross Section]. Assim como utilizamos dados amostrais para fazer inferências sobre uma população, em séries temporais usamos a realização para fazer inferências sobre o processo estocástico subjacente".

Sobre séries estacionárias, Pindyck e Rubinfeld (2004) comentam que se o processo estocástico que gerou a série não varia em relação ao tempo pode-se modelar o processo através de uma equação com coeficientes estimados com base na suposição que sua relação estrutural não muda com o tempo. Se tal relação muda com o tempo não se pode usar o modelo para fazer previsões.

Porém sabe-se, que a maioria das séries temporais econômicas são não estacionárias, entretanto pode-se através de processos de diferenciação da variável em relação a períodos defasados, transformá-las em séries estacionárias, sobre o tema, Pindyck e Rubinfeld (2004) comentam que provavelmente poucas séries temporais são estacionárias, mas a maioria delas têm a propriedade desejável de que quando as diferenciamos uma ou mais vezes, as séries resultantes são estacionárias. Tais séries não-estacionárias são chamadas de homogenias.

Fica evidente a importância da habilidade na detecção da estacionariedade ou não de uma série temporal. Para este fim (detecção de estacionariedade), comumente faz-se o uso da *função autocorrelação (FAC)*. Gujarati (2000), relata que a *função autocorrelação amostral* $\hat{\rho}_k$ (rô) na defasagem *k*, é definida como:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (1) \qquad \hat{\gamma}_k = \frac{\sum (Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{n} \quad (2) \qquad \hat{\gamma}_0 = \frac{\sum (Y_t - \overline{Y})^2}{n} \quad (3)$$

Sendo que $\hat{\gamma}_k$ (gama), representa a covariância amostral na defasagem k e $\hat{\gamma}_0$, a variância amostral e que $\overline{Y}$ é a média da amostra (série) e *n* o tamanho da amostra (série).

Uma representação gráfica de $\hat{\rho}_k$ contra sua defasagem *k* é conhecida como *correlograma amostral* (comumente encontrado nos pacotes estatísticos) e é de grande utilidade na detecção da estacionariedade ou não de uma série temporal.

Quando no *correlograma* o coeficiente de correlação inicial se mostra elevado e com o crescimento das defasagens *k* este comportamento declina lentamente, é característico de uma série não estacionária como mostra a Figura 1. Em contraste, quando no *correlograma* o coeficiente de correlação cai abruptamente logo após as primeiras defasagens é característico que o processo estocástico é aleatório ou estacionário, Figura 2.

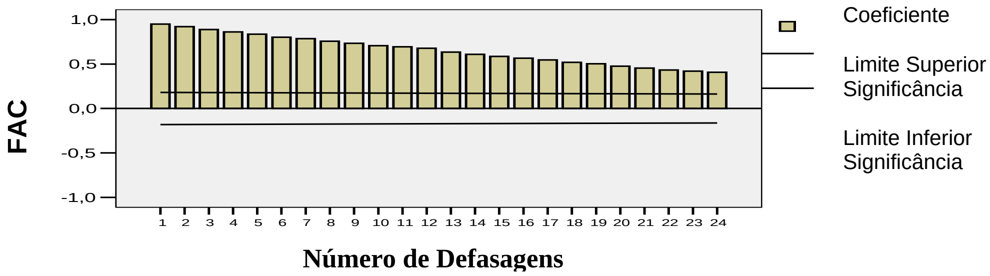

Figura 1 - Exemplo de correlograma amostral de uma série temporal não estacionária

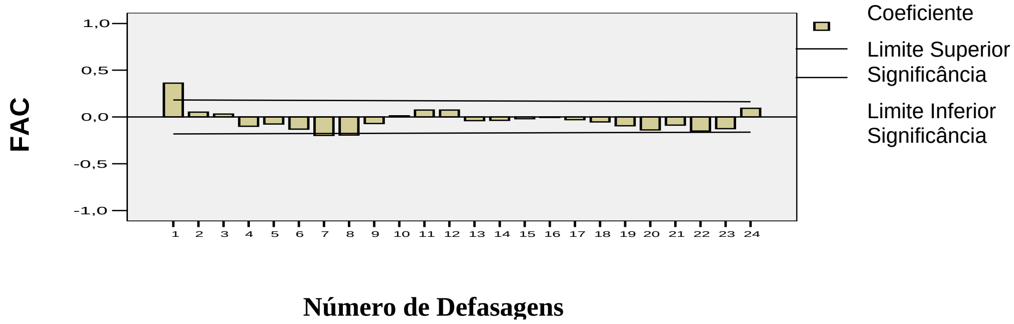

Figura 2 - Exemplo de um correlograma amostral de uma série temporal estacionária

Sobre o uso da função autocorrelação e sua representação gráfica o correlograma, Pindyck e Rubinfeld (2004) observam que a função de autocorrelação para uma série estacionária declina à medida que k, o número de defasagens, se torna maior, mas em geral o mesmo não acontece com séries não-estacionarias. Se após a diferenciação de uma série não-estacionária, a série resultante se mostra estacionária, diz-se que a série original é *integrada de ordem 1* ou representada com *I(1)*, se apenas na segunda diferenciação a série se torna estacionária a série original é denominada como *I(2)*, generalizando se uma série para alcançar a estacionariedade for diferenciada *n* vezes, é considerada integrada ou homogenia de ordem *n I(n)*.

Uma alternativa para detectar estacionariedade em séries temporais é o teste de raiz unitária. Gujarati (2000) coloca que se após a regressão da expressão:

$$Y_t = \rho Y_{t-1} + u_t \tag{4}$$

se verificar que $\hat{\rho}$ (rô) é significativamente igual a 1, diz-se que a variável *Y* tem uma raiz unitária, sendo que uma série temporal com raiz unitária é conhecida como *passeio aleatório* e um passeio aleatório é um exemplo de série não-estacionária. A estatística *t* para testar raiz unitária é conhecida como estatística τ (tau), cujo valores críticos foram tabulados por Dickey e Fuller. O teste *tau* é conhecido na literatura especializada como teste de *Dickey-Fuller (DF)*.

