

### 3

## Dados e metodologia

### 3.1

#### Apresentação de Dados

Para a realização dessa pesquisa foram utilizados os dados da série histórica dos preços da soja (em grão) do Estado do Paraná, obtidos da base de dados do Centro de Estudos Avançados em Economia Aplicada (CEPEA) da faculdade ESALQ/USP<sup>7</sup>, cuja cotação é feita em reais por saca de 60 Kg.

É importante destacar os critérios utilizados pelo CEPEA na seleção de seus dados. A região de referência, o Estado do Paraná, foi dividida em cinco sub-regiões: Paranaguá, Ponta Grossa, Norte, Oeste e Sudeste. O percentual de participação de cada região foi determinado pela capacidade instalada de esmagamento, segundo informações da Abiove<sup>8</sup> (1997), assim sendo:

Porto de Paranaguá – 19,13%

Ponta Grossa – 25,49 %

Norte – 34,72%

Oeste – 9,73%

Sudoeste – 10,93 %

A série de dados coletados contempla cotações diárias desde 29/07/1997 até 19/03/2009. No entanto, para os cálculos dessa pesquisa os dados referentes a julho de 1997 e os dados referentes ao ano de 2009 foram expurgados. Assim sendo, a série utilizada foi de Agosto de 1997 até Dezembro de 2008. É importante destacar, também, que para efeito dos cálculos e dos testes, os dados foram trabalhados de forma a agrupar a média dos preços mês a mês, transformando, assim, as 2837 observações diárias dos dados brutos em 120

---

<sup>7</sup><http://www.cepea.esalq.usp.br/soja/>

<sup>8</sup> Associação Brasileira das Indústrias de óleos vegetais (Abiove). <http://www.abiove.com.br/>

observações mensais. Além disso, optou-se por trabalhar com a série em escala logarítmica.

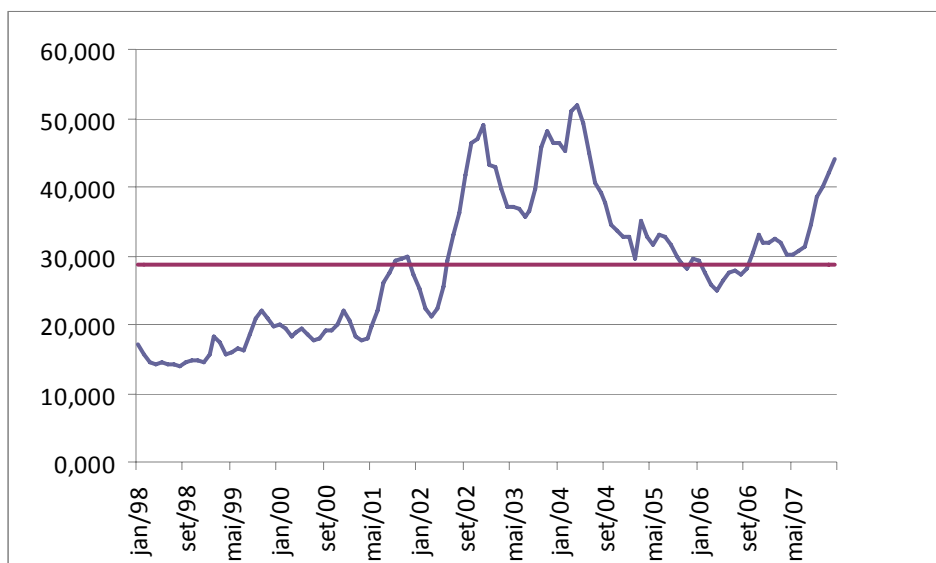


Figura 3.1: Série histórica do logaritmo dos preços mensais da Soja (Grão).

Fonte: CEPEA (2009).

A metodologia utilizada foi fundamentada na modelagem utilizada por Das (1998), conforme foi descrito, de forma detalhada, no primeiro capítulo. A proposta é testar a adequação dos modelos de reversão à média com saltos ao histórico de preços da *commodity* soja.

A análise do comportamento dos preços das commodities já foi objetos de vários estudos. Dixit e Pindyck (1994) descrevem o preço das commodities com uma característica particular relacionada ao custo marginal de produção no longo prazo, o que significa que no curto prazo podemos encontrar algumas variações no preço de commodities, mas no longo prazo, em geral, elas tendem a voltar para o custo marginal de produção, logo possuem uma característica marcante do modelo de reversão à média.

