

# Identificação, Estimação e Validação em Modelos ARIMA

Pedro Valls\*  
EESP-FGV

## Sumário

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Identificação</b>  | <b>5</b>  |
| 1.1      | Exemplo de um $ARMA(1,1)$ - dados gerados . . . . .         | 5         |
| 1.2      | Exemplo - dados bolsa.xls . . . . .                         | 7         |
| 1.2.1    | Exercícios . . . . .  | 11        |
| <b>2</b> | <b>Estimação de Modelos ARMA</b>                            | <b>11</b> |
| 2.1      | Introdução . . . . .  | 11        |
| 2.2      | Estimador de MV exato e MQC para o $AR(1)$ . . . . .        | 11        |
| 2.3      | Procedimento Iterativo para Maximizar a Verossimilhança . . | 14        |
| 2.4      | Programa no R para estimar um $AR(1)$ por MQO e MV . . .    | 17        |
| <b>3</b> | <b>Estimação de um <math>AR(p)</math></b>                   | <b>18</b> |
| 3.1      | Estimador de Máxima Verossimilhança Condicional . . . . .   | 20        |
| <b>4</b> | <b>Estimação no Modelo <math>MA(1)</math></b>               | <b>21</b> |
| 4.1      | Programa no R para estimar um $MA(1)$ por MV e MQC. . . .   | 22        |
| 4.2      | Estimação no Modelo $ARMA(1,1)$ . . . . .                   | 24        |
| 4.3      | Estimação do $ARMA$ para $\Delta LDolar$ . . . . .          | 25        |

---

\*EESP-FGV, Rua Doutor Plínio Barreto 365, s.1319, 01313-905, São Paulo, SP, Brasil.  
E-mail: pedro.valls@fgv.br. ©-2024-Pedro Valls

|           |   |           |
|-----------|---|-----------|
| 4.3.1     | Exercícios . . . . .  | 29        |
| <b>5</b>  | <b>Validação</b>  | <b>31</b> |
| 5.1       | Introdução . . . . .  | 31        |
| 5.2       | Mínimos Quadrados Recursivos . . . . .  | 32        |
| 5.2.1     | Testes Baseados na Estimação Recursiva . . . . .  | 35        |
| 5.2.2     | Exercícios . . . . .  | 41        |
| <b>6</b>  | <b>Previsão em Modelos ARMA</b>   | <b>42</b> |
| 6.1       | Previsão de um $AR(1)$ . . . . .  | 42        |
| 6.2       | Previsão de um $MA(1)$ . . . . .  | 44        |
| 6.3       | Previsão de um $ARMA(1, 1)$ . . . . .   | 46        |
| 6.4       | Previsão de um $ARMA(p, q)$ . . . . .   | 47        |
| 6.5       | Previsão para um $ARIMA(0, 1, 0)$ com constante . . . . .   | 48        |
| 6.6       | Previsão para Transformações de Variáveis . . . . .   | 50        |
| 6.7       | Previsão para a série <i>DLDolar</i> . . . . .  | 51        |
| 6.8       | Exercícios . . . . .  | 52        |
| <b>7</b>  | <b>Avaliação de Previsões</b>   | <b>52</b> |
| <b>8</b>  | <b>Testes para Comparação de Previsões</b>  | <b>53</b> |
| 8.1       | Testes para Comparação de Previsão - Morgan-Granger-Newbold   | 53        |
| 8.2       | Testes para Comparação de Previsão - Diebold & Mariano . .  | 54        |
| 8.3       | Testes para Comparação de Previsão - Conjunto de modelos<br>com confiança ( <i>model confidence set</i> ) (MCS) . . . . . | 55        |
| 8.4       | Testes de Comparação de Previsões - Classificando os modelos:<br>valores-p para MCS . . . . .                             | 57        |
| <b>9</b>  | <b>A Combinação de Previsões</b>  | <b>58</b> |
| <b>10</b> | <b>Abrangente em Previsão</b>   | <b>60</b> |
| <b>11</b> | <b>Apêndice I - Critérios de Informação</b>   | <b>60</b> |
| <b>12</b> | <b>Apêndice II - Inferência para Estimadores de Máxima Ve-</b>  |           |
|           | <b>rossimilhança</b>  | <b>61</b> |

## Lista de Tabelas

|   |  |    |
|---|--|----|
| 1 | Modelling AR(1) by Exact MLE . . . . .   | 18 |
| 2 | Modelling AR(1) by CMLE . . . . .  | 18 |
| 3 | Modelling MA(1) by Exact MLE . . . . .   | 23 |
| 4 | Modelling MA(1) by Conditional OLS . . . . .                                   | 24 |
| 5 | Modelling ARMA(1,1) by Conditional LSE . . . . .                               | 25 |
| 6 | Modelling ARMA(12,12) by Exact MLE for DiDolar . . . . .                       | 26 |
| 7 | Critério de Schwarz para todos $ARMA(p, q)$ com $p, q \leq 12$ . .             | 28 |
| 8 | Best Model for DiDolar MA(1) with intercept estimate by<br>Exact MLE . . . . . | 29 |

## Lista de Figuras

|    |  |    |
|----|--|----|
| 1  | ARMA(1,1): $X_t = 0.5 * X_{t-1} + u_t - 0.49 * u_{t-1}$ . . . . .  | 6  |
| 2  | FAC e FACP para ARMA(1,1): $X_t = 0.5 * X_{t-1} + u_t - 0.49 * u_{t-1}$                                  | 6  |
| 3  | Séries Câmbio, IBV, IBRX, S&P500 . . . . .   | 7  |
| 4  | Janela de Valls para Câmbio . . . . .  | 8  |
| 5  | FAC e FACP para Câmbio . . . . .   | 9  |
| 6  | Janela de Valls para Dlog(Câmbio) . . . . .  | 10 |
| 7  | FAC e FACP para Dlog(Câmbio) . . . . .   | 10 |
| 8  | Estimação por Máxima Verossimilhança Exata do AR(1): $X_t =$<br>$0.85 * X_{t-1} + u_t$ . . . . .         | 17 |
| 9  | Estimação por Máxima Verossimilhança Exata e MQC para o<br>MA(1): $X_t = u_t + 0.45 * X_{t-1}$ . . . . . | 23 |
| 10 | Resíduos Recursivos . . . . .  | 36 |
| 11 | Resíduos MQO . . . . .   | 36 |
| 12 | Histograma dos Resíduos MQO . . . . .  | 37 |
| 13 | QQPlot dos Resíduos MQO . . . . .  | 37 |
| 14 | Histograma dos Resíduos Recursivos . . . . .   | 38 |
| 15 | QQPlot dos Resíduos MQO . . . . .  | 38 |
| 16 | Estimativa Recursiva de $\phi$ . . . . .   | 39 |
| 17 | CUSUM para o AR(1 estimado para DiDolar . . . . .  | 40 |
| 18 | CUSUMSQ para o AR(1 estimado para DiDolar . . . . .  | 41 |

|    |  |    |
|----|--|----|
| 19 | Previsões para as últimas 7 observações para $DlDolar$ . . . . | 51 |
|----|--|----|

# 1 Identificação

Na literatura de Séries Temporais o "procedimento de Box & Jenkins" (veja Box and Jenkins (1976)) ficou associado a sequência: Identificação, Estimção e Validação.

A identificação de uma série de tempo na classe *ARIMA*, corresponde a identificar a ordem da parte autorregressiva e da parte média móvel assim como a ordem de diferenciação.

Já vimos que primeiro temos que determinar a ordem de integração do processo para depois identificar as ordens dos polinômios autorregressivos e de médias-móveis. Esta identificação, em geral, é feita usando-se as FAC e FACP.

Observe que como um *ARMA* pode ser aproximado por um  $AR(\infty)$  podemos identificar mais de um processo quando se utiliza as FAC e FACP. A escolha do modelo, neste estágio, pode ser feito usando-se o princípio da parsimônia - modelo mais simples.

Vamos apresentar alguns exemplos deste processo de identificação.

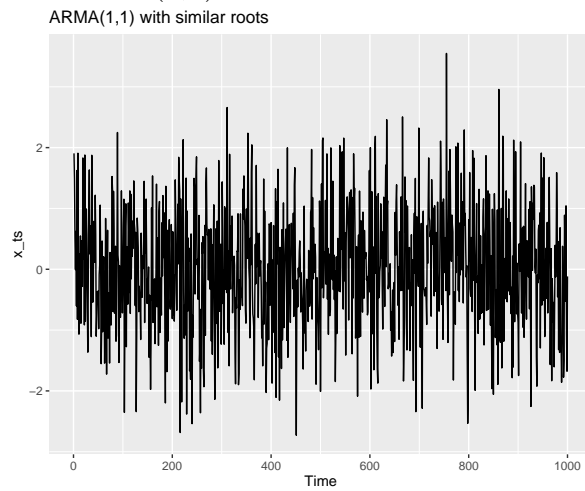
## 1.1 Exemplo de um $ARMA(1, 1)$ - dados gerados

O programa a seguir gera um  $ARMA(1, 1)$  com parâmetros da parte *AR* e da parte *MA* muito próximos.

$ARMA(1,1)$  com raízes comuns

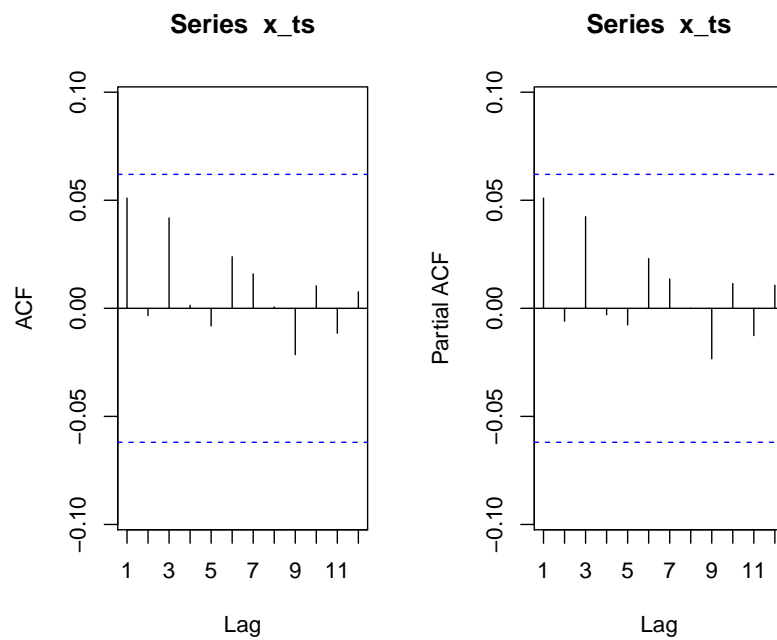
O gráfico abaixo apresenta a série  $x$  e podemos observar que a mesma é claramente estacionária.