Gujarati (2000) acrescenta que normalmente a equação (4) é expressa de uma forma alternativa como: $\Delta Y_t = (\rho - 1)Y_{t-1} + u_t$ ou $\Delta Y_t = \gamma Y_{t-1} + u_t$ (5)

Neste caso se testa a significância de γ (gama) = 0 para verificar se a série possui raiz unitária.

Segundo Hill, Griffiths e Judge (2003), Dickey e Fuller também desenvolveram valores críticos em presença de constante, equação (6) e com tendência, equação (7), como segue: $\Delta Y_t = \alpha_0 + \gamma Y_{1-1} + u_t$ (6) $\Delta Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{1-1} + u_t$ (7)

No Quadro 1, apresenta-se os valores críticos τ (tau) para o teste de raiz unitária.

Modelo / Significância	**1%**	**5%**	**10%**
$\Delta Y_t = \gamma Y_{t-1} + u_t$	-2,56	-1,94	-1,62
$\Delta Y_t = \alpha_0 + \gamma Y_{1-1} + u_t$	-3,43	-2,86	-2,57
$\Delta Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{1-1} + u_t$	-3,96	-3,41	-3,13

Fonte: adaptado de Davidson e Mackinnon (1993, apud HILL, GRIFFITHS e JUDGE 2003, p. 399)

Quadro 1 – Valores críticos para o teste de Dickey-Fuller

O Quadro 1, mostra os valores críticos para a estatística τ (tau) que são válidos em grandes amostras para um teste unilateral, para a hipótese nula γ = 0, um processo não estacionário de raiz unitária, ou que seja rejeitada em favor da alternativa γ < 0, um processo estacionário Hill, Griffiths e Judge (2003). Fava (2000c) acrescenta que o teste de Dickey-Fuller se destina às séries que possuem no máximo uma raiz unitária, ou seja, séries que são estacionárias ou são estacionarizadas com a aplicação de uma diferença. Entretanto, com base em inúmeros trabalhos empíricos, observa-se que grande parte das séries macroeconômicas são integradas de ordem 1.

2.3 Modelo ARIMA de Previsões de Séries Temporais

O modelo ARIMA (Auto Regressivo Integrado de Média Móvel), conhecido também como metodologia de Box-Jenkins, é próprio para previsões de séries temporais. Fava (2000b, p. 205) observa que “os modelos ARIMA resultam da combinação de três componentes também denominados “filtros”: o componente Auto-regressivo (AR), o filtro de Integração (I) e o componente de Médias Móveis (MA)”. Sobre o modelo ARIMA, Maddala (2003) relata que a abordagem Box-Jenkins é uma das metodologias mais usadas para a análise de dados em séries temporais. Ela é popular em conseqüência de sua generalidade; ela pode lidar com qualquer série, estacionária ou não, com ou sem elementos sazonais.

Este modelo é comumente representado pela notação ARIMA(p,d,q), sendo (p,d,q) a representação da ordem do modelo. Uma representação ARIMA(1,2,0), indica um modelo de ordem 1 para o componente AR (Auto-Regressivo), ordem 2 para o componente I (Integração ou diferenciação) e o último 0 para o componente MA (Média Móvel).

Para a utilização do modelo ARIMA, faz-se necessário o uso de recurso computacional específico encontrado em pacotes estatísticos especializados em previsões de séries temporais. Segundo Pindyck e Rubinfeld (2004) a aplicação do modelo ARIMA é composta de quatro etapas, *identificação, estimação, verificação e previsão*.

Identificação

Consiste em determinar os filtros (p, d, q) e a ordem que melhor representa a série temporal. Sobre esta etapa da metodologia, Morettin e Toloi (2004), propõem a seguinte seqüência: primeiro se determina se a série é ou não estacionária, se for estacionária, o filtro *d* ou ordem de integração é zero e representa que a série não precisa de diferenciação para torná-la estacionária, neste caso tem-se um modelo ARIMA(p,0,q). Se a série não for estacionária, faz-se necessário diferenciações para torná-la estacionária, se com uma diferenciação a série resultante se mostra estacionária o filtro *d* é de ordem 1, se não, a série resultante não se mostrar estacionária, deve-se efetuar outra diferenciação, e assim sucessivamente, até torná-la estacionária. Na prática, a maioria das séries econômicas torna-se estacionária com apenas uma ou no máximo duas diferenciações.

No processo de identificação de estacionariedade, faz-se o uso da função autocorrelação FAC e sua representação gráfica (correlograma) ou alternativamente o teste de raiz unitária, entre os quais o mais conhecido é o de *Dickey-Fuller (DF).*

Às vezes apenas diferenciações não se mostram suficientes para "estacionar" uma série temporal. Sobre isto, Fava (2000b) observa que, quando uma série possui tendência determinística, as diferenças não são suficientes para estacioná-la, antes a tendência deve ser retirada.

Estimação

Após a determinação dos valores *p, d* e *q*, passa-se para a estimação dos *p* parâmetros ϕ (Fi) e dos *q* parâmetros θ (Teta) e da variância σ_ε^2 (Sigma) do modelo de regressão. (FAVA, 2000b):

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q} \qquad (8)$$

Fava (2000b) coloca que a estimação pode ser feita por mínimos quadrados ou por máxima verossimilhança; porém se há um componente MA, o modelo será não-linear, o que exigirá a utilização do método de mínimos quadrados não-lineares. Contudo, qualquer que seja o método, o processo de estimação é extremamente trabalhoso e requer o uso software específico. Experimentos realizados indicam que o método de máxima verossimilhança é superior ao de mínimos quadrados, quando o tamanho da série é pequena.

Verificação

Esta etapa consiste em avaliar a adequação do modelo quanto à consistência, verificar se há parâmetros em excesso e se são significativos, e também, se os erros resultantes não são autocorrelacionados.

Para verificar se o modelo está superespecificado, contém parâmetros em excesso, ou está subespecificado por falta de parâmetros; Fava (2000b) propõe os seguintes procedimentos:

- Se o valor de um coeficiente estimado for pequeno em relação a seu erro-padrão, indicando sua não significância estatística, é provável que haja superespecificação. Se for o coeficiente de maior ordem, deve-se suprimi-lo, se for um de menor ordem, convém analisar sua correlação com os demais coeficientes do modelo, se há alta correlação indica que um dos dois pode ser redundante.
- Para verificar se há subespecificação, deve-se introduzir parâmetros adicionais e analisar sua significância estatística. A cada nova estimação do modelo apenas um parâmetro deve ser introduzido.
- Se após a verificação pelas formas indicadas anteriormente houver empate entre dois modelos, o desempate pode ser feito escolhendo o modelo que tenha a menor estimativa da variância do ε_t, $\hat{\sigma}_\varepsilon^2$, e os menores valores para os critérios AIC (Akaike's Information Criteria) e BIC (Bayesian Information Criteria).