Se observarmos o histograma da figura 3.2, verificaremos que há no histórico de preços que estamos estudando uma curva de distribuição próxima a curva Normal.

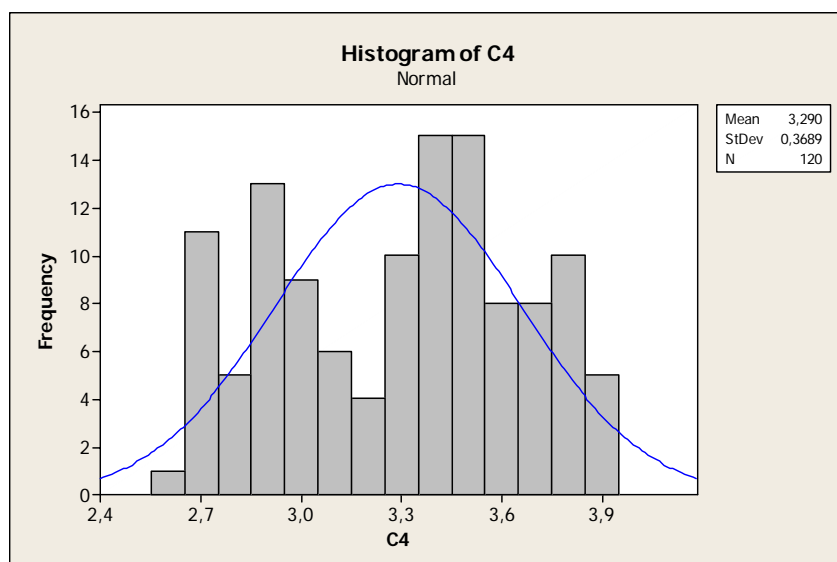


Figura 3.2: Histograma da série histórica do logaritmo dos preços mensais da soja.

Fonte: CEPEA (2009), Minitab

No entanto, pelos resultados apresentados pela análise da Estatística Descritiva dos dados, que se encontram na Tabela 3.1, podemos notar que o resultado da curtose é menor que zero, o que indica uma tendência platicúrtica à série, deixando um forte indício de que na série em questão não deve ser relevante a presença de saltos.

Tabela 3.1: Estatísticas Descritivas.

Média	3,26
Mediana	3,32
Mínimo	2,63
Máximo	3,95
Desvio padrão	0,37
Curva de Simetria	0,11
Enviesamento	0,01
Curtose Ex.	-1,08

Fonte: Grelt.

Com os resultados de simetria e curtose apresentados na tabela 3.1, pode-se realizar a análise de normalidade da série, através do teste Jarque Bera. Conforme podemos observar os resultados na Tabela 3.2.

Tabela 3.2: Teste de Normalidade.

<b>Teste da normalidade de LN_P_:</b>
Doornik-Hansen test = 8,71214, com p-valor 0,0128287
Shapiro-Wilk W = 0,958568, com p-valor 0,00097241
Lilliefors test = 0,100679, com p-valor ≈ 0
<b>Jarque-Bera test = 5,8478, com p-valor 0,0537239</b>

Fonte: Grelt.

Pelo resultado do teste *Jarque-Bera* apresentado na tabela 3.2, ao nível de significância de 5%, não podemos rejeitar a hipótese nula de normalidade. O teste de normalidade *Jarque-Bera* é baseado nas diferenças entre os coeficientes de assimetria e curtose da distribuição amostral da série e da distribuição teórica Normal e serve para testar a hipótese nula de que a amostra foi extraída de uma distribuição normal.

No entanto, o resultado apresentado, está muito próximo do limite para a rejeição da hipótese nula de normalidade da série. Tal resultado é muito importante, pois a normalidade é pressuposto base para a aplicação da metodologia adotada nesta pesquisa.

Por último realizou-se o teste de Dickey-Fuller Aumentado (ADF) a fim de se descobrir se a série analisada poderia ser considerada estacionária ou não, que tem como base a seguinte expressão:

$$\Delta Y_t = \beta + \delta T + \gamma Y_{t-1} + \alpha_i \sum_{i=1}^m \Delta Y_{t-i} + \varepsilon_t$$

onde:  $\beta$  é uma constante; T é a tendência;  $\Delta$  é o operador diferença, o qual é representado como:  $\Delta_t = Y_t - Y_{t-1}$ .

