

# TRANSMISSÃO DE PREÇOS NO MERCADO INTERNACIONAL DA SOJA: UMA ABORDAGEM PELOS MODELOS ARMAX E VAR

Oswaldo Cândido da Silva Filho  
Mestrando do Programa de Pós-Graduação em Economia  
Pesquisador do Núcleo de Economia Aplicada - UFPB

Bruno Ferreira Frascaroli  
Mestrando do Programa de Pós-Graduação em Economia  
Pesquisador do Núcleo de Economia Aplicada - UFPB

Sinézio Fernandes Maia  
Doutor em Economia pela Universidade Federal de Pernambuco - PIMES  
Professor do Programa de Pós-Graduação em Economia  
Curso de Mestrado em Economia – CCSA UFPB  
Pesquisador do Núcleo de Economia Aplicada da UFPB

## RESUMO

Este trabalho analisa o mercado internacional de soja através das relações entre os preços. Estimou-se um modelo VAR avaliando a transmissão do preço da *commodity* da soja americana para a soja em grão doméstica e seus derivados: óleo e farelo de soja. Adicionalmente estimou-se um ARMAX para captar a transmissão entre os preços domésticos da soja para seus derivados. Foram realizados testes de raiz unitária, causalidade de *GRANGER*, testes de cointegração entre os preços, função de impulso resposta e a decomposição da variância dos resíduos para se verificar o grau de influência das relações propostas. Tanto os modelos VAR como ARMAX parecem se ajustar bem à explicação da dinâmica dos preços da soja. As significâncias estatísticas dos parâmetros das estimações do modelo ARMAX confirmam a transmissão dos preços da soja em grão doméstica para seus derivados. Do ponto de vista do modelo VAR, os preços da soja e do farelo domésticos se alinham muito bem com o preço da *commodity* de soja em grão americana, já o preço do óleo de soja, possivelmente por estar integrado numa rede mais ampla de mercados correlacionados tanto pra trás quanto pra frente, apresenta características distintas dos demais segmentos do complexo soja.

**Palavras-chave:** Soja, *Commodities*, Função de Transferência, Vetores Auto-Regressivos.

## ABSTRACT

This work focuses the international market of soybean through the price relationships kept by these. Was chosen two models of time series: ARMAX and VAR. These models have a good performance in explaining simultaneity relationships among phenomena because they are multivariate models. The justification for use this instrumentals comes from the own concentration of the international market, where, in agreement with previous works, it is taken individually into account the market slice corresponding to each producing country. Was considered each model evaluating the parameters of transmission of prices among the price of commodity of american soybean for each category of product of domestic market: the domestic soybean and yours derived acted by the soybean oil and the soybean bran. The estimates like indicate a positive correlation among the prices in different magnitude to each individual market. Tests of unitary root were accomplished, of causality of *GRANGER* and cointegration tests in order to verify if there is a long run relationship among the variables. The response impulse function and the decomposition of residues variance were also estimated, for the verification of price of commodity of american soybean influence degree for each category of product of the domestic market.

**Key-words:** Soybean, *Commodities*, Transfer Function, Vectorial Autoregressive Model.

Código JEL: G15, G28, Q10

## 1. INTRODUÇÃO

Dentro de infinitos estudos econômicos, o interesse em analisar a dinâmica do mercado mundial da soja sob a ótica dos modelos ARMAX e VAR, originou-se de dois principais motivos. O primeiro motivo seriam fatos e características inerentes e específicas do mercado mundial da soja, e, o segundo motivo seria o poder e a eficiência em descrever relações econômicas dos modelos de estimação escolhidos.

Os caminhos desta análise não deixam de estarem ligados a fatores onde o pesquisador faz escolhas fundamentais, como relação de causalidade entre as variáveis analisadas ou restrições *a priori* sobre a mecânica deste funcionamento que podem levar ao melhor esclarecimento do problema ou investigação proposta ou podem obscurecer a análise.

Desta maneira, nos deparamos com preocupações de forma que, acumulando-se menos hipóteses, podemos descrever melhor o objeto de análise escolhido. Contudo, isto é uma forma de se atentar para a utilização apropriada do instrumental econométrico sempre em acordo com o suporte dado pelos modelos teóricos econômicos.

Um dos objetivos do presente trabalho foi analisar do mercado brasileiro de soja e seus derivados a partir de um dos motivos que levam às flutuações destes: o preço da *commodity* americana de soja em grão. Para isso, utilizou-se um modelo de vetores auto-regressivos (VAR). O outro objetivo se concentra na utilização de estimações de funções de transferência (ARMAX) para captar a transmissão do preço da soja em grão doméstica para o preço de seus derivados, o preço do óleo e do farelo de soja. Dividem-se os produtores internacionais em tomadores de preço e formadores do preço internacional da soja. Mais especificamente, construímos, identificamos e escolhemos corretamente cada modelo; Estimamos e avaliamos os parâmetros da transmissão dos preços que compõem este mercado e, desta forma, buscamos explicar o funcionamento desta dinâmica.

### 1.1. A Soja

Em meados da década de 1960, a soja chega ao Brasil, mais precisamente a região de Santa Rosa no Rio Grande do Sul. Com a chamada Revolução Verde<sup>1</sup>, o governo brasileiro muda a política de apoio à agricultura familiar ou de subsistência adotando os moldes das políticas praticadas internacionalmente de apoio à exportação.

Dessa forma, o Rio Grande do Sul passa a ser o estado pioneiro na produção da soja para exportação. Seguindo na mesma linha, ente a década de 1960 e 1970, outros estados brasileiros com características distintas de clima e solo passam a cultivar a soja.

Até a segunda metade da década de 60, o mercado de soja, que abrange grão, farelo e óleo, foi amplamente dominado pelos Estados Unidos, que produziam mais de 80% da soja mundial. Com o aumento das cotações internacionais, países como Brasil e Argentina passaram a exercer um importante papel nas exportações mundiais. Já década de 1970, fica consolidado o aumento extraordinário da produção da soja no Brasil acompanhada por aumentos de produtividade significantes.

Segundo Brum (2004), em uma década, o Brasil eleva a produção de 206.000 toneladas de soja sobre uma área 171.500 hectares para 1,5 milhões de toneladas sobre uma área cultivada de pouco mais de 1,3 milhões de hectares.

Várias características, sejam no âmbito sócio-político interno ou macroeconômico internacional, fizeram com que, na década de 1980, as altas taxas de desenvolvimento do mercado da soja fossem diminuindo acompanhadas por aumento de incerteza e risco associado à atividade.

Embora seja necessário levar em consideração na década de 1980 que, mesmo com a economia brasileira e a produção de soja doméstica experimentando baixas taxas de crescimento, o mercado financeiro globalmente se desenvolveu muito diminuindo deficiências de mercado no tocante a contratos.

Marcada por profundas mudanças político-sociais, abertura econômica e estabilização econômica, a década de 1990 foi significativa em evoluções no mercado da soja. Neste período, a produção nacional

---

<sup>1</sup> Refere-se à invenção de novas práticas agrícolas que permitiram um vasto aumento na produção agrícola em países menos desenvolvidos durante as décadas de 60 e 70.

de soja retoma o crescimento em linha com avanços tecnológicos importantes, iniciando pelo plantio direto, muito bem aceito pelos produtores. A produção nacional salta de 20,4 milhões de toneladas em 1990 para 31,4 milhões de toneladas em 1999.

Nos anos subseqüentes, esta tendência de crescimento da cultura continua acompanhada por desenvolvimentos de novas tecnologias que permitiram o plantio da soja com qualidade na região centro-oeste, norte e nordeste num movimento de expansão da cultura.

Atualmente, a principal *commodity* de exportação brasileira é a soja, sendo o Brasil é o segundo país individual em termos de volume produzido de soja no mundo ficando atrás somente dos EUA.

*De acordo com dados de USDA (2003), 79.87% da produção mundial total no período entre 1994/1995 e 2002/2003 aconteceram em três países, Estados Unidos (44.85%), Brasil (21.71%) e Argentina (13.30%). Estes três países também eram os maiores os exportadores conduzindo o mercador exportador mundial, contabilizando juntos cerca de 90.50% dos 44.32 milhões de toneladas no mesmo período. Os Estados Unidos eram o exportador principal com 56.85% do total, seguido pelo Brasil (24.10%) e Argentina (9.54%) (Margarido, Turolla & Bueno – 2004).*

## **1.2. Os Mercados Futuros**

Os mercados de contratos futuros existem a mais de 150 anos, e foi nas negociações da *Board of Trade of the City of Chicago* (CBOT) que o mercado ganhou volume e, nos últimos dez anos, explodiu em crescimento. Atualmente, há mais de cinquenta tipos de *commodities* agrícolas diferentes e produtos que são comercializados, incluindo soja em grão, milho, trigo, algodão, gado de corte, café, cacau, açúcar, suco de laranja concentrado congelado etc.

