

## 4. Modelagem

Em um modelo estatístico adequado para se evidenciar a existência de uma relação “lead-lag” entre as variáveis à vista e futura de um índice é necessário primeiramente verificar se as variáveis logaritmo neperiano dos preços à vista e futuro possuem raízes unitárias, conseqüentemente, as séries seriam não estacionárias. Em seguida, verificar se as variáveis logaritmo neperiano dos retornos à vista e futuro não possuem raízes unitárias, conseqüentemente, as séries seriam estacionárias. As variáveis logaritmo neperiano dos retornos à vista e futuro são dadas, respectivamente, por:

$$s_t = \ln(S_t / S_{t-1}) \quad (3)$$

$$f_t = \ln(F_t / F_{t-1}) \quad (4)$$

### 4.1. Estacionaridade

Uma série é dita estacionária quando não há comportamento de tendência definido e ela freqüentemente cruza sua média (zero). Uma série com essas características também é conhecida como ruído branco. Já uma série não estacionária exhibe movimentos que a faz se distanciar de sua média, que ela cruza raramente ou nunca, dependendo do tipo de não estacionaridade (Brooks 2002). Uma série dita não estacionária é caracterizada pelo não decaimento dos erros passados, isto é, os valores dos ruídos de períodos anteriores vão ter um efeito não descendente nos valores correntes da variável dependente.

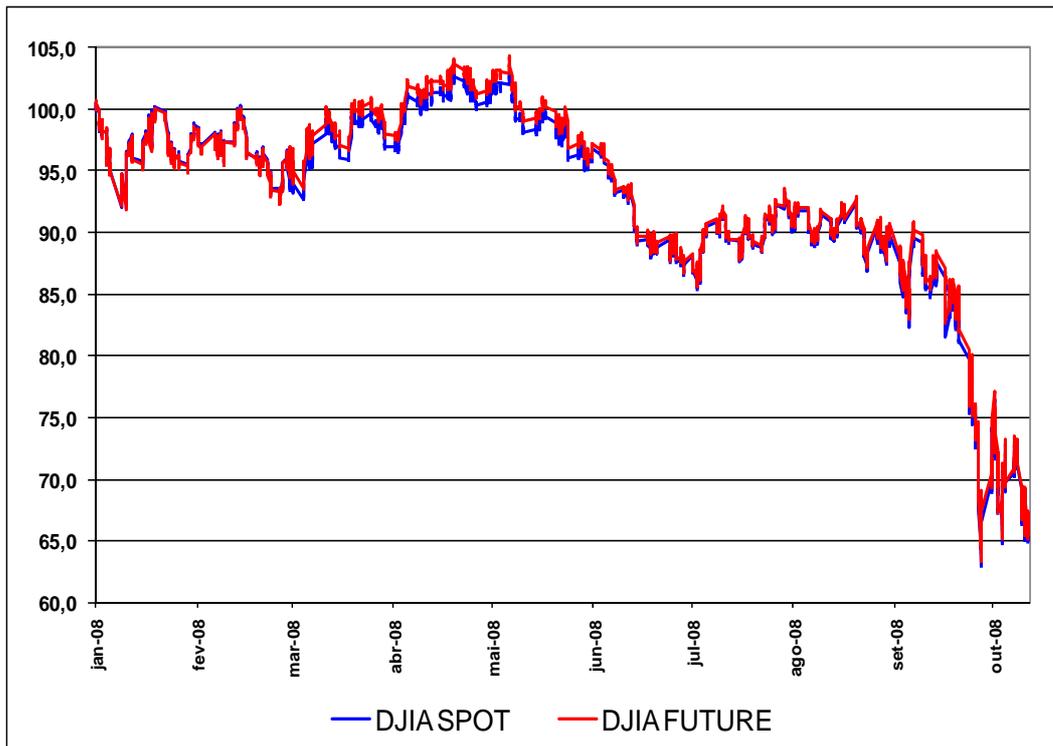


Figura 1 – Exemplo de séries não estacionárias. Dow Jones Industrial Average à vista e futuro.

