

## 5 Aplicação empírica dos modelos

A aplicação de modelos de difusão com saltos tem sido proeminente no estudo da estrutura a termo da taxa de juros. No entanto, visto que o processo de preços de commodities, notadamente dos produtos agrícolas, é marcado pela evidência de picos e saltos, torna-se extremamente interessante avaliar a consideração destas características do mercado. Sendo assim, o presente estudo justifica-se como um ferramental alternativo e moderno na modelagem de tais processos, uma vez que a literatura de finanças parece ser escassa no que diz respeito à modelagem de commodities agrícolas.

Este capítulo apresenta a metodologia adotada, os conceitos necessários ao seu entendimento, assim como descreve a base de dados utilizada e os resultados obtidos.

### 5.1. O contexto dos processos Poisson-Gaussiano de Das: Síntese do artigo

A metodologia utilizada fundamenta-se na modelagem proposta por Das (1998), quando estimou uma classe de processos Poisson–Gaussiano para o comportamento estocástico da taxa de juros norte-americana. Sendo assim, nesta seção será feita uma breve apresentação a respeito das contribuições do artigo como ferramental para a aplicação dos métodos de difusão com saltos. Neste estudo, Das baseou-se no processo de reversão à média com a inclusão de saltos (*jumps*). A justificativa para a abordagem dos saltos sustentou-se pela observação de fatos estilizados, a destacar o excesso de curtose, do mercado de títulos e, por conseguinte, da estrutura a termo da taxa de juros. De acordo com Das, este excesso de curtose poderia alternativamente ser capturado por meio dos processos Poisson-Gaussiano ou pelos modelos de volatilidade estocástica ou ainda pelos

modelos mais simples com a volatilidade variando no tempo, tais como os processos ARCH.

A análise teórica de Das foi estruturada essencialmente em quatro etapas, sendo elas: **(i)** Derivação da função característica para os modelos propostos; **(ii)** Obtenção das funções de densidade de probabilidade; **(iii)** Cálculo dos quatro primeiros momentos dos processos estocásticos, e por fim, **(iv)** Determinação das expressões analíticas para os preços dos títulos. Em seguida, foi realizada uma inspeção empírica que contou com dados diários de janeiro de 1988 a dezembro de 1997, num total de 2609 observações.

O processo de estimação dos parâmetros dos modelos foi realizado a princípio em tempo contínuo através de procedimentos numéricos. De forma análoga, posteriormente, a estimação foi implementada de modo bem mais simples, em tempo discreto. A discretização do processo estocástico possibilitou tratar a volatilidade através de modelos do tipo ARCH aninhados a difusão de saltos. O autor estendeu esta aproximação do modelo de forma a analisar os efeitos dos dias da semana e das atividades do Banco Central americano (Federal Reserve) para a ocorrência de saltos. Outro processo de estimação empreendido por Das deu-se através da utilização do método dos momentos.

Os resultados de Das corroboraram o desempenho dos modelos de difusão de saltos através de uma melhoria substancial no ajuste aos dados empíricos. Além disso, foi possível identificar através dos saltos evidências de que as reuniões do Fed de fato revelam informação ao mercado e, conseqüentemente, têm influência nos preços dos títulos. Foi constatado também que os efeitos dos dias da semana são bastante significativos, de forma que a ocorrência de saltos é mais provável nos dias de vencimento das opções, influenciando assim na forma dos movimentos do mercado.

## **5.2.**

### **Descrição dos modelos**

A estimação em tempo contínuo constitui-se num processo numérico excessivamente intensivo no qual requer que a função de densidade, a ser submetida a otimização numérica, seja obtida pela inversão de Fourier da função

característica. Sendo assim, dada a sua maior simplicidade, nesta pesquisa optou-se por adotar apenas a aproximação discreta como método de estimação.

Com o intuito de se obter o melhor modelo para representar o comportamento dos preços da commodity café arábica, além do processo estocástico de reversão à média utilizado por Das, aqui a modelagem será estendida de forma a considerar também o processo geométrico browniano bem como sua variante com saltos. Outro acréscimo a abordagem de Das, será admitir que o termo de variância dos processos de difusão siga também estruturas GARCH.

Em tempo discreto, os processos Poisson-Gaussiano adotados nesta pesquisa, tendo por base o movimento de reversão à média e o movimento geométrico browniano, podem ser representados, respectivamente, pelas equações a seguir.

### **Processo Poisson-Gaussiano de Reversão à Média:**

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

onde:

- $\eta$  é a velocidade anual de reversão;
- $\theta$  é a média de longo prazo do processo de reversão;
- $\sigma^2$  é a variância anual do choque Gaussiano;
- $\Delta z$  é o termo de choque normal padrão do processo de difusão. Tal como discutido no capítulo 2, trata-se de um processo de Wiener;
- $J(\mu, \gamma^2)$  é o termo de choque do processo de salto, que é normalmente distribuído com média  $\mu$  e variância  $\gamma^2$ . É interpretado como o tamanho do salto, e
- $\Delta \Pi(q)$  é o incremento de Poisson em tempo discreto, aproximado de uma distribuição Bernoulli com parâmetro  $q$ . Em outras palavras, a chegada de saltos é governada por um processo Poisson  $\Pi$  com parâmetro de frequência de chegada  $q$ , que denota o número de saltos por ano.

### Processo Poisson-Gaussiano Geométrico Browniano:

$$\Delta \ln(P_t) = \alpha \Delta t + \sigma \Delta z + J(\mu, \gamma^2) \Delta \Pi(q), \quad (2)$$

onde  $\alpha$  é o termo de *drift* do processo.

No entanto, ao admitir os efeitos ARCH e GARCH na modelagem da variância dos processos de difusão, as equações (1) e (2) são reformuladas de forma que  $\sigma^2$  seja dado pelas expressões a seguir.

#### Variância condicional do modelo ARCH:

$$\sigma^2_{t+\Delta t} = a_0 + a_1 [\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 \quad (3)$$

#### Variância condicional do modelo GARCH

$$\sigma^2_{t+\Delta t} = a_0 + a_1 [\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 + a_2 \sigma^2_{t-1} \quad (4)$$

Dada a importância do entendimento dos modelos ARCH e GARCH para o processo de estimação, a seguir será apresentada uma descrição sucinta a respeito dos conceitos envolvendo tais modelos.

### Modelos ARCH

Introduzidos por Engle (1982), os modelos da família ARCH (*Autoregressive Conditional Heteroskedasticity*) são modelos não lineares utilizados na determinação da volatilidade de séries temporais, principalmente financeiras, que apresentam a variância condicional evoluindo no tempo.

Seja  $P_t$  o preço de uma commodity no período  $t$  e  $r_t = \ln \frac{P_t}{P_{t-1}}$  o retorno

composto continuamente gerado pela commodity no período entre  $t-1$  e  $t$ . Suponha-se que os retornos gerados possam ser descritos pelo seguinte processo ARMA ( $p, q$ ):

$$r_t = c_t + u_t, \quad c_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} - \sum_{i=1}^q \theta_i u_{t-i}, \quad (5)$$

sendo  $p$  e  $q$  inteiros não negativos e  $u_t$  uma série ruído branco com média zero e variância  $\sigma_u^2$ .

A idéia básica dos modelos ARCH é que a variância incondicional  $u_t$  é serialmente descorrelacionada, porém dependente, e a dependência de  $u_t$  pode ser descrita por uma simples função quadrática de seus valores defasados. Matematicamente, o modelo ARCH de ordem  $m$ , ARCH ( $m$ ), é definido por:

$$u_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \text{Var}[r_t | \Omega_{t-1}] = E[u_t^2 | \Omega_{t-1}] = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_m u_{t-m}^2, \quad (6)$$

onde  $\sigma_t^2$  denota a variância condicional dado o conjunto de informações disponíveis em  $t-1$ ,  $\Omega_{t-1}$ . As variáveis aleatórias  $\varepsilon_t$  são independentes e identicamente distribuídas (iid) com média zero e variância 1. Os coeficientes  $\alpha_i$  devem satisfazer certas condições de regularidade de forma a assegurar que a variância incondicional  $u_t$  seja positiva e fracamente estacionária. Sendo assim, tem-se que  $\alpha_0 > 0, \alpha_i \geq 0$  para  $i > 0$  e  $\sum_{i=1}^m \alpha_i < 1$ . Na prática, é comum assumir que  $\varepsilon_t$  segue a distribuição normal padrão ou uma distribuição t-Student.