Figura 1: ARMA(1,1):  $X_t = 0.5 * X_{t-1} + u_t - 0.49 * u_{t-1}$



A seguir é apresentada as FAC e FACP

Figura 2: FAC e FACP para ARMA(1,1):  $X_t = 0.5 * X_{t-1} + u_t - 0.49 * u_{t-1}$



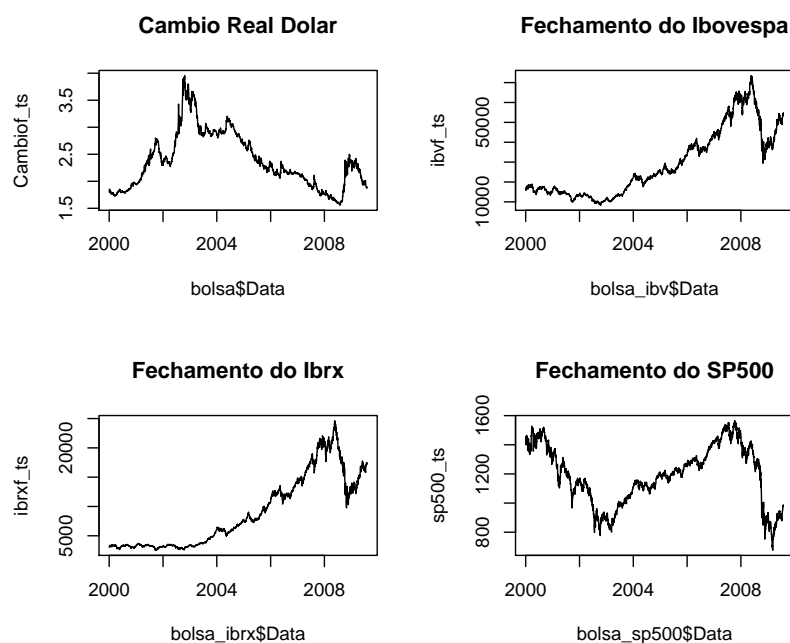
Observe que pela FAC e FACP a ordem tanto da parte  $AR$  quanto da parte  $MA$  seria zero. Este resultado indica que embora o modelo gerado tenha sido um  $ARMA(1, 1)$ , mas como o parâmetro da parte  $AR$  é quase igual ao da parte  $MA$  teremos fatores comuns nestes dois polinômios e portanto a FAC e FACP parece de um ruído branco.

## 1.2 Exemplo - dados bolsa.xls

No arquivo bolsa.xls temos, para o período 3/1/2000 até 31/7/2009, dados do índice Bovespa (ibv), do índice IBX (ibx), da taxa de câmbio real-dólar (dólar) e do índice da bolsa americana S&P500.

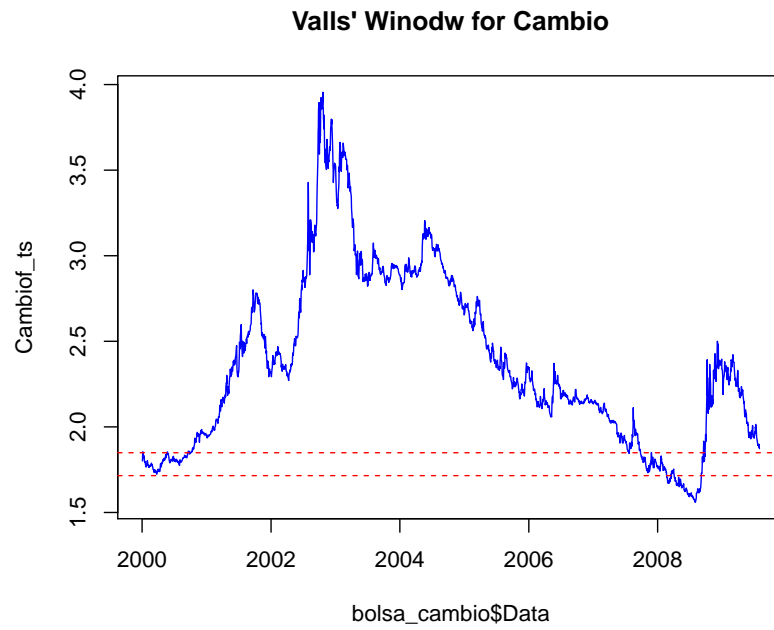
Abaixo estão os gráficos destas variáveis contra o tempo.

Figura 3: Séries Câmbio, IBV, IBRX, S&P500



Todas as séries parecem ter tendência estocástica. Por exemplo usando a Janela de Valls para a série do Câmbio temos o seguinte gráfico:

Figura 4: Janela de Valls para Câmbio

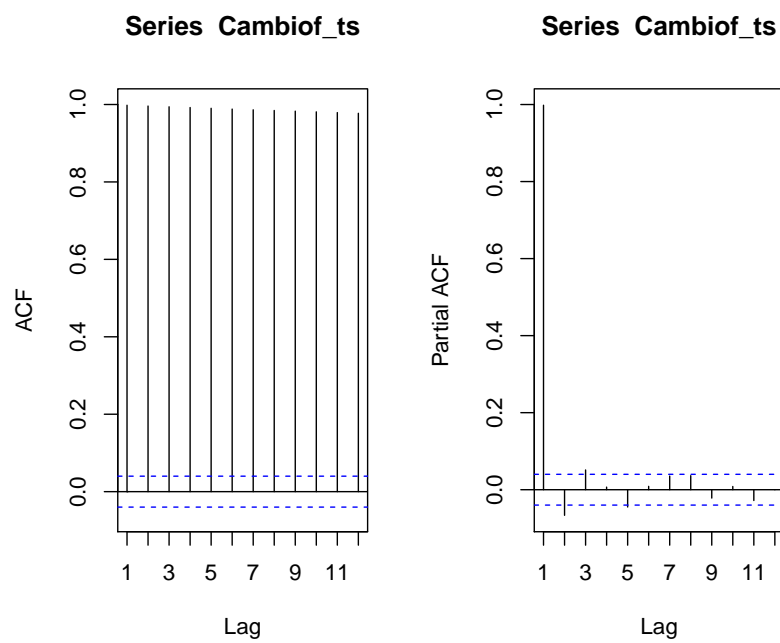


Observe que a série saiu fora da banda em 2001 e só volta em 2008, voltando a sair em 2009, indicando que tanto a média quanto a variância não são constantes. Violando, portanto as duas primeiras condições de estacionaridade.

Para a série de Dolar temos a seguinte FAC e FACP:

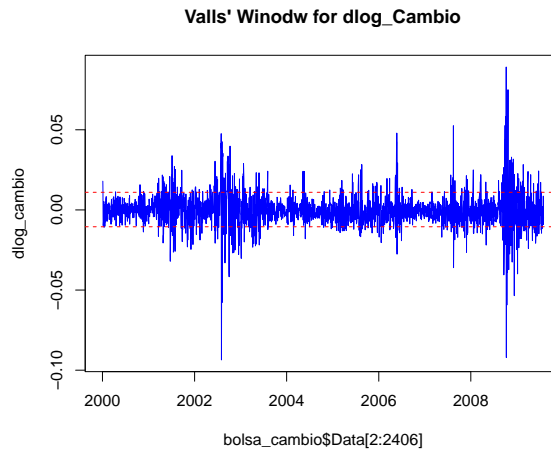


Figura 5: FAC e FACP para Câmbio



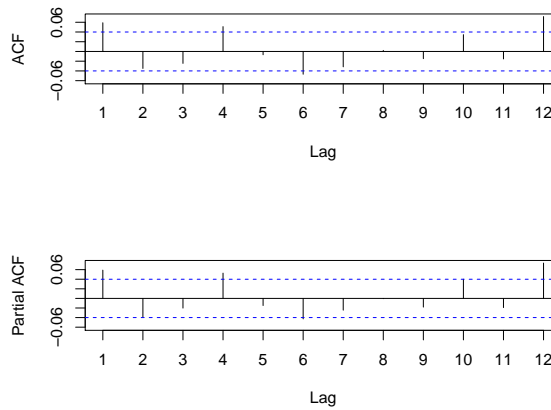
Pela FAC e FACP parece que a série tem uma raíz unitária, fazendo a transformação  $\Delta \text{Log}(\text{Dolar})$  a série se torna estacionária. Abaixo temos o gráfico contra o tempo desta série e a janela de Valls.

Figura 6: Janela de Valls para Dlog(Câmbio)



Agora pela FAC e FACP podemos observar que esta série tem estrutura de autocorrelação. Podendo ser um modelo do tipo  $ARMA(p, q)$  com  $p, q \leq 12$ , uma vez que a décima segunda autocorrelação é diferente de zero.

Figura 7: FAC e FACP para Dlog(Câmbio)



Observe que provavelmente um  $ARMA(12, 12)$  estará super-parametrizado e os polinômios da parte  $AR$  e  $MA$  terão raízes comuns.

### 1.2.1 Exercícios

1. Para a série do IBV determine a ordem do melhor *ARIMA*.
2. Para a série do IBX determine a ordem do melhor *ARIMA*.
3. Para a série do S&P500 determine a ordem do melhor *ARIMA*.

## 2 Estimação de Modelos ARMA

### 2.1 Introdução

A estimação de modelos na classe *AR* pode ser feita de duas formas:

- (i) Mínimos Quadrados Ordinários
- (ii) Máxima Verossimilhança.

No primeiro caso o estimador obtido será uma aproximação uma vez que trataremos algumas observações como fixas e por isto este método é também chamado de **mínimos quadrados condicionais**.

Já o segundo método se a maximização for de toda a verossimilhança o estimador obtido será o estimador de **máxima verossimilhança exato**.

A estimação de modelos na classe *MA* será feita por procedimentos iterativos.

Vamos considerar primeiro o caso de um *AR*(1).

### 2.2 Estimador de MV exato e MQC para o *AR*(1)

Para definir o modelo gaussiano *AR*(1), partiu-se das seguintes condições:

$$E[X_t | \mathbf{I}_{t-1}^{t-1}] = \phi X_{t-1} \quad (1)$$

$$Var[X_t | \mathbf{I}_{t-1}^{t-1}] = \sigma^2 \quad (2)$$

onde  $\mathbf{I}_{t-1}^{t-1}$  representa o conjunto de informação formado sómente pela informação no instante  $t - 1$  e tem-se então o seguinte modelo:

$$X_t = E[X_t | \mathbf{I}_{t-1}^{t-1}] + u_t = \phi X_{t-1} + u_t \quad (3)$$

onde  $u_t$  por construção é um ruído branco gaussiano, isto é,  $u_t \sim NI(0, \sigma^2)$ .

Observe que a distribuição conjunta da amostra  $\{X_1, \dots, X_T\}$  é dada por:

$$D(X_1, \dots, X_T | \phi : \sigma^2) \quad (4)$$

mas por ser uma amostra dependente, esta distribuição conjunta, não pode ser fatorada como o produto das marginais. Mas podemos dividir esta distribuição conjunta como o produto da distribuição de  $X_T$  dada a informação passada e a distribuição marginal do passado, isto é:

$$D(X_1, \dots, X_T | \phi : \sigma^2) = D(X_T | X_1, \dots, X_{T-1} : \phi : \sigma^2) D(X_1, \dots, X_{T-1} | \phi : \sigma^2) \quad (5)$$

e usando o mesmo argumento recursivamente temos que (4) pode ser escrito da seguinte forma:

$$D(X_1, \dots, X_T | \phi : \sigma^2) = \prod_{t=2}^T D(X_t | X_1, \dots, X_{t-1} : \phi : \sigma^2) D(X_1 | \phi : \sigma^2) \quad (6)$$

onde a última expressão do lado direito de (6) é a distribuição não condicional de  $X_1 \sim N(0, \frac{\sigma^2}{1-\phi^2})$  e cada parcela do produtório, isto é,  $D(X_t | X_1, \dots, X_{t-1} : \phi : \sigma^2) \sim N(\phi X_{t-1}, \sigma^2)$ .