Fonte: Bloomberg

## 4.2. Teste de Dickey-Fuller

Para se verificar as estacionaridades das séries log-preço e log-retorno dos índices à vista e futuro, basta rodar um teste de Dickey-Fuller Aumentado, também conhecido como Augmented Dickey-Fuller Test ou ADF test, que tem como hipótese nula a existência de raiz unitária ( $H_0: \varphi = 1$ ), assim, se o p-valor resultante do teste de Dickey-Fuller Aumentado for menor que 5%, rejeita-se a hipótese nula de existência de raiz unitária com um grau de confiança de 95% e com isso conclui-se que as séries em questão são de fato estacionárias (Brooks 2002). A equação da teoria de Dickey-Fuller pode ser expressa por:

$$y_t = \varphi y_{t-1} + u_t \quad (5)$$

$$H_0 : \varphi = 1$$

### 4.3. Teste de Duas Etapas de Engle-Granger

A teoria por trás da relação de “lead-lag” exige que haja um equilíbrio de longo prazo e para isso não basta que a formulação contenha somente diferenciais de primeira ordem, é necessário que haja uma relação intertemporal entre as variáveis à vista e futura. A solução, então, é verificar se existe uma relação de co-integração entre as séries de log-preço dos índices futuro e à vista (Brooks, Rew e Ritson 2001) e para isso, utilizamos a metodologia de Duas Etapas de Engle-Granger. Esta consiste em rodar uma regressão simples utilizando o modelo de Mínimos Quadrados Ordinários que se pode expressar por:

$$\ln S_t = \gamma_0 + \gamma_1 \ln F_t \quad (6)$$

E em seguida rodar o teste de Dickey-Fuller Aumentado para o resíduo dessa regressão, caso a série do resíduo se verifique estacionária, fica comprovado a relação de co-integração entre as séries de log-preço dos índices futuro e à vista.

Comprovada a relação de co-integração entre as séries futuro e à vista do índice, o Teorema de Representação Granger anuncia que há um Modelo de Correção de Erro correspondente. O Modelo de Correção de Erro para o log-preço à vista e futuro pode ser expresso por:

$$\Delta \ln S_t = \beta_0 + \delta \hat{z}_{t-1} + \sum_{i=1}^r \beta_i \Delta \ln S_{t-i} + \sum_{j=1}^s \alpha_j \Delta \ln F_t + \varepsilon_t \quad (7)$$

Onde  $\hat{z} = \ln S_t - \gamma_0 - \gamma_1 \ln F_t$  são os resíduos da regressão de primeira etapa dos log-preços (o termo de correção de equilíbrio).

#### 4.4. Teste de Causalidade de Granger

Todavia, existe a possibilidade que as séries de log-preço dos índices futuro e à vista possuam uma relação de causalidade bi-direcional, ou seja, uma série implica na outra e vice-versa. Para testarmos essa possibilidade, optamos por rodar o Teste de Causalidade de Granger que consiste em usar um teste-F para verificar se dados passados de uma variável Y possuem alguma informação estatisticamente significativa sobre outra variável X na presença de dados passados de X. Se possuir, dizemos que Y causa X.

Existem várias maneiras de se implementar o Teste de Causalidade de Granger. Um método simples seria usar a especificação autoregressiva de uma autoregressão vetorial bivariada (SAS Institute Inc 2009). Assuma um período passado  $p$  e estime a equação sem restrição que se segue pela metodologia de Mínimos Quadrados Ordinários:

$$x_t = c_1 + \sum_{i=1}^p \alpha_i x_{t-i} + \sum_{i=1}^p \beta_i y_{t-i} + u_t \quad (8)$$

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Conduza um teste-F da hipótese nula estimando a equação restritiva a seguir também pela metodologia dos Mínimos Quadrados Ordinários:

$$x_t = c_t + \sum_{i=1}^p \gamma_i x_{t-i} + \varepsilon_t \quad (9)$$

Compare suas respectivas somas dos resíduos quadrados e se a estatística do teste for maior que um valor crítico específico, então rejeitamos a hipótese nula de Y não causar X.

#### 4.5. Metodologia dos Mínimos Quadrados em Dois Estágios

Caso haja comprovação de uma relação de causalidade bi-direcional das séries testadas, teremos que refazer o teste de Duas Etapas de Engle-Granger e rodar uma regressão utilizando a metodologia dos Mínimos Quadrados em Dois Estágios, pois a regressão com a metodologia dos Mínimos Quadrados Ordinários fica comprometida devido ao problema de simultaneidade.

A metodologia dos Mínimos Quadrados em Dois Estágios é um método de regressão para cobrir modelos que violem as hipóteses de recursividade da regressão pela metodologia dos Mínimos Quadrados Ordinários, especificamente modelos onde o termo de distúrbio da variável dependente é correlacionado com as causas das variáveis independentes (Oliveira 2008).

A metodologia dos Mínimos Quadrados em Dois Estágios é definida por (1) um estágio onde variáveis dependentes ou endógenas novas são criadas para substituir as variáveis originais, e (2) um estágio onde a regressão é operacionalizada pela metodologia dos Mínimos Quadrados Ordinários, no entanto, usando as variáveis dependentes novas. O propósito do primeiro estágio é criar variáveis dependentes novas que não violem as hipóteses de recursividade da regressão pela metodologia dos Mínimos Quadrados Ordinários.