Previsão

Após identificação, estimação e verificação, o modelo ou modelos selecionados devem ser testados por meio de simulações. Pindyck e Rubinfeld (2004) observam que o objetivo é prever valores futuros de uma série temporal que estejam sujeitos ao menor erro possível, por isso considera-se como a melhor previsão a que apresenta o mínimo erro quadrado médio.

Sobre previsões de séries temporais com o modelo ARIMA, Gujarati (2000 p. 745) observa que: "Uma das razões para a popularidade da modelagem ARIMA é seu sucesso em fazer previsão. Em muitos casos, as previsões obtidas com este método são mais confiáveis do que as obtidas com a modelagem econométrica tradicional especialmente para previsões de curto prazo. Naturalmente é preciso checar cada caso".

As previsões são elaboradas com a aplicação do modelo ou equação (8) estimada nos passos anteriores. Na prática, felizmente, o processo de previsão é realizado quase totalmente com o uso de software especializado. Isso permite testar para diferentes períodos e formatos de previsão. Fava (2000b) alerta que os modelos ARIMA têm sua capacidade de previsão comprometida no longo prazo, então, sempre que possível, deve-se atualizar as previsões já realizadas.

Modelo ARIMA sazonal ou SARIMA

A modelagem ARIMA também é útil na análise de séries temporais com características sazonais. Um grande número de séries econômicas tem esta característica, tais como; preço, produção, venda, etc.

O modelo ARIMA sazonal ou SARIMA, segue o seguinte modo de apresentação ARIMA(p,d,q)(P,D,Q), onde os parâmetros (P,D,Q) são os equivalentes sazonais de (p,d,q).

Fava (2000b) argumenta que os instrumentos utilizados na identificação de modelos sazonais continuam sendo a *FAC* (Função Autocorrelação) e *FACP* (Função Autocorrelação Parcial). O primeiro passo, agora, consiste em determinar os parâmetros *d* e *D.* A análise da *FAC* da série original e de suas diferenças, consecutivas, sazonais ou ambas, auxilia neste processo. Após as diferenciações *d* e *D*, segue-se para a determinação dos parâmetros *p*, *P*, *q* e *Q*. Para modelos que contêm filtros sazonais e não sazonais a *FAC* e *FACP* são mais complexas. Para facilitar, pode-se dizer que elas são uma mistura das funções dos modelos puramente sazonais e não sazonais. O comportamento dos coeficientes de autocorrelação de ordem baixa fornece subsídios para determinação de *p* e *q*, e os de ordem alta, múltiplas de *s* (período sazonal) ajudam a definir *P* e *Q*.

Modelo ARIMA com variáveis explicativas

A metodologia de Box-Jenkhins permite a inclusão de uma ou várias variáveis explicativas ao modelo, aproximando-o dos modelos econométricos tradicionais. Fava (2000d) relata que, quando as variáveis explicativas são também séries econômicas temporais, o modelo é conhecido como Função de Transferência. Quando são variáveis binárias (*dummies*), destinadas a captar eventos não mensuráveis, o modelo é chamado de Análise de Intervenção.

A inclusão de variáveis explicativas ao modelo ARIMA potencializa a metodologia, porque aproxima as qualidades intrínsecas dos modelos univariados, que captam os padrões de comportamento passado da própria variável e os unificam aos padrões de comportamento de outras variáveis (explicativas) que podem contribuir para prever o seu comportamento futuro. Contudo, as dificuldades encontradas nos modelos econométricos são transmitidas a essa metodologia: com a inclusão de variáveis explicativas, o modelo agora precisa também da projeção destas variáveis para o período requerido de previsão e tais projeções podem ser tão problemáticas quanto a projeção da própria variável em estudo.

2.4 Avaliação da performance de previsões

O objetivo principal da metodologia é encontrar um modelo que melhor descreva o comportamento da série no futuro. Um modelo pode (e deve) apresentar todas as evidências estatísticas que o tornem consistente, porém de nada adianta se não efetuar boas previsões. Esta checagem é o teste definitivo do modelo. Modelos com ótimas evidências estatísticas, muitas vezes podem ter fraco desempenho preditivo, e outros não tão ajustados estatisticamente podem ter bom desempenho em suas previsões. Bacchi e Hoffmann (1995, p. 21) em estudo anterior com aplicação do modelo ARIMA, observaram que: “Os modelos ajustados para a série de preços de bovinos (ΔY_t) com enfoque univariado não apresentaram bom desempenho em termos de previsão (apesar de muitos deles apresentarem-se adequados do ponto de vista dos testes estatísticos), não conseguindo captar a variação estacional da série”.

Uma metodologia útil para aferir o desempenho de um modelo de série temporal, consiste em previsões *ex post*, na qual são feitas simulações para períodos passados. Sobre previsões *ex post*, Pindyck e Rubinfeld (2004) consideram que neste tipo de previsão as variáveis já são conhecidas com certeza para o período de previsão, assim previsões *ex post* podem ser verificadas com os dados existentes e oferecem um meio de avaliar um modelo de previsão. Entretanto, os dados do período determinado para os testes de previsão não devem fazer parte da amostra na qual o modelo é rodado, os dados deste período devem ser reservados para a avaliação do desempenho de previsão.

Pindyck e Rubinfeld (2004) sugerem dois métodos de avaliação de previsões: *Raiz do Erro de previsão Quadrático Médio - REQM (Root Mean Square Erro)* e C*oeficiente de desigualdade de Theil - U de Theil*, definidos como:

$$REQM = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2} \qquad (11)$$

$$U = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2}}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s)^2} + \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^a)^2}} \qquad (12)$$

Onde: Y_t^s = valor previsto, Y_t^a = valor efetivo e T = número de períodos

A *REQM* é uma medida de desvio da variável simulada em comparação com sua evolução temporal. Pindyck e Rubinfeld (2004) observam que a magnitude desse erro apenas pode ser comparada com a média da variável em questão. O *U de Theil* mede a raiz do quadrado médio do erro de previsão em termos relativos e terá sempre um valor entre *0* e *1*: *0* para ajustamento perfeito e *1* para a pior previsão possível.