Existem três mercados independentes e, ao mesmo tempo relacionados, para estes produtos além de muitos outros produtos agrícolas. Os três mercados são: mercado à vista (*spot market*), o mercado para contratos futuros, e o mercado para contratos de opções. Os mercados à vista são mercados onde *commodities* e produtos mudam de mãos em transações financeiras. Os compradores e vendedores vão a um tipo mercado organizado, reservadamente frente a frente, através de telefone, fax, ou mesmo usando a internet, negociar um preço. A maioria das transações no complexo da soja, em outras *commodities* agrícolas ou *commodities* de produtos não-agrícolas acontecem no mercado à vista.

Características básicas dos contratos futuros elaboradas pela *Futures Industry Association* (FIA).

- Os contratos de futuros são altamente alavancados;
- O investimento inicial é pequeno em relação ao compromisso;
- A duração de um contrato de futuro é sempre limitada;
- No mercado de futuro utilizam-se limites de preços e de posições;
- Os investidores e os corretores precisam estar mais atentos nas suas operações.

O real crescimento nos mercados futuros foram das *commodities* não-agrícolas como petróleo, metais e os contratos financeiros. Segundo Woolverton (2003), em um recente período de cerca de dez anos, dos contratos futuros mais ativamente comercializados, só três eram contratos agrícolas. Os contratos de soja em grão se classificaram em sétimo lugar em volume de transações. Vinte anos atrás, os Estados Unidos dominaram os mercados futuros, centrados nos dois maiores e principais mercados de Chicago. Enquanto mais da metade do volume de contratos ainda são comercializados nos EUA, dos dez maiores mercados de futuros no mundo, cinco estão fora dos EUA.

Castro e Rossi Júnior (2000), observam que como o grão e o farelo da soja são utilizados em países desenvolvidos na pecuária como ração para animais, eles têm seu mercado caracterizado por uma demanda mais estável no tempo. Diferentemente, o óleo de soja apresenta demanda com maior instabilidade, pois têm como principais consumidores no mercado internacional as nações subdesenvolvidas que o utilizam como fonte energética.

## 2. REFERENCIAL TEÓRICO

Freqüentemente, os fenômenos sejam naturais ou econômicos, acontecem simultaneamente, logo uma metodologia mais condizente com esta ótica, leva em consideração observações em duas ou mais séries de tempo. Dessa maneira, análises multivariadas podem ser úteis para desenvolver um modelo de análise para descrever a inter-relação entre as séries.

O modelo econômico a ser utilizado no presente trabalho é um simples modelo clássico de comércio entre dois países ou duas regiões, que envolve a transação de um produto homogêneo, portanto substituto perfeito, feita com uma moeda em comum, respeitando hipóteses e axiomas microeconômicos sobre a maneira e o mercado no qual este produto é transacionado, como quantidade e simetria de informações, por exemplo. Numa economia fechada, o preço e a quantidade de equilíbrio de um determinado produto são determinados pelo equilíbrio no mercado entre a demanda e a oferta.

Desta maneira podemos representar da seguinte forma matemática este mercado:

$$q^d = \alpha_0 + \alpha_1 p + \alpha_2 y \quad \alpha_1 < 0, \alpha_2 > 0 \quad (1)$$

$$q^s = \beta_0 + \beta_1 p \quad \beta_1 > 0 \quad (2)$$

Onde, a equação (1) representa o lado da demanda, e a equação (2) representa o lado da oferta. Onde  $q^d$  é a quantidade demandada,  $q^s$  é a quantidade ofertada,  $p$  é o preço e  $y$  é a renda.

Igualando as equações (1) e (2) e resolvendo para encontrar o preço de equilíbrio  $\bar{p}$ , e a quantidade de equilíbrio  $\bar{q}$ , temos:

$$\bar{p} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2}{\alpha_1 - \beta_1} y \quad (3)$$

$$\bar{q} = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} - \frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1} y \quad (4)$$

Tomando duas regiões e seus dois conjuntos de demanda e oferta correspondentes:

$$q^d_1 = \alpha_0 + \alpha_1 p_1 + \alpha_2 y_1 \quad (5)$$

$$q^s_1 = \beta_0 + \beta_1 p_1 \quad (6)$$

e

$$q^d_2 = \lambda_0 + \lambda_1 p_2 + \lambda_2 y_2 \quad (7)$$

$$q^s_2 = \pi_0 + \pi_1 p_2 \quad (8)$$

Com economias fechadas, isto é, com cada região quando isolada, cada uma delas encontrará o preço e sua quantidade de equilíbrio. Quando essas regiões são integradas por uma relação de comércio, a de menor preço  $p_1$  tenderá a vender para a de maior preço  $p_2$  se:

$$\bar{p}_1 > \bar{p}_2 + k \quad (9)$$

Onde,

$\bar{p}_1$  e  $\bar{p}_2$  são, respectivamente, os preços de equilíbrios de cada região;

$k = \ln(1 + K)$ , em que  $K$  é um parâmetro qualquer de diferencial de preços de curto prazo.

Então, quando o equilíbrio se estabelece entre as duas regiões obtêm-se:

$$p_2 = p_1 + k \quad (10)$$

## 3. PROCEDIMENTOS METODOLÓGICOS

Para se ilustrar a dificuldade de se expressar modelos com regressores endógenos, considere o exemplo tomado em Hamilton (1994). Dado  $p_t$  como o logaritmo do preço de um bem em um dado ano qualquer e  $q_t^d$  o logaritmo da quantidade deste bem que o público está disposto a comprar. Para simplificar, suponha uma versão estocástica mais simples do modelo apresentado na seção de referencial

teórico, e além do mais, que a covariância entre preço e quantidade são estacionárias no tempo e possuem ruído-branco.

A curva de demanda é dada por:

$$q_t^d = \beta P_t + \varepsilon_t^d \quad (11)$$

Com  $\beta < 0$ .

O termo  $\varepsilon_t^d$  neste modelo representa outros fatores que influenciam a demanda diferentes do preço. Além disso, ele é independente e identicamente distribuído com média zero e variância  $\sigma_d^2$ .

A oferta do bem é dada por:

$$q_t^s = \gamma P_t + \varepsilon_t^s \quad (12)$$

Com  $\gamma > 0$ .

Analogamente,  $\varepsilon_t^s$  representa outros fatores que influenciam a oferta que não sejam o preço. Além disso, ele é também independente e identicamente distribuído com média zero e variância  $\sigma_s^2$ . Adicionalmente, tem-se que  $\varepsilon_t^s$  e  $\varepsilon_t^d$  não são correlacionados.

Igualando as equações (11) e (12) com o propósito de se encontrar o equilíbrio no mercado do bem, obtemos:

$$\beta P_t + \varepsilon_t^d = \gamma P_t + \varepsilon_t^s \quad (13)$$

Rearranjando os termos de (13):

$$P_t = \frac{\varepsilon_t^d - \varepsilon_t^s}{\gamma - \beta} \quad (14)$$

Substituindo (14) em (12):

$$q_t = \gamma \frac{\varepsilon_t^d - \varepsilon_t^s}{\gamma - \beta} + \varepsilon_t^s = \frac{\gamma}{\gamma - \beta} \varepsilon_t^d - \frac{\beta}{\gamma - \beta} \varepsilon_t^s \quad (15)$$

Considere a consequência de se estimar (11) pelo método MQO. A regressão será:

$$b_t = \frac{\left(\frac{1}{T}\right) \sum_{t=1}^T P_t q_t}{\left(\frac{1}{T}\right) \sum_{t=1}^T P_t^2} \quad (16)$$

Substituindo (14) e (15) em (16):

$$\begin{aligned} \left(\frac{1}{T}\right) \sum_{t=1}^T P_t q_t &= \left(\frac{1}{T}\right) \sum_{t=1}^T \left[ \frac{1}{\gamma - \beta} \varepsilon_t^d - \frac{1}{\gamma - \beta} \varepsilon_t^s \right] \left[ \frac{\gamma}{\gamma - \beta} \varepsilon_t^d - \frac{\beta}{\gamma - \beta} \varepsilon_t^s \right] \\ \left(\frac{1}{T}\right) \sum_{t=1}^T P_t q_t &= \left(\frac{1}{T}\right) \sum_{t=1}^T \left[ \frac{\gamma}{(\gamma - \beta)^2} (\varepsilon_t^d)^2 + \frac{\beta}{(\gamma - \beta)^2} (\varepsilon_t^s)^2 - \frac{\gamma + \beta}{(\gamma - \beta)^2} \varepsilon_t^d \varepsilon_t^s \right] \\ &\xrightarrow{p} \frac{\gamma \sigma_d^2 + \beta \sigma_s^2}{(\gamma - \beta)^2} \end{aligned} \quad (17)$$

Onde a notação  $\xrightarrow{p}$  quer dizer que a série converge em probabilidade, logo o efeito de

$\frac{\gamma + \beta}{(\gamma - \beta)^2} \varepsilon_t^d \varepsilon_t^s$  pode ser desprezível.

