

3 Metodologia

3.1. Tipo de pesquisa

Este trabalho pode ser classificado como uma pesquisa descritiva, quantitativa, bibliográfica, experimental, exploratória e explicativa.

Este estudo é descritivo, pois é feita uma análise das características de uma determinada população, os preços da commodity cobre listado na LME.

Quantitativo devido ao seu caráter numérico, sendo quase tudo exemplificado, mostrado e concluído com a ajuda de números e de recursos estatísticos, como: percentagem, moda, média, mediana, desvio-padrão, variância, coeficiente de correlação, análise de gráficos e de regressões,...).

Bibliográfico, pois todo o trabalho está baseado em livros, artigos, teses e trabalhos científicos.

A pesquisa exploratória constitui-se em um elemento chave para este trabalho, uma vez que proporcionou maiores informações sobre o tema e serviu como uma etapa preliminar para a motivação da realização da pesquisa. Segundo Andrade (2005), a pesquisa exploratória complementada pela bibliográfica é aquela que visa descobrir novas idéias, e dar um novo enfoque para o trabalho pesquisado. Tem a finalidade de definir os objetivos de um trabalho, facilitar a delimitação de um tema e se constitui em uma fase preparatória para outro tipo de pesquisa.

Ainda, de acordo com os objetivos propostos por este trabalho, percebe-se seu caráter explicativo, a qual tem por finalidade aprofundar o conhecimento da realidade, procurando a razão, o porquê das coisas (ANDRADE, 2005).

Neste trabalho, portanto, as pesquisas descritiva, exploratória e experimental realizadas inicialmente motivaram o aprofundamento na pesquisa sobre o tema. Quando complementada pela revisão bibliográfica, subsidiam a realização da análise empírica dos dados, a partir do que, é possível alcançar os objetivos gerais e específicos propostos inicialmente.

3.2. Os dados

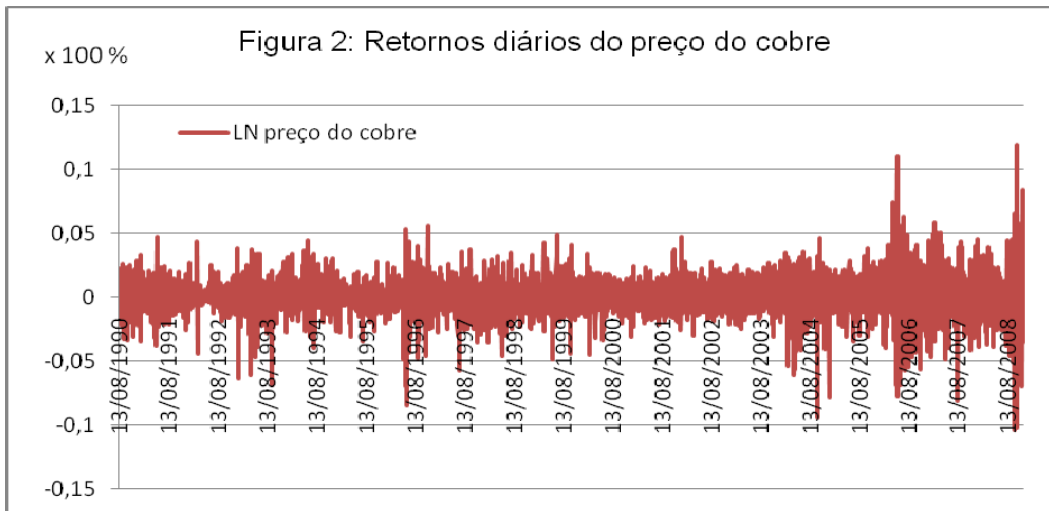
O trabalho está baseado na análise das séries temporais dos preços spot da commodity cobre listados na LME entre os dias 13/08/1990 e 10/12/2008, ou seja, 4635 observações. Utilizou-se a série de preços spot devido a sua grande liquidez, diminuindo as distorções especulativas e facilitando o estudo com cotações em todos os dias úteis da amostra.