Para Gluckstern (2003) uma variante da equação (11) que mede a magnitude do erro é a *Raiz do Erro Quadrático Porcentual Médio - REQPM (Root Mean Square Percent Error) e* o *ErroPercentual Absoluto Médio - EPAM (Mean Absolute Percent Error)*, definidos como:

$$REQPM = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{(Y_t^s - Y_t^a)}{Y_t^a}\right)^2} \qquad (13)$$

$$EPAM = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{Y_t^s - Y_t^a}{Y_t^a}\right| \qquad (14)$$

3 Metodologia de análise adotada

3.1 Caracterização da pesquisa

Quanto aos procedimentos metodológicos, optou-se pelo que propõem Raupp e Beuren (2004), que enfocam as tipologias de delineamento de pesquisa, agrupadas em três categorias: quanto aos objetivos, quanto aos procedimentos e quanto à abordagem do problema.

Quanto aos objetivos, esta pesquisa se enquadra como descritiva, pois se pretende estudar e relatar o comportamento de uma série temporal econômica utilizando uma técnica estatística. Segundo Raupp e Beuren (2004), vários estudos utilizam a pesquisa descritiva para análise e descrição de problemas e neste tipo de pesquisa, normalmente ocorre o emprego de técnicas estatísticas, desde as mais simples até as mais sofisticadas.

Quanto aos procedimentos, esta pesquisa pode ser enquadrada como um estudo de caso, por concentrar-se na análise e projeção de um preço específico, em um mercado específico, com base em uma amostra de dados específica. Raupp e Beuren (2004) observam que a pesquisa do tipo estudo de caso caracteriza-se pelo estudo aprofundado e concentrado de um único caso e seus resultados não podem ser generalizados para outros fenômenos. Esta pesquisa também possui aspecto bibliográfico, já, que se buscou na literatura de referência o embasamento metodológico para a ferramenta estatística utilizada na análise e interpretação dos dados.

Quanto à abordagem do problema, a pesquisa se enquadra como quantitativa, uma vez que se pretende uma análise e interpretação dos dados com a utilização de ferramental estatístico. Segundo Raupp e Beuren (2004) diferentemente da pesquisa qualitativa, a abordagem quantitativa caracteriza-se pelo emprego de instrumentos estatísticos, tanto na coleta quanto no tratamento dos dados.

3.2 Método utilizado para a coleta e análise dos dados

Para aplicação do modelo ARIMA de séries temporais, utiliza-se a série histórica de janeiro de 1996 a dezembro de 2005 do preço médio em R$/Kg do frango inteiro resfriado no grande atacado do estado de São Paulo. Os dados foram coletados do banco de dados da empresa de consultoria JOX Assessoria Agropecuária no site http://www.jox.com.br/. Sendo os preços atualizados monetariamente para o período de dezembro de 2005 pelo Índice Geral de Preços Disponibilidade Interna – IGP-DI da Fundação Getúlio Vargas. A abordagem da metodologia ARIMA adotada é a forma univariada (sem variáveis explicativas). Utiliza-se o software estatístico SPSS como ferramenta computacional de suporte a aplicação do modelo.

Como forma de aferição do desempenho preditivo dos modelos desenvolvidos, faz-se o uso de 3 indicadores de mensuração de erros de previsão, *Erro Percentual Absoluto Médio – EPAM*, *Raiz do Erro Quadrático Médio – REQM* e *U de Theil*.

São elaboradas previsões trimestrais e anuais para os anos de 2004 e 2005, com os anos de previsão excluídos da amostra de análise. Por exemplo: para a previsão do 2º. trimestre de 2005, são utilizados os dados de janeiro de 1996 até março de 2005. Para testar a eficiência de previsão de longo prazo, são realizadas previsões diretas para 24 meses, 2004 e 2005, com os dados destes anos excluídos da amostra de análise.

Para a determinação dos modelos, é utilizada a metodologia de Box e Jenkins para modelo ARIMA sazonal, que compreende os seguintes passos: *Identificação*, *Estimação*, *Verificação e Previsão*. Para testar a estacionariedade da série, são utilizadas a Função Autocorrelação – FAC e Função Autocorrelação Parcial – FACP, e também o teste de Raiz Unitária de Dickey-Fuller. A escolha dos modelos se dá por sua consistência estatística dos parâmetros e pelo seu desempenho de predição.

4 Identificação e aplicação do modelo

Para o desenvolvimento e aplicação do modelo ARIMA, utilizou-se a série temporal do preço médio mensal do frango resfriado no grande atacado do estado de São Paulo, de janeiro de 1996 a dezembro de 2005. Para a realização de todos os testes, os preços foram corrigidos monetariamente para o mês de dezembro de 2005, pelo IGP-DI. Testes empíricos também foram feitos com a série nominal, sem atualização monetária e os resultados ficaram aquém das projeções com a série corrigida, principalmente para períodos mais longos.

Como determina a metodologia ARIMA, a série temporal deve ser estacionária, se não o for, deve-se diferenciá-la seqüencialmente, sazonalmente ou de ambos os modos, para torná-la estacionária. Um primeiro teste para detectar estacionariedade de uma série, consiste no exame de sua representação gráfica ao longo do tempo como mostra a Figura 3.