Desenvolvendo os cálculos para o denominador tem-se:

$$\left(\frac{1}{T}\right) \sum_{t=1}^T P_t^2 = \left(\frac{1}{T}\right) \sum_{t=1}^T \left[ \frac{1}{\gamma - \beta} \varepsilon_t^d - \frac{1}{\gamma - \beta} \varepsilon_t^s \right]^2 \xrightarrow{p} \frac{\sigma_d^2 + \sigma_s^2}{(\gamma - \beta)^2}$$

A equação (16) converge em probabilidade para:

$$b_i \xrightarrow{P} \left[ \frac{\gamma\sigma_d^2 + \beta\sigma_s^2}{(\gamma - \beta)^2} \right] \left[ \frac{\sigma_d^2 + \sigma_s^2}{(\gamma - \beta)^2} \right]^{-1} = \frac{\gamma\sigma_d^2 + \beta\sigma_s^2}{\sigma_d^2 + \sigma_s^2} \quad (18)$$

O resultado do estimador  $\beta$  da estimação dado por (18) não é a elasticidade preço-demanda do bem, mas sim uma média de  $\beta$  e a elasticidade da oferta  $\gamma$  com estimações dependendo do tamanho das variâncias  $\sigma_d^2$  e  $\sigma_s^2$ .

Observe a hipótese da consistência do estimador (18). Se  $\sigma_d^2 \rightarrow 0$  ou  $\sigma_s^2 \rightarrow \infty$ , obtém-se um estimador consistente da elasticidade preço-demanda  $\beta$ . Na outra mão, se  $\sigma_d^2 \rightarrow \infty$  ou  $\sigma_s^2 \rightarrow 0$ , obtém-se um estimador consistente da elasticidade da oferta  $\gamma$ . Este fenômeno é conhecido na literatura como viés de equações simultâneas. Logo, os parâmetros não poderiam ser estimados por MQO, pois há violação de pressupostos tornando o estimador não-confiável.

Para se solucionar o problema descrito acima da forma de estimação do modelo por MQO são feitas restrições. Estas restrições consistem em tornar o sistema de variáveis estimável por MQO.

Por uma manipulação algébrica pode-se chegar à forma reduzida do sistema e obter um modelo VAR. No entanto, como esta análise pode ser feita somente com os preços que compõem este mercado sem que haja perda de poder na metodologia proposta, apenas utilizaremos a relação entre os vetores de preços (equação (10)) que compõem o mercado internacional da soja e a relação com o mercado doméstico, bem como as relações internas de preços que compõem este último.

A questão importante aqui, que foi discutida por Sims (1980), passa a ser as questões que envolvem a identificação do modelo. A partir da forma reduzida é desejável que possa se retornar à forma primitiva sem perder parâmetros. Podem ser obtidos modelos:

- Exatamente identificados, onde o número de parâmetros recuperados da forma reduzida é igual ao número de parâmetros da forma primitiva;
- Subidentificados, onde o número de parâmetros recuperados é menor que o número de parâmetros da forma primitiva, isto é, há mais incógnitas que equações (logo, infinitas soluções);
- Sobreidentificados, onde o número de parâmetros recuperados é maior que o número de parâmetros da forma primitiva, isto é, o número de equações é maior que o de incógnitas.

A forma prática de se resolver problemas de identificação é adicionar variáveis predeterminadas em algumas equações.

Sims (Op. Cit.), ressalta a importância do tratamento padronizado e simultâneo das séries, ponderando mais pela constatação empírica da teoria e ponderando menos por hipóteses teóricas do tipo *ad hoc*.

Para se resolver problemas de identificação do modelo, é possível a adoção de restrições ao modelo, como pelo exemplo utilizado por Hamilton (1994), igualar  $\beta$  a zero. Isso significa dizer que  $P_t$  não tem efeito contemporâneo sobre  $q_t^d$  e a curva de demanda por laranja é inelástica a  $P_t$  no mesmo período, mas  $P_t$  tem efeito contemporâneo sobre  $q_t^s$ . Choques dados por  $\varepsilon_t^s$  e  $\varepsilon_t^d$  têm efeito em  $q_t^s$ . Onde é de suma importância que restrições como esta seja devidamente acompanhada por teorias econômicas.

Para se que seja possível estimar os modelos de série de tempo propostos por este trabalho se faz necessário decompor uma série temporal, isto é, extrair dela os componentes sazonalidade, ciclo, tendência e irregularidade. Uma série pode ser representada de forma geral como:  $y_t = f(S_t, C_t, T_t, E_t)$ , onde  $S_t$  é a sazonalidade,  $C_t$  o ciclo,  $T_t$  a tendência e  $E_t$  a irregularidade.

Um processo é estritamente estacionário se a função densidade de probabilidade conjunta das observações  $y_1, y_2, \dots, y_n$  não muda com a origem do eixo do tempo.

Proposição: O processo é estacionário se  $\lim \psi(B) = |B|$ , em que  $|B| < 1$ , pode-se concluir que o processo está dentro do círculo unitário incluindo sua fronteira. E é inversível se  $\lim \pi(B) = (B)$ , onde  $|B| \leq 1$ .

De acordo com Enders (1995), um vetor de processos estocásticos é estacionário se e somente se:

a) Se todos os vetores aleatórios têm o mesmo vetor médio  $\boldsymbol{\varepsilon}$ ,  $E(\mathbf{y}_t) = E(\mathbf{y}_{t-s}) = \boldsymbol{\varepsilon}$ , para qualquer  $t$ ;

b) As variâncias de todas as variáveis aleatórias são finitas,  $\text{var}(\mathbf{y}_{mt}) < \infty$  para  $m=1, \dots, M$  e todo  $t$

$$E[(\mathbf{y}_t - \boldsymbol{\varepsilon})^2] = E[(\mathbf{y}_{t-s} - \boldsymbol{\varepsilon})^2] = \boldsymbol{\sigma}_y^2;$$

$$[\text{var}(\mathbf{y}_t) = \text{Var}(\mathbf{y}_{t-s}) = \boldsymbol{\sigma}_y^2]$$

c) A matriz de covariâncias dos vetores  $\mathbf{y}_t$  e  $\mathbf{y}_{t+1}$  possui  $s$  períodos distintos que não dependem de  $t$ , mas apenas de  $s$ .

$$E[(\mathbf{y}_t - \boldsymbol{\varepsilon})(\mathbf{y}_{t-s} - \boldsymbol{\varepsilon})] = E[(\mathbf{y}_{t-j} - \boldsymbol{\varepsilon})(\mathbf{y}_{t-j-s} - \boldsymbol{\varepsilon})] = \boldsymbol{\gamma}_s$$

$$\text{cov}[(\mathbf{y}_t, \mathbf{y}_{t-s})] = \text{cov}[(\mathbf{y}_{t-j}, \mathbf{y}_{t-j-s})]$$

Para se detectar a estacionariedade da série, isto é, se  $\mathbf{y}_t$  possui ou não raiz unitária devem-se fazer testes robustos, como o teste Dickey-Fuller<sup>2</sup>

### 3.1. Relação Causal

Granger (1969), citado por Enders (1995), introduz um conceito de causalidade desenvolvendo uma técnica que permite que sejam escolhidas as variáveis dependente e independente. Numa situação em que a variável  $x_t$  ajuda na previsão da variável  $y_t$  melhorando a estimação, denomina-se  $y_t$  causado-de-GANGER por  $x_t$ . O procedimento consiste em estimar ambas variáveis na posição de variável dependente, uma de cada vez como especificadas a seguir:

$$Y_t = \sum_{i=0}^{\infty} \phi_i Y_{t-i} + \sum_{i=1}^{\infty} \theta_i X_{t-i} + \boldsymbol{\varepsilon}_{yt} \quad (19)$$

$$X_t = \sum_{i=1}^{\infty} \delta_i X_{t-i} + \sum_{i=0}^{\infty} \rho_i Y_{t-i} + \boldsymbol{\varepsilon}_{xt} \quad (20)$$

Perceba que os testes são efetuados nos parâmetros defasados  $\theta_i$  e  $\rho_i$ , equações (19) e (20) respectivamente. Se os parâmetros  $\theta_i$  forem estatisticamente iguais a zero a variável  $x$  não causa  $y$  na equação (19). Analogamente, se os parâmetros  $\rho_i$  forem estatisticamente iguais a zero a variável  $y$  não causa  $x$  na equação (20).