Como uma série qualquer de preços de commodity só tem valores positivos, essa série não pode ter uma distribuição normal que varia de menos infinito até mais infinito. Para se tentar deixar a série com uma distribuição normal calculamos o log dos retornos diários.

Foi utilizada a função logarítmica da série de retornos diários do preço do cobre. Segundo o Teorema Central do Limite, para uma amostra grande o suficiente, uma série log(retornos diários) tende a uma distribuição normal. Para isso, foi usada a equação a seguir:

$$r_t = \ln \frac{x_t}{x_{t-1}} \quad (1)$$

Onde x_t é o preço no dia t e x_{t-1} o preço no dia anterior. Como podemos analisar, trabalhando com a função logarítmica dos retornos diários, perde-se a primeira observação, passando a ter 4634 observações. Contudo essa perda é compensada com a maior estabilidade da variância e com a melhor semelhança da nova amostra com uma normal. Essa melhora será mostrada mais a frente. A seguir, vemos o gráfico da distribuição normal:



Fonte: Elaboração Própria

O trabalho foi realizado com os preços nominais, ou seja, sem descontar a inflação. Isto devido à dificuldade de se encontrar um índice de inflação mundial, existindo basicamente índices regionais, além disso, saber qual índice melhor influência o mercado de cobre. Mesmo que se escolhesse um índice qualquer, esses índices são mensais, então a série de dados diários teria que ser transformada em mensal através de preços médios ou finais, com isso o tamanho da amostra seria dividido por aproximadamente 22 (dias úteis em um mês). Como o trabalho deseja apenas comprovar a melhoria de eficiência utilizando o modelo Box & Jenkins, não haverá grandes perdas em se usar os preços nominais.

3.3. Teste de normalidade

O teste de normalidade mais utilizado segundo Brooks (2002) é o teste de Jarque-Bera (JB). Ele consiste na comparação da assimetria e curtose da série testada com a assimetria e curtose de uma normal, através da seguinte estatística:

$$JB = n \left[\frac{A^2}{6} + \frac{(C-3)^2}{24} \right] \quad (2)$$

Onde A é a assimetria, C a curtose e n o tamanho da amostra.

O teste JB mede o excesso de assimetria e curtose com relação à normal. A assimetria permite medir a extensão a qual a distribuição possui simetria em relação à própria média da série. A curtose compara a distribuição de frequências da série com a distribuição normal (LAPPONI, 2000). Nesse sentido, a distribuição normal é simétrica e mesocúrtica. A estatística JB segue uma distribuição qui-quadrada com dois graus de liberdade. Para cada valor de JB temos uma probabilidade p associada à estatística qui-quadrada. Se o valor p for demasiadamente baixo, rejeitamos a hipótese nula de que a amostra é normal. Resumidamente, pode-se assumir que o p valor é a probabilidade da amostra ser normal.

3.4. Teste da estacionaridade

De acordo com Brooks (2002), um processo é considerado estacionário se sua média, variância e autovariância se mantêm constantes ao longo do tempo. Na prática, o que acontece é que os dados flutuam ao longo de uma média constante, independente do tempo (PINTO, 2006).

Para se verificar se uma série é estacionária ou não se pode aplicar o teste da raiz unitária. Se for encontrada uma raiz unitária em uma série, rejeita-se a hipótese nula de que a amostra é estacionária.

H_0 : comportamento estacionário; não existem raízes unitárias;

H_1 : comportamento não-estacionário; existem raízes unitárias.

Suponha que os valores de y possam ser encontrados a partir da seguinte regressão:

$$y_t = \rho \ln(x_t) + \varepsilon_t \quad (3)$$

Para a qual se tem: $\rho \geq 1$ comportamento não-estacionário;

$\rho < 1$ comportamento estacionário.

O teste de raiz unitária utilizado no trabalho é o teste aumentado de Dickey-Fuller e com ele é possível estimar o valor de ρ .