O número de defasagens é determinado pelo menor valor do Critério de Schwarz, visando eliminar a auto-correlação dos resíduos, conforme podemos verificar na figura 3.3.

A hipótese nula é de que  $\gamma = 0$  ou  $\rho = 1$ . Ou seja, existe raiz unitária, o que significa que a série é não-estacionária.

Tabela 3.3: Teste ADF.

```

teste Aumentado de Dickey-Fuller, para LN_P_
incluindo uma defasagem de (1-L)LN_P_
dimensão de amostragem 118
hipótese nula de raiz unitária: a = 1

teste com constante (GLS)
modelo: (1-L)y = b0 + (a-1)*y(-1) + ... + e
coeficiente de 1ª ordem para e: 0,020
valor estimado de (a - 1): -0,0155901
estatística de teste: tau = -1,14832
p-valor assintótico 0,2292

```

Fonte: Grelt.

Para se testar a estacionariedade da série, o teste ADF verifica a ordem de integração da variável de interesse, ou seja, verifica a existência ou não de raízes unitárias.

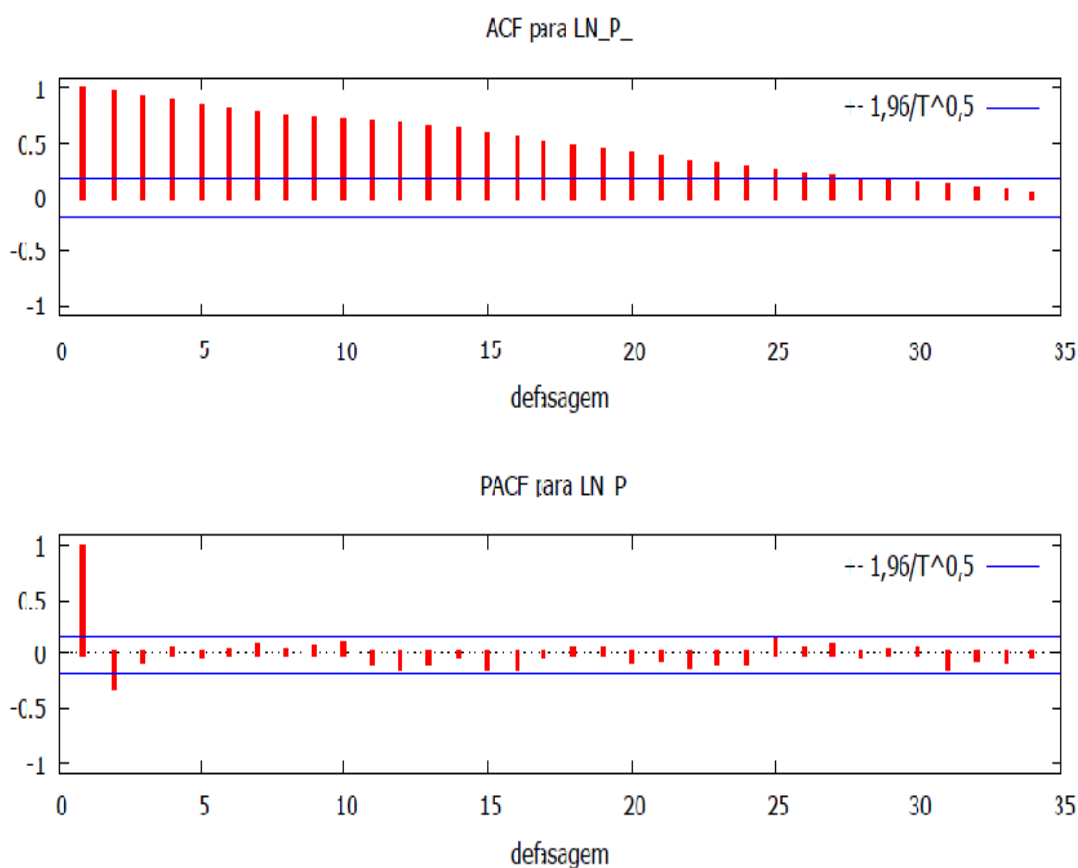


Figura 3.3: Gráfico de Auto-correlação da Série de Preços.