Da estrutura do modelo, percebe-se que a volatilidade aparece em grupos, sejam eles de maior ou menor variabilidade. Ou seja, grandes choques quadráticos passados  $\{u_{t-i}^2\}_{i=1}^m$  implicam em grandes variâncias condicionais. Conseqüentemente,  $u_t$  tende a assumir valores grandes (em módulo). Em outras palavras, sob a estrutura ARCH, grandes choques tendem a ser seguidos por outros grandes choques. Esta propriedade explica a aglomeração da volatilidade observada nos retornos dos ativos.

### **Modelos GARCH**

Embora o modelo ARCH seja simples, em geral, dado que existe alta persistência de choques na volatilidade das séries de retornos, a ordem  $m$  do modelo costuma ser elevada. Conseqüentemente, tal fato requer a estimação de um grande número de parâmetros. Sendo assim, Bollerslev (1986) propôs uma extensão do modelo ARCH, na tentativa de expressar de forma mais parcimoniosa a dependência temporal da variância condicional. Trata-se do modelo ARCH

generalizado, ou ainda, modelo GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*).

O modelo GARCH estabelece que a volatilidade de uma série de retornos depende não somente dos quadrados das inovações, mas também das variâncias condicionais anteriores. Como anteriormente, seja  $u_t = r_t - c_t$ . Então  $u_t$  segue um processo GARCH ( $m, s$ ) se:

$$u_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \text{Var}[r_t | \Omega_{t-1}] = E[u_t^2 | \Omega_{t-1}] = \alpha_0 + \sum_{i=1}^m \alpha_i u_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \quad (7)$$

onde  $m$  representa a ordem do componente ARCH e  $s$  a ordem do componente GARCH. Novamente as variáveis aleatórias  $\varepsilon_t$  são independentes e identicamente distribuídas (iid) com média zero e variância 1, sendo assumida a distribuição normal padrão ou t-Student. As restrições impostas aos parâmetros para que a variância do processo seja positiva e fracamente estacionária são:

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \text{ e } \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1, \quad (8)$$

de forma que  $\alpha_i = 0$  para  $i > m$  e  $\beta_j = 0$  para  $j > s$ . A persistência de choques na volatilidade da série de retornos é medida pela soma de  $\alpha_m$  e  $\beta_s$ . Assim, quanto mais próxima de um for a soma, maior o tempo que um choque na série levará para dissipar-se. Na equação (7), fazendo  $s = 0$  o processo reduz-se ao modelo ARCH ( $m$ ) puro. No entanto, de acordo com Bollerslev *et al.* a especificação mais robusta verificada nas aplicações é a do modelo GARCH (1, 1).

### 5.3. Estimação

A viabilidade do processo de estimação do modelo Poisson-Gaussiano em tempo discreto resulta da aproximação de Bernoulli, introduzida por Ball & Torous (1983), na qual é adotado o pressuposto de que em cada intervalo de tempo apenas um salto ocorre ou nenhum salto ocorre. Esta hipótese torna o procedimento de estimação altamente tratável, estável e convergente. Sob condições regulares, as estimativas dos parâmetros são assintoticamente não

viesadas, consistentes e eficientes sendo obtidas de maneira econômica usando a estimação por verossimilhança.

A distribuição de Bernoulli consiste num alicerce para a construção de muitas outras distribuições em probabilidade, inclusive da distribuição de Poisson. Esta relação será apresentada a seguir.

Seja uma variável estocástica que segue a distribuição de Bernoulli; onde  $X = 1$  se um salto ocorre, ou  $X = 0$  caso contrário. Defina-se  $O(\Delta t)$  como o símbolo de ordem assintótica usado para denotar uma função  $\zeta$ , tal que  $\lim_{\Delta t \rightarrow 0} \zeta = \frac{\zeta(\Delta t)}{\Delta t} = 0$ . Seja ainda a variável aleatória  $N(t)$  que representa o número de eventos (saltos) que ocorrem em um intervalo de tempo de tamanho  $t$ . Denote  $\Delta t = t/n$  para um inteiro positivo e arbitrário  $n$  de forma que a subdivisão do intervalo  $(0, t)$  seja dada por  $n$  intervalos iguais de tamanho  $\Delta t$ . A variável  $X_i$  representa o número de eventos que ocorrem em um subintervalo  $i$ . Sob a hipótese de incrementos independentes estacionários, tem-se que  $N(t) = \sum_{i=1}^n X_i$ . Este é um processo de contagem dado pela soma de  $n$  variáveis aleatórias independentes e identicamente distribuídas com as seguintes propriedades:

- $N(0) = 0$
- $Prob[X_i = 0] = 1 - \lambda\Delta t + O(\Delta t)$
- $Prob[X_i = 1] = \lambda\Delta t + O(\Delta t)$  para  $i = 1, 2, \dots, n$ .
- $Prob[X_i > 1] = O(\Delta t)$ ,

onde  $\lambda$  é a taxa de chegada dos saltos.

Para  $n$  grande, a probabilidade de um salto ocorrer num subintervalo  $i$  é desprezível de modo que  $X_i$  é uma variável de Bernoulli com parâmetro  $\lambda\Delta t = \lambda t/n$ . Logo, sendo a soma de  $n$  variáveis de Bernoulli i.i.d.,  $N$  é, aproximadamente, uma variável Binomial com a seguinte distribuição:

$$\text{Prob}[N = k] \equiv \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \quad k = 0, 1, 2, \dots, n. \quad (9)$$

Conseqüentemente, para um subintervalo de tempo muito pequeno, ou seja,  $n$  tendendo a infinito, a distribuição binomial aproxima-se da distribuição de Poisson, conforme a equação a seguir.

$$\lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} = \frac{\exp(-\lambda t)(\lambda t)^k}{k!} \quad k = 0, 1, 2, \dots \quad (10)$$

No entanto, considere agora que o intervalo de tempo  $t$  é muito pequeno. Neste caso,  $N$  pode ser satisfatoriamente aproximado pela variável aleatória de Bernoulli  $X$  com parâmetro  $\lambda t$  definida por:

$$\text{Prob}[X = 0] = 1 - \lambda t \text{ e } \text{Prob}[X = 1] = \lambda t \quad (11)$$

Sendo assim, visto que o limite do processo Binomial é governado por uma distribuição de Poisson, a função de densidade do processo Poisson-Gaussiano pode ser estabelecido através de uma mistura Bernoulli de distribuições Normais, conduzindo a difusão e os choques de salto. As probabilidades de transição para o logaritmo natural do preço, seguindo um processo de difusão com saltos (do tipo MRM) de acordo com a equação (1), são escritas como (para  $s > t$ ):

$$f[\ln(P_s) | \ln(P_t)] = q \exp\left(\frac{-[\ln(P_s) - \ln(P_t) - \eta(\theta - \ln(P_t))\Delta t - \mu]^2}{2(\sigma^2 \Delta t + \gamma^2)}\right) \frac{1}{\sqrt{2\pi(\sigma^2 \Delta t + \gamma^2)}} + \\ (1-q) \exp\left(\frac{-[\ln(P_s) - \ln(P_t) - \eta(\theta - \ln(P_t))\Delta t]^2}{2\sigma^2 \Delta t}\right) \frac{1}{\sqrt{2\pi\sigma^2 \Delta t}} \quad (12)$$

Enquanto que para o processo de difusão de saltos baseado no MGB, apresentado na equação (2), a expressão da densidade fica:

$$f[\ln(P_s) | \ln(P_t)] = q \exp\left(\frac{-[\ln(P_s) - \ln(P_t) - \alpha\Delta t - \mu]^2}{2(\sigma^2 \Delta t + \gamma^2)}\right) \frac{1}{\sqrt{2\pi(\sigma^2 \Delta t + \gamma^2)}} + \\ (1-q) \exp\left(\frac{-[\ln(P_s) - \ln(P_t) - \alpha\Delta t]^2}{2\sigma^2 \Delta t}\right) \frac{1}{\sqrt{2\pi\sigma^2 \Delta t}} \quad (13)$$

A variável  $q$ , probabilidade de um salto ocorrer no intervalo  $\Delta t$ , é análoga ao parâmetro  $\lambda$  para a intensidade de chegada do salto, pela relação  $q \approx \lambda \Delta t$ , onde



$q = \lambda \Delta t + O(\Delta t)$ . Esta relação é responsável por aproximar a verdadeira densidade Poisson-Gaussiana a uma mistura de distribuições Normais. As expressões (12) e (13) podem ser interpretadas como uma combinação convexa de duas densidades Normais, sendo que uma delas tem o termo de variância aumentado pela variância do tamanho do salto.