3.5. Metodologia Box & Jenkins

Conforme Moretti & Toloi (1987), uma série temporal é qualquer conjunto de observações ordenadas no tempo.

De acordo com DeLurgio (1998) e Makridakis et al (1998), nenhum outro modelo de previsão univariável tem sido tão amplamente discutido quanto à construção dos modelos ARIMA, também sendo expresso como modelo auto-regressivo integrado de médias móveis. Este tipo de modelo econométrico foi desenvolvido inicialmente por G. P. Box e G. M. Jenkins. Segundo Box & Jenkins (1974) sua metodologia consiste em um modelo univariado, ou seja, que baseia em uma única série temporal. A metodologia Box & Jenkins parte do princípio de que os modelos podem ser construídos a partir da informação contida nos próprios dados.

O modelo auto-regressivo de ordem p – AR(p) é usado quando há autocorrelação entre as observações, ou seja, de acordo com Gujarati (2000), o processo auto-regressivo é usado quando o valor de uma variável n o período t depende de seu valor no período anterior ($t-1$) e de um termo aleatório. A forma geral do modelo auto-regressivo de p -ésima ordem, AR(p) pode ser definida como:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + u_t \quad (4)$$

Onde u_t é um erro aleatório do tipo ruído branco. Sendo assim, o exemplo mais simples de um processo auto-regressivo é o chamado AR(1):

$$y_t = \alpha_1 y_{t-1} + u_t \quad (5)$$

O modelo de média móvel de ordem q – MA(q) é usado quando há autocorrelação entre os resíduos, ou seja, existe uma relação de dependência entre o conjunto de erros em períodos passados (DELURGIO, 1998). O caso geral de um processo MA(q) é definido como:

$$y_t = \mu - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \beta_3 u_{t-3} - \dots - \beta_q u_{t-q} + u_t \quad (6)$$

Onde μ é uma constante. O exemplo mais simples de um processo de média móvel é o chamado MA(1):

$$y_t = \mu + u_t - \beta_1 u_{t-1} \quad (7)$$

O modelo auto-regressivo de média móvel – ARMA(p,q) é usado quando há auto-correlação entre as observações e auto-correlação entre os resíduos. A forma geral de um modelo ARMA(p,q) é:

$$y_t = \theta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \dots - \beta_q u_{t-q} \quad (8)$$

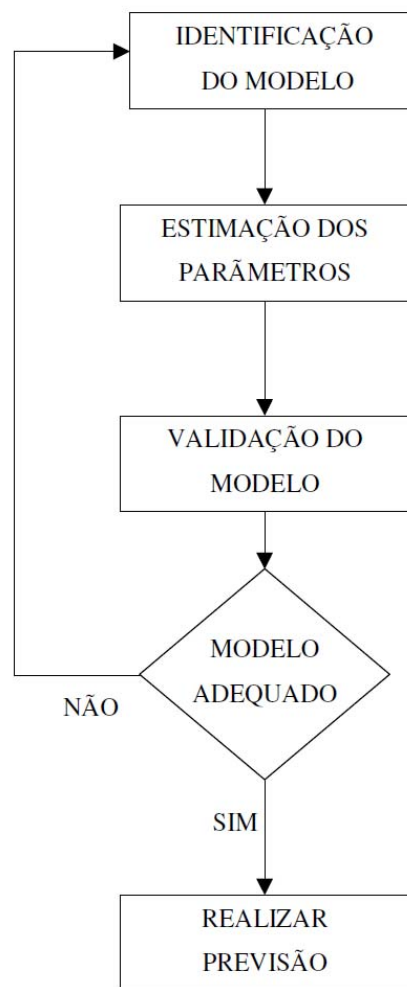
Onde θ é uma constante. O modelo mais simples é o ARMA(1,1):

$$y_t = \theta + \alpha_1 y_{t-1} + u_t - \beta_1 u_{t-1} \quad (9)$$

Os modelos AR, MA e ARMA são utilizados apenas para séries temporais estacionárias, que não tem raiz unitária. Porém, caso uma série seja não estacionária (tem raiz unitária), é possível diferenciá-lo d vezes, até torná-la estacionária, então, aplicar o modelo ARIMA (p,q). Esse é o chamado modelo ARIMA(p,d,q), que significa auto-regressivo integrado de média móvel.