Observe que podemos escrever (6) da seguinte forma:

$$\begin{aligned} D(X_1, \dots, X_T | \phi : \sigma^2) &= \prod_{t=2}^T D(X_t - E(X_t | \mathbf{I}_{t-1}) | X_1, \dots, X_{t-1} : \phi : \sigma^2) D(X_1 - E(X_1) | \phi : \sigma^2) \\ &= \prod_{t=2}^T D(u_t | X_1, \dots, X_{t-1} : \phi : \sigma^2) D(u_1 | \phi : \sigma^2) \end{aligned} \quad (7)$$

que é chamado de **decomposição do erro de previsão**, uma vez que  $u_t = X_t - E(X_t | \mathbf{I}_{t-1}^{t-1})$ .

Observe que (7) visto como função dos parâmetros é a função de verossimilhança para o modelo  $AR(1)$ . Tomando-se a transformação logarítmica temos:

$$\log l(\phi : \sigma^2) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^T (X_t - \phi X_{t-1})^2 + \frac{1}{2} \ln(1-\phi^2) - \frac{(1-\phi^2)}{2\sigma^2} X_1^2 \quad (8)$$

e maximizando-se esta função obtemos os Estimadores de Máxima Verossimilhança Exatos para os parâmetros.

Observe que as condições de primeira ordem são dadas por:

$$\frac{\partial \log l}{\partial \phi} = \frac{1}{\sigma^2} \sum_{t=2}^T (X_t - \phi X_{t-1}) X_{t-1} - \frac{\phi}{1-\phi^2} + \frac{\phi}{\sigma^2} X_1^2 \quad (9)$$

$$\frac{\partial \log l}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=2}^T (X_t - \phi X_{t-1})^2 + \frac{(1-\phi^2)}{2\sigma^4} X_1^2 \quad (10)$$

que é um sistema não linear nos parâmetros, cuja solução não é fechada e para se obter os estimadores exatos é melhor maximizar diretamente a função de verossimilhança. Para se fazer isto é necessário um procedimento iterativo que será apresentado na próxima seção.

Observe também se a distribuição não condicional de  $X_1$  não for usada, isto é, tratarmos a primeira observação como fixa, (8) é simplificado para:

$$\log l(\phi : \sigma^2 \mid X_1) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^T (X_t - \phi X_{t-1})^2 \quad (11)$$

e as condições de primeira ordem ficam:

$$\frac{\partial \log l}{\partial \phi} = \frac{1}{\sigma^2} \sum_{t=2}^T (X_t - \phi X_{t-1}) X_{t-1} \quad (12)$$

$$\frac{\partial \log l}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=2}^T (X_t - \phi X_{t-1})^2 \quad (13)$$

que agora é um sistema linear nos parâmetros cuja solução é:

$$\hat{\phi} = \frac{\sum_{t=2}^T X_t X_{t-1}}{\sum_{t=2}^T X_{t-1}^2} \quad (14)$$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=2}^T \left( X_t - \hat{\phi} X_{t-1} \right)^2 \quad (15)$$

que é o estimador de mínimos quadrados ordinários condicional para o modelo  $AR(1)$ .

Para garantir que as soluções (14) e (15) são os máximos de (11) temos que mostrar que a matriz de derivadas segundas é negativa definida.

Temos então

$$\begin{aligned} \left. \frac{\partial^2 \log l(\theta)}{\partial \theta \partial \theta'} \right|_{\hat{\phi}, \hat{\sigma}^2} &= \left[ \begin{array}{cc} \frac{\partial^2 \log l(\theta)}{\partial \phi \partial \phi} & \frac{\partial^2 \log l(\theta)}{\partial \sigma^2 \partial \phi} \\ \frac{\partial^2 \log l(\theta)}{\partial \phi \partial \sigma^2} & \frac{\partial^2 \log l(\theta)}{\partial \sigma^2 \partial \sigma^2} \end{array} \right] \bigg|_{\hat{\phi}, \hat{\sigma}^2} \\ &= \left[ \begin{array}{cc} -\frac{1}{\sigma^2} \sum_{t=2}^T X_{t-1}^2 & -\frac{1}{\sigma^4} \sum_{t=2}^T (X_t - \phi X_{t-1}) X_{t-1} \\ -\frac{1}{\sigma^4} \sum_{t=2}^T (X_t - \phi X_{t-1}) X_{t-1} & \frac{T}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{t=2}^T (X_t - \phi X_{t-1})^2 \end{array} \right] \bigg|_{\hat{\phi}, \hat{\sigma}^2} \\ &= \left[ \begin{array}{cc} -\frac{1}{\hat{\sigma}^2} \sum_{t=2}^T X_{t-1}^2 & 0 \\ 0 & -\frac{T}{2\hat{\sigma}^4} \end{array} \right] \quad (16) \end{aligned}$$

que é negativa definida.

## 2.3 Procedimento Iterativo para Maximizar a Verossimilhança

Para se obter o EMV para os parâmetros  $\theta = (\phi : \sigma^2)$  devemos maximizar  $\log l(\cdot)$  e as condições de primeira ordem, também chamadas de equações de escore, são dadas por:

$$\frac{\partial \log l(\theta)}{\partial \theta} = q(\theta) \quad (17)$$

e as condições de primeira ordem para o máximo são dadas por:

$$q(\hat{\theta}) = 0 \quad (18)$$

e as condições de segunda ordem dadas por:

$$\frac{\partial^2 \log l(\theta)}{\partial \theta \partial \theta'} = \frac{\partial q(\theta)}{\partial \theta'} = H(\theta) \quad (19)$$

que deve ser negativa definida.

Observe que em geral as equações (18), como no caso do  $AR(1)$ , são difíceis de serem resolvidas, mas é sempre possível fazer uma aproximação de primeira ordem na vizinhança de um ponto  $\theta_i$ , isto é,

$$q(\theta) \simeq q(\theta_i) + H(\theta_i)(\theta - \theta_i) = 0 \quad (20)$$

cujas aproximação para o próximo valor de  $\hat{\theta}$  será dada por:

$$\theta_{i+1} = \theta_i - H^{-1}(\theta_i)q(\theta_i) \quad (21)$$

que nos dá um procedimento iterativo para obter o máximo da função.

Em geral as derivadas da verossimilhança são calculadas numericamente e os critérios de parada deste procedimento são dados por uma (ou mais de uma) das condições abaixo:

$$| l(\theta_{i+1}) - l(\theta_i) | < \varepsilon_1 \quad (22)$$

$$| q(\theta_{i+1}) | < \varepsilon_2 \quad (23)$$

$$| \theta_{i+1} - \theta_i | < \varepsilon_3 \quad (24)$$

onde  $\varepsilon_i$  para  $i = 1, 2, 3$  são menores do que  $10^{-n}$ , onde  $n$  é a precisão que se deseja.

No caso em que a primeira observação do  $AR(1)$  não é usada a função a ser maximizada pode ser escrita da seguinte forma:

$$\begin{aligned} \log l(\theta | X_1) &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} S(\theta) \\ S(\theta) &= \sum_{t=2}^T u_t^2(\theta) \end{aligned} \quad (25)$$

e as condições de primeira ordem são dadas por:

$$\frac{\partial \log l(\theta)}{\partial \theta} = -\frac{1}{\sigma^2} \sum_{t=2}^T z_t(\theta) u_t(\theta) \quad (26)$$

com

$$z_t(\theta) = \frac{\partial u_t(\theta)}{\partial \theta} \quad (27)$$

e as condições de segunda ordem são dadas por:

$$\frac{\partial^2 \log l(\theta)}{\partial \theta \partial \theta'} = -\frac{1}{\sigma^2} \sum_{t=2}^T z_t(\theta) z_t'(\theta) \quad (28)$$

e usando (26) e (28) em (21) temos:

$$\theta_{i+1} = \theta_i - \left[ \sum_{t=2}^T z_t(\theta_i) z_t'(\theta_i) \right]^{-1} \left[ \sum_{t=2}^T z_t(\theta_i) u_t(\theta_i) \right] \quad (29)$$

que no caso do  $AR(1)$  temos:

$$u_t(\theta_i) = (X_t - \phi_i X_{t-1}) \quad (30)$$

$$z_t(\theta_i) = -X_{t-1} \quad (31)$$

e substituindo-se (126) e (127) em (29) temos:

$$\phi_{i+1} = \phi_i + \frac{\sum_{t=2}^T (X_t - \phi_i X_{t-1}) X_{t-1}}{T \sum_{t=2}^T X_{t-1}^2} \quad (32)$$

O procedimento começa com um valor inicial para  $\phi_0$  e usando (128) obtém-se a sequência de valores de  $\phi_i$ . A condição inicial é muito importante para que o procedimento iterativo tenha uma convergência rápida e para um máximo global. No caso de MQC podemos usar  $\phi_0 = 0$  e a convergência se dará em uma iteração. Para o caso em que não se exclua a primeira observação podemos usar  $\phi_0 = \hat{\phi}_{MQC}$  e como este estimador é consistente a convergência também será rápida.