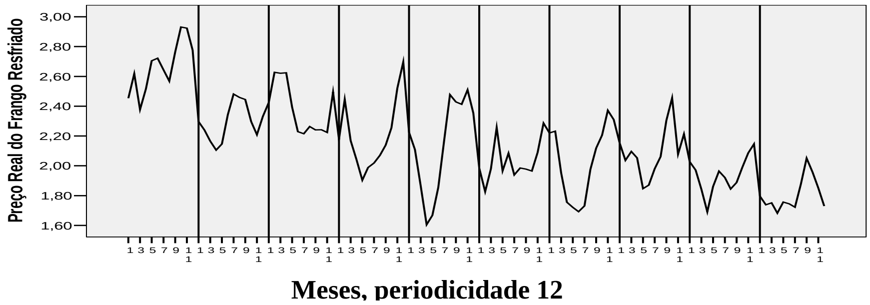

Figura 3 – Gráfico seqüencial da série do preço real do frango

Examinando a Figura 3, nota-se uma tendência sistemática de queda ao longo do tempo, contudo, parece menos acentuada nas datas mais recentes. Esta tendência de queda sistemática é típica de séries não estacionárias. Também se percebem picos seguidos de quedas, ao longo do tempo, caracterizando uma possível sazonalidade, comum em séries econômicas mensais. A Função Autocorrelação (FAC) auxilia na verificação da estacionariedade da série e na determinação da sazonalidade, como mostra a Figura 4.

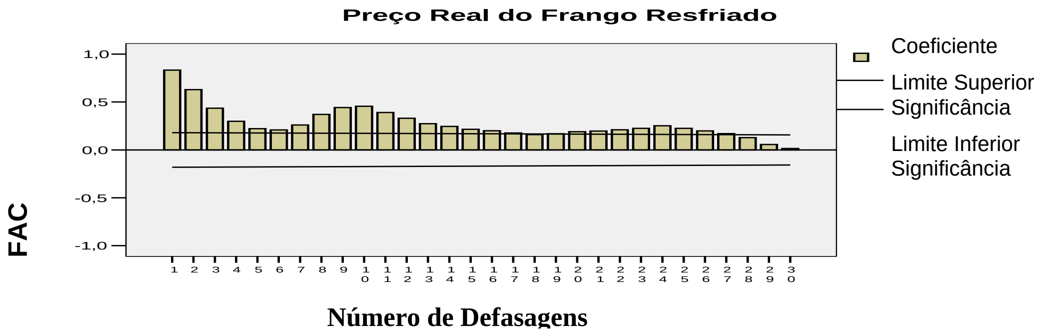

Figura 4 – Correlograma amostral da série do preço real do frango

Com a análise do correlograma, Figura 4, confirma-se o diagnóstico da não estacionariedade da série, pois quando uma série é estacionária logo nas primeiras defasagens o coeficiente de autocorrelação tende a cair abruptamente para zero, como na Figura 2. Também nota-se no correlograma, uma característica sazonal da série, com uma oscilação da FAC, típica de séries sazonais.

Com base neste diagnóstico, deve-se aplicar uma diferença seqüencial e também uma sazonal para conseguir a estacionariedade da série. O gráfico seqüencial e o correlograma da série diferenciada resultante são apresentados nas Figuras 5 e 6, a seguir.

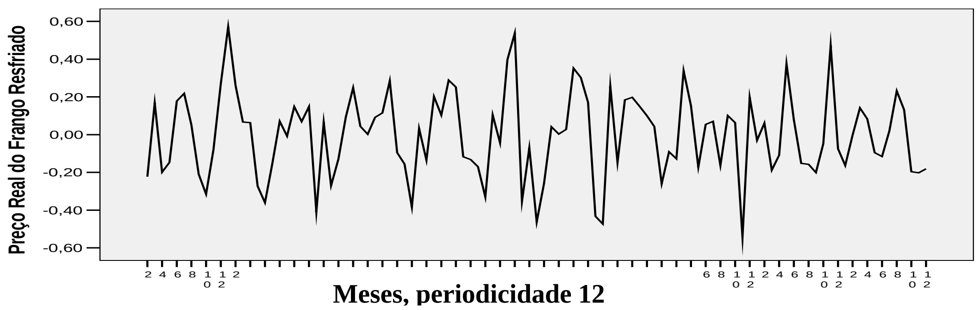

Figura 5 – Gráfico seqüencial da série diferenciada do preço real do frango

Com as diferenciações, a série se mostra sem tendência, Figura 5, e estabilizada em torno de sua média de valor zero, o que caracteriza um modelo sem constante. A análise do correlograma, Figura 6 fornece outras informações sobre a série diferenciada.

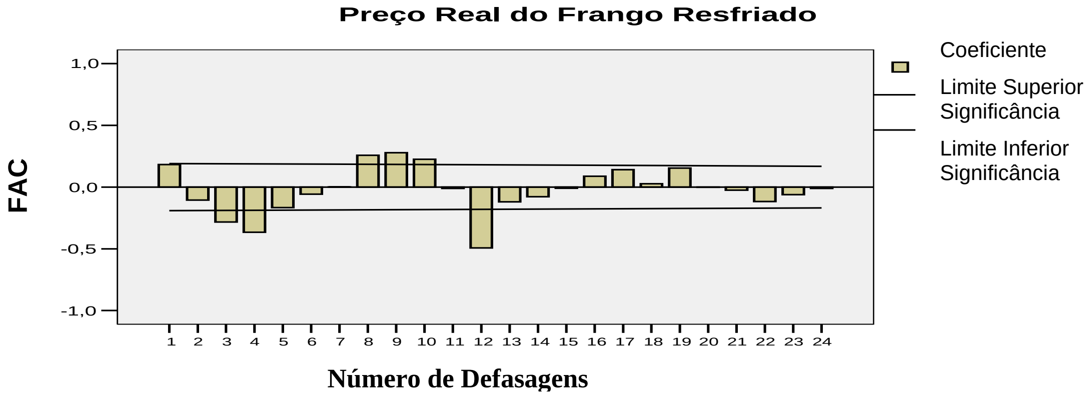

Figura 6 – Correlograma amostral da série diferenciada do preço real do frango

Com a análise do correlograma, Figura 6, não fica claro se a série diferenciada, seqüencial e sazonal resultou estacionária. Alguns coeficientes se mostraram não significativos (3, 4, 8, 9, 10, 12) e somente após a décima segunda defasagem convergiram para o limite de significância. A literatura de referência sugere que apenas alguns poucos coeficientes podem se mostrar não significativos para indicar estacionariedade, se em muitas defasagens ainda se encontram fora da faixa de significância, deve-se aplicar novas diferenças até torná-los significativos. Na segunda diferenciação seqüencial o correlograma, Figura 7, mostrou um bom ajustamento para a série, com apenas as defasagens 1, 12 e 13 não significativas. Contudo, os testes empíricos dos modelos testados com duas diferenciações seqüenciais se mostraram inadequados estatisticamente, em termos de *testes t* para todos os parâmetros, embora alguns modelos apresentarem boas previsões.