Box e Jenkins (1976) formalizam a teoria da utilização de componentes auto-regressivos e de médias móveis na modelagem de séries temporais utilizando-se de duas idéias básicas na criação de sua metodologia de construção de modelos: a parcimônia que é a utilização do menor número possível de parâmetros para se obter uma representação adequada do fenômeno em estudo e a interatividade, ou seja, a informação empírica é analisada teoricamente e o resultado deste estágio é confrontado com a prática sucessivas vezes, até que o modelo obtido seja satisfatório. O resumo do fluxo dessa metodologia é apresentado a seguir:

Figura 3: Fluxograma da metodologia Box & Jenkins (PINTO, 2006)



3.5.1. Identificação do modelo

Primeiramente faz-se a preparação dos dados para atingir a estacionaridade, Makridakis et al (1998) e DeLugio (1998) sugerem a utilização dos seguintes procedimentos: projeção dos dados em gráficos para verificar a existência de algum padrão; ajustes como deflacionar ou logaritimizar, estabilizando a variância; usar a Função de Auto-correlação (FAC) e a Função Parcial de Auto-correlação (FACP) para verificar a existência de algum padrão nos dados da série; diferenciação dos dados para obter estacionaridade; examinar os FAC e FACP para identificar potenciais modelos, podendo ter ajuda de softwares especialistas.

A função auto-correlação deriva da função auto-variância e é definida como:

$$\rho_k = \gamma_k / \text{var}(x_t) \quad (10)$$

Onde $k = 0, 1, 2, \dots$ e ρ_k é uma medida padrão de dependência entre os intervalos de tempo (defasagem), onde, para todo k , $|\rho_k| \leq 1$.

A função auto-correlação parcial é definida como sendo o último termo auto-regressivo de um modelo AR(p):

$$AR(1) = y_t = \theta_1 y_{t-1} + u_t \quad (11)$$

$$AR(2) = y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + u_t \quad (12)$$

$$AR(p) = y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + u_t \quad (13)$$

Sendo assim, $\theta_1, \theta_2, \theta_p$ são os coeficientes de auto-correlação parcial. Para facilitar os cálculos para estimar-los, utilizam-se os próprios coeficientes de auto-correlação ρ_k . Considerando $K=1, 2, 3, \dots, p$ e $\rho_t = \rho_{-t}$, obtemos assim um sistema de equações conhecido como equações de Yule Walker:

$$\rho_1 = \theta_1 + \theta_2 \rho_1 + \dots + \theta_p \rho_{p-1} \quad (14)$$

$$\rho_2 = \theta_1 \rho_1 + \theta_2 + \dots + \theta_p \rho_{p-2} \quad (15)$$

$$\rho_p = \theta_1 \rho_{p-1} + \theta_2 \rho_{p-2} + \dots + \theta_p \quad (16)$$

Portanto, conhecendo o valor de p e as auto-correlações $\rho_1, \rho_2, \dots, \rho_p$, pode-se calcular os coeficientes das auto-correlação parcial através da matriz a seguir.

$$\begin{pmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{pmatrix} \times \begin{pmatrix} \theta_{k1} \\ \theta_{k2} \\ \vdots \\ \theta_{kk} \end{pmatrix} = \begin{pmatrix} \rho_k \\ \rho_{k-1} \\ \vdots \\ \rho_1 \end{pmatrix} \quad (17)$$