A hipótese nula é que não causalidade no modelo em nenhum dos dois sentidos. O próximo passo após estimar ambas as regressões, é verificar o sentido da causalidade utilizando o teste  $F$ . As hipóteses podem ser descritas da seguinte forma:

$$\begin{cases} H_0: \theta_i = 0 \\ H_0: \rho_i = 0 \end{cases}$$

Se  $\theta_i$  e  $\rho_i$  forem nulos, não se rejeita as hipóteses nulas, logo a consequência imediata é que as variáveis com os valores defasados não influenciam o modelo. Obviamente, se as hipóteses nulas forem rejeitadas, têm-se uma relação de causalidade, e pode-se observar como os preços dos mercados se ajustam.

### 3.2. Estrutura do Modelo VAR

Para representar a relação dos vetores de preços, tomamos um exemplo de sistemas de equações apresentado por Enders (Op. Cit.) que será utilizado na estimação do modelo VAR. Este exemplo é adaptado a uma extensão natural da análise de transferência para o mercado de soja, o caso de duas

<sup>2</sup> Ver mais em DICKEY & FULLER (1979).

variáveis, nesse caso os preços da soja, onde o trajeto no tempo  $\{P_{1t}\}$  é afetado por  $\{P_{2t}\}$ , e a trajetória no tempo de  $\{P_{2t}\}$  é afetado por  $\{P_{1t}\}$ . Supõe-se que as seqüências  $\{P_{1t}\}$  e  $\{P_{2t}\}$  sejam estacionárias, isto é, seus resíduos são caracterizados por serem ruído-branco, com variância constante e não correlacionados:

$$P_{1t} = b_{10} - b_{12}P_{2t} + \gamma_{11}P_{1t-1} + \gamma_{22}P_{2t-1} + \varepsilon_{P_{1t}} \quad (21)$$

$$P_{2t} = b_{20} - b_{21}P_{1t} + \gamma_{21}P_{1t-1} + \gamma_{12}P_{2t-1} + \varepsilon_{P_{2t}} \quad (22)$$

Com um exercício de álgebra, podemos transformar o sistema de equações (21) e (22) num sistema de equações na forma reduzida estimável:

$$\begin{bmatrix} 1 & b_{12} \\ b_{12} & 1 \end{bmatrix} \begin{bmatrix} P_{1t} \\ P_{2t} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} P_{1t-1} \\ P_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{P_{1t}} \\ \varepsilon_{P_{2t}} \end{bmatrix} \quad (23)$$

ou

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t \quad (24)$$

Em que:

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{12} & 1 \end{bmatrix} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \quad x_t = \begin{bmatrix} P_{1t} \\ P_{2t} \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{P_{1t}} \\ \varepsilon_{P_{2t}} \end{bmatrix}$$

Multiplicando a equação (24) por  $B^{-1}$ , permite que encontremos o Vetor Auto-Regressivo em sua forma padrão:

$$x_t = A_0 + A_1 x_{t-1} + \varepsilon_t \quad (25)$$

onde

$$A_0 = B^{-1}\Gamma_0 \quad A_1 = B^{-1}\Gamma_1 \quad \varepsilon_t = B^{-1}\varepsilon_t$$

Usando a nova notação desenvolvida, podemos reescrever (21) e (22):

$$P_{1t} = a_{10} + a_{11}P_{1t-1} + a_{12}P_{2t-1} + \varepsilon_{1t} \quad (26)$$

$$P_{2t} = a_{20} + a_{21}P_{1t-1} + a_{22}P_{2t-1} + \varepsilon_{2t} \quad (27)$$

O sistema composto pelas equações (26) e (27) é chamado de VAR padrão. Onde seu principal problema é verificar se este sistema permite recuperar o sistema de equações (21) e (22), isto é, se é possível identificar o modelo com equações simultâneas.

Enders (Op. Cit.), chama a atenção para os erros  $\varepsilon_{1t}$  e  $\varepsilon_{2t}$ , note que o mesmo é decomposto por dois choques  $\varepsilon_{P_{1t}}$  e  $\varepsilon_{P_{2t}}$ . Desde que  $\varepsilon_t = B^{-1}\varepsilon_t$  podemos escrever os erros como:

$$\varepsilon_{1t} = \frac{(\varepsilon_{P_{1t}} - b_{12}\varepsilon_{P_{2t}})}{(1 - b_{12}b_{21})} \quad (28)$$

$$\varepsilon_{2t} = \frac{(\varepsilon_{P_{2t}} - b_{12}\varepsilon_{P_{1t}})}{(1 - b_{12}b_{21})} \quad (29)$$

Desde que, os  $\varepsilon_{P_{1t}}$  e  $\varepsilon_{P_{2t}}$  sejam processos do tipo ruído-branco, ou seja, tenham média zero, variância constante e não sejam autocorrelacionados. Derivando as propriedades de,  $\varepsilon_{1t}$  primeiro tomamos sua esperança matemática:

$$E\varepsilon_{1t} = E\frac{(\varepsilon_{P_{1t}} - b_{12}\varepsilon_{P_{2t}})}{(1 - b_{12}b_{21})} = 0 \quad (30)$$

$$E\varepsilon_{1t}^2 = E\left[\frac{(\varepsilon_{P_{1t}} - b_{12}\varepsilon_{P_{2t}})}{(1 - b_{12}b_{21})}\right]^2 = \frac{(\sigma_{P_1}^2 + b_{12}^2 b_{21}^2 \sigma_{P_2}^2)}{(1 - b_{12}b_{21})^2} \quad (31)$$

Assim, a variância de  $\varepsilon_{1t}$  é independente do tempo e a autocorrelação de  $\varepsilon_{1t}$  e  $\varepsilon_{1t-i}$  é:

$$E\varepsilon_{1t}\varepsilon_{1t-i} = E\left\{\frac{[(\varepsilon_{P_{1t}} - b_{12}\varepsilon_{P_{2t}})(\varepsilon_{P_{1t-i}} - b_{12}\varepsilon_{P_{2t-i}})]}{(1 - b_{12}b_{21})^2}\right\} = 0 \quad \forall i \neq 0. \quad (32)$$



Similarmente,  $\mathcal{E}_{2t}$  também é um processo estacionário com média zero, variância constante e correlação entre os erros nula. Enders (Op. Cit.), chama a atenção para o fato dos erros  $\mathcal{E}_{1t}$  e  $\mathcal{E}_{2t}$  serem correlacionados. Sua covariância pode ser escrita como:

$$E \mathcal{E}_{1t} \mathcal{E}_{2t} = E \left\{ \frac{[(\mathcal{E}_{P_{1t}} - b_{12} \mathcal{E}_{P_{2t}})(\mathcal{E}_{P_{2t}} - b_{21} \mathcal{E}_{P_{1t}})]}{(1 - b_{12} b_{21})^2} \right\} = - \frac{(b_{21} \sigma_{P_1}^2 + b_{12} \sigma_{P_2}^2)}{(1 - b_{12} b_{21})^2} \quad (33)$$

Normalmente, a equação (24) não é igual a zero a menos que não haja efeitos contemporâneos dos choques entre  $\{P_{1t}\}$  sobre  $\{P_{2t}\}$  e  $\{P_{2t}\}$  sobre  $\{P_{1t}\}$ . Logo, temos o caso utilizado no presente trabalho em que a matriz de variância e covariância é dada por:

$$\Sigma = \begin{bmatrix} \text{var}(\mathcal{E}_{1t}) & 0 \\ \text{cov}(\mathcal{E}_{1t}, \mathcal{E}_{2t}) & \text{var}(\mathcal{E}_{1t}) \end{bmatrix} \quad (34)$$

Apesar dos testes de *GRANGER* apontarem para uma relação de causalidade simultânea entre as séries utilizadas na estimação dos modelos, denotamos a relação de causalidade pelo suporte dado pela teoria e, desde que todos os elementos da matriz dada por (34) sejam independentes do tempo, podemos reescrevê-la como:

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ \sigma_{21} & \sigma^2 \end{bmatrix} \quad (35)$$

Onde  $\text{var}(\mathcal{E}_{1t}) = \sigma_1^2$  e  $\sigma_{21} = \text{cov}(\mathcal{E}_{1t}, \mathcal{E}_{2t})$ .

### 3.2.1. Estabilidade

Nos modelos auto-regressivos de primeira ordem como  $P_t = a_0 + a_1 P_{t-1} + \mathcal{E}_t$ , a condição de estabilidade é  $a_1 < |1|$ , onde esta situação também é válida para a condição de estabilidade da matriz  $A_1$ . Para resolver o sistema, reiteramos a equação (25) uma vez obtendo:

$$x_t = A_0 + A_1(A_0 + A_1 x_{t-2} + \mathcal{E}_{t-1}) + \mathcal{E}_t = (I + A_1)A_0 + A_1^2 x_{t-2} + A_1 \mathcal{E}_{t-1} + \mathcal{E}_t \quad (36)$$

Onde  $I$  é uma matriz  $2 \times 2$ .