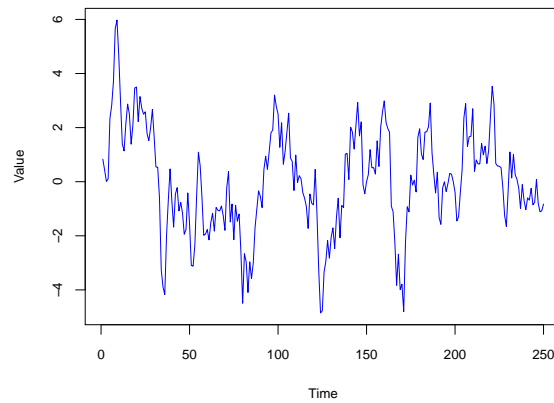
## 2.4 Programa no R para estimar um $AR(1)$ por MQO e MV

A seguir é apresentado um programa no Eviews para estimar um  $AR(1)$  por MQO e MV

Estimadores de Máxima Verossimilhança Exato para  $AR(1)$

Os resultados para o estimador exato são apresentados abaixo:

Figura 8: Estimação por Máxima Verossimilhança Exata do  $AR(1)$ :  $X_t = 0.85 * X_{t-1} + u_t$



e os resultados para a estimação por Máxima Verossimilhança Exata são dados abaixo:

Tabela 1: Modelling  $AR(1)$  by Exact MLE

| EQ( 1) Modelling $AR(1)$ by Exact MLE |             |            |                |         |
|---------------------------------------|-------------|------------|----------------|---------|
| The estimation sample is: 1 - 250     |             |            |                |         |
|                                       | Coefficient | Std. Error | t-Statitistics | p-value |
| $AR(1)$                               | 0.8657      | 0.0314     | 27.5566        | 0.000   |
| $R^2 = 0.7513$                        |             |            |                |         |
| no of observations: 250               |             |            |                |         |
| no of parameters: 1                   |             |            |                |         |
| $AIC = 2.7946$                        |             |            |                |         |
| $BIC = 2.8307$                        |             |            |                |         |
| log-likelihood = -348.3203            |             |            |                |         |

e os resultados para a estimação por MQO são dados abaixo:

Tabela 2: Modelling  $AR(1)$  by CMLE

| EQ( 2) Modelling $AR(1)$ by OLS   |             |            |                |         |
|-----------------------------------|-------------|------------|----------------|---------|
| The estimation sample is: 2 - 250 |             |            |                |         |
|                                   | Coefficient | Std. Error | t-Statitistics | p-value |
| $AR(1)$                           | 0.8668      | 0.0317     | 27.3800        | 0.000   |
| $R^2 = 0.7513$                    |             |            |                |         |
| no of observations: 249           |             |            |                |         |
| no of parameters: 1               |             |            |                |         |
| $AIC = 2.7602$                    |             |            |                |         |
| $BIC = 2.7884$                    |             |            |                |         |
| log-likelihood = -348.5666        |             |            |                |         |

Observe que neste caso a diferença entre os dois é pequena isto porque o tamanho da amostra é grande.

### 3 Estimação de um $AR(p)$

Considere o  $AR(p)$  dado por:

$$X_t = E[X_t | \mathbf{I}_{t-1}^{t-p}] + u_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + u_t \quad (33)$$

onde por construção  $u_t \sim NI(0, \sigma^2)$  e denotando os parâmetros a serem estimados por  $\theta = (\phi_1, \dots, \phi_p, \sigma^2)$

Para obter a verossimilhança deste modelo vamos usar a decomposição do erro de previsão, isto é

$$D(X_1, \dots, X_T | \theta) = \prod_{t=p+1}^T D(X_t | \mathbf{I}_{t-1}^{t-p} : \theta) D(X_1, \dots, X_p | \theta) \quad (34)$$

onde

$$D(X_t | \mathbf{I}_{t-1}^{t-p} : \theta) \sim N(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}, \sigma^2) \quad (35)$$

e

$$D(X_1, \dots, X_p | \theta) \sim N(0, \sigma^2 V_p) \quad (36)$$

onde  $\sigma^2 V_p$  é a matriz de variância-covariância para as primeiras  $p$  observações que é dada por

$$V_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdot & \cdot & \cdot & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdot & \cdot & \cdot & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdot & \cdot & \cdot & \gamma_{p-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdot & \cdot & \cdot & 1 \end{bmatrix} \quad (37)$$

Então o logarítmo da função de verossimilhança, usando  $\mathbf{X}_p = (X_1, \dots, X_p)$  será dado por:

$$\begin{aligned}
\log l(\theta) &= -\frac{p}{2} \log(2\pi) - \frac{p}{2} \log(\sigma^2) + \frac{1}{2} \log |V_p^{-1}| - \frac{1}{2\sigma^2} \mathbf{X}'_p(V_p^{-1})\mathbf{X}_p \\
&\quad - \frac{T-p}{2} \log(2\pi) - \frac{T-p}{2} \log(\sigma^2) \\
&\quad - \frac{1}{2\sigma^2} \sum_{t=p+1}^T (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p})^2 \\
&= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) + \frac{1}{2} \log |V_p^{-1}| - \frac{1}{2\sigma^2} \mathbf{X}'_p(V_p^{-1})\mathbf{X}_p \\
&\quad - \frac{1}{2\sigma^2} \sum_{t=p+1}^T (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p})^2 \quad (38)
\end{aligned}$$

A maximização de (38) resulta nos estimadores de máxima verossimilhança exata. Observe que é necessário inverter a matriz  $V_p$ , mas esta inversa tem uma expressão fechada. Galbraith and Galbraith (1974) mostraram que o elemento  $(i, j)$  da matriz inversa é dado por

$$v^{ij} = \left[ \sum_{k=0}^{i-1} \phi_k \phi_{k+j-i} - \sum_{k=p+1-j}^{p+i-j} \phi_k \phi_{k+j-i} \right] \quad \text{para } 1 \leq i \leq j \leq p \quad (39)$$

onde  $\phi_0 = -1$ .

### 3.1 Estimador de Máxima Verossimilhança Condicional

A função de verossimilhança condicional nas primeiras  $p$  observações é (38) sem os termos  $\frac{1}{2} \log |V_p^{-1}| - \frac{1}{2\sigma^2} \mathbf{X}'_p(V_p^{-1})\mathbf{X}_p$  tem-se então:

$$\begin{aligned}
\log l(\theta) &= -\frac{T-p}{2} \log(2\pi) - \frac{T-p}{2} \log(\sigma^2) \\
&\quad - \frac{1}{2\sigma^2} \sum_{t=p+1}^T (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p})^2 \quad (40)
\end{aligned}$$

e maximizar esta expressão é equivalente a minimizar

$$\sum_{t=p+1}^T (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p})^2 \quad (41)$$

que resulta nos estimadores de mínimos quadrados ordinários dos parâmetros  $\phi_i$ 's e o estimador da variância residual será dado por:

$$\hat{\sigma}^2 = \frac{1}{T-p} \sum_{t=p+1}^T (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} - \dots - \hat{\phi}_p X_{t-p})^2$$

## 4 Estimação no Modelo $MA(1)$

Considere o modelo  $MA(1)$  dado por:

$$X_t = u_t + \theta u_{t-1} \quad (42)$$

e para estimar  $\theta$  neste modelo poderíamos usar MQC, isto é poderíamos minimizar a soma dos quadrados dos resíduos que é dada por:

$$\sum_{t=1}^T \hat{u}_t^2 = \sum_{t=1}^T (X_t - \theta \hat{u}_{t-1})^2 \quad (43)$$

Observe que para se obter  $\hat{u}_t$  o procedimento usado é o chamado "back-casting" (veja Box and Jenkins (1976)) que é dado por:

- (i) construa a série  $\hat{u}_t = X_t - \theta \hat{u}_{t-1}$  para uma valor inicial de  $\theta$  e assumindo que  $\hat{u}_0 = 0$ .
- (ii) minimize a soma dos quadrados dos resíduos dada em (130).

Uma outra forma, semelhante a apresentada acima, é usar um procedimento iterativo observando que

$$u_t(\theta) = X_t - \theta u_{t-1}(\theta) \quad (44)$$

pode ser construindo assumindo um dado  $\theta$  e  $u_0(\theta) = 0$ . Deste modo temos as "observações"  $u_t(\theta)$ . Derivando (131) em relação a  $\theta$ , temos:

$$q(\theta) = \frac{\partial u_t(\theta)}{\partial \theta} = -u_{t-1}(\theta) - \theta \frac{\partial u_{t-1}(\theta)}{\partial \theta} \quad (45)$$

que para um dado  $\theta$  e assumindo que  $\frac{\partial u_0(\theta)}{\partial \theta} = 0$  podemos construir  $q(\theta)$ .

Agora derivando mais uma vez (138) temos:

$$H(\theta) = \frac{\partial^2 u_t(\theta)}{\partial \theta \partial \theta} = -2 \frac{\partial u_{t-1}(\theta)}{\partial \theta} - \theta \frac{\partial^2 u_{t-1}(\theta)}{\partial \theta \partial \theta} \quad (46)$$

que para um dado  $\theta$  e assumindo que  $\frac{\partial^2 u_0(\theta)}{\partial \theta \partial \theta} = 0$  podemos construir  $H(\theta)$ .

Usando (138) e (139) em (21) temos uma forma iterativa de obter o estimador de  $\theta$ .

Outra forma de obter um estimador consistente para  $\theta$  é inverter o modelo  $MA(1)$  para obter um  $AR(\infty)$ . Podemos truncar a ordem na primeira ordem e estimar por MQC o  $\theta$  através no coeficiente do  $AR(1)$ .

Uma outra alternativa é usar o procedimento proposto por Valls Pereira (1987).

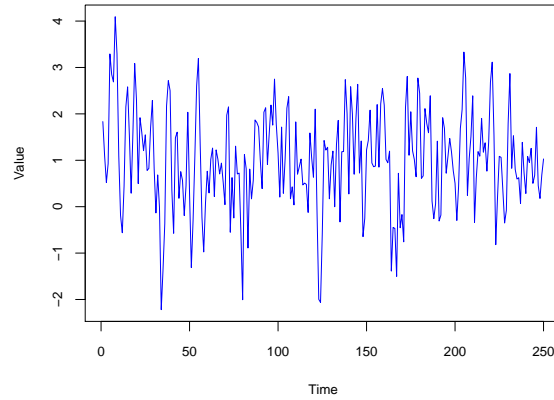
#### 4.1 Programa no R para estimar um $MA(1)$ por MV e MQC.

O programa abaixo estima o parâmetro de um  $MA(1)$  por MV e MQC.

Estimadores de Máxima Verossimilhança Exato para  $MA(1)$

Os resultados para o estimador exato são apresentados abaixo:

Figura 9: Estimação por Máxima Verossimilhança Exata e MQC para o MA(1):  $X_t = u_t + 0.45 * X_{t-1}$



e os resultados para a estimação por Máxima Verossimilhança Exata são dados abaixo:

| Tabela 3: Modelling MA(1) by Exact MLE |             |            |              |         |
|--|-------------|------------|--------------|---------|
| EQ( 3) Modelling $MA(1)$ by Exact MLE  |             |            |              |         |
| The estimation sample is: 1 - 250      |             |            |              |         |
|  | Coefficient | Std. Error | t-Statistics | p-value |
| $MA(1)$                                | 0.5090      | 0.0669     | 7.6063       | 0.0000  |
| $const$                                | 0.9944      | 0.1841     | 5.4324       | 0.0000  |
| $R^2 = 0.1734$                         |             |            |              |         |
| no of observations: 250                |             |            |              |         |
| no of parameters: 2                    |             |            |              |         |
| $AIC = 2.7481$                         |             |            |              |         |
| $BIC = 2.7843$                         |             |            |              |         |
| log-likelihood = -342.5138             |             |            |              |         |

Temos os seguintes resultados para MQC

Tabela 4: Modelling  $MA(1)$  by Conditional OLS

| EQ( 4) Modelling $MA(1)$ by Conditional OLS |             |            |              |         |
|---|-------------|------------|--------------|---------|
| The estimation sample is: 1 - 250           |             |            |              |         |
|   | Coefficient | Std. Error | t-Statistics | p-value |
| $MA(1)$                                     | 0.5094      | 0.0547     | 9.3126       | 0.0000  |
| $const$                                     | 1.0016      | 0.0907     | 11.0430      | 0.0000  |
| $R^2 = 0.2140$                              |             |            |              |         |
| no of observations: 250                     |             |            |              |         |
| no of parameters: 2                         |             |            |              |         |
| $AIC = 2.7559$                              |             |            |              |         |
| $BIC = 2.8282$                              |             |            |              |         |
| log-likelihood = -342.4824                  |             |            |              |         |

Observe que a diferença entre o estimador exato e o aproximado é pequena uma vez que a amostra é grande.