Fonte: Grelt.

O teste ADF é um teste robusto à não normalidade dos resíduos e a presença de heteroscedasticidade. Assim os resíduos não devem ser auto-correlacionados.

Tabela 3.4: Resultados do teste de Auto-correlação.

Função de autocorrelação para LN_P_					
DEFASAGEM	ACF		PACF		Q-stat. [p-valor]
1	0,9770	***	0,9770	***	117,4284 [0,000]
2	0,9412	***	-0,2922	***	227,3376 [0,000]
3	0,9006	***	-0,0524		328,8245 [0,000]
4	0,8613	***	0,0503		422,4515 [0,000]
5	0,8235	***	-0,0102		508,7778 [0,000]
6	0,7879	***	0,0123		588,5004 [0,000]
7	0,7578	***	0,0885		662,8993 [0,000]
8	0,7316	***	0,0096		732,8563 [0,000]
9	0,7110	***	0,0759		799,5306 [0,000]
10	0,6975	***	0,1058		864,2854 [0,000]
11	0,6850	***	-0,0693		927,3028 [0,000]
12	0,6680	***	-0,1124		987,7911 [0,000]
13	0,6449	***	-0,0722		1044,7004 [0,000]
14	0,6188	***	-0,0072		1097,5789 [0,000]
15	0,5867	***	-0,1178		1145,5736 [0,000]
16	0,5475	***	-0,1156		1187,7683 [0,000]
17	0,5051	***	-0,0121		1224,0316 [0,000]
18	0,4639	***	0,0274		1254,9171 [0,000]
19	0,4278	***	0,0647		1281,4463 [0,000]
20	0,3939	***	-0,0635		1304,1645 [0,000]
21	0,3626	***	-0,0412		1323,6097 [0,000]
22	0,3305	***	-0,1017		1339,9255 [0,000]
23	0,2965	***	-0,0699		1353,1953 [0,000]
24	0,2605	***	-0,0739		1363,5471 [0,000]
25	0,2309	**	0,1503		1371,7655 [0,000]
26	0,2059	**	0,0389		1378,3654 [0,000]
27	0,1849	**	0,0889		1383,7463 [0,000]
28	0,1636	*	-0,0180		1388,0037 [0,000]
29	0,1435		0,0134		1391,3163 [0,000]
30	0,1267		0,0431		1393,9269 [0,000]
31	0,1076		-0,1211		1395,8331 [0,000]
32	0,0871		-0,0468		1397,0950 [0,000]
33	0,0638		-0,0592		1397,7801 [0,000]
34	0,0380		-0,0106		1398,0259 [0,000]

Fonte: Grelt.

Na tabela 3.3 temos o resultado do teste de raiz unitária e o gráfico da função de auto-correlação da série, que nos leva a crer que a série é não estacionária.

### 3.2

#### Aplicação dos Modelos

A modelagem utilizada nessa pesquisa adota a aproximação em tempo discreto para a estimação dos parâmetros, partindo da suposição de que os dados são normalmente distribuídos, conforme pode ser confirmado através dos resultados do teste Jarque Bare, apresentado na Tabela 3.2.

Para o tratamento dos processos com saltos (Poisson Gaussiano), utilizou-se a aproximação de Bernoulli, onde se admite a possibilidade de dois tipos de eventos por intervalo de tempo, a ocorrência de salto ou a não ocorrência de salto. O limite da distribuição de Bernoulli é governado por uma distribuição de Poisson, logo a função do processo de Poisson Gaussiano pode ser considerada como uma ocorrência de Bernoulli de distribuições normais, o que nos possibilita aproximar a função de verossimilhança para o modelo de Poisson Gaussiano.

$$\Delta \ln(P) = k(\theta - \ln(P)) \Delta t + \sigma \Delta z + J(\mu, \gamma^2) \Delta \Pi(q)$$