O trabalho empírico da estimação foi realizado através do método da máxima verossimilhança, que consiste em maximizar o logaritmo da função de densidade de transição para os vetores de parâmetros dos processos, conforme expressão a seguir.

$$\max \sum_{t=1}^T [\log(f[\ln P_s | \ln P_t])] \quad (14)$$

para o MRM :  $(\eta, \theta, \mu, \sigma, \gamma^2, q)$   
para o MGB :  $(\alpha, \mu, \sigma, \gamma^2, q)$

Primeiramente, as estimativas dos parâmetros dos processos foram obtidas utilizando-se a ferramenta *Solver* do Excel. Em seguida, os valores alcançados foram empregados como solução inicial para o desenvolvimento da estimação no software *Eviews* (versão 4.1), permitindo assim o conhecimento das estatísticas de testes dos parâmetros.

Sabe-se, no entanto, que os mecanismos tradicionais de otimização por vezes não são suficientemente robustos, embora tal fato não implique que na prática eles não sejam utilizados com sucesso em certas situações. Dentre as barreiras enfrentadas por estes mecanismos, podem-se destacar os seguintes aspectos: Utilização de um único ponto no espaço de soluções, que é manipulado iterativamente sob algumas heurísticas (regras de transição determinística); além das limitações sobre o espaço de busca no que se refere a condições de continuidade e existência de derivadas. Estes fatores podem afetar os resultados de forma que não se tem garantias da convergência para soluções ótimas em níveis globais.

Deste modo, visto que os Algoritmos Genéticos são muito eficientes para busca de soluções ótimas, ou aproximadamente ótimas em uma grande variedade de problemas, optou-se por utilizá-los como uma alternativa ao processo de estimação dos parâmetros. Para tal tarefa, foi utilizado o software *RiskOptimizer for Excel* (versão 1.0). Os resultados da estimação e a comparação entre o método

tradicional (do *Solver* e do *Eviews*) e os Algoritmos Genéticos (do *RiskOptimizer*) serão examinados mais a frente (seção 5.5). Porém, a seguir serão apresentados os conceitos básicos a respeito dos Algoritmos Genéticos.

### **Algoritmos Genéticos (AG's)**

Os Algoritmos Genéticos, desenvolvidos a princípio por John Holland (1975), são métodos computacionais de busca e otimização baseados na teoria da evolução natural e da genética. Seu procedimento de busca difere-se das técnicas de otimização tradicional, uma vez que operam paralelamente sobre uma população de soluções e utilizam regras de transição probabilísticas.

A aplicação dos Algoritmos Genéticos é recomendada em problemas de otimização com espaços de busca muito complexos, que envolvem um grande número de variáveis e, por conseguinte espaços de soluções de dimensões elevadas. Além disso, os Algoritmos Genéticos produzem resultados satisfatórios em termos de convergência (possibilidade de escapar de máximos ou mínimos locais), mesmo quando outras estratégias de otimização falham. Outra vantagem dos AG's deve-se a sua robustez, uma vez que não são um método sensível a erros de arredondamento.

O ponto de partida para a utilização de Algoritmos Genéticos em um problema qualquer é a representação das possíveis soluções em uma estrutura de dados que possa ser manipulada computacionalmente. Em geral, estas representações são formadas por uma seqüência de símbolos gerados a partir de um alfabeto finito. Tradicionalmente, o alfabeto binário é o mais utilizado, onde os elementos da cadeia de bits denotam a presença (1) ou ausência (0) de uma característica.

Em analogia a genética biológica, tem-se que a seqüência de caracteres é um cromossomo (ou genótipo) e cada elemento equivale a um gene. Os cromossomos representam os indivíduos e o conjunto de indivíduos uma população. Estes indivíduos correspondem a um ponto no espaço de soluções do problema de otimização. O processo adotado nos Algoritmos Genéticos consiste em gerar, através de regras específicas, um grande número de indivíduos, população, de forma a promover uma varredura tão extensa quanto necessária do espaço de soluções.

Em linhas gerais o funcionamento dos Algoritmos Genéticos pode ser explicado em termos da representação do problema, do uso de três operadores (seleção, cruzamento e mutação), e de uma função de avaliação de aptidão. A estrutura básica dos Algoritmos Genéticos consiste nas seguintes etapas:

- 1) **Inicialização:** Gera-se de maneira aleatória uma população de  $N$  cromossomos (soluções candidatas para o problema). O tamanho da população ( $N$ ) é um dos parâmetros de um AG, sendo importante que a população inicial cubra a maior área possível do espaço de busca, sem interessar se são boas ou não.
- 2) **Avaliação:** Durante o processo evolutivo esta população é avaliada de modo a refletir o seu grau de adaptação a função objetivo. O intuito é descobrir quais indivíduos serão escolhidos para reproduzirem-se e sobreviver. Deste modo, ainda nesta fase os indivíduos são ordenados conforme sua aptidão.
- 3) **Nova população:** Cria-se uma nova população através da aplicação dos operadores genéticos. Estes operadores garantem a transformação da população através de sucessivas gerações, estendendo a busca até chegar num resultado satisfatório. Eles são necessários para que a população se diversifique e preserve características de adaptação adquiridas pelas gerações anteriores. Os operadores genéticos são:
  - **Seleção** – Baseado no princípio de seleção natural Darwiniano, o operador seleção privilegia a sobrevivência dos mais aptos. A aptidão dos indivíduos é diretamente proporcional à probabilidade destes serem selecionados e contribuir para a criação de indivíduos em uma próxima geração. Este operador tem como característica básica a sua concepção probabilística.
  - **Cruzamento (Cross-over)** – Após o processo de seleção, os indivíduos são escolhidos aleatoriamente aos pares para serem recombinados. O operador de cross-over executa uma troca de

informações entre dois indivíduos a partir do conteúdo de seus cromossomos, originando novos descendentes com material genético de seus progenitores. A aplicação ou não do cross-over após a escolha de dois cromossomos é probabilística, isto é, dois novos híbridos podem ser produzidos, ou então os dois indivíduos inicialmente selecionados são preservados. Esta probabilidade, denotada por probabilidade de cross-over, é especificada pelo usuário. Na literatura recomenda-se que a taxa de cross-over esteja na faixa de 60 a 65%.

- **Mutação:** A operação de mutação é utilizada para garantir uma maior varredura do espaço de estados e evitar que o Algoritmo Genético convirja muito cedo para ótimos locais. A mutação é realizada alternando-se o valor de um gene de um indivíduo sorteado aleatoriamente com uma determinada probabilidade, conhecida por probabilidade de mutação. Deste modo, vários indivíduos da nova população podem ter seus genes alterados aleatoriamente. Geralmente, a taxa de mutação fica em torno de 0,1 a 5%.

Por fim, outro aspecto relevante refere-se à substituição da população antiga pela nova geração, o que pode ser feita total ou parcialmente. No último caso, apenas os piores indivíduos são substituídos, recebendo o nome de Steady-State Reproduction.

#### 5.4. Dados

A commodity escolhida para testar a metodologia proposta refere-se ao café arábica<sup>6</sup>. A série histórica de preços à vista do café foi obtida no site do Centro de Estudos Avançados em Economia Aplicada da Escola Superior de Agricultura Luiz de Queiroz (CEPEA/ ESALQ), da Universidade de São Paulo. O parágrafo a seguir descreve a metodologia utilizada pelo CEPEA no cômputo deste indicador.