## 4.2 Estimação no Modelo $ARMA(1, 1)$

Para estimar os parâmetros de um  $ARMA(1, 1)$  tanto podemos usar MQC quanto MV exata. A apresentação de MV exata será feita pelo Filtro de Kalman que será apresentado em modelos estruturais de séries temporais.

Quanto a estimação por MQC o comando no R para fazê-lo é:

```
arma11model <- arima(x, order = c(1,0,1),method=('CSS'), include.mean
= FALSE)
```

e usando este comando no programa da seção (1.1) temos:



Tabela 5: Modelling ARMA(1,1) by Conditional LSE

| Modelling $ARMA(1, 1)$ by CLSE    |             |            |              |         |
|-----------------------------------|-------------|------------|--------------|---------|
| The estimation sample is: 1 - 250 |             |            |              |         |
|                                   | Coefficient | Std. Error | t-Statistics | p-value |
| $AR(1)$                           | 0.1499      | 0.3214     | 0.4664       | 0.6413  |
| $MA(1)$                           | -0.0821     | 0.3214     | -0.2555      | 0.7985  |
| $R^2 = 0.0052$                    |             |            |              |         |
| no of observations: 250           |             |            |              |         |
| no of parameters: 2               |             |            |              |         |
| $AIC = 11.2868$                   |             |            |              |         |
| $BIC = 11.3592$                   |             |            |              |         |
| log-likelihood = -1408.8480       |             |            |              |         |

Observe que as raízes dos dois polinômios são muito próximas implicando que o modelo tem fatores comuns nos dois polinômios e pode ser simplificado para um  $ARMA(0, 0)$ . Observe também que os dois coeficientes são não significativos indicando, mais uma vez que o modelo pode ser simplificado.

### 4.3 Estimação do $ARMA$ para $\Delta LDolar$

Como na seção (1.2) o modelo identificado foi um  $ARMA(12,12)$  para a série  $\Delta LDolar$ , tal modelo foi estimado usando-se o seguinte programa no R:

Melhor  $ARMA(12,12)$  para  $\Delta LDolar$

obtendo-se os seguintes resultados:

Tabela 6: Modelling ARMA(12,12) by Exact MLE for DiDolar

| Modelling $ARMA(12, 12)$ by MLE for DiDolar |             |            |              |         |
|---|-------------|------------|--------------|---------|
| The estimation sample is: 1 - 250           |             |            |              |         |
|   | Coefficient | Std. Error | t-Statistics | p-value |
| $AR(1)$                                     | -0.064      | 0.3214     | 0.4664       | 0.6413  |
| $AR(3)$                                     | 0.1466      | 0.1290     | 1.1364       | 0.2558  |
| $AR(4)$                                     | 0.4085      | 0.0670     | 6.0991       | 0.0000  |
| $AR(5)$                                     | -0.0725     | 0.2121     | -0.3419      | 0.7325  |
| $AR(6)$                                     | 0.1051      | 0.0819     | 1.2832       | 0.1995  |
| $AR(7)$                                     | 0.4264      | 0.1325     | 3.2189       | 0.0000  |
| $AR(8)$                                     | -0.2561     | 0.3124     | -0.9196      | 0.7934  |
| $AR(9)$                                     | -0.1315     | 0.0713     | -1.8450      | 0.0652  |
| $AR(10)$                                    | -0.4044     | 0.0980     | -5.0468      | 0.0000  |
| $AR(11)$                                    | -0.4827     | 0.3221     | -1.5143      | 0.1301  |
| $AR(12)$                                    | 0.4233      | 0.3919     | 1.0721       | 0.2838  |
| $MA(1)$                                     | 0.1247      | 0.4893     | 0.2548       | 0.7989  |
| $MA(2)$                                     | -0.1117     | 0.3315     | -0.3369      | 0.7362  |
| $MA(3)$                                     | -0.1917     | 0.1245     | -1.5401      | 0.1237  |
| $MA(4)$                                     | -0.3833     | 0.0768     | -4.9920      | 0.0000  |
| $MA(5)$                                     | 0.0618      | 0.2341     | 0.2642       | 0.7916  |
| $MA(6)$                                     | -0.1246     | 0.0872     | -1.4295      | 0.1530  |
| $MA(7)$                                     | -0.4492     | 0.1393     | 3.2236       | 0.0013  |
| $MA(8)$                                     | 0.2597      | 0.3205     | 0.8087       | 0.4188  |
| $MA(9)$                                     | 0.1519      | 0.0664     | 2.2893       | 0.0211  |
| $MA(10)$                                    | 0.5544      | 0.1197     | 4.6320       | 0.0000  |
| $MA(11)$                                    | 0.4920      | 0.3735     | 1.3173       | 0.1877  |
| $MA(12)$                                    | -0.3685     | 0.4185     | -0.8805      | 0.3784  |
| $R^2 = 0.0469$                              |             |            |              |         |
| no of observations: 2405                    |             |            |              |         |
| no of parameters: 24                        |             |            |              |         |
| $AIC = -6.2957$                             |             |            |              |         |
| $BIC = -6,1605$                             |             |            |              |         |
| log-likelihood = -7594.59                   |             |            |              |         |

Através das raízes e também da não significância de alguns coeficientes, este modelo pode ser simplificado.

Observe que o procedimento a ser adotado para simplificar o modelo não é único e uma possível estratégia é estimar "todos" os modelos desta classe e escolher aquele que tem o menor Critério de Schwarz<sup>1</sup> tendo o cuidado de manter a amostra de estimação constante.

A tabela abaixo apresenta todos os modelos na classe  $ARMA(p, q)$  com  $p, q \leq 12$

---

<sup>1</sup>Vide apêndice para uma definição.

Tabela 7: Critério de Schwarz para todos  $ARMA(p, q)$  com  $p, q \leq 12$

| $\frac{MA}{AR}$ | 0                 | 1                 | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         | 11         | 12         |
|-----------------|-------------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 0               | -15083.480        | <b>-15084.710</b> | -15079.083 | -15073.832 | -15071.399 | -15063.615 | -15060.972 | -15054.838 | -15047.053 | -15039.891 | -15038.067 | -15031.390 | -15032.419 |
| 1               | <b>-15084.077</b> | -15077.879        | -15071.879 | -15068.589 | -15063.391 | -15055.841 | -15054.457 | -15047.080 | -15039.267 | -15040.232 | -15034.271 | -15026.502 | -15029.225 |
| 2               | -15079.862        | -15072.354        | -15081.923 | -15064.854 | -15070.330 | -15063.577 | -15047.316 | -15039.289 | -15045.117 | -15034.094 | -15033.892 | -15025.771 | -15021.761 |
| 3               | -15073.065        | -15072.308        | -15065.527 | -15057.907 | -15062.621 | -15058.821 | -15039.664 | -15038.251 | -15038.386 | -15034.622 | -15026.937 | -15017.927 | -15022.740 |
| 4               | -15071.973        | -15064.397        | -15070.382 | -15063.944 | -15051.777 | -15052.487 | -15037.421 | -15026.669 | -15035.387 | -15028.352 | -15019.250 | -15017.601 | -15019.721 |
| 5               | -15064.717        | -15059.802        | -15063.245 | -15055.182 | -15051.345 | -15048.831 | -15025.759 | -15020.864 | -15032.305 | -15016.177 | -15017.636 | -15024.539 | -15012.838 |
| 6               | -15061.336        | -15054.772        | -15047.473 | -15040.114 | -15039.203 | -15032.980 | -15039.571 | -15031.999 | -15031.065 | -15014.002 | -15020.280 | -15016.555 | -15005.029 |
| 7               | -15055.001        | -15047.215        | -15046.082 | -15036.232 | -15026.947 | -15030.695 | -15024.117 | -15031.261 | -15009.945 | -15018.491 | -15020.494 | -15008.299 | -15002.775 |
| 8               | -15047.217        | -15039.431        | -15044.292 | -15036.583 | -15038.995 | -15038.235 | -15029.196 | -15009.212 | -15002.871 | -15018.342 | -15010.764 | -15000.402 | -15008.327 |
| 9               | -15040.215        | -15040.577        | -15038.432 | -15035.982 | -15030.134 | -15022.818 | -15006.675 | -15012.421 | -15011.538 | -15002.802 | -15010.775 | -15002.496 | -14995.958 |
| 10              | -15036.312        | -15034.447        | -15035.213 | -15029.542 | -15030.126 | -15022.968 | -15014.587 | -15014.728 | -15012.441 | -15008.145 | -14993.733 | -14995.324 | -14994.436 |
| 11              | -15029.414        | -15026.825        | -15027.481 | -15021.144 | -15017.975 | -15020.356 | -15018.072 | -15012.504 | -14998.063 | -14995.118 | -14994.490 | -14992.660 | -14999.941 |
| 12              | -15034.697        | -15026.968        | -15023.238 | -15024.123 | -15022.678 | -15015.766 | -15007.874 | -15004.685 | -15002.238 | -14999.829 | -14999.071 | -15000.235 | -14986.682 |

O melhor modelo é portanto um  $MA(1)$  e o segundo melhor modelo é um  $AR(1)$ . Para o melhor modelo os parâmetros estimados estão abaixo

Tabela 8: Best Model for DIIDolar  $MA(1)$  with intercept estimate by Exact MLE

| Best Model for DIIDolar            |             |            |                |         |
|------------------------------------|-------------|------------|----------------|---------|
| The estimation sample is: 1 - 2405 |             |            |                |         |
|                                    | Coefficient | Std. Error | t-Statitistics | p-value |
| <i>Intercept</i>                   | 0.0000      | 0.0000     | 0.0712         | 0.9432  |
| <i>MA(1)</i>                       | 0.0633      | 0.0210     | 3.0200         | 0.0025  |
| $R^2 = 0.0045$                     |             |            |                |         |
| no of observations: 2405           |             |            |                |         |
| no of parameters: 2                |             |            |                |         |
| $AIC = -15104.07$                  |             |            |                |         |
| $BIC = -15076.92$                  |             |            |                |         |
| log-likelihood = -7554.033         |             |            |                |         |

#### 4.3.1 Exercícios

1. Para a série do  $\log(\text{IBV})$  determine o melhor modelo  $ARIMA$ .
2. Para a série do  $\log(\text{IBX})$  determine o melhor modelo  $ARIMA$ .
3. Para a série do  $\log(\text{S\&P500})$  determine o melhor modelo  $ARIMA$ .
4. Construa um experimento de Monte Carlo para comparar o estimador de MQC e MV exato para um modelo  $AR(1)$  com  $\phi = -0,9; 0; +0,9$
5. Construa um experimento de Monte Carlo para comparar o estimador de MQC e MV exato para um modelo  $MA(1)$  com  $\phi = -0,9; 0; +0,9$
6. Mostre que as duas primeiras autocorrelação parciais para um modelo  $AR(2)$  são estimadores dos coeficientes  $\phi_1$  e  $\phi_2$