Após  $n$  iterações,

$$x_t = (I + A_1 + \dots + A_1^n) A_0 + \sum_{i=0}^n A_1^i \mathcal{E}_{t-1} + A_1^{n+1} x_{t-n-1} \quad (37)$$

Reiterando continuamente o sistema está claro que o termo  $A_1^n$  irá desaparecer, pois com o passar do tempo, quando o número de iterações se aproxima de infinito,  $A_1^n$  tende a zero. Assim, assumindo que a condição de estabilidade é conhecida, podemos escrever a solução particular para  $x_t$ :

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i \mathcal{E}_{t-i} \quad (38)$$

Onde  $\mu = \begin{bmatrix} \bar{P}_1 \\ \bar{P}_2 \end{bmatrix}$

Se a condição de estabilidade é mantida, as seqüências  $\{P_{1t}\}$  e  $\{P_{2t}\}$  possuirão conjuntamente uma covariância estacionária e cada seqüência possuirá suas médias e variâncias constantes ao longo do tempo.

### 3.2.2. Função Impulso Resposta

Segundo Sims (1980), o modelo VAR pode ser representado por um modelo VMA, desde que sejam respeitadas as condições de estacionariedade e de invertibilidade. Logo, a representação VMA pode estabelecer o comportamento das variáveis do sistema a choques ou inovações. A partir da equação (38), obtemos a equação (39) na forma matricial:

$$\begin{bmatrix} P_{1t} \\ P_{2t} \end{bmatrix} = \begin{bmatrix} \bar{P}_1 \\ \bar{P}_2 \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} \mathcal{E}_{1t-i} \\ \mathcal{E}_{2t-i} \end{bmatrix} \quad (39)$$

A equação (39) expressa  $P_{1t}$  e  $P_{2t}$  em termos das seqüências de resíduos  $\{\varepsilon_{1t}\}$  e  $\{\varepsilon_{2t}\}$ . Das equações (28) e (29), o vetor de resíduos pode ser escrito da seguinte forma:

$$\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ (1 - b_{12}b_{21}) \end{bmatrix} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{P_{1t}} \\ \varepsilon_{P_{2t}} \end{bmatrix} \quad (40)$$

Combinando as equações (39) e (40):

$$\begin{bmatrix} P_{1t} \\ P_{2t} \end{bmatrix} = \begin{bmatrix} \bar{P}_1 \\ \bar{P}_2 \end{bmatrix} + \begin{bmatrix} 1 \\ (1 - b_{12}b_{21}) \end{bmatrix} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-i} \\ \varepsilon_{2t-i} \end{bmatrix} \quad (41)$$

Para tornar a notação mais acessível, Enders (1995) faz uma simplificação definindo a matriz  $\phi_i$  2 x 2 como:

$$\phi_i = \begin{bmatrix} A_1^i \\ 1 - b_{12}b_{21} \end{bmatrix} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \quad (42)$$

Tomando essa notação, podemos representar as equações (39) e (40) por médias móveis (MA), escritas nos termos das seqüências dos resíduos  $\{\varepsilon_{P_{1t}}\}$  e  $\{\varepsilon_{P_{2t}}\}$  como segue:

$$\begin{bmatrix} P_{1t} \\ P_{2t} \end{bmatrix} = \begin{bmatrix} \bar{P}_1 \\ \bar{P}_2 \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{P_{1t-i}} \\ \varepsilon_{P_{2t-i}} \end{bmatrix} \quad (43)$$

De forma mais simples:

$$x_t = \mu + \sum_{i=0}^{\infty} \phi_i \varepsilon_{t-i} \quad (44)$$

A partir desta representação, é possível mensurar a relação entre as seqüências  $\{P_{1t}\}$  e  $\{P_{2t}\}$ . Através dos coeficientes  $\phi_i$  podem-se gerar efeitos de choques aplicados aos resíduos sobre as seqüências  $\{P_{1t}\}$  e  $\{P_{2t}\}$ . Assim cada elemento da matriz de coeficientes da equação (43) pode ser chamado de multiplicador de impactos.

### 3.2.3. Decomposição da Variância dos Erros de Previsão

Enders (Op. Cit.), sugere que se pode admitir que os parâmetros  $\Gamma_o$  e  $\Gamma_s$  da equação (24) do modelo VAR, sejam conhecidos. Além disso, se desejam prever os vários valores de  $x_{t+1}$  condicionados ao valor de  $x_t$ . Logo, trazendo a equação (25) para  $x_{t+1}$ , temos:

$$x_{t+1} = A_0 + A_1 x_t + \varepsilon_{t+1} \quad (45)$$

Aplicando a esperança matemática condicionada na equação (45):

$$E_t x_{t+1} = A_0 + A_1 x_t \quad (46)$$

O erro de previsão no período seguinte é representado por:

$$x_{t+1} - E_t x_{t+1} = e_{t+1} \quad (47)$$

Para dois períodos seguintes, temos:

$$x_{t+2} = A_0 + A_1 x_{t+1} + \varepsilon_{t+2} \quad (48)$$

Analogamente ao raciocínio para se encontrar o erro de previsão para um período seguinte, podemos encontrar o erro de previsão para dois períodos seguintes e para n períodos seguintes. Adicionalmente, utilizando o mesmo raciocínio aplicado à equação (38) para transformá-la num processo VMA, podemos transformar o erro de previsão num processo em função de seus próprios resíduos como expressado a seguir:

$$x_{t+n} - E_t x_{t+n} = \sum_{i=0}^{n-1} \phi_i \varepsilon_{t+n+i} \quad (49)$$

Representando a variância do erro de previsão para  $P_{1t+n}$  como  $\sigma_R(n)^2$ , obtém-se:

$$\sigma_R(n)^2 = E_t \left[ (P_{1t+n} - E_t P_{1t+n})^2 \right] \quad (50)$$

Note que como os valores de  $\phi_{jk}^{(i)^2}$  são não-negativos, a variância do erro de previsão aumenta com diretamente com o aumento do horizonte de previsão  $n$ .

### 3.3. Estrutura do Modelo ARMAX

#### 3.3.1. Função de Correlação Cruzada

Como o próprio nome diz, a Função de Correlação Cruzada estima a correlação entre as séries de tempo em diferentes momentos, além de ser um importante instrumental para se identificar os modelos de função de transferência. Este instrumental testa se existe uma relação de causalidade entre as variáveis, de forma que as séries precisam ser devidamente tratadas antes de serem testadas a fim de se eliminar o problema de se obter relações espúrias.

Define – se:

$$\delta_{xy}(k) = E[(P_{1t} - \mu_x)(P_{2t+k} - \mu_y)] \quad (51)$$

Onde:

$\delta_{xy}(k)$  é a covariância entre as variáveis  $P_{1t+k}$  e  $P_{2t}$ ; sendo  $k = 0, \pm 1, \pm 2, \dots$ ;

$\mu_x, \mu_y$  são as médias dos parâmetros  $P_1$  e  $P_2$  respectivamente.

O coeficiente de correlação ou função de correlação cruzada  $\rho_{xy}(k)$  entre as variáveis  $P_{1t+k}$  e  $P_{2t}$  é dada por:

$$\rho_{xy}(k) = \frac{\delta_{xy}(k)}{\sigma_x \sigma_y} \quad (52)$$

Onde:

$\sigma_x, \sigma_y$  os desvios-padrões das séries de tempo  $P_1$  e  $P_2$  respectivamente.

Quando os valores de  $k$  são positivos demonstram períodos passados ou *lags*; e quando negativos demonstram períodos futuros ou *leads*. Esta função  $\delta_{xy}(k)$  é estimada para valores positivos e negativos de  $k$ , pois ela não é uma função simétrica, isto é, quando  $k > 0$ , as correlações estatisticamente significativas demonstram que a variável  $P_1$  é a preditora da variável  $P_2$ , ocorrendo o contrário para o caso de  $k < 0$ .