A identificação do modelo deriva do comportamento da FAC e da FACP. Segundo Pinto (2006), a FAC decai exponencialmente para os modelos AR, sofre um corte brusco após a defasagem q para os modelos MA e decai exponencialmente após a defasagem q para o modelo ARMA, ou seja, analisando a FAC é possível identificar apenas o parâmetro q . a definição do parâmetro p é feita através da análise da FACP, que alias, só foi definida para essa finalidade. A FACP decai exponencialmente para os modelos MA, sofre um corte brusco após a defasagem p para os modelos AR e decai exponencialmente após a defasagem p para os modelos ARMA. A seguir, tem-se um quadro resumido das características das funções para os diversos processos:

Tabela 3: Comportamento da FAC e FACP nos modelos AR, MA e ARMA

	AR	MA	ARMA
Comportamento FAC	Decai exponencialmente	Corte brusco após a defasagem q	Decai exponencialmente após a defasagem q
Comportamento FACP	Corte brusco após a defasagem p	Decai exponencialmente	Decai exponencialmente após a defasagem p

Fonte: Pinto, 2006

Para melhor entendimento do quadro, temos em seguida exemplos dos gráficos das funções teóricas das FAC's, FACP's dos modelos AR(1), MA(1) e ARMA(1,1) extraídos de Brooks (2002).

Figura 4: Exemplo 1 de FAC e FACP do AR(1)

Modelo AR(1): exemplo 1

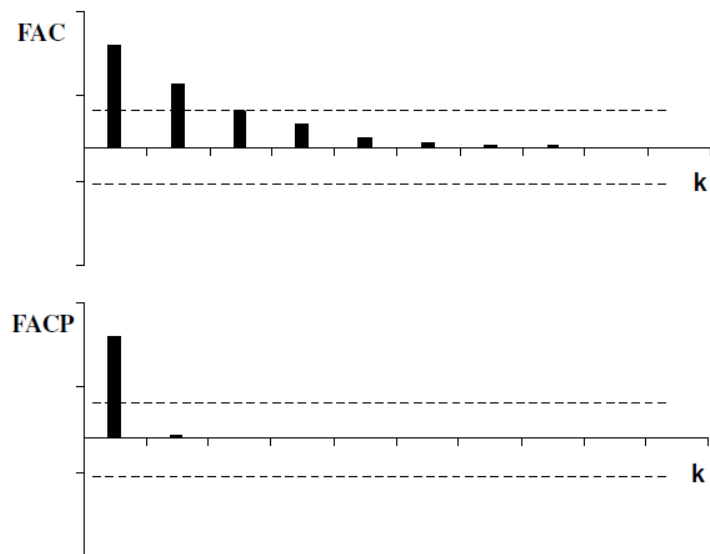


Figura 5: Exemplo 2 de FAC e FACP de AR(1)