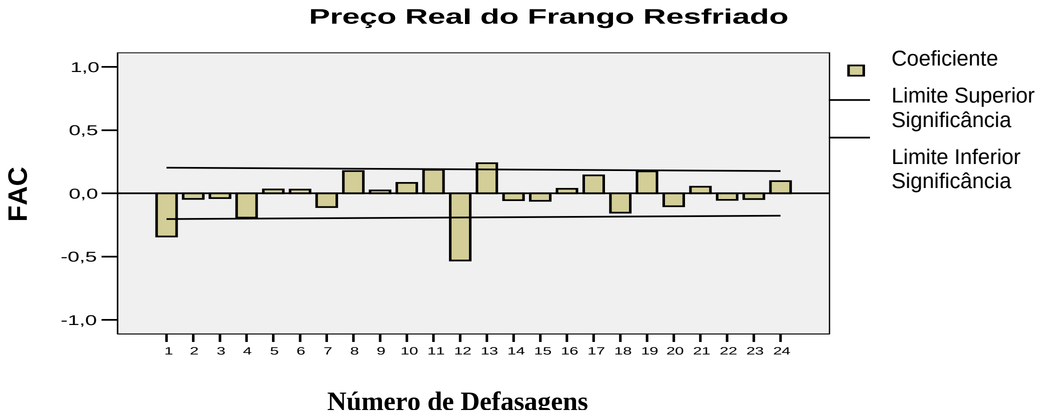

Figura 7 – Correlograma amostral da série com duas diferenciações seqüenciais e uma sazonal do preço real do frango

Além da análise da função autocorrelação, utilizar-se-á, para verificação da estacionariedade da série, o *teste de Dickey-Fuller (DF), de Raiz Unitária*. Para este teste, roda-se a regressão (5) $\Delta Y_t = \gamma Y_{t-1} + u_t$ com a série original diferenciada. Nas obras consultadas não se encontrou menção sobre o teste *DF* em séries também diferenciadas sazonalmente, sendo assim, se testará a série das duas formas, com apenas a diferença seqüencial e com a diferença seqüencial e sazonal.

- Teste 1 - Série diferenciada seqüencial e sazonalmente: $\Delta\Delta Y_t = -0{,}816\Delta Y_{t-1}$ (tau) (-8,495)
- Teste 2 - Série diferenciada apenas sequencialmente: $\Delta\Delta Y_t = -0{,}875\Delta Y_{t-1}$ (tau) (-9,522)

Com base na superioridade em módulo da estatística *tau* em relação a seus valores críticos, Quadro 1, (-2,56 para um nível de significância de 1%), rejeita-se para ambos os testes, a hipótese nula de γ = 0, série não estacionária, em favor da alternativa γ < 0, série estacionária. Após o teste de Dickey-Fuller, têm-se maiores evidências de que a série em estudo é estacionária de primeira ordem I(1), podendo, ser-lhe aplicado o modelo *ARIMA*.

Após a identificação das ordens de diferenciação seqüencial e sazonal, tem-se um modelo com a seguinte característica *ARIMA(p,1,q)(P,1,Q)*. O correlograma da Figura 6 pode auxiliar na identificação dos parâmetros AR e MA. A forma como se apresenta o correlograma da série diferenciada, com picos alternados, positivos e negativos, semelhante a uma curva senoidal, sugere um processo misto, auto-regressivo e de médias móveis, porém a determinação do valor do parâmetro é de difícil verificação.

Com os testes até agora realizados, identificou-se um modelo inicial básico com a seguinte configuração, ARIMA(1,1,1)(0,1,0). Iniciando com a configuração básica, foi se adicionando parâmetros, testando significâncias e desempenho preditivo *ex post*, trimestral e anual para 2004. Utilizou-se a mesma sistemática com previsões *ex post* para 2005, incluindo neste caso, os dados de 2004. Este procedimento, resultou em 4 modelos que melhor se ajustaram à série, pelo desempenho e grau de significância dos parâmetros. Para medir a performance preditiva dos modelos, utilizou-se os seguintes índices: *EPAM* (14), *REQM* (11) e U de *Theil* (12). Os quatro melhores modelos resultantes, com os respectivos índices de performance, estão resumidos no Quadro 2. A transformação logarítmica da série não foi adotada por não apresentarem melhoria de performance dos modelos.

Em geral todos os quatro modelos tiveram bom desempenho preditivo. As previsões para o ano de 2004, com um EPAM em torno de 4%, se mostraram superiores às de 2005, com um EPAM em torno de 5% ou 6%. O resultado inferior para 2005, explica-se pelos erros de previsão no último trimestre, que sazonalmente são meses de maior consumo e tendência de aumento. Porém, 2005 foi um ano atípico, com queda nos preços do último trimestre e nenhum dos modelos captou tal mudança. E nem poderiam, já, que em nenhum período da amostra houve queda sistemática no último trimestre do ano. Esta queda atípica para o último trimestre de 2005, foi resultado dos focos de febre aftosa no rebanho bovino no sul do país, que provocou uma retração em todo o mercado exportador de carnes, inclusive de frango. Com a retração no mercado externo, a produção foi direcionada para o mercado interno, provocando queda de preço por excesso de oferta. Nas Figuras 8 e 9 estão as representações gráficas do Modelo 2 para os anos de 2004 e 2005.

ARIMA	**MOD 1 (2,1,1)(0,1,1)**	**MOD 2 (2,1,1)(1,1,1)**	**MOD 3 (2,1,2)(0,1,1)**	**MOD 4 (2,1,2)(1,1,1)**
Previsão Anual – 2004				
EPAM	4,13%	3,93%	4,63%	3,99%
REQM	0,0937	0,0884	0,1095	0,0952
U	0,0240	0,0227	0,0278	0,0243
Previsão Anual – 2005				
EPAM	5,62%	5,42%	5,67%	6,01%
REQM	0,1354	0,1299	0,1344	0,1390
U	0,0367	0,0353	0,0365	0,0377

Quadro 2 – Resultados dos melhores modelos ARIMA testados.