O desvio-padrão da correlação cruzada das séries de variáveis  $DP[r_{P_1 P_2}(k)]$ , é definido por:

$$DP[r_{P_1 P_2}(k)] = \frac{1}{\sqrt{N}} \quad (53)$$

Adotando o nível de significância de 5% obtém-se:

$$|r_{P_1 P_2}(k)| \geq \frac{1,96}{\sqrt{N}}$$

#### 3.3.2. Modelo ARMAX (r,s,b)(p,q)

O modelo ARMAX ou função de transferência pode ser descrito como uma adaptação do modelo descrito em NARAYAN, NARAYAN & PRASAD (2004).

$$P_{1t} - \alpha_1 P_{1t-1} - \alpha_2 P_{1t-2} = \lambda + \pi_0 P_{2t} + \pi_1 P_{2t-1} + \pi_2 P_{2t-2} + \dots + \pi_q P_{2t-q} + \varepsilon_t - \rho_1 \varepsilon_{t-1} - \dots - \rho_q \varepsilon_{t-q} \quad (54)$$

Onde  $p$  e  $q$ : são os comprimentos dos *lags* dos termos auto-regressivos e termos de erro por médias-móveis respectivamente;  $\varepsilon$  é o resíduo caracterizado por ser um processo com ruído-branco.

Rearranjando os termos e utilizando a notação de *lags* podemos reescrever a equação (54) como segue:

$$P_{1t}(1 - \alpha_1 L - \alpha_2 L^2) = \lambda + (\pi_0 + \pi_1 L + \pi_2 L^2) P_{2t} + (1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_q L^q) \varepsilon_t \quad (55)$$

Escrito dessa forma, pode-se agora derivar os impactos da variável explicativa na variável dependente baseado na razão dos *lags* dos polinômios das duas variáveis.

$$P_1(1-\alpha_1L - \alpha_2L^2) = (\pi_0 + \pi_1L + \pi_2L^2)P_2, \quad (56)$$

Denotemos:

$$A(L) = (1-\alpha_1L - \alpha_2L^2) \quad (57)$$

$$B(L) = (\pi_0 + \pi_1L + \pi_2L^2)$$

Substituindo a primeira equação do sistema (57) em (56), temos:

$$A(L)P_1 = B(L)P_2 \quad (58)$$

Observa-se agora que os impactos da variável  $P_1$  na variável  $P_2$  são descritos pela razão dos polinômios  $A(L)$  e  $B(L)$ . Substituindo a razão  $B(L)/A(L)$  por  $Z(L)$ , onde  $Z(L) = (Z_0 + Z_1L + Z_2L^2 + Z_3L^3 + \dots)$ , temos uma série infinita de distribuição de *lags*. A informação dos pesos dos *lags* em  $Z(L)$  vem da expressão da relação na sua forma expandida:

$$Z(L)A(L) = B(L) \quad (59)$$

Expandindo o produto do lado esquerdo da equação (59), temos:

$$\begin{aligned} Z_0 + Z_1L + Z_2L^2 + Z_3L^3 + \dots - Z_0\alpha_1L - Z_0\alpha_2L^2 - Z_1\alpha_1L^2 - Z_1\alpha_2L^3 - Z_2\alpha_1L^3 - Z_2\alpha_2L^4 - Z_3\alpha_1L^4 - Z_3\alpha_2L^5 - \dots \\ = \pi_0 + \pi_1L + \pi_2L^2 \end{aligned} \quad (60)$$

Montando o sistema combinando os termos em  $L$  pode-se resolver  $Z_0, Z_1, Z_2, Z_3, \dots$

$$Z_0 = \pi_0$$

$$Z_1 = Z_0\alpha_1 + \pi_1$$

$$Z_2 = Z_1\alpha_1 + Z_0\alpha_2 + \pi_2 \quad (61)$$

$$Z_3 = Z_2\alpha_1 + Z_1\alpha_2$$

O resto dos termos  $Z$  seguem repetidamente a ordem:

$$Z_k = Z_{k-1}\alpha_1 + Z_{k-2}\alpha_2 \quad (62)$$

Logo, quando os coeficientes  $\alpha_1, \alpha_2; \pi_0, \pi_1$  e  $\pi_2$  são estimados, as ponderações dos *lags* dos  $Z$ 's podem ser calculadas. O impacto total de uma mudança na variável  $P_1$  em  $P_2$  é dado por:

$$\sum_{i=0}^{\infty} Z_i \quad (63)$$

Dito de outra forma calcula-se os parâmetros  $(r, s, b)$  aos quais se referem à parametrização da função de transferência ligando  $P_1$  e  $P_2$  em  $(p, q)$ . Identificando os termos ARMAX $(r, s, b)(p, q)$ , isto é, os parâmetros  $\alpha_1, \alpha_2; \pi_0, \pi_1$  e  $\pi_2$ . Onde os três últimos parâmetros  $b, r$  e  $s$  são obtidos no Correlograma das Correlações Cruzadas. Contudo é necessário observar que nem sempre a correlação cruzada fornece toda a informação para o cálculo dos parâmetros  $(r, s, b)$ .

#### 4. ANÁLISE DOS RESULTADOS

A série preço da *commodity* americana de soja em grão foi coletada na Fundação Getúlio Vargas. A série *spot price* da soja doméstica, foi coletada na BM&F em observações diárias e, para se transformar estas observações em observações mensais, utilizou-se a média geométrica dos dias de transação no mercado. As séries preço do farelo e do óleo de soja domésticos foram gentilmente cedidas pela ESALQ a partir de dados obtidos no CEPEA. O preço do óleo de soja doméstico é o preço médio mensal do óleo degomado bruto com 12/% de ICMS colocado em São Paulo. O preço do farelo da soja doméstico é a média mensal dos preços de duas praças: Oeste do Paraná e Campinas – SP.

As observações são mensais e compreendem o período de Janeiro de 1999 a Fevereiro de 2005. Utilizou-se na estimação os logaritmos naturais desses preços. Foram testadas possíveis sazonalidades das séries e não foram encontradas qualquer evidência da presença destas.

A Figura 1<sup>3</sup> compreende o conjunto de gráficos que denotam o comportamento das séries de preços utilizadas neste trabalho.

Na Tabela 1 temos o conjunto das estatísticas das séries utilizadas para o modelo ARMAX. Os valores entre parênteses apresentados juntamente com as estatísticas são os respectivos níveis de significâncias destas. Observam-se pelo teste de normalidade da distribuição *Jarque Bera* e sua respectiva significância, que todas as séries possuem distribuição aproximadamente normal.

A Tabela 2 demonstra o teste de causalidade de *GRANGER*, onde pela estatística do teste *F* podemos observar que a  $F(1,73)$  graus de liberdade rejeitam-se as hipóteses nulas de não causalidade, pois valor crítico de *F* é aproximadamente 4,00.

Através da Tabela 3 podemos observar as estatísticas dos três modelos ARMAX estimados escolhidos pelos critérios *AIC* e *SBC*<sup>4</sup> onde quanto melhor o modelo, menor o valor calculado destes critérios. Na mesma tabela temos os parâmetros resultantes estimados, onde entre parênteses estão suas respectivas estatísticas *t* de *Student*. Podemos concluir que a transmissão ou impulso resposta do preço da soja em grão doméstica para o preço do farelo de soja doméstico (modelo 1) e para o preço do óleo de soja doméstico (modelo 2) é significativa ao nível de significância de 5% denotada pelas estatísticas *t* de *Student* dos parâmetros que captam esta transmissão. Embora podemos denotar uma diferença na forma de transmissão entre estes preços, ou seja, temos que no modelo 1 a transmissão é instantânea captada pelo parâmetro  $B_0$ , ao passo que, no modelo 2 a transmissão não é instantânea captada pelo parâmetro  $B_1$ . Caso a transmissão no modelo 2 fosse instantânea teríamos assim, como no modelo 1, o parâmetro  $B_0$  que captaria a transmissão instantânea do preço da soja em grão doméstica para o preço do óleo doméstico. Podemos inferir, portanto, que o farelo de soja, talvez por ser um produto em grande parte destinado à exportação, possua mais dinâmica com o mercado da soja em grão doméstica.

Na Tabela 4 é apresentado o teste Dickey - Fuller ampliado ou ADF, que testa todos os parâmetros da série temporal conjuntamente. Para que a série seja estacionária, temos que o valor de  $\tau$  calculado deve ser maior em módulo que o valor de  $\tau$  tabelado. Denotamos por *k* o número de defasagens e *I*(d) o número de diferenças realizadas. O critério de escolha dos modelos, respeitados as significâncias de  $Q(16)$  que indica se o resíduo da série é ruído-branco, foram os critérios menores *AIC* e *SBC*.

Os testes de cointegração entre os pares de variáveis estimadas nos três modelos propostos seguem na Tabela 5, onde podemos observar que nos dois primeiros modelos: transmissão do preço da *commodity* americana de soja em grão para o *spot price* da soja doméstica e transmissão do preço da *commodity* americana de soja em grão para o preço do óleo de soja doméstico, não existe cointegração entre as séries. Porém, no terceiro modelo, transmissão do preço da *commodity* americana de soja em grão para o preço do farelo de soja doméstico existe cointegração entre as séries e, portanto, para este modelo será estimado um VEC diferentemente dos dois primeiros modelos que serão estimados pela metodologia VAR padrão.