Modelo AR(1): exemplo 2

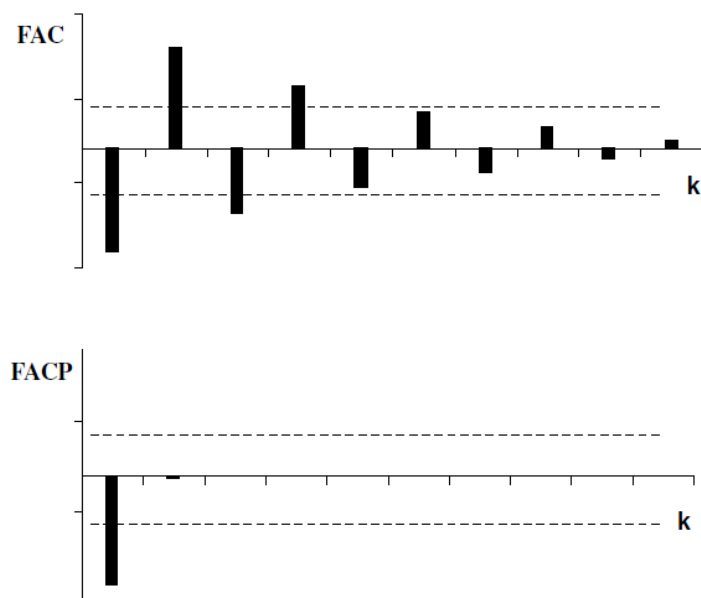


Figura 6: Exemplo 1 de FAC e FACP de MA(1)
Modelo MA(1): exemplo 1

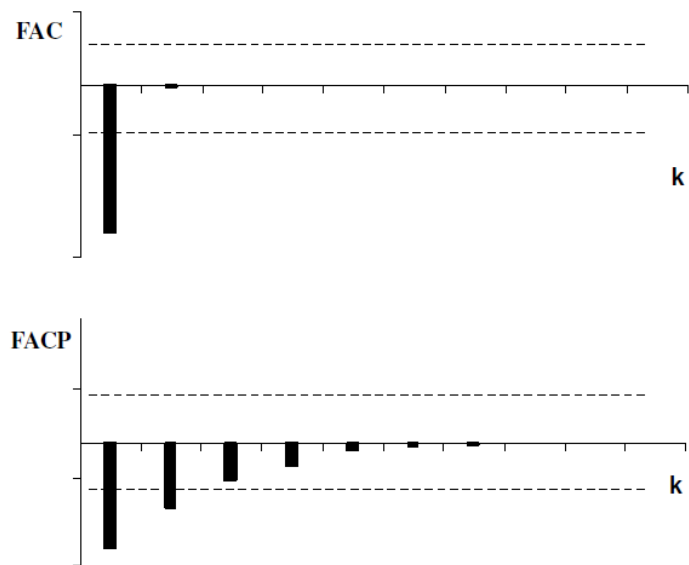


Figura 7: Exemplo 2 de FAC e FACP do MA(1)

Modelo MA(1): exemplo 2

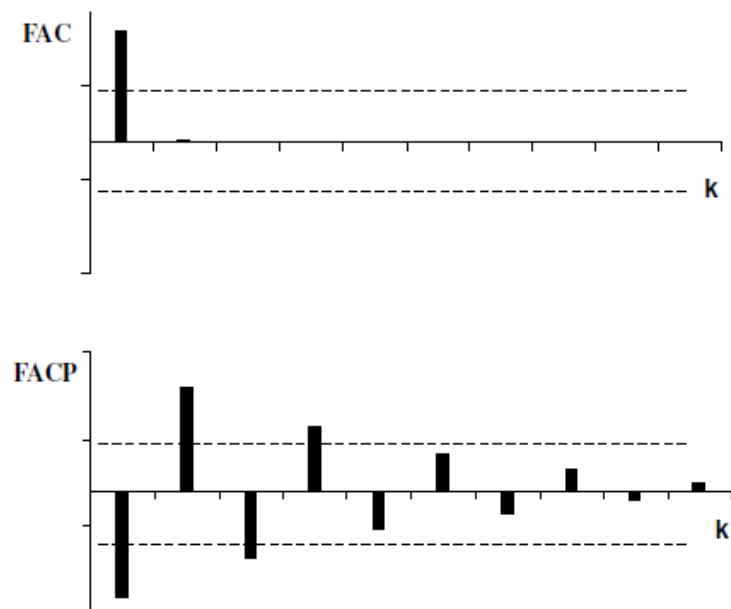


Figura 8: Exemplo 1 de FAC e FACP do ARMA(1,1)