Onde  $\sigma^2$  é a variância dos saltos Gaussianos,  $\Delta z$  é o termo padrão normal de choque,  $\Delta \Pi(q)$  é o incremento de Wiener.  $J(\mu, \gamma^2)$  é o termo que representa os saltos do processo,  $\Delta \Pi(q)$  é o incremento de Poisson para tempo discreto, aproximado pela distribuição de Bernoulli e com parâmetro  $q = h \Delta t + O(\Delta t)$ . Observe que a variável  $q$  pode ser interpretada como a probabilidade de um salto ocorrer no intervalo de tempo  $\Delta t$ . Com isso podemos escrever a seguinte equação de probabilidade de transição (para  $s > t$ ):

$$f[\ln(P_s) | \ln(P_t)] = q \exp\left(\frac{-[\ln(P_s) - \ln(P_t) - \eta(\theta - \ln(P_t))\Delta t - \mu]^2}{2(\sigma^2 \Delta t + \gamma^2)}\right) \frac{1}{\sqrt{2\pi(\sigma^2 \Delta t + \gamma^2)}} \\ + (1 - q) \exp\left(\frac{-[\ln(P_s) - \ln(P_t) - \eta(\theta - \ln(P_t))\Delta t]^2}{2\sigma^2 \Delta t}\right) \frac{1}{\sqrt{2\pi\sigma^2 \Delta t}}$$

Esta equação aproxima a relação verdadeira de densidade de Poisson - Gaussiana a uma mistura de distribuições Normais e pode ser interpretada como uma combinação convexa de duas densidades Normais, sendo que uma delas possui o termo de variância aumentado pela variância do tamanho do salto.<sup>9</sup>

O processo de estimação utilizado nesta pesquisa envolve a maximização da equação acima:

$$\max_{(k, \theta, \mu, \sigma, \gamma^2, q)} \sum_{t=1}^T [\log(f[\ln P_s | \ln P_t])]$$

<sup>9</sup>Rocha, Ana (2008)

Este processo de estimação é conhecido como método de máxima verossimilhança que consiste em adotar, como estimativas dos parâmetros, os valores que maximizam a probabilidade, no caso das variáveis aleatórias discretas ou a densidade de probabilidade para variáveis aleatórias contínuas de ocorrência das observações.

Os modelos estimados foram:

### 3.2.1

#### Modelo Gaussiano puro (MRM puro)

$$\Delta \ln (P_t) = \eta (\theta - \ln (P_{t-1})) \Delta t + \sigma \Delta z$$

$$E(P_t) = \theta - (\ln (P_t) - \theta) e^{-\eta \Delta t}$$

$$\text{Var} (P_t) = \sigma^2/2\eta (1 - e^{-2\eta \Delta t})$$

Onde  $\eta$  é a velocidade de reversão à média;  $\theta$  é média de longo prazo que no caso das commodities agrícolas tendem ao custo marginal de produção;  $\sigma^2$  é a variância do processo;  $\Delta z$  é o incremento de Wiener.

### 3.2.2

#### Modelo de Poisson –Gaussiano (MRM com saltos)

$$\Delta \ln (P_t) = \eta (\theta - \ln (P_{t-1})) \Delta t + \sigma \Delta z + J (\mu, \gamma^2) + \Delta \Pi (q)$$

Nessa equação acrescentasse os termos  $J (\mu, \gamma^2)$ , que representa os processos de saltos que são normalmente distribuídos com média  $\mu$  e variância  $\gamma^2$ , é o tamanho do salto e  $\Delta \Pi (q)$  é o incremento de Poisson em tempo discreto, governado por uma distribuição de Bernoulli com parâmetro  $q$ .

### 3.2.3

#### Modelo Gaussiano puro ARCH

$$\Delta \ln (P_t) = \eta (\theta - \ln (P_{t-1})) \Delta t + \sigma \Delta z$$

$$\text{Var} (P_t) = a_0 + a_1 [ \Delta \ln (P_t) - E [ \Delta \ln (P_t) ] ]^2$$



A equação da variância do modelo Gaussiano puro é alterada pela variância condicional do modelo ARCH.

### 3.2.4

#### Modelo Gaussiano puro GARCH

$$\Delta \ln (P_t) = \eta (\theta - \ln (P_{t-1})) \Delta t + \sigma \Delta z$$

$$\text{Var} (P_t) = a_0 + a_1 [ \Delta \ln (P_t) - E [\Delta (\ln (P_t)) ] ]^2 + a_2 \sigma^2_{t-1}$$

A equação da variância do modelo Gaussiano puro é alterada pela variância condicional do modelo GARCH.