O indicador CEPEA/ESALQ de café arábica é apurado diariamente através das cotações em reais de preços praticados por saca de 60 quilos líquidos (R\$/ saca) entre corretores, cooperativas, torrefadores e exportadores, referentes às negociações no mercado físico de lotes entre empresas. Estes preços são coletados nas principais praças de comercialização do país (Cerrado, Mogiana (SP), Sul de Minas Gerais, Paulista (SP) e noroeste do Paraná), sendo que os valores são posto São Paulo, sempre com referência a Bolsa de Nova York. Os preços fornecidos pelos agentes atuantes em diferentes segmentos da cadeia são os faturados para pagamento em 7 dias, a retirar. Uma vez coletados, os preços são descontados pelo período de pagamento através da NPR (Nota Promissória Rural), e acrescidos dos custos de frete da região de origem até a praça de São Paulo, local de entrega. O indicador de preços é calculado por uma média ponderada das médias regionais, sendo o fator de ponderação proporcional à participação de cada região, anteriormente mencionada, na produção de café. O valor do indicador não considera impostos. O indicador em dólares americanos é calculado pela taxa de câmbio comercial para venda cotado às 16h30.

---

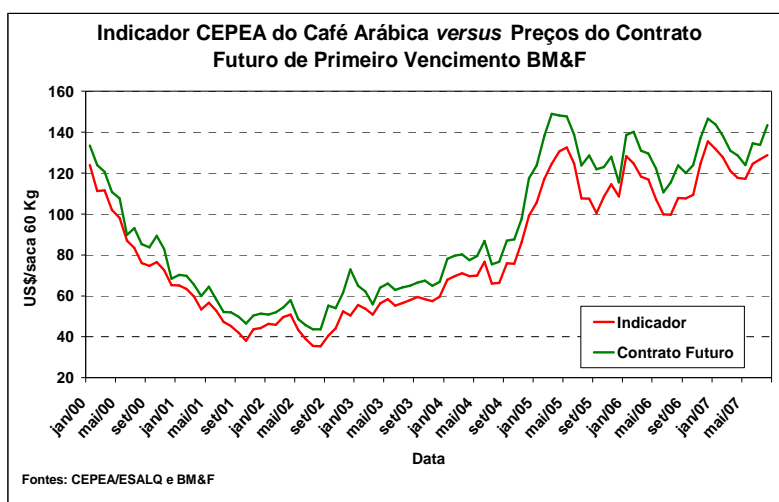
<sup>6</sup> A classificação do café é uma fase muito importante para seu processo de comercialização, pois através da análise dos grãos é que são definidos os valores financeiros do produto. Essencialmente, existem duas formas de classificação: (i) por tipo ou defeito (com base na contagem de grãos defeituosos e impurezas contidas numa amostra de 300g, de acordo com a Tabela Oficial Brasileira de Classificação); e (ii) pela qualidade, na qual são levados em consideração fatores como cor, torração, descrição, peneira e, principalmente, a bebida (avaliação das características aromáticas e gustativas).

O padrão do café arábica objeto de pesquisa do indicador CEPEA/ESALQ atende às especificações do café negociado na BM&F, ou seja, trata-se do café arábica tipo 6, bebida dura para melhor. Tal descrição denota, em termos da classificação mencionada acima, um café com 86 grãos defeituosos numa amostra de 300g (tipo 6) e que apresenta gosto acre, adstringente e áspero, podendo exibir sabor superior (bebida dura para melhor) um pouco mais suave, com adstringência ou aspereza mais leves, mas não podendo ter sabor de iodofórmio (características de cafés rio/riado).

O indicador CEPEA/ESALQ, utilizado para registrar os preços praticáveis no mercado físico brasileiro, representa um mecanismo de apoio e elemento referencial para análises de mercado, refletindo indiretamente as oscilações dos preços no mercado futuro. No caso do café arábica, a comparação entre físico e futuro é favorecida pela similaridade entre o padrão do produto coletado pelo CEPEA e o negociado pela BM&F (Ribeiro, Sousa e Rogers, 2006).

Sendo assim, embora o mercado físico alterne períodos de menor liquidez com momentos onde a liquidez é maior (quando cafeicultores precisam fazer caixa para custear a colheita, por exemplo), pelas razões explicitadas acima, justifica-se a utilização da série à vista nesta dissertação. A figura 5.1 ilustra o comportamento dos preços à vista e do contrato futuro de primeiro vencimento negociado na BM&F.

**Figura 5.1** – Comparação entre os preços à vista e futuro do café arábica (2000 a 2007)



Os dados do indicador CEPEA/ESALQ são disponibilizados diariamente a partir de 02/09/1996 em reais (R\$/ saca). No entanto, foi calculada a média dos preços de cada mês a fim de se trabalhar com observações mensais. Para a estimação dos modelos em estudo foi utilizada uma amostra que compreende o período de setembro/1996 a dezembro/2004, compondo uma base de dados de 100 meses (ou equivalentemente, 2068 dias). Para efeito de comparação da previsão fora da amostra, foram separadas algumas observações, que abrangem o período

de janeiro/2005 a agosto/2007, constituindo um total de 32 meses (ou semelhantemente, 667 dias).

Com o objetivo de que os dados refletissem apenas flutuações de seu valor intrínseco e não do processo econômico como um todo, a série de preços foi deflacionada para agosto/2007. O deflator adotado foi o Índice de Preços ao Atacado- Disponibilidade Interna (IPA-DI- Média Geral), obtido no site da Fundação Getúlio Vargas (FGV Dados). Na tabela 5.1 são apresentadas algumas estatísticas descritivas básicas da série de preços.

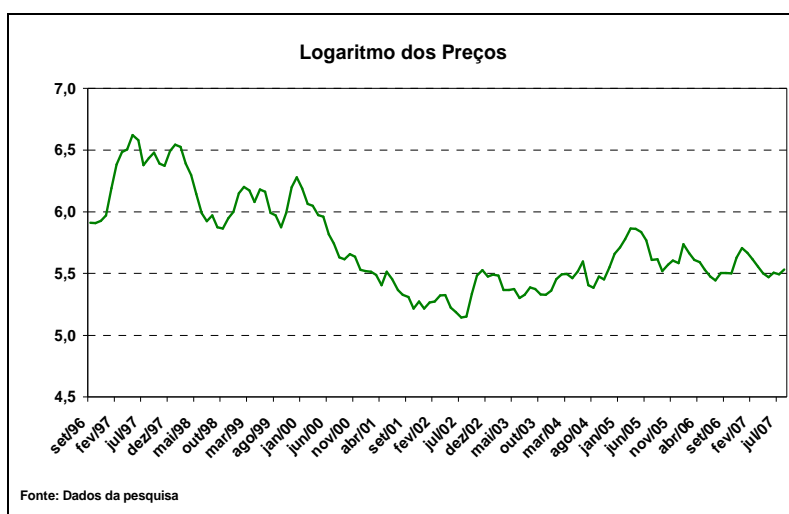
**Tabela 5.1** – Estatísticas descritivas da série de preços deflacionada do café no período de 09/1996 a 08/2007 (132 meses)

Estatísticas	
Média	331,0896
Mediana	273,8188
Máximo	751,3885
Mínimo	171,2582
Desvio padrão	140,1879
Assimetria	1,2472
Curtose	3,6425

Fonte: Dados da pesquisa

Devido à grande variabilidade das observações, optou-se ainda em trabalhar com a série em escala logarítmica, cujo comportamento é ilustrado na figura 5.2.