7. As seguintes FAC e FACP foram observadas para uma série temporal com 1000 observações. Qual o modelo da classe *ARIMA* voce sugere. Como voce poderia obter uma estimativa para os parâmetros deste modelo.

|             | 1     | 2      | 3     | 4      | 5     | 6     | 7      | 8      | 9     | 10    |
|-------------|-------|--------|-------|--------|-------|-------|--------|--------|-------|-------|
| <i>FAC</i>  | 0.825 | 0.587  | 0.388 | 0.241  | 0.151 | 0.110 | 0.084  | 0.057  | 0.039 | 0.048 |
| <i>FACP</i> | 0.825 | -0.295 | 0.016 | -0.018 | 0.035 | 0.036 | -0.028 | -0.017 | 0.017 | 0.072 |

8. As seguintes FAC e FACP foram observadas para uma série temporal com 1000 observações. Qual o modelo da classe *ARIMA* voce sugere. Como voce poderia obter uma estimativa para os parâmetros deste modelo.

|             | 1     | 2      | 3     | 4     | 5      | 6      | 7      | 8      | 9      | 10    |
|-------------|-------|--------|-------|-------|--------|--------|--------|--------|--------|-------|
| <i>FAC</i>  | 0.033 | -0.036 | 0.014 | 0.019 | -0.058 | -0.020 | -0.024 | -0.020 | -0.060 | 0.038 |
| <i>FACP</i> | 0.033 | -0.037 | 0.016 | 0.017 | -0.058 | -0.015 | -0.028 | -0.018 | -0.059 | 0.039 |

9. As questões a seguir são do tipo falsas, verdadeiras ou ambíguas:

- (a) No modelo *AR*(1) a primeira autocorrelação é um estimador para o parâmetro  $\phi$
- (b) Um possível estimador para o parâmetro  $\theta$  é a primeira autocorrelação da série.
- (c) Qualquer modelo *ARIMA*(0, 1, 1) pode representar um modelo de nível local.
- (d) Qualquer modelo *ARIMA*(0, 2, 2) pode representar um modelo de tendência local.
- (e) No procedimento iterativo para obter o estimador do parâmetro  $\phi$  de um modelo *AR*(1) é sempre convergente.
- (f) No procedimento iterativo para obter o estimador do parâmetro  $\theta$  de um modelo *MA*(1) é sempre convergente.

## 5 Validação

### 5.1 Introdução

Para verificar a adequação do modelo estimado devemos testar:

- (i) hipóteses sobre os resíduos, se são ruídos brancos gaussianos
- (ii) hipóteses sobre constância dos parâmetros.

Vamos considerar o caso de um  $AR(1)$  estacionário, estimado por MQC. Suponha que se deseja testar hipóteses sobre os resíduos, que são os resíduos de MQO. Sabemos que estes resíduos são dados por:

$$\hat{u}_t = X_t - \hat{\phi}X_{t-1}$$

que em forma matricial será dado por:

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{X} - \mathbf{X}'_{-1}\hat{\phi} \\ &= \mathbf{X} - \mathbf{X}'_{-1}(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{X}'_{-1}\mathbf{X} \\ &= [\mathbf{I} - \mathbf{X}'_{-1}(\mathbf{X}'_{-1}\mathbf{X}_{-1})^{-1}\mathbf{X}'_{-1}]\mathbf{X} \\ &= \mathbf{M}_{\mathbf{X}_{-1}}\mathbf{X}\end{aligned}\tag{47}$$

onde  $\mathbf{X}_{-1}$  é a matriz  $\mathbf{X}$  defasada um período e  $\mathbf{M}_{\mathbf{X}_{-1}}$  é a matriz, idempotente, que dá a projeção ortogonal no espaço de  $\mathbf{X}_{-1}$ . Temos então:

$$\hat{\mathbf{u}} \sim N(0, \sigma^2\mathbf{M}_{\mathbf{X}_{-1}})\tag{48}$$

e como a matriz  $\mathbf{M}_{\mathbf{X}_{-1}}$  é diferente da identidade, temos que os resíduos de MQO são heteroscedásticos e tem correlação serial. Para se testar se estes resíduos são ruídos branco não devemos usar os resíduos de MQO uma vez que estes por construção são heteroscedasticos e apresentam correlação serial. Uma solução para este problema é apresenta na próxima seção.

## 5.2 Mínimos Quadrados Recursivos

Vamos apresentar este método de estimação para o caso geral de uma regressão múltipla, como foi feito em Brown et al. (1975) ou Harvey. Considere então o seguinte modelo de regressão:

$$y_t^1 = X_t^1 \beta + u_t^1 \quad \text{para } t = 1, \dots, T \quad (49)$$

onde  $y_t^1$  é um vetor  $t \times 1$ ,  $X_t^1$  é uma matriz  $t \times k$ ,  $\beta$  é um vetor  $k \times 1$  e  $u_t^1$ , por construção, tem distribuição  $NI(0, \sigma_y^2 I_t)$  e desejamos estimar  $\beta$  sequencialmente, isto é, para  $J = M > k, \dots, T$ .

Podemos fazer isto através da sequência de  $T - M + 1$  estimadores de  $\beta$  dados por:

$$\hat{\beta}_t = (X_t^{1'} X_t^1)^{-1} X_t^{1'} y_t^1 \quad \text{para } t = M, \dots, T \quad (50)$$

que é computacionalmente ineficiente uma vez que temos que inverter  $T - M + 1$  matrizes  $X_t^{1'} X_t^1$ .

Mas podemos usar a estrutura recursiva de  $X_{t+1}^{1'} X_{t+1}^1$  dado que:

$$X_{t+1}^{1'} X_{t+1}^1 = X_t^{1'} X_t^1 + x_{t+1} x_{t+1}' \quad (51)$$

O Lema de inversão de matrizes dado abaixo permite obter a inversa de (145)

**Lema 1 :** *Considere a matriz  $F$  não singular de ordem  $n + m$  dada por:*

$$F = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (52)$$

onde  $A$  é uma matriz de ordem  $n$ ,  $B$  e  $C'$  são matriz  $n \times m$  e  $D$  é de ordem  $m$ .

Quando  $A$  e  $H = (D - CA^{-1}B)$  são não singulares temos:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ -H^{-1}CA^{-1} & H^{-1} \end{bmatrix} \quad (53)$$

com  $|F| = |D - CA^{-1}B||A|$ .



Alternativamente quando  $D$  e  $G = (A - BD^{-1}C)$  são não singulares temos:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} G^{-1} & -G^{-1}BD^{-1} \\ -D^{-1}CG^{-1} & D^{-1} + D^{-1}CG^{-1}BD^{-1} \end{bmatrix} \quad (54)$$

com  $|F| = |A - BD^{-1}C||D|$ .

E por (147) e (54) temos que:

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} \quad (55)$$

e quando  $F$  é simétrica, isto é,  $A = A'$ ,  $D = D'$ ,  $C = B'$  temos

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}B')^{-1} & -(A - BD^{-1}B')^{-1}BD^{-1} \\ -(A - BD^{-1}B')^{-1}B'A^{-1} & (D - B'A^{-1}B)^{-1} \end{bmatrix} \quad (56)$$

Fazendo  $A = X_t^{1'}X_t^1$ ,  $B = C' = x_{t+1}$  e  $D = -1$  e por (56) temos:

$$\begin{aligned} (X_{t+1}^{1'}X_{t+1}^1)^{-1} &= (X_t^{1'}X_t^1 + x_{t+1}x_{t+1}')^{-1} \\ &= (X_t^{1'}X_t^1)^{-1} - \lambda_{t+1}\lambda_{t+1}'(1 + x_{t+1}'\lambda_{t+1})^{-1} \end{aligned} \quad (57)$$

onde  $\lambda_{t+1} = (X_t^{1'}X_t^1)^{-1}x_{t+1}$  e  $f_{t+1} = (1 + x_{t+1}'\lambda_{t+1}) = (1 + x_{t+1}'(X_t^{1'}X_t^1)^{-1}x_{t+1})$  que é um escalar.

Usando (57) em (50) temos:

$$\begin{aligned} \widehat{\beta}_{t+1} &= (X_{t+1}^{1'}X_{t+1}^1)^{-1}X_{t+1}^{1'}y_{t+1}^1 \\ &= (X_t^{1'}X_t^1)^{-1} - \lambda_{t+1}\lambda_{t+1}'(1 + x_{t+1}'\lambda_{t+1})^{-1}(X_t^{1'}y_t^1 + x_{t+1}y_{t+1}) \\ &= \widehat{\beta}_t + (X_t^{1'}X_t^1)^{-1}x_{t+1}y_{t+1} - (\lambda_{t+1}\lambda_{t+1}'X_t^{1'}y_t^1 + \lambda_{t+1}\lambda_{t+1}'x_{t+1}y_{t+1}) f_{t+1}^{-1} \\ &= \widehat{\beta}_t + \lambda_{t+1}y_{t+1} - \lambda_{t+1}[x_{t+1}'\widehat{\beta}_t + (f_{t+1} - 1)y_{t+1}] f_{t+1}^{-1} \\ &= \widehat{\beta}_t + \lambda_{t+1}v_{t+1}(1 + x_{t+1}'\lambda_{t+1})^{-1} \end{aligned} \quad (58)$$

onde  $v_{t+1} = y_{t+1} - \widehat{\beta}_t' x_{t+1}$  é chamado de erro de previsão um passo à frente ou resíduo recursivo. Observe que o erro de previsão pode ser reescrito da seguinte forma:

$$\begin{aligned} v_{t+1} &= y_{t+1} - \widehat{\beta}_t' x_{t+1} \\ &= u_{t+1} - (\beta - \widehat{\beta}_t)' x_{t+1} \end{aligned} \quad (59)$$

e vamos mostrar que a variância deste resíduo é dada por  $\sigma_y^2 f_{t+1}$  e que são serialmente não correlacionados e portanto os resíduos padronizados, isto é,  $\tilde{v}_{t+1} = \frac{v_{t+1}}{\sqrt{f_{t+1}}} \sim NI(0, \sigma_y^2)$ .