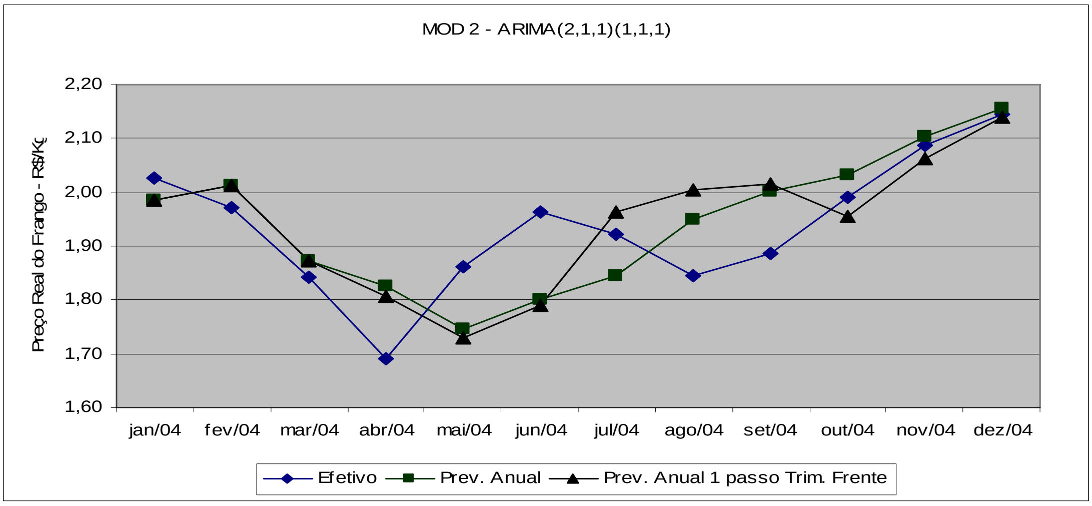

Figura 8 – Gráfico das previsões do preço real do frango para 2004, Mod. 2

Analisando a Figura 8, nota-se que as previsões apresentaram um bom ajustamento aos dados. Esperava-se um melhor desempenho das previsões trimestrais um passo a frente, já que são incluídos os dados do trimestre anterior para o cálculo do próximo trimestre. Assim, o modelo poderia captar as tendências de curto prazo daquele ano.

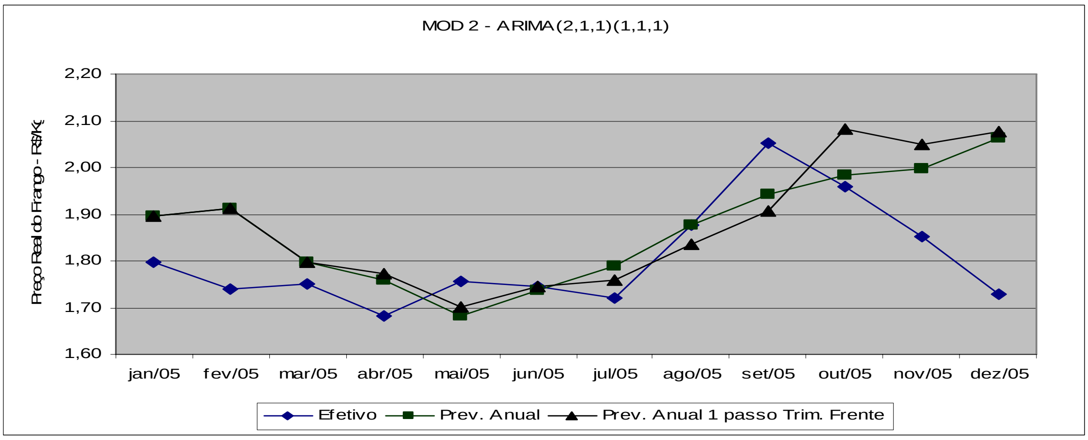

Figura 9 – Gráfico das previsões do preço real do frango para 2005, Mod. 2

As previsões para o ano de 2005 com o modelo 2, Figura 9, apresentaram um excelente ajustamento exceto no último trimestre, quando o modelo não captou a queda dos preços. Também, para este ano, o desempenho da previsão trimestral não se mostrou superior às previsões de prazos mais longos.

Como teste de capacidade preditiva de longo prazo para os modelos, foram realizadas previsões diretas para os anos de 2004 e 2005 com base nos dados de 1996 até 2003. (Quadro 3 e Figura 10).

ARIMA	MOD 1 (2,1,1)(0,1,1)	MOD 2 (2,1,1)(1,1,1)	MOD 3 (2,1,2)(0,1,1)	MOD 4 (2,1,2)(1,1,1)
EPAM	4,97%	5,00%	6,01%	5,67%

Quadro 3 – Resultados dos melhores modelos ARIMA testados, previsões para 2 anos

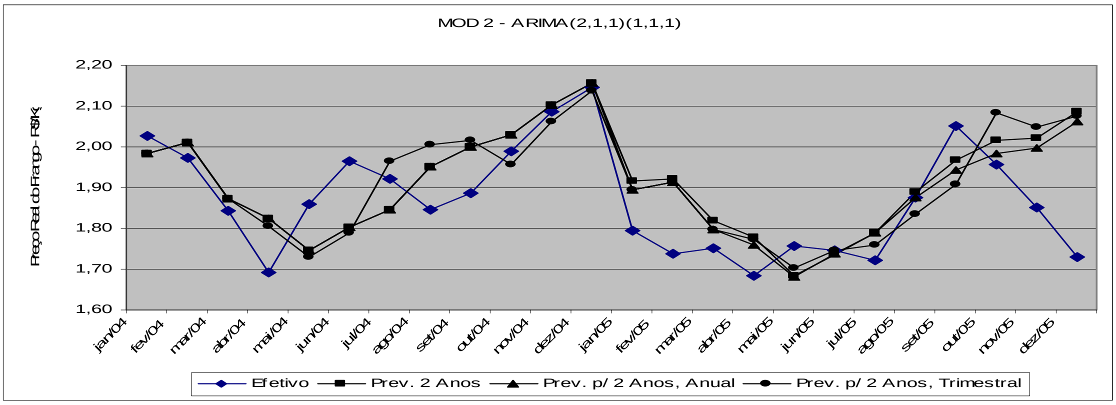

Figura 10 – Gráfico das previsões do preço real do frango para 2004 e 2005, Mod. 2

Como mostram o Quadro 3 e a Figura 10, os modelos tiveram um ótimo desempenho preditivo, as previsões para 2 anos à frente foram equivalentes às para um ano à frente e até para um trimestre à frente. Estes resultados contrariam a literatura referenciada, para qual os modelos ARIMA têm bom desempenho apenas em previsões de curto prazo.