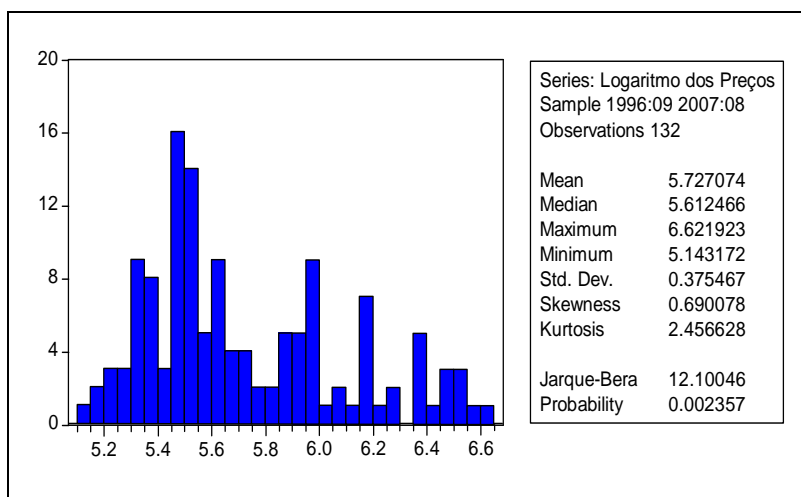
**Figura 5.2** – Logaritmo dos preços da commodity café arábica (1996 a 2007)



Por uma simples inspeção visual dos dados, percebe-se uma tendência de decréscimo dos preços no período de maio de 1997 a julho de 2002, a partir do qual a variável apresenta uma leve recuperação.

As estatísticas descritivas da série logarítmica são apresentadas na figura 5.3. Ao analisar a forma da distribuição dos dados em relação à distribuição Normal, os coeficientes de assimetria e curtose revelam, respectivamente, uma distribuição levemente assimétrica à direita (assimetria positiva) e platicúrtica (mais achatada em sua parte superior). O teste de Jarque-Bera mostra que a série não é normalmente distribuída, pois ao nível de 5% de significância a hipótese nula de normalidade é rejeitada.

**Figura 5.3** – Histograma e estatísticas descritivas da série do logaritmo dos preços do café no período de 09/1996 a 08/2007: Out put do software Eviews



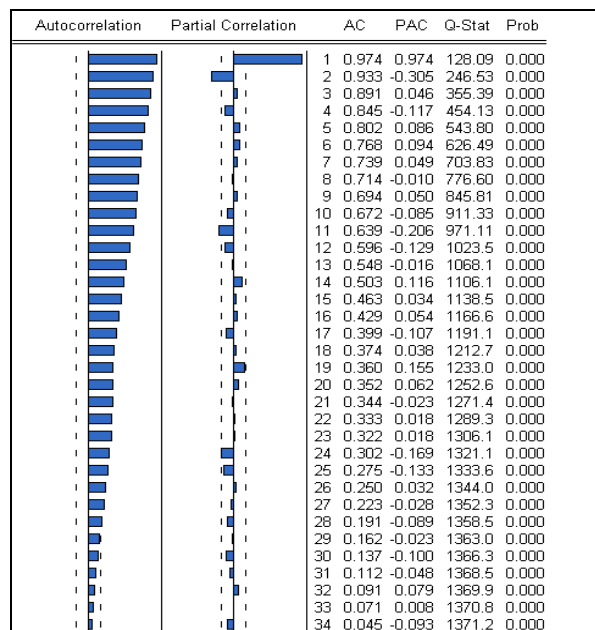
Fonte: Dados da pesquisa

Antes de se proceder com a modelagem propriamente dita, realizou-se uma análise a fim de saber se a série pode ser considerada estacionária ou não. Para esta análise, foi utilizado o gráfico da função de autocorrelação da série e a aplicação de um método formal de teste de raiz unitária. O teste aplicado foi o de Dickey Fuller aumentado (ADF - Augmented Dickey-Fuller). O decaimento lento da função na figura 5.4 representa um indício de não estacionariedade. Tal comportamento, portanto, é constatado ao se observar a magnitude do p-valor atribuído ao teste ADF na figura 5.5, o que leva fortemente a não rejeitar a hipótese de não estacionariedade da série em questão. No entanto, cabe ressaltar que este resultado não apresenta nenhuma implicação para a modelagem



empregada. A análise da estacionariedade foi procedida apenas como um adicional a investigação do comportamento da série, realizada anteriormente.

**Figura 5.4** – Função de autocorrelação da série de preços do café



Fonte: Dados da pesquisa

**Figura 5.5** – Resultado do teste de Dickey-Fuller: Out put do Eviews

Null Hypothesis: In do Preço has a unit root

Exogenous: Constant

Lag Length: 1 (Automatic based on HQ, MAXLAG=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.725856	0.4159
Test critical values: 1% level	-3.481217	
5% level	-2.883753	
10% level	-2.578694	

\*MacKinnon (1996) one-sided p-values.

Fonte: Dados da pesquisa

## 5.5.Resultados

Nesta seção serão reportados e analisados os resultados da estimação pelo método da máxima verossimilhança das variações consideradas a partir do modelo de difusão de saltos de Das. Estas variações consistem em seis modelos, que foram estimados tanto para o processo de reversão à média quanto para o movimento geométrico browniano, decorrentes das equações gerais dadas em (1) e (2) misturadas aos processos ARCH e GARCH descritos em (3) e (4),

respectivamente. Os resultados da estimação para o período de setembro de 1996 a dezembro de 2004 são apresentados nas tabelas 5.2 (para o MRM) e 5.3 (para o MGB). No entanto, para um melhor entendimento destes resultados, antes de apresentá-los será feita uma breve exposição de cada modelo, destacando os parâmetros a serem estimados em cada um.

### **Processo Gaussiano Puro de Reversão à Média:**

$$\Delta \ln(P_t) = \eta[\theta - \ln(P_{t-1})]\Delta t + \sigma\Delta z, \quad (1.1)$$

onde os parâmetros a serem estimados são:  $\eta$  (velocidade anual de reversão);  $\theta$  (média de longo prazo do processo de reversão) e  $\sigma^2$  (variância anual do choque Gaussiano).

### **Processo ARCH Gaussiano de Reversão à Média:**

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

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1[\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 \quad (1.2)$$

De forma que os parâmetros a serem estimados são os mesmos do modelo (1.1) acrescidos dos parâmetros  $a_0$  e  $a_1$ , constantes da variância do processo ARCH.

### **Processo GARCH Gaussiano de Reversão à Média:**

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

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1[\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 + a_2\sigma^2_{t-1} \quad (1.3)$$

Sendo a constante  $a_2$  adicionada ao conjunto de parâmetros a serem estimados pelo modelo (1.2)

### **Processo Poisson-Gaussiano de Reversão à Média:**

$$\Delta \ln(P_t) = \eta[\theta - \ln(P_{t-1})]\Delta t + \sigma\Delta z + J(\mu, \gamma^2)\Delta \Pi(q), \quad (1.4)$$

Além dos parâmetros do processo de difusão  $\eta$ ,  $\theta$  e  $\sigma^2$ ; no processo Poisson-Gaussiano são ainda estimados os parâmetros  $\mu$ ,  $\gamma^2$  (respectivamente, média e variância da variável tamanho do salto) e  $q$  (frequência de chegada dos saltos).

**Processo ARCH Poisson-Gaussiano de Reversão à Média:**

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

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1[\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 \quad (1.5)$$

Os parâmetros a serem estimados são:  $\eta$ ,  $\theta$ ,  $\sigma^2$ ,  $\mu$ ,  $\gamma^2$ ,  $q$ ,  $a_0$  e  $a_1$ .

**Processo GARCH Poisson-Gaussiano de Reversão à Média:**

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

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1[\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 + a_2\sigma^2_{t-1} \quad (1.6)$$

Os parâmetros a serem estimados são:  $\eta$ ,  $\theta$ ,  $\sigma^2$ ,  $\mu$ ,  $\gamma^2$ ,  $q$ ,  $a_0$ ,  $a_1$  e  $a_2$ .

**Processo Gaussiano Puro Geométrico Browniano:**

$$\Delta \ln(P_t) = \alpha\Delta t + \sigma\Delta z, \quad (2.1)$$

onde os parâmetros a estimar são:  $\alpha$  e  $\sigma^2$  (termo de *drift* e variância do processo, respectivamente).

**Processo ARCH Gaussiano Geométrico Browniano:**

$$\Delta \ln(P_t) = \alpha\Delta t + \sigma\Delta z,$$

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1[\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 \quad (2.2)$$

De forma que os parâmetros a serem estimados são os mesmos do modelo (2.1) acrescidos dos parâmetros  $a_0$  e  $a_1$ , constantes da variância do processo ARCH.