Observando a Tabela 6 temos os testes para a especificação dos resíduos dos modelos onde, pelo teste *Jarque-Bera*, testamos a normalidade dos resíduos e pelo teste ARCH LM testamos conjuntamente heterocedasticidade e autocorrelação dos resíduos do modelo. Temos que todos os modelos obedecem aos requisitos desejáveis para que os parâmetros sejam eficientes, consistentes e não-viesados.

Para se escolher o critério de seleção de variáveis para os modelos VAR padrão 1 e 2: transmissão do preço da *commodity* americana de soja em grão para o *spot price* da soja doméstica e transmissão do preço da *commodity* americana de soja em grão para o preço do óleo de soja doméstico, foram feitos os testes de seleção defasagens apresentados na Tabela 7, onde o critério de escolha desta foi pelo teste *HQ*.

Na Tabela 8 temos a decomposição da variância pela metodologia de *CHOLESKY*, onde se observam em conjunto com a Figura 2 a influência do preço da *commodity* americana de soja em grão sobre o *spot price* da soja doméstica, o preço do óleo de soja doméstico e o preço do farelo de soja doméstico. Nas colunas da Tabela 9 estão as percentagens destas influências nas explicações dos preços domésticos com o passar dos meses. Pela função de impulso resposta, observamos um efeito transitório em todos os três modelos. Assim, temos que aproximadamente no décimo mês, o *spot price* da soja em

<sup>3</sup> Todas as figuras e tabelas referidas nesta seção se encontram no apêndice do trabalho.

<sup>4</sup>  $AIC = T \ln(SQR) + 2n$   
 $SBC = T \ln(SQR) + n \ln(T)$

Onde *n* é o número de parâmetros estimados e *T* é o número de observações utilizáveis.

grão doméstica é explicado em torno de 32,97% pelo preço da *commodity* americana de soja em grão convergindo para uma percentagem de explicação em torno desta. No caso do preço do óleo de soja doméstico, esta percentagem converge para aproximadamente 9,05% no décimo mês e para o preço do farelo de soja doméstico esta percentagem de explicação converge para aproximadamente 35,20% também no décimo mês. Contudo, no modelo 3, existe um efeito inicial crescente dos preços do farelo de soja doméstico dado um choque positivo nos resíduos do preço da *commodity* da soja americana, diferentemente dos outros dois modelos onde o ajuste dos preços é simultâneo. Isso sugere uma margem de arbitragem neste mercado, ou seja, o ajuste do preço doméstico do farelo de soja às flutuações no preço da *commodity* da soja americana não é simultâneo demorando aproximadamente três meses para se reajustar. Em contrapartida a este *lag* de ajuste de preços, podemos inferir que o farelo de soja, talvez por ser um produto em grande parte destinado à exportação, e desta maneira mais ligado ao mercado internacional da soja, possua mais inter-relação com este. Isso pode ser observado na decomposição da variância através da percentagem de explicação do preço do farelo de soja doméstico pelo preço da *commodity* americana de soja em grão, de 35,20%.

## 5. CONCLUSÃO

Na amostra utilizada foi possível identificar alguns pontos sobre a dinâmica mercado mundial da soja. Embora os modelos ARMAX estimados perdessem graus de liberdade porque necessitaram de um grande número de parâmetros no processo de pré-branqueamento dos ruídos para que estes fossem *white-noise*, identificamos uma relação de transmissão entre os preços domésticos, isto é, flutuações do preço da soja em grão doméstica se transmitem para os preços do óleo e do farelo da soja.

Nos modelos VAR padrão e VEC estimados para captar transmissão de preços no mercado internacional, identificamos três transmissões. Como existe uma diferença na magnitude de transmissão entre os preços, talvez isso nos leva a pensar em possíveis erros de especificação dos modelos, embora todos os critérios sejam eles estatísticos, econométricos ou econômicos foram seguidos de acordo com o rigor dos trabalhos referenciais neste assunto. Assim, mais especificamente, como o mercado do óleo de soja, possivelmente por estar integrado numa rede mais ampla de mercados correlacionados tanto pra trás quanto pra frente, apresenta características distintas dos demais segmentos do complexo soja possuindo a menor transmissão de preços, exige mais concentração e esforço na designação das variáveis que determinam seu preço.

Num estudo mais profundo sobre este mercado o trabalho deixa, como sugestão, a busca de outros determinantes para a explicação dos preços domésticos do mercado da soja. Determinantes, tanto do lado da demanda, como os estoques de soja do principal consumidor mundial de soja - China, e variáveis do lado da oferta, como por exemplo, os fortes subsídios dados pelo governo americano aos produtores de soja, onde estes produtores são os principais beneficiados dentre os demais produtores de outras culturas americanas. Além disso, temos que levar em consideração, em detrimento a fatores externos, fatores domésticos como, por exemplo, fatores tecnológicos, onde temos que lembrar que o Brasil é um dos líderes em pesquisas neste campo.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

BOLSA DE MERCADORIAS E FUTUROS. **Séries Históricas de Cotações**. [s.l.]: [s.n], [2005]. Disponível em: < <http://www3.bmf.com.br/datastore1/IndexSeries1.asp> >. Acesso em: 26.03.2005.

BRUM, Argemiro Luís. **Economia da Soja: História e Futuro** - Uma visão desde o Rio Grande do Sul - Disponível em: < <http://www.agromil.com.br/econosoja.html> > Acesso em: 22/03/2005.

CASTRO, Alexandre Samy de, JÚNIOR, José Luiz Rossi, **Modelos de Previsão para a Exportação das Principais Commodities Brasileiras** - Texto para Discussão N° 716 - Rio de Janeiro, abril de 2000.

ENDERS, Walter. *Applied Econometric Time Series*. New York: John Wiley & Sons, Inc, 1995a.

ENDERS, Walter. *Rats Handbook for Econometric Time Series*. New York: John Wileyand Sons, Inc, 1995b.

Fundação Getúlio Vargas. *Séries Gratuitas*. [s.l.]: [s.n], [2005]. Disponível em: <[http://www.fgvdados.com.br/dsp\\_frs\\_pai\\_ferramentas.asp](http://www.fgvdados.com.br/dsp_frs_pai_ferramentas.asp)> Acesso em: 03/04/2005.

HAMILTON, James D. *Time series analysis*. New Jersey: Princeton University Press, 1994.

MARGARIDO, Mario A., TUROLLA, Frederico A. & BUENO, Carlos R.F., *The World Market for Soybeans: Price Transmission Into Brazil And Effects from the Timing of Crop And Trade*, ANPEC 2004.

NARAYAN, PARESH KUMAR, NARAYAN , SEEMA, PRASAD C. BIMAN *Forecasting Fiji's Exports And Imports, 2003-2020*, USPEC Working Paper, October, 2004.

SIMS, C. *Macroeconomics and reality*. *Econometrica*, v.48, n.1, p.1-48, 1980.

WOOLVERTON, Michael W., *Price Risk Management in the Soy Complex*, ASA Dubai Conference - February 2003

## 7. APÊNDICE

**TABELA 1 – Estatísticas das Séries de Preços do Mercado de Soja  
(Jan/1999 – Fev/2005)**

Estatísticas	SÉRIES			
	SOJA USA	SOJA BR	ÓLEO BR	FARELO BR
<b>Média</b>	0,000186	0,00896	0,005032	0,008886
<b>Desvio-Padrão</b>	0,060955	0,07678	0,081239	0,0787576
<b>Variância</b>	0,003716	0,00590	0,0066	0,006203
<b>Curtose</b>	-0,15492 (0,596637)	0,29038 (0,321187)	0,56921 (0,05182)	-0,08863 (0,76206)
<b>Assimetria</b>	0,76881 (0,201509)	-0,76815 (0,201895)	0,35179 (0,55892)	-0,59071 (0,326408)
<b>Jarque Bera</b>	2,08982 (0,351724)	2,82063 (0,244065)	4,31845 (0,115414)	1,15691 (0,56076)

Fonte: Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

**TABELA 2 – Teste de Causalidade de GRANGER**

<b>Hipótese Nula</b>	<b>F (1,73) <math>\equiv</math> (4,00)</b>
Soja BR não causa Soja USA	19,8402 (0,00003067)
Soja USA não causa Soja BR	19,8402 (0,00003067)
Óleo BR não causa Soja USA	5,8051 (0,01857493)
Soja USA não causa Óleo BR	5,8051 (0,01857493)
Farelo BR não causa Soja USA	15,8497 (0,00016370)
Soja USA não causa Farelo BR	15,8497 (0,00016370)