#### 4.6. Teste de Co-integração de Johansen e Juselius

Outro modelo para se evidenciar a relação de co-integração entre as séries futuro e à vista do índice, como o objetivo detectar se há relacionamento de longo prazo entre as variáveis, seria o Teste de Co-integração de Johansen e Juselius (Johansen e Juselius 1990). Sendo a hipótese nula de que existem pelo menos  $r$  vetores de co-integração representada matematicamente por:

$$H_0: \lambda_i = 0 \quad i = r + 1, \dots, n \quad (10)$$

Ou seja, somente os primeiros  $r$  autovalores ( $\lambda$ ) são diferentes de zero. Essa restrição pode ser imposta para diferentes valores de  $r$ . O próximo passo consiste

na comparação do valor do logaritmo da função de verossimilhança do modelo com restrição relativamente ao logaritmo da função de verossimilhança do modelo sem restrição. Esse teste é denominado de estatística traço e é representado em termos algébricos como:

$$\lambda_{\text{traço}} = -2\log(Q) \quad (11)$$

Onde  $Q$  = (função de verossimilhança restrita maximizada dividida pela função de verossimilhança sem restrição maximizada).

Apesar do Teste de Co-integração de Johansen apresentar também a estatística Máximo Autovalor, focaremos apenas na primeira estatística nesse trabalho.

#### 4.7. Modelo Vetorial de Correção de Erro

Um Modelo Vetorial de Correção de Erro ou Vector Error Correction Model (VECM) pode vir a ser necessário caso o teste de Co-integração de Johansen evidencie a presença de um vetor de co-integração. Esse modelo pode ser representado por:

$$\nabla z_t = \Gamma_1 \nabla z_{t-1} + \dots + \Gamma_{k-1} \nabla z_{t-k+1} + \Pi z_{t-k} + \Phi D_t + u_t \quad (12)$$

Onde,  $\Gamma_i = -(I - A_1 - \dots - A_i)$ , ( $i = 1, \dots, k - 1$ ), e  $\Pi = -(I - A_1 - \dots - A_k)$ .

O Modelo Vetorial de Correção de Erro especifica as dinâmicas de curto prazo para cada variável no sistema em um arcabouço teórico em que direciona as dinâmicas para a relação de equilíbrio de longo prazo sugerida pela teoria econômica. A existência de um equilíbrio de longo prazo não previne a existência de desvios no curto prazo. Phillips (Phillips 1998) mostrou que projeções baseadas em um Modelo Vetorial de Correção de Erro que explicitamente estimam a relação de co-integração e raízes unitárias são consistentes e assintoticamente ótimas. Empiricamente, a literatura de projeções (forecasting) tende a respaldar a superioridade dos Modelos Vetoriais de Correção de Erro para

projeções de um horizonte de longo prazo, embora essa vantagem não seja tão clara para o curto prazo.

#### 4.8. Teste de Heterocedasticidade de White

Um teste e suma importância que deve ser implementado neste trabalho é o Teste de Heterocedasticidade de White para verificar a existência de autocorrelação dos resíduos das regressões. Segundo Brooks (Brooks 2002), o teste de White é obtido através de uma regressão auxiliar, onde a variável dependente é composta pelo vetor dos resíduos, e as independentes são as variáveis explicativas, seus valores elevados ao quadrado e os produtos cruzados destas mesmas variáveis. Por fim, a estatística é calculada multiplicando o  $R^2$  da regressão auxiliar pelo tamanho da amostra. Esta estatística segue assintoticamente distribuição Qui-quadrada, com grau de liberdade igual ao número de regressores (excluindo o termo constante) da regressão auxiliar. O Teste de Heterocedasticidade de White tem como hipótese nula a homocedasticidade dos resíduos.

#### 4.9. Modelo GARCH

Caso venhamos a verificar a existência de resíduos heterocedásticos, teremos de rodar um teste do modelo GARCH (Generalised Autoregressive Conditionally Heteroscedastic Model) que trata de variáveis não-lineares. Com o modelo GARCH é possível interpretar a variância corrente como uma função ponderada de um valor médio de longo prazo, informação de volatilidade de períodos passados e a variância do modelo durante períodos passados (Brooks 2002). Podemos representar esse modelo por:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (13)$$

Onde  $\sigma_t^2$  é a variância condicional. Usando o modelo GARCH é possível interpretar a variância corrente como uma função ponderada de um valor médio de longo prazo (dependente de  $\alpha_0$ ), informação sobre a volatilidade de períodos passados ( $\alpha_i u_{t-i}^2$ ) e a variância do modelo durante períodos passados ( $\beta_j \sigma_{t-j}^2$ ).