**Processo GARCH Gaussiano Geométrico Browniano:**

$$\Delta \ln(P_t) = \alpha \Delta t + \sigma \Delta z,$$

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1 [\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 + a_2 \sigma^2_{t-1} \quad (2.3)$$

Os parâmetros a serem estimados são:  $\alpha$ ,  $\sigma^2$ ,  $a_0$ ,  $a_1$  e  $a_2$ .

**Processo Poisson-Gaussiano Geométrico Browniano:**

$$\Delta \ln(P_t) = \alpha \Delta t + \sigma \Delta z + J(\mu, \gamma^2) \Delta \Pi(q), \quad (2.4)$$

Os parâmetros a serem estimados são  $\alpha$  e  $\sigma^2$ , pertinentes ao processo de difusão; além dos parâmetros do processo de saltos  $\mu$ ,  $\gamma^2$  e  $q$ .

**Processo ARCH Poisson-Gaussiano Geométrico Browniano:**

$$\Delta \ln(P_t) = \alpha \Delta t + \sigma \Delta z + J(\mu, \gamma^2) \Delta \Pi(q),$$

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1 [\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 \quad (2.5)$$

Os parâmetros a serem estimados são os mesmos da equação (2.4) juntamente com as constantes  $a_0$  e  $a_1$  do processo ARCH.

**Processo GARCH Poisson-Gaussiano Geométrico Browniano:**

$$\Delta \ln(P_t) = \alpha \Delta t + \sigma \Delta z + J(\mu, \gamma^2) \Delta \Pi(q),$$

$$\text{onde } \sigma^2_{t+\Delta t} = a_0 + a_1 [\Delta \ln(P_t) - E[\Delta \ln(P_t)]]^2 + a_2 \sigma^2_{t-1} \quad (2.6)$$

Os parâmetros a serem estimados são:  $\alpha$ ,  $\sigma^2$ ,  $\mu$ ,  $\gamma^2$ ,  $q$ ,  $a_0$ ,  $a_1$  e  $a_2$ .

**Tabela 5.2 – Resultados da estimação (MRM)**

Parâmetros	Gaussiano Puro			ARCH Gaussiano			GARCH Gaussiano			Poisson Gaussiano			ARCH Poisson Gaussiano			GARCH Poisson Gaussiano		
	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG
$\eta$	0,2618	0,2618	0,2618	0,3095	0,3100	0,3095	0,3095	0,3101	0,3095	0,4680	0,4680	0,4675	0,4857	0,4982	0,4807	0,4801	0,4908	0,4596
t-Statistic	1,0117			1,2077			1,2080			1,8568			1,9832			1,8035		
p-value	0,3117			0,2272			0,2271			0,0633			0,0473			0,0713		
$\theta$	5,6489	5,6489	5,6489	5,5081	5,5066	5,5082	5,5081	5,5065	5,5082	5,0024	5,0024	5,0004	5,0011	4,9706	4,7390	5,0066	4,9816	4,8333
t-Statistic	11,6458			11,7589			11,7587			8,0261			7,9720			7,8659		
p-value	0,0000			0,0000			0,0000			0,0001			0,0000			0,0000		
$\sigma^2$	0,0900	0,0900	0,0900							0,0551	0,0551	0,0551						
t-Statistic	13,4559									3,1623								
p-value	0,0000									0,0016								
$\mu$										0,1359	0,1359	0,1358	0,1293	0,1289	0,0967	0,1308	0,1304	0,1000
t-Statistic										4,7247			3,8168			3,7496		
p-value										0,0000			0,0001			0,0002		
$\gamma^2$										0,0000	0,0000	0,0000	0,0000	0,0000	0,0021	0,0000	0,0000	0,0015
t-Statistic										*			*			*		
p-value										*			*			*		
$q$										0,2001	0,2001	0,2007	0,1840	0,1976	0,3513	0,1803	0,1915	0,2870
t-Statistic										-1,6122			-1,3029			-1,3524		
p-value										0,1069			0,1926			0,1762		
$a_0$				0,0817	0,0827	0,0817	0,0817	0,0827	0,0817				0,0523	0,0513	0,0478	0,0459	0,0450	0,0503
t-Statistic				-15,5476			-15,5476						-7,7503			-2,1887		
p-value				0,0000			0,0000						0,0000			0,0286		
$a_1$				1,0000	1,0000	1,0000	1,0000	1,0000	1,0000				1,0000	1,0000	1,0000	0,8834	0,8806	0,9685
t-Statistic				*			*						*			0,1785		
p-value				*			*						*			0,8584		
$a_2$							0,0000	0,0000	0,0000							0,1166	0,1194	0,0315
t-Statistic							*									**		
p-value							*									**		
<b>Log-likelihood</b>	101,7246	101,7246	101,7246	102,3310	102,3281	102,3310	102,3310	102,3281	102,3310	102,9029	102,9029	102,9029	103,1685	103,1613	103,1296	103,1922	103,1870	103,1387

Fonte: Dados da pesquisa

\*Os parâmetros foram fixados conforme os valores apresentados na tabela/ \*\*O valor do parâmetro saiu por diferença, uma vez que no modelo GARCH  $\sum a_1 + a_2 = 1$

**Tabela 5.3 – Resultados da estimação (MGB)**

Parâmetros	Gaussiano Puro			ARCH Gaussiano			GARCH Gaussiano			Poisson Gaussiano			ARCH Poisson Gaussiano			GARCH Poisson Gaussiano		
	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG	Solver	Eviews	AG
$\alpha$	-0,0304	-0,0304	-0,0304	-0,0656	-0,0656	-0,0656	-0,0656	-0,0656	-0,0656	-0,2561	-0,2562	-0,2576	-0,2378	-0,2684	-0,2289	-0,2310	-0,2316	-0,2257
t-Statistic	-0,2872			-0,6031			-0,6034			-0,4713			-0,3345			-0,3289		
p-value	0,7740			0,5464			0,5463			0,6374			0,7380			0,7422		
$\sigma^2$	0,0910	0,0910	0,0910							0,0696	0,0696	0,0695						
t-Statistic	13,6648									-4,5534								
p-value	0,0000									0,0000								
$\mu$										0,1135	0,1135	0,1134	0,1016	0,0984	0,1029	0,1046	0,1060	0,1042
t-Statistic										1,5643			1,2135			0,8978		
p-value										0,1177			0,2249			0,3693		
$\gamma^2$										0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
t-Statistic										*			*			*		
p-value										*			*			*		
$\eta$										0,1657	0,1657	0,1670	0,1438	0,1754	0,1349	0,1356	0,1374	0,1312
t-Statistic										-0,4798			-0,2919			-0,3238		
p-value										0,6374			0,7704			0,7461		
$a_0$				0,0832	0,0841	0,0832	0,0832	0,0841	0,0832				0,0682	0,0680	0,0686	0,0621	0,0543	0,0660
t-Statistic				-15,7268			-15,7270						-3,8133			-1,5344		
p-value				0,0000			0,0000						0,0001			0,1249		
$a_1$				1,0000	1,0000	1,0000	1,0000	1,0000	0,9998				0,9809	0,9170	0,9814	0,9144	0,7962	0,9644
t-Statistic				*			*						0,1074			0,1682		
p-value				*			*						0,9145			0,8664		
$a_2$							0,0000	0,0000	0,0000							0,0856	0,2038	0,0356
t-Statistic							*									**		
p-value							*									**		
<b>Log-likelihood</b>	101,1737	101,1737	101,1737	101,6235	101,6206	101,6235	101,6235	101,6206	101,6235	101,5033	101,5033	101,5033	101,7811	101,7748	101,7809	101,7893	101,7742	101,7863

Fonte: Dados da pesquisa

\*Os parâmetros foram fixados conforme os valores apresentados na tabela/ \*\*O valor do parâmetro saiu por diferença, uma vez que no modelo GARCH  $\Sigma_{a_1+a_2}=1$

A validação das variáveis dos modelos estimados foi empregada através do teste estatístico *t* de Student e de seu *p*-valor associado, conforme as tabelas 5.2 e 5.3. Através da análise do *p*-valor, ao nível de significância de 1% e 5%, pôde-se constatar que o único modelo, dentre os 12 modelos estimados, a apresentar todas as estimativas dos parâmetros estatisticamente significativas foi o modelo GARCH Poisson Gaussiano Geométrico Browniano.