5 Conclusões

Na série de testes empíricos realizados, a busca foi por configurações de modelos que apresentassem bom desempenho preditivo e parâmetros consistentes em suas significâncias estatísticas. Modelos com parâmetros não significativos até podem gerar boas previsões para um período específico, mas quando testados para outros períodos não mostram bom desempenho. Porém nesta pesquisa, constatou-se que alguns modelos mesmo com alguns parâmetros não significativos tiveram boa performance preditiva em todos os períodos testados. Observou-se também que alguns modelos quando são testados nos seus parâmetros em um período da amostra, apresentam significância em todos os parâmetros e quando são testados em outros períodos alguns parâmetros deixam de apresentar significância..

Foram realizadas projeções trimestrais, anuais e bienais. Constatou-se um bom desempenho para todas as periodicidades, contrariando a literatura de referência que atesta bom desempenho preditivo ao modelo ARIMA para apenas previsões de curto prazo. Com esta pesquisa percebe-se que a metodologia ARIMA, para determinadas séries temporais, pode ser utilizada para previsões de médio e longo prazo. Para tanto, necessita-se uma determinação ideal ou calibragem adequada dos parâmetros do modelo e também a realização de vários testes de previsão *ex post* para verificar de sua adequação.

Em concordância com a literatura de referência, observou-se dificuldade na determinação dos melhores modelos, principalmente quando a série se ajusta a modelos com várias defasagens nos parâmetros. Para identificar a ordem de Integração (Diferenciação) do modelo, o parâmetro d, há testes claros e facilmente aplicáveis, mas para a Identificação do parâmetro de integração sazonal D os testes não se mostram tão claros. Porém, logo após as primeiras simulações de predição, torna-se clara a sua necessidade, se for o caso. Nos testes realizados neste trabalho, quando sem a presença do parâmetro D, os modelos apresentavam resultados muito insatisfatórios. Para a determinação dos parâmetros Autorregressivos p, P, e os de Médias Moveis, q e Q, os testes de identificação com o uso do correlograma se mostraram não conclusivos, conduzindo a indecisões na sua interpretação. Observou-se que a identificação destes últimos parâmetros requer experiência e esforço tentativo do usuário na busca do melhor modelo.

Pelos bons desempenhos encontrados, constata-se, grandes possibilidades no uso de modelos de séries temporais no meio empresarial, fornecendo subsídio ferramental ao processo de tomada de decisão gerencial e estratégica. Previsões apoiadas em um modelo, podem facilitar o processo, detectando o comportamento passado da série e extrapolando-o para o futuro, cabendo ao gestor a crítica da adequação dos resultados e acréscimos pessoais advindos de informações conjunturais ou internas da própria empresa, que naturalmente, nenhum modelo estatístico será capaz de absorver.

Como sugestões para futuras pesquisas, propõe-se a aplicação desta metodologia em outras séries econômicas para avaliar o seu grau de adaptação. Estudos podem ser realizados com outras metodologias de previsão de séries temporais univariadas ou não. Um complemento a este trabalho seria a inclusão de variáveis explicativas ao modelo aqui estudado para verificar se há ganhos significativos de predição.

Referências

BACCHI, Mirian R. P.; Hoffmann, Rodolfo. Previsão de preços de bovino e frango com modelos de séries temporais. **Revista de Economia e Sociologia Rural**, Brasília, v.33, n. 4, p. 9-28, out./dez. 1995.

FAVA, Vera Lúcia. Análise de séries de tempo. In: VASCONSELOS, M. A. S.; ALVES, Denisard (Org.). **Manual de Econometria:** nível intermediário. São Paulo: Atlas, 2000a. p. 199-203.

______. Metodologia de Box-Jenkins para modelos univariados. In: VASCONSELOS, M. A. S.; ALVES, Denisard (Org.). **Manual de Econometria:** nível intermediário. São Paulo:

Atlas, 2000b. p. 205-231.

______. Teste de raízes unitárias e co-integração. In: VASCONSELOS, M. A. S.; ALVES, Denisard (Org.). **Manual de Econometria:** nível intermediário. São Paulo: Atlas, 2000c. p. 233-243.

______. Modelos de função de transferência e de análise de intervenção. In: VASCONSELOS, M. A. S.; ALVES, Denisard (Org.). **Manual de Econometria:** nível intermediário. São Paulo: Atlas, 2000d. p. 244-252.

FGVDADOS Informação Econômica Online. <http://fgvdados.fgv.br/index.htm>

GLUCKSTERN, Marco César. Aplicação do modelo Hull-White a precificação de opções sobre IDI. **Invest Sul**, junho 2003. Disponível em: <http://www.investsul.com.br/textos_academicos.asp> . Acesso em: 10 fev. 2006.

GUJARATI, Damodar N.. **Econometria Básica**. São Paulo: Makron Books, 2000.

HILL, R. Carter; GRIFFITHS, William E.; JUDGE, George G. **Econometria**. 2ª. Ed. São

Paulo: Saraiva, 2003.

JOX Assessoria Agropecuária. <http://www.jox.com.br/>

MADDALA, G.S.. **Introdução à econometria**. 3ª. Ed. Rio de Janeiro: LTC, 2003.

MORETTIN, Pedro A.; TOLOI, Clélia M. C.. **Análise de séries temporais**. São Paulo:

Edgard Blücher, 2004.

RAUPP, Fabiano Maury; BEUREN, Ilse Maria. Metodologia da pesquisa aplicada às ciências sociais. In: BEUREN, Ilse Maria (Org.). **Como elaborar trabalhos monográficos em contabilidade:** teoria e prática. 2ª. ed. São Paulo: Atlas, 2004. p. 76-97.

PINDYCK, Robert S.; RUBINFELD, Daniel L.. **Econometria**: modelos & previsões. Rio de Janeiro: Elsevier, 2004.