**Fonte:** Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

**TABELA 3 – Estatísticas dos Modelos de Transferência Estimados**

<b>TRANSFERÊNCIA DO PREÇO DA SOJA BR PARA:</b>		
	<b>FARELO BR</b>	<b>ÓLEO BR</b>
	<b>Estatísticas</b>	
<b>R<sup>2</sup> Ajustado</b>	0,384910	0,239909
<b>Q(16)</b>	25,6604 (0,10776556)	20,2630 (0,20860785)
<b>AIC</b>	-87,81832	-67,52944
<b>SBC</b>	-74,50128	-54,30128
	<b>Coefficientes</b>	
$\alpha_1$	-0,079673927 (-0,48583)	0,460068137 (3,55859)
$\alpha_2$	-0,168650501 (-1,20715)	-0,405285894 (-2,92258)
$\alpha_3$	-0,009407944 (-0,06517)	0,336724375 (2,37393)
<b>B<sub>0</sub></b>	0,739124898 (5,69262)	-
<b>B<sub>1</sub></b>	0,726968050 (5,42437)	0,366853208 (2,35288)
<b>B<sub>2</sub></b>	-	-0,398357848 (-2,91391)
<b>A<sub>1</sub></b>	-1,018557137 (33,34673)	0,796292422 (2,88515)

**Fonte:** Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV



**TABELA 4 – Testes Dickey - Fuller Ampliado**

Séries	I(d)	k	estatística $\tau$ +3,52	Q(16)	AIC	SBC
LSOY	0	1	-2,148034	0,995	-2,977944	-2,883083
DLSOY	1	1	-5,367783	0,978	-2,940992	-2,877751
LSOJA	0	1	-1,517282	0,305	-2,612979	-2,518118
DLSOJA	1	1	-4,353579	0,208	-2,663825	-2,568219
LOLEO	0	3	-1,502274	0,319	-2,369209	-2,208602
DLOLEO	1	1	-4,905746	0,031	-2,342271	-2,246665
LFARELO	0	1	-1,802001	0,319	-2,435991	-2,341130
DLFARELO	1	1	-5,005530	0,149	-2,431800	-2,336194

Fonte: Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

**TABELA 5 – Teste de JOHANSEN - JUSELIUS para Cointegração**

$\lambda_{\text{traco}}$		$\lambda_{\text{máximo}}$		Teste do maior autovalor		Teste do traço	
H <sub>0</sub>	H <sub>1</sub>	H <sub>0</sub>	H <sub>1</sub>	-TSl <sub>n</sub> (1-l <sub>i+1</sub> )	Valor Crítico a 5% de signif.	-Tln(1-l <sub>i</sub> )	Valor Crítico a 5% de signif.
<b>Modelo 1: Transmissão do preço da soja em grão americana para o preço da soja doméstica</b>							
r = 0	r = 1	r = 0	r = 1	8,160424	14,07	11,62886	15,41
r ≤ 1	r > 2	r = 1	r = 2	3,468432	3,76	3,468432	3,76
<b>Modelo 2: Transmissão do preço da soja em grão americana para o preço do óleo de soja doméstico</b>							
r = 0	r = 1	r = 0	r = 1	7,298694	14,07	8,659219	15,41
r ≤ 1	r > 2	r = 1	r = 2	1,360525	3,76	1,360525	3,76
<b>Modelo 3: Transmissão do preço da soja em grão americana para o preço do farelo de soja doméstico</b>							
r = 0	r = 1	r = 0	r = 1	6,507923	14,07	10,65079	15,41
r ≤ 1	r > 2	r = 1	r = 2	4,142870	3,76	4,142870	3,76

Fonte: Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

**TABELA 6 - Testes para especificação dos resíduos do modelo**

TESTES	Modelo 1	Resíduo Soja USA	Resíduo Soja BR
Jarque - Bera	0,3792	0,5556	0,2200
ARCH LM	0,2762		
LB Q (16)	0,3223		
	Modelo 2	Resíduo Soja USA	Resíduo Óleo BR
Jarque - Bera	0,7497	0,5264	0,7258
ARCH LM	0,8753 <sup>5</sup>		
LB Q (16)	0,1558		
	Modelo 3	Resíduo Soja USA	Resíduo Farelo BR
Jarque - Bera	0,3405	0,3203	0,3262
ARCH LM	0,1616		
LB Q (16)	0,7231		

Fonte: Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

<sup>5</sup> O critério de significância adotado para os resíduos deste modelo foi Q (8).

**TABELA 7 – Teste para Seleção de Defasagens**

	<i>k</i> defasagens	<i>AIC</i>	<i>SBC</i>	<i>HQ</i>
<b>Modelo 1</b>	1	-5,722776	-5,507718	-5,639197
<b>Modelo 2</b>	1	-5,094965	-4,879907	-5,011386

Fonte: Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

**TABELA 8 – Decomposição da Variância**

<b>Decomposição da Variância da SOJA BR:</b>			
Meses	Desvio-Padrão	SOJA USA	SOJA BR
1	0,06463	25,42213	74,57787
2	0,07289	30,24458	69,75542
3	0,07497	32,11772	67,88228
10	0,07569	32,97096	67,02904

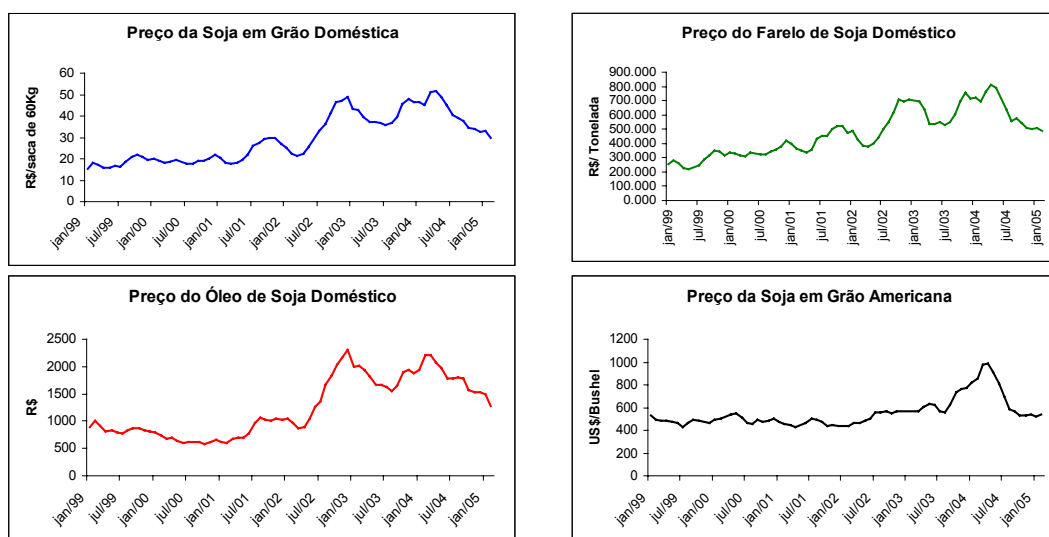
<b>Decomposição da Variância do ÓLEO BR:</b>			
Meses	Desvio-Padrão	SOJA USA	ÓLEO BR
1	0,0749	4,574426	95,2557
2	0,08048	7,516748	92,48325
3	0,08161	8,609761	91,39024
10	0,08197	9,053488	90,94651

<b>Decomposição da Variância do FARELO BR:</b>			
Meses	Desvio-Padrão	SOJA USA	FARELO BR
1	0,07689	30,26307	69,73693
2	0,09687	43,05715	56,94285
3	0,11298	54,73689	45,26311
4	0,12657	61,79803	38,20197
10	0,18816	74,38623	25,61377

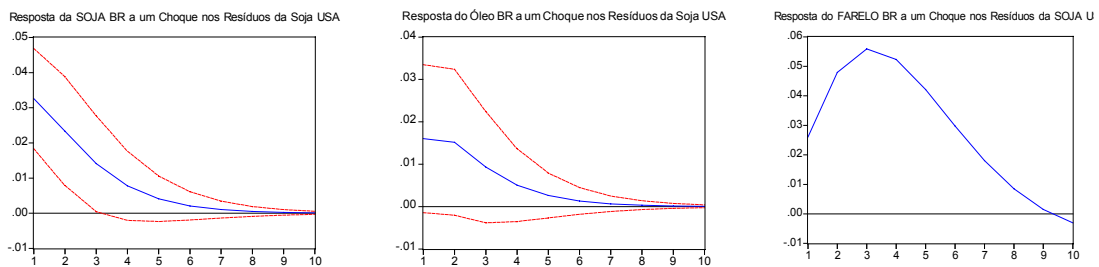
Fonte: Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

**FIGURA 1 – Gráficos das Séries de Preços do Mercado de Soja (Jan/1999–Fev/2005)**



Fonte: Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV

## FIGURA 2 – Funções de Impulso Resposta



**Fonte:** Elaboração própria a partir de dados obtidos no CEPEA, BM&F e FGV