Conseqüentemente, verificou-se que a não significância dos parâmetros faz com que um modelo decaia sobre o outro. Sendo assim, observou-se que:

#### **Para a classe de modelos Reversão à Média**

- Os modelos ARCH Gaussiano Puro e GARCH Gaussiano Puro, onde o único parâmetro significativo (tanto ao nível de 1% como a 5% de significância) é a velocidade de reversão ( $\eta$ ), reduzem-se ao modelo Gaussiano Puro. Já para os processos de difusão com saltos, o modelo ARCH Poisson Gaussiano, cujos parâmetros significativos (ao nível de 1% de significância) são a velocidade de reversão ( $\eta$ ) e a frequência dos saltos ( $q$ ), reduz-se ao modelo Poisson Gaussiano. No entanto, o modelo GARCH Poisson Gaussiano é um caso a parte, no qual ao nível de 1% de significância, apresenta os seguintes parâmetros significativos:  $\eta$ ,  $q$ ,  $a_0$  e  $a_1$ . Logo, este modelo não recai em nenhum outro anterior.

#### **Para a classe de modelos Geométrico Browniano**

- Assim como os modelos de Reversão à Média, a não significância de todos os parâmetros dos modelos implica na redução a um modelo mais simples. Deste modo, para os processos de difusão puro (sem saltos), os modelos ARCH Gaussiano e GARCH Gaussiano apresentam, ao nível de 1% e 5% de significância, apenas o termo de drift ( $\alpha$ ) como significativo; e assim tais modelos reduzem-se ao modelo Gaussiano Puro. Porém, o mesmo não acontece para os modelos de difusão com saltos. O modelo ARCH Poisson Gaussiano apresenta, tanto ao nível de 1% como a 5% de significância, os parâmetros  $\alpha$ ,  $\mu$ ,  $q$ , e  $a_1$  como significativos. Enquanto, que o modelo GARCH Poisson Gaussiano

tem todos os parâmetros significativos. Sendo assim, tais modelos não sobrepõem nenhum outro.

Ainda analisando os resultados obtidos, constatou-se que o procedimento de estimação realizado através da ferramenta *Solver* do Excel apresentou os melhores resultados para a função de verossimilhança, enquanto que o *Eviews* e os Algoritmos genéticos por ora mostraram-se iguais ou ligeiramente inferiores. Por conseguinte, a comparação do desempenho do ajuste dos modelos será baseada nos resultados do *Solver*.

Embora os Algoritmos genéticos apresentem-se como uma alternativa de caráter genérico aos problemas de otimização, com soluções satisfatórias; não há como assegurar seu sucesso. Os AG's tratam-se de métodos de busca "cega", uma vez que não têm conhecimento específico do problema a ser resolvido e guiam-se apenas pela função objetivo. São métodos codificados que não trabalham diretamente com o domínio do problema e sim com representações dos seus elementos. Apresentam ainda características estocásticas (aleatórias), que combinam em proporção variável regras probabilísticas e determinísticas.

Deste modo, tais fatos podem ser uma justificativa para o fato dos AG's não terem encontrado solução melhor que o método tradicional. É provável que os AG's tenham convergido para um máximo local, visto que para aquela região ele constitui-se na melhor solução; apesar de não ser a melhor solução dentro do espaço de busca.

Para avaliar o poder de aderência dos modelos aos dados (*goodness of fit*), foram calculadas três diferentes medidas de erro.

- $$\text{Root Mean Squared Error (RMSE)} = \sqrt{\frac{\sum_{t=1}^T (\text{série}_t - \text{previsão}_t)^2}{T}} \quad (15)$$

- $$\text{Mean Absolute Deviation (MAD)} = \frac{\sum_{t=1}^T |\text{série}_t - \text{previsão}_t|}{T} \quad (16)$$

- $$\text{Mean Absolute Percentage Error (MAPE)} = \frac{1}{T} \sum_{t=1}^T \left| \frac{\text{série}_t - \text{previsão}_t}{\text{série}_t} \right| * 100\% \quad (17)$$

Cabe ressaltar que os erros (RMSE, MAD e MAPE) foram analisados fora da amostra, considerando tanto a série de preços como a série de seus logaritmos.



Porém, para uma validação mais adequada dos modelos é relevante a adoção de critérios que considerem as quantidades de parâmetros de acordo com o princípio da parcimônia. Em geral, à medida que um número maior de parâmetros pode melhorar o ajuste aos dados, o modelo pode apresentar um elevado grau de complexidade e, conseqüentemente, a probabilidade de se obter previsões menos precisas torna-se maior. Logo, um bom modelo mantém o equilíbrio entre o ajuste e o número de preditores. Os critérios aplicados de acordo com este princípio foram:

- *Akaike Information Criterion (AIC)* =  $-2(\ell/T) + 2(k/T)$  (18)

- *Bayesian Information Criterion (BIC)* =  $-2(\ell/T) + k \ln(T)/T$  (19)

- *Hannan–Quinn Criterion (HQ)* =  $-2(\ell/T) + 2k \ln[\ln(T)]/T$  (20),

onde  $\ell$  significa o valor da função de verossimilhança com  $k$  parâmetros estimados usando  $T$  observações. Estes critérios de informação foram calculados de acordo com as definições adotadas pelo *Eviews*, pelas quais são estabelecidas que -2 vezes a função de verossimilhança deve ser ajustada por uma função de penalização. Os resultados obtidos, tanto para as medidas de erro como para os critérios, encontram-se nas tabelas 5.4 e 5.5.

Os valores encontrados para os erros (Mean Absolute Deviation - MAD) da série de preços do café (para o período de jan/05 a ago/07) ficaram em torno de R\$14,89/saca para os modelos MGB e R\$15,00/saca para os modelos MRM. Embora estes valores sejam bem maiores quando comparados aos erros obtidos pelo logaritmo da série; tais resultados mostram-se plausíveis, visto que a série dos preços para o período em questão fica em torno de R\$275,57/saca de 60 kg.

O modelo selecionado será aquele que minimiza os valores dos critérios. Para ambos os casos (MRM e MGB), embora os modelos tenham apresentado um desempenho bastante semelhante, pode-se verificar uma tênue melhoria para o processo Gaussiano Puro. Além disso, percebe-se também que o processo oriundo do movimento geométrico browniano tem um ajuste um pouco melhor. No entanto, através de uma simples inspeção visual (figura 5.6) da comparação out-of-sample entre os modelos do MRM e MGB não se consegue apontar tais diferenças.

Por fim, o modelo a ser adotado para representar o processo de preços do café arábica no período em questão é o processo Gaussiano Puro baseado no movimento geométrico browniano.

Conforme relatado na literatura, inclusive em Das, o comportamento leptocúrtico da distribuição dos dados consiste num indício da presença de saltos, justificando a necessidade da aplicação dos modelos de difusão de saltos. Todavia, na análise descritiva realizada anteriormente, viu-se que os dados desta pesquisa comportam-se de acordo com uma distribuição platicúrtica. Sendo assim, o fato de nenhum modelo com salto ter se mostrado melhor que o processo Gaussiano Puro consiste numa conclusão plausível. Outro resultado da modelagem que vai de acordo com as análises anteriores refere-se a não estacionariedade da série, que condiz com a característica do movimento geométrico browniano.