A variância de  $v_{t+1}$  é dada por:

$$\begin{aligned} Var(v_{t+1}) &= Var(u_{t+1} - (\beta - \widehat{\beta}_t)' x_{t+1}) \\ &= Var(u_{t+1}) + Var(x_{t+1}'(\beta - \widehat{\beta}_t)) - 2Cov(u_{t+1}, x_{t+1}'(\beta - \widehat{\beta}_t)) \\ &= \sigma_y^2 + x_{t+1}' Var(\beta - \widehat{\beta}_t) x_{t+1} \\ &= \sigma_y^2 + x_{t+1}' \sigma_y^2 (X_t^{1'} X_t^1)^{-1} x_{t+1} \\ &= \sigma_y^2 (1 + x_{t+1}' (X_t^{1'} X_t^1)^{-1} x_{t+1}) \\ &= \sigma_y^2 f_{t+1} \end{aligned}$$

Agora vamos mostrar que  $E(v_t v_s) = 0$  para  $t \neq s$

$$\begin{aligned} E(v_t v_s) &= E[(u_t - (\beta - \widehat{\beta}_{t-1})' x_t)(u_s - (\beta - \widehat{\beta}_{s-1})' x_s)] \\ &= E[u_t u_s - u_t (\beta - \widehat{\beta}_{s-1})' x_s - x_t' (\beta - \widehat{\beta}_{t-1}) u_s \\ &\quad + x_s' (\beta - \widehat{\beta}_{s-1}) (\beta - \widehat{\beta}_{t-1})' x_t] \\ &= E(u_t u_s) + E[u_t u_{s-1} X_{s-1} (X_{s-1}' X_{s-1})^{-1} x_s] + \\ &\quad E[x_t' (X_{t-1}' X_{t-1})^{-1} X_{t-1}' u_{t-1} u_s] + \\ &\quad E[x_s' (X_{s-1}' X_{s-1})^{-1} X_{s-1}' u_{s-1} u_{t-1} X_{t-1} (X_{t-1}' X_{t-1})^{-1} x_t] \\ &= 0 \end{aligned}$$

A Soma dos Quadrados dos Resíduos Padronizadas também pode ser escrita de forma recursiva um vez que satisfaz:

$$SQRer_{t+1} = SQRer_t + \frac{v_{t+1}^2}{f_{t+1}} \quad (60)$$

### 5.2.1 Testes Baseados na Estimação Recursiva

Observe que devido a natureza recursiva podemos construir intervalos de confiança para  $\hat{\beta}_t$  em cada instante de tempo uma vez que:

$$\hat{\beta}_t \sim N(\beta_t, \sigma_y^2 (X_t' X_t)^{-1}) \quad (61)$$

e podemos então contruir gráficos da estimativa do parâmetro contra o tempo e testar a constância do mesmo verificando se o intervalo de confiança do parâmetro inicial cobre os IC para os parâmetros nos outros instantes do tempo.

O comando no R para estimar por Mínimos Quadrados Recursivos o modelo  $AR(1)$  para a série `DiDolar`<sup>2</sup>, e apresentar o gráfico dos resíduos recursivos é dado por:

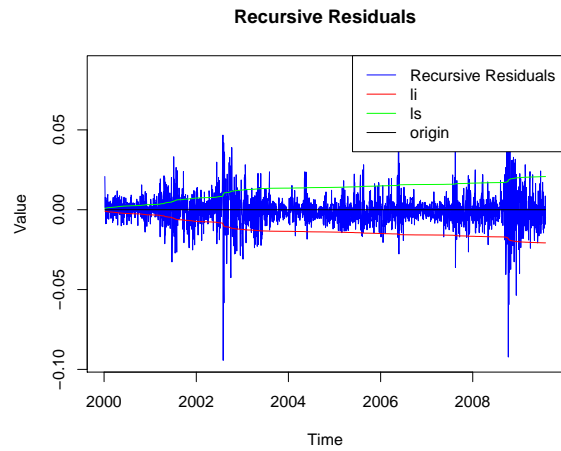
Estimação Recursiva para o  $AR(1)$  para `DiDolar`

tem-se então o seguinte gráfico

---

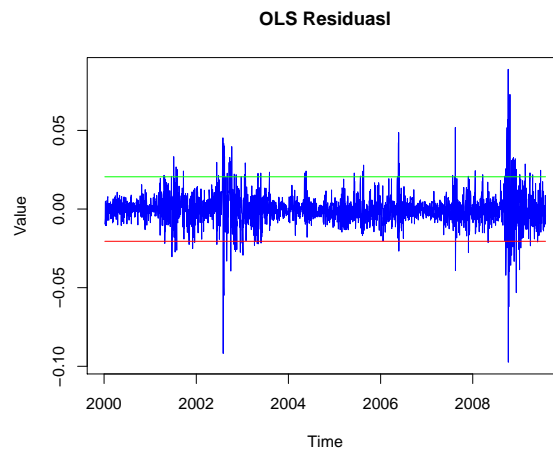
<sup>2</sup>Embora o R não tenha um comando para estimar por mínimos quadrados recursivos foi feito um programa no R para fazer esta estimação usando um *loop*. Poderíamos ter usado o melhor modelo que foi um  $MA(1)$  com constante. Mas optou-se por usar um  $AR(1)$ , que foi o segundo melhor modelo. Isto foi feito dado que alguns softwares só usam RLS quando a estimação é feita por MQO, que não é o caso para o modelo  $MA(1)$ .

Figura 10: Resíduos Recursivos



e abaixo é apresentado os resíduos de MQO

Figura 11: Resíduos MQO

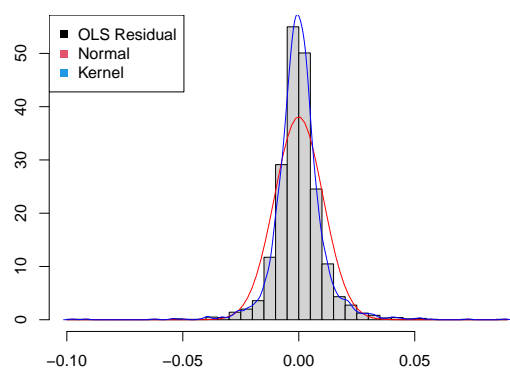


Observe que os resíduos de MQO saem mais frequentemente das bandas. Por outro lado pelo Gráfico (10) fica claro que a variância dos resíduos recursivos não é constante.

Vamos verificar se os resíduos de mínimos quadrados ordinários e os recursivos tem distribuição normal. Podemos testar normalidade comparando a distribuição empírica com a normal equivalente, ou através do QQplot ou usando o teste de Jarque & Bera (veja Jarque and Bera (1980).

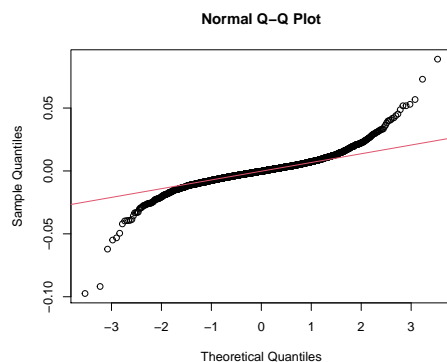
Primeiro o histograma para os resíduos de MQO.

Figura 12: Histograma dos Resíduos MQO



agora o QQPlot dos Resíduos de MQO

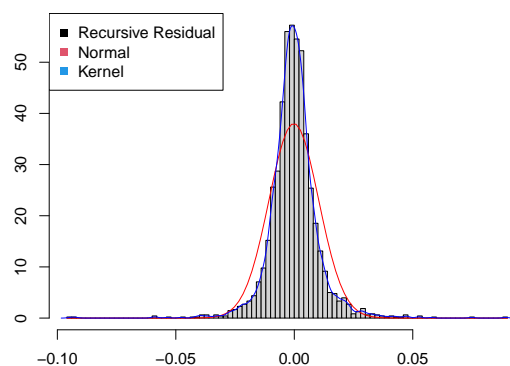
Figura 13: QQPlot dos Resíduos MQO



O teste de Jarque & Bera tem o seguinte resultado  $\chi^2(2) = 14838$  cujo p-valor é menor do que  $10^{-4}$ , logo normalidade é rejeita por este teste e também pelo histograma e QQPlot.

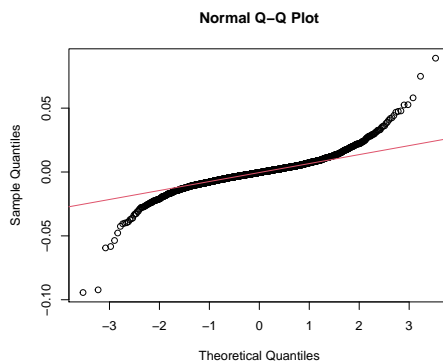
Agora para os Resíduos Recursivos temos primeiro o histograma

Figura 14: Histograma dos Resíduos Recursivos



A seguir é apresentado o QQ Plot para os resíduos recursivos.

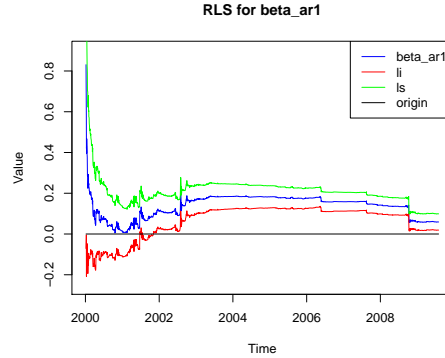
Figura 15: QQPlot dos Resíduos MQO



E por último o teste de Jarque & Bera cujo resultado  $\chi^2(2) = 13972$ , cujo p-valor é menor do que  $10^{-4}$ , logo normalidade é rejeita por este teste e também pelo histograma e QQPlot.

O gráfico abaixo apresenta a estimativa recursiva do parâmetro do  $AR(1)$

Figura 16: Estimativa Recursiva de  $\phi$



Observe que o IC do início da amostra cobre o IC final logo podemos considerar que este parâmetro é invariante no tempo. Embora em 2002 e 2008 existam evidências de quebras estruturais. Em 2002 foi o receio da vitória no Lula nas eleições e em 2008 a crise nos EUA.

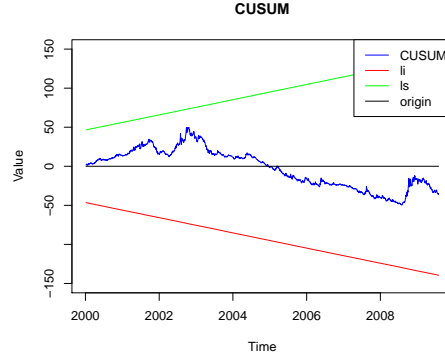
Como o erro de previsão padronizado tem distribuição  $NI(0, \sigma_y^2)$ , um teste proposto na literatura compara a soma dos resíduos padronizados, isto é,

$$W_t = \frac{\sum_{j=k+1}^t \tilde{v}_j}{s} \text{ para } t = k+1, \dots, T, \text{ com zero.}$$

Se os parâmetros são constantes então  $E(W_t) = 0$  e se os parâmetros não forem constantes  $W_t$  tenderá a divergir de zero.

Abaixo é apresentada este gráfico do CUSUM. Observe que os intervalos de confiança são obtidos observando que  $Var[W_t] = t$  para  $t = k+1, \dots, T$ . Observe que não se pode rejeitar a nula de que  $E[W_t] = 0$  implicando que os parâmetros são constantes.