Modelo ARMA (1,1): exemplo 1

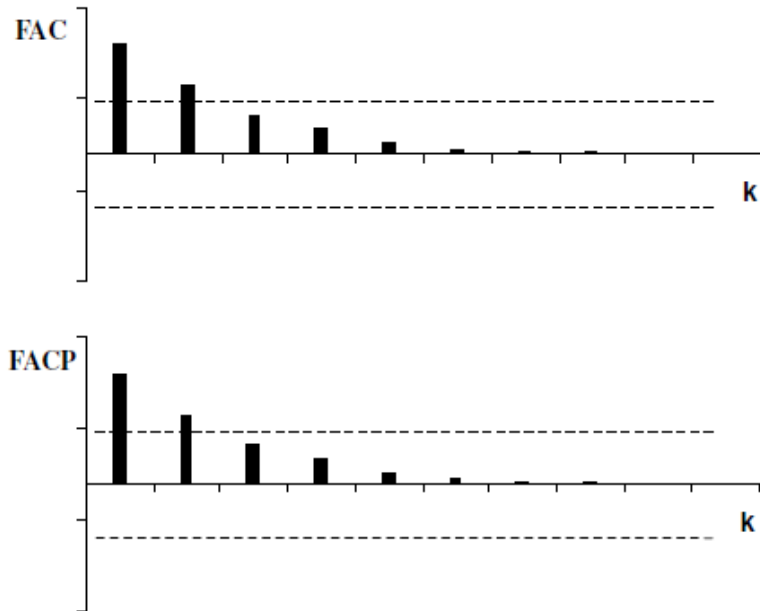
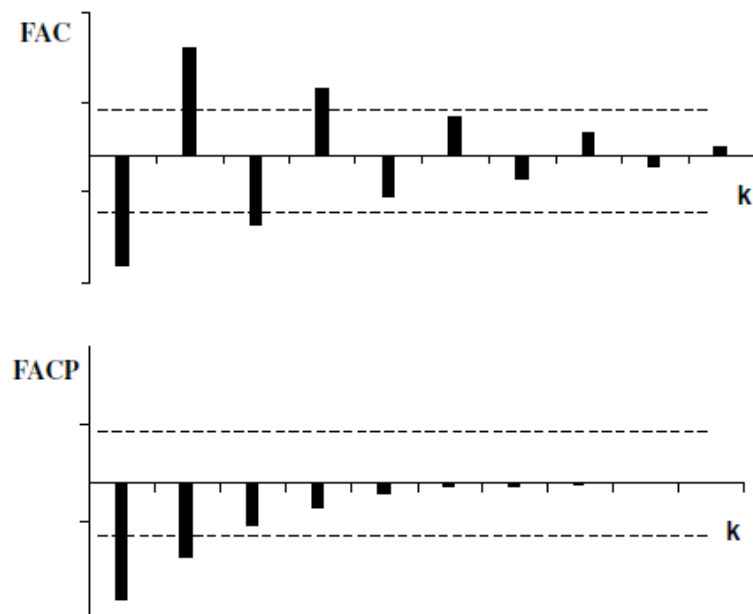


Figura 9: Exemplo 2 de FAC e FACP do ARMA(1,1)

Modelo ARMA (1,1): exemplo 2



No modelo ARIMA temos que definir o parâmetro d . Ele nada mais é do que o grau de diferenciação necessário para tornar os dados estacionários. Pela análise da FAC e da FACP a série é diferenciada até o grau d , de forma que os comportamentos das funções mencionadas sejam semelhantes a um dos pares de comportamento apresentados nas figuras anteriores. Uma vez conhecido o parâmetro d , podemos definir os parâmetros p e q do modelo ARIMA (p,d,q).

3.5.2. Estimação dos parâmetros

Segundo Bolfarine e Sandoval (2001), nessa fase é necessário definir os parâmetros α 's e β 's da equação de regressão. Os parâmetros auto-regressivos (α 's) são lineares, sendo mais simples de se estimar. Pode-se utilizar o método dos mínimos quadrados ordinários que consiste na obtenção dos parâmetros que minimize a soma das diferenças entre os pontos observados na amostra e os pontos estimados pela equação de regressão com os respectivos parâmetros ao quadrado. Já os parâmetros de média móvel β 's não são lineares, tornando a sua estimação mais difícil. Utiliza-se nesse caso o método da verossimilhança que consiste em encontrar estimadores que gerem valores que mais se assemelhem aos valores da amostra.