**Tabela 5.4** – Resumo do desempenho dos modelos no ajuste aos dados (MRM)

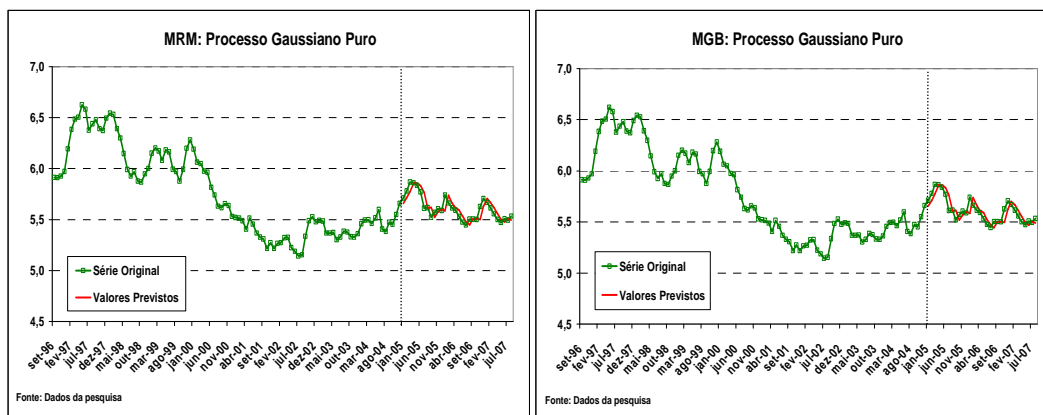
Série dos preços do café						
Erro	Gaussiano Puro	ARCH Gaussiano	GARCH Gaussiano	Poisson Gaussiano	ARCH Poisson Gaussiano	GARCH Poisson Gaussiano
RMSE	19,1188	19,3916	19,3916	20,6797	20,8599	20,8031
MAD	14,4005	14,5004	14,5004	15,4299	15,5713	15,5273
MAPE (%)	5,3234	5,3946	5,3946	5,8512	5,9172	5,8966
Série do logaritmo dos preços						
Erro	Gaussiano Puro	ARCH Gaussiano	GARCH Gaussiano	Poisson Gaussiano	ARCH Poisson Gaussiano	GARCH Poisson Gaussiano
RMSE	0,0651	0,0648	0,0648	0,0649	0,0645	0,0645
MAD	0,0527	0,0520	0,0520	0,0530	0,0522	0,0523
MAPE (%)	0,9375	0,9253	0,9253	0,9421	0,9282	0,9292
Critérios de validação baseados no princípio da parcimônia						
Critério	Gaussiano Puro	ARCH Gaussiano	GARCH Gaussiano	Poisson Gaussiano	ARCH Poisson Gaussiano	GARCH Poisson Gaussiano
AIC	<b>-1,9745</b>	-1,9666	-1,9466	-1,9381	-1,9234	-1,9038
BIC	<b>-1,8963</b>	-1,8624	-1,8164	-1,7817	-1,7410	-1,6954
HQ	<b>-1,9429</b>	-1,9244	-1,8939	-1,8748	-1,8496	-1,8195

Fonte: Dados da pesquisa

**Tabela 5.5** – Resumo do desempenho dos modelos no ajuste aos dados (MGB)

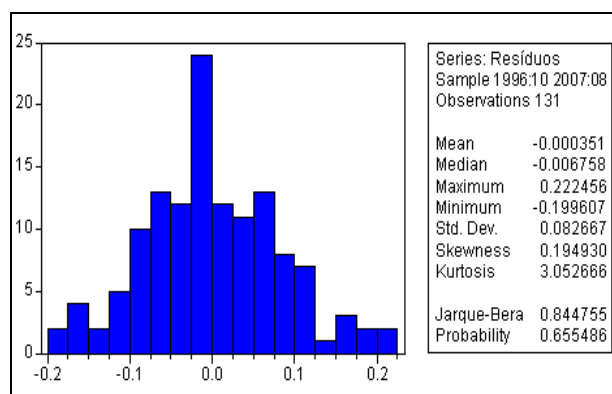
Medidas de erro calculadas para a série dos preços do café						
Erro	Gaussiano Puro	ARCH Gaussiano	GARCH Gaussiano	Poisson Gaussiano	ARCH Poisson Gaussiano	GARCH Poisson Gaussiano
RMSE	18,7091	18,7089	18,7089	18,7091	18,7089	18,7089
MAD	14,8903	14,8898	14,8898	14,8903	14,8898	14,8899
MAPE (%)	5,3250	5,3249	5,3249	5,3250	5,3249	5,3249
Medidas de erro calculadas para a série do logaritmo dos preços						
Erro	Gaussiano Puro	ARCH Gaussiano	GARCH Gaussiano	Poisson Gaussiano	ARCH Poisson Gaussiano	GARCH Poisson Gaussiano
RMSE	0,0656	0,0656	0,0656	0,0656	0,0656	0,0656
MAD	0,0526	0,0522	0,0522	0,0526	0,0522	0,0522
MAPE (%)	0,9362	0,9285	0,9285	0,9362	0,9288	0,9290
Critérios de validação baseados no princípio da parcimônia						
Critério	Gaussiano Puro	ARCH Gaussiano	GARCH Gaussiano	Poisson Gaussiano	ARCH Poisson Gaussiano	GARCH Poisson Gaussiano
AIC	<b>-1,9835</b>	-1,9725	-1,9525	-1,9301	-1,9156	-1,8958
BIC	<b>-1,9314</b>	-1,8943	-1,8483	-1,7998	-1,7593	-1,7134
HQ	<b>-1,9624</b>	-1,9408	-1,9103	-1,8773	-1,8524	-1,8220

Fonte: Dados da pesquisa

**Figura 5.6** – Avaliação do poder preditivo dos modelos selecionados (Processos Gaussiano Puro): Comparação out-of-sample entre os modelos MRM e MGB

Na figura 5.7 é apresentado o resultado do teste Jarque-Bera aplicado aos resíduos, calculados tanto dentro como fora da amostra (131 observações), da série do logaritmo dos preços ajustada pelo modelo Gaussiano Puro (MGB). Com um p-valor de 0,6555 associado a estatística Jarque-Bera, verifica-se que a hipótese de normalidade é sustentada aos níveis de 1% e 5% de significância.

**Figura 5.7** – Histograma e estatísticas descritivas dos resíduos da série



Fonte: Dados da pesquisa

Para as correlações seriais também foram considerados os resíduos da série dos logaritmos dos preços, dentro e fora da amostra. Percebe-se pela figura 5.8 que as autocorrelações são não significativas, e assim conclui-se que os erros não apresentam qualquer tipo de estrutura que o modelo não tenha captado.

**Figura 5.8** – Função de autocorrelação dos resíduos

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.302	0.302	12.223	0.000	
2	0.019	-0.080	12.269	0.002	
3	0.058	0.085	12.732	0.005	
4	-0.099	-0.158	14.067	0.007	
5	-0.199	-0.128	19.527	0.002	
6	-0.148	-0.070	22.597	0.001	
7	-0.062	0.006	23.134	0.002	
8	-0.114	-0.108	24.981	0.002	
9	0.010	0.068	24.995	0.003	
10	0.245	0.202	33.628	0.000	
11	0.211	0.082	40.078	0.000	
12	0.077	-0.026	40.946	0.000	
13	-0.063	-0.158	41.534	0.000	
14	-0.081	-0.029	42.515	0.000	
15	-0.164	-0.084	46.551	0.000	
16	-0.064	0.109	47.174	0.000	
17	-0.094	-0.104	48.518	0.000	
18	-0.201	-0.137	54.722	0.000	
19	-0.109	-0.069	56.570	0.000	
20	0.023	0.001	56.652	0.000	
21	0.073	-0.007	57.503	0.000	
22	0.026	-0.071	57.610	0.000	
23	0.148	0.151	61.157	0.000	
24	0.198	0.175	67.535	0.000	
25	-0.036	-0.082	67.743	0.000	
26	0.045	0.041	68.080	0.000	
27	0.104	0.060	69.881	0.000	
28	-0.067	-0.032	71.149	0.000	
29	-0.092	0.080	72.586	0.000	
30	-0.018	-0.025	72.644	0.000	
31	-0.064	-0.085	73.367	0.000	
32	-0.074	-0.046	74.334	0.000	
33	0.106	0.079	76.332	0.000	
34	0.140	0.007	79.833	0.000	
35	0.063	-0.009	80.562	0.000	
36	0.022	-0.062	80.654	0.000	

Fonte: Dados da pesquisa

Finalmente, apresenta-se na figura 5.9 os valores previstos num horizonte de previsão de 12 meses (setembro de 2007 a agosto de 2008) e os intervalos de confiança para estas previsões com 95% de confiança.

**Figura 5.9** – Previsão do logaritmo dos preços 12 passos à frente e o intervalo de confiança para os valores previstos