3.5.3. Validação do modelo

Nessa etapa estabelecemos critérios para escolher o melhor modelo dentre os possíveis. Considerando-se o princípio da parcimônia, já citado anteriormente, leva-se em consideração a menor quantidade de parâmetros a serem adicionados ao modelo, reduzindo-se os seus graus de liberdade. Entretanto ao se adicionar novos parâmetros, pode-se reduzir a soma do seu erro quadrático e melhorar o R^2 . Para melhor decidir qual modelo usar, utilizamos o critério do AIC (Akaike Information Criteria) e ou o SBC (Schawartz Bayesian Criterion), representados a seguir:

$$\text{AIC} = T \ln(\text{soma do quadrado dos resíduos}) + 2n \quad (18)$$

$$\text{SBC} = T \ln(\text{soma do quadrado dos resíduos}) + n \ln(T) \quad (19)$$

Onde T é o número total de observações e n é o número de parâmetros estimados (p, q, d, \dots). A regra de escolha é que o melhor modelo é aquela que apresentar o menor AIC e o menor SBC entre os candidatos (PINTO et al, 2008).

Outros testes são utilizados para saber a qualidade do modelo. Eles consistem em analisar as auto-correlações entre os resíduos encontrados. A estatística Q calculada por meio do teste de Box-Pierce (Portmanteau) ou pelo teste de Ljung-Box é verificada para testar se um determinado conjunto de auto-correlações de resíduos é estatisticamente diferente de zero. Essa estatística possui distribuição qui-quadrada, k graus de liberdade e são definidas como:

$$Q(k) = T \sum_{i=1}^k \rho_i^2 \quad (\text{Box-Pierce}) \quad (20)$$

$$Q(k) = T(T + 2) \sum_{i=1}^k \frac{\rho_i^2}{T-1} \quad (\text{Ljung e Box}) \quad (21)$$

Onde T é o número total de observações e as hipóteses são:

$$H_0 = \rho_1 = \dots = \rho_k = 0 \quad (22)$$

$$H_a = \rho_i \neq 0 \quad (23)$$

A interpretação de ambos os modelos é a mesma. Se o Q calculado for maior do que o Q tabelado para determinado nível de significância, não se aceita a hipótese nula H_0 de que os resíduos não são auto-correlacionados. Porém, isto implica em aceitar a hipótese alternativa H_a de que pelo menos uma auto-correlação é estatisticamente diferente de zero, ou seja, o erro não é ruído branco. No caso da aceitação da hipótese alternativa, rejeita-se o modelo.

3.5.4. Previsão

Utilizaremos o modelo Box & Jenkins para realizar previsões. Como a maior preocupação das previsões é a acurácia dos valores estimados, utilizaremos a medida do erro para testar a sua eficácia.

Uma medida bastante utilizada é a representação dos desvios em termos relativos. A forma mais direta desse tipo de representação, de acordo com Makridakis et al (1998), é apresentar os erros percentuais:

$$PE_t = \left[\frac{y_t - F_t}{F_t} \right] \times 100 \quad (24)$$

Onde F_t é o valor previsto de y no tempo t .

Essa equação não consegue medir o erro ao longo do tempo, apenas mede o erro percentual ao longo do tempo. Se os erros de cada instante forem simplesmente somados, os erros negativos irão compensar os erros positivos e o erro total seria muito baixo. Segundo Makridakis et al (1998) a solução é calcular a média dos valores absolutos dos erros percentuais (MAPE) com relação ao tamanho da amostra n :

$$MAPE = \frac{\sum_{t=1}^n |PE_t|}{n} \quad (25)$$

3.6. Ferramentas utilizadas

A coleta de dados foi feita através do programa Bloomberg.

O tratamento dos dados coletados e o cálculo da função log-normal dos retornos diários foram feitos com o auxílio da planilha eletrônica EXCEL.

Toda parte estatística, da metodologia Box & Jenkins e de aplicação dos testes foi feita com a ferramenta EVIEWS 5.0. No apêndice pode ser verificado o roteiro, passo a passo, dos procedimentos e testes realizados.