Figura 17: CUSUM para o  $AR(1)$  estimado para D1Dolar



Outro teste, chamado de CUSUMSQ, é baseado na comparação da soma dos quadrados dos resíduos até um dado instante de tempo com a soma dos quadrados dos resíduos para toda a amostra, isto é

$$S_t = \frac{\sum_{j=k+1}^t \tilde{v}_j^2}{\sum_{j=k+1}^T \tilde{v}_j^2}$$

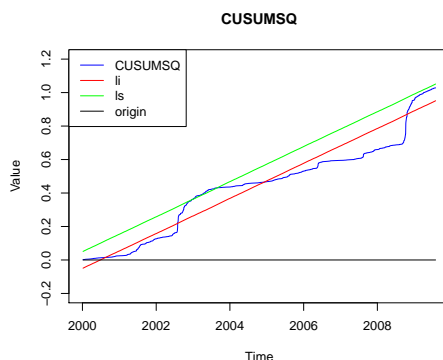
Sob a nula o quadrado dos resíduos tem distribuição  $\mathcal{N}^2(1)$  então o numerador tem valor esperado igual a  $t - k$  e o denominador de  $T - k$ . Logo em média  $E[S_t] = \frac{t-k}{T-k}$  que é uma reta dado que  $t$  varia de  $k$  até  $T$ .

Afastamentos desta reta indicam uma rejeição da nula de variância constante.

Temos o seguinte resultado



Figura 18: CUSUMSQ para o  $AR(1)$  estimado para DiDolar



e não se pode considerar a variância constante.

Estes resultados implicam que este modelo não é de todo adequado para esta série uma vez que é violada a hipótese de constância dos parâmetros e da variância.

### 5.2.2 Exercícios

1. Para a série  $\text{Log}(\text{Ibx})$  determine o melhor modelo da classe ARIMA. Estime o melhor modelo usando a estimação de mínimos quadrados recursiva e teste se os resíduos recursivos são normais, tem variância constante e o modelo não tem quebras estruturais na média condicional.
2. Refaça o exercício anterior para a série  $\text{Log}(\text{SP500})$ .
3. Refaça o exercício anterior para a série  $\text{Log}(\text{Ibv})$ .
4. Para a série ipca do arquivo ipca.xls, determine a ordem de integração usando toda a amostra. Usando só o período de 1994:08 até 2003:02 (pos-real) determine a ordem de integração. Estime o melhor modelo usando a amostra toda e a amostra pos-real. Usando uma aproximação AR adequada use estimação por mínimos quadrados recursivos e teste se os resíduos recursivos são normais, tem variância constante e o modelo não tem quebras estruturais na média condicional.

## 6 Previsão em Modelos ARMA

Vamos considerar primeiro a previsão para os modelos  $AR(1)$ ,  $MA(1)$  e  $ARMA(1, 1)$  e a seguir apresentamos o caso geral

### 6.1 Previsão de um $AR(1)$

Considere um modelo  $AR(1)$  dado por:

$$X_t = \phi_1 X_{t-1} + u_t \quad \text{para } t = 1, \dots, T \quad (62)$$

A previsão um passo à frente será dada por:

$$\begin{aligned} E[X_{T+1}|I_T] &= E[\phi_1 X_T + u_{T+1}|I_T] \\ &= \phi_1 X_T + E[u_{T+1}|I_T] \\ &= \phi_1 r_T \end{aligned} \quad (63)$$

e o erro de previsão é dado por:

$$\begin{aligned} v_{T+1} &= X_{T+1} - E[X_{T+1}|I_T] \\ &= \phi_1 X_T + u_{T+1} - \phi_1 X_T \\ &= u_{T+1} \end{aligned} \quad (64)$$

cuja variância, que corresponde ao Erro Médio Quadrático da Previsão um passo à frente, será dada por:

$$\begin{aligned} Var(v_{T+1}) &= Var(u_{T+1}) \\ &= \sigma^2 \end{aligned} \quad (65)$$

A previsão dois passos à frente será dada por:

$$\begin{aligned} E[X_{T+2}|I_T] &= E[\phi_1 X_{T+1} + u_{T+2}|I_T] \\ &= \phi_1 E[X_{T+1}|I_T] + E[u_{T+2}|I_T] \\ &= \phi_1 E[X_{T+1}|I_T] \end{aligned} \quad (66)$$

que pode ser reescrita da seguinte forma:

$$X_{T+2|T} = \phi_1 X_{T+1|T} \quad (67)$$

ou

$$X_{T+2|T} = \phi_1^2 X_T \quad (68)$$

Agora o erro de previsão dois passos à frente é dado por:

$$\begin{aligned} v_{T+2} &= X_{T+2} - E[X_{T+2}|I_T] \\ &= \phi_1 X_{T+1} + u_{T+2} - \phi_1 E[X_{T+1}|I_T] \\ &= u_{T+2} + \phi_1 [X_{T+1} - E[X_{T+1}|I_T]] \\ &= u_{T+2} + \phi_1 v_{T+1} \\ &= u_{T+2} + \phi_1 u_{T+1} \end{aligned} \quad (69)$$

e a variância é dada por:

$$\begin{aligned} Var(v_{T+2}) &= Var(u_{T+2} + \phi_1 u_{T+1}) \\ &= \sigma^2(1 + \phi_1^2) \end{aligned} \quad (70)$$

Por (67) ou (68) podemos obter a função de previsão para o modelo  $AR(1)$  que será dada por:

$$X_{T+l|T} = \phi_1 X_{T+l-1|T} \quad \text{para } l = 1, 2, \dots \quad (71)$$

ou

$$X_{T+l|T} = \phi_1^l X_T \quad \text{para } l = 1, 2, \dots \quad (72)$$

O erro de previsão  $l$  passos à frente é dado por:

$$\begin{aligned} v_{T+l} &= X_{T+l} - E[X_{T+l}|I_T] \\ &= \phi_1 X_{T+l-1} + u_{T+l} - \phi_1 E[X_{T+l-1}|I_T] \\ &= u_{T+l} + \phi_1 v_{T+l-1} \\ &= u_{T+l} + \phi_1 u_{T+l-1} + \dots + \phi_1^l u_{T+1} \end{aligned} \quad (73)$$

e a variância do erro de previsão é dada por:

$$\begin{aligned} Var(v_{T+l}) &= Var(u_{T+l} + \phi_1 u_{T+l-1} + \dots + \phi_1^l u_{T+1}) \\ &= \sigma^2(1 + \phi_1^2 + \dots + \phi_1^{2l}) \end{aligned} \quad (74)$$

Assumindo normalidade temos que a distribuição de  $X_{T+l}$  dado o passado é dada por

$$X_{T+l} | I_T \sim N(X_{T+l|T}, Var(X_{T+l} | I_T)) \quad (75)$$

e portanto podemos construir intervalos de confiança para as previsões que serão dados por

$$(X_{T+l|T} - z_{\alpha/2} \widehat{Var}(X_{T+l} | I_T), X_{T+l|T} + z_{\alpha/2} \widehat{Var}(X_{T+l} | I_T)) \quad (76)$$

onde

$$\widehat{Var}(X_{T+l} | I_T) = \hat{\sigma}^2(1 + \hat{\phi}_1^2 + \dots + \hat{\phi}_1^{2l}) \quad (77)$$

e  $\hat{\sigma}^2$  é a estimativa de  $\sigma^2$  assim como  $\hat{\phi}_1$  é a estimativa de  $\phi_1$ .

## 6.2 Previsão de um $MA(1)$

Considere um modelo  $MA(1)$  dado por:

$$X_t = u_t - \theta_1 u_{t-1} \quad \text{para } t = 1, \dots, T \quad (78)$$

A previsão um passo à frente será dada por:

$$\begin{aligned} E[X_{T+1} | I_T] &= E[u_{T+1} - \theta_1 u_T | I_T] \\ &= E[u_{T+1} | I_T] - \theta_1 E[u_T | I_T] \\ &= -\theta_1 u_T \end{aligned} \quad (79)$$

O erro de previsão é dado por:

$$\begin{aligned} v_{T+1} &= X_{T+1} - E[X_{T+1} | I_T] \\ &= u_{T+1} - \theta_1 u_T + \theta_1 u_T \\ &= u_{T+1} \end{aligned} \quad (80)$$

Variância será dada por:

$$\begin{aligned} Var(v_{T+1}) &= Var(u_{T+1}) \\ &= \sigma^2 \end{aligned} \quad (81)$$

A previsão dois passos à frente é dada por:

$$\begin{aligned} E[X_{T+2}|I_T] &= E[u_{T+2} - \theta_1 u_{T+1}|I_T] \\ &= E[u_{T+2}|I_T] - \theta_1 E[u_{T+1}|I_T] \\ &= 0 \end{aligned} \quad (82)$$

Agora o erro de previsão dois passos à frente é dado por:

$$\begin{aligned} v_{T+2} &= X_{T+2} - E[X_{T+2}|I_T] \\ &= u_{T+2} - \theta_1 u_{T+1} \end{aligned} \quad (83)$$

Variância será dada por:

$$\begin{aligned} Var(v_{T+2}) &= Var(u_{T+2} - \theta_1 u_{T+1}) \\ &= (1 + \theta_1^2)\sigma^2 \end{aligned} \quad (84)$$

A função de previsão para este modelo é dado por:

$$X_{T+l|T} = \begin{cases} -\theta_1 u_T & \text{para } l = 1 \\ 0 & \text{para } l \geq 2 \end{cases} \quad (85)$$

O erro de previsão  $l$  passos à frente é dado por:

$$v_{T+l} = \begin{cases} u_{T+1} & \text{para } l = 1 \\ u_{T+l} - \theta_1 u_{T+l-1} & \text{para } l \geq 2 \end{cases} \quad (86)$$

e a variância do erro de previsão  $l$  passos a frente dada por:

$$Var(v_{T+l}) = \begin{cases} \sigma^2 & \text{para } l = 1 \\ (1 + \theta_1^2)\sigma^2 & \text{para } l \geq 2 \end{cases}$$

### 6.3 Previsão de um $ARMA(1, 1)$

Considere um modelo  $ARMA(1, 1)$  dado por:

$$X_t = \phi_1 X_{t-1} + u_t - \theta_1 u_{t-1} \quad \text{para } t = 1, \dots, T \quad (87)$$

A previsão um passo à frente será dada por:

$$\begin{aligned} E[X_{T+1}|I_T] &= E[\phi_1 X_T + u_{T+1} - \theta_1 u_T | I_T] \\ &= \phi_1 E[X_T | I_T] + E[u_{T+1} | I_T] - \theta_1 E[u_T | I_T] \\ &= \phi_1 X_T - \theta_1 u_T \end{aligned} \quad (88)$$

o erro de previsão é dado por:

$$\begin{aligned} v_{T+1} &= X_{T+1} - E[X_{T+1} | I_T] \\ &= \phi_1 X_T + u_{T+1} - \theta_1 u_T - \phi_1 X_T + \theta_1 u_T \\ &= u_{T+1} \end{aligned} \quad (89)$$

A variância será dada por:

$$\begin{aligned} Var(v_{T+1}) &= Var(u_{T+1}) \\ &= \sigma^2 \end{aligned} \quad (90)$$

A previsão dois passos à frente é dada por:

$$\begin{aligned} E[X_{T+2}|I_T] &= E[\phi_1 X_{T+1} + u_{T+2} - \theta_1 u_{T+1} | I_T] \\ &= \phi_1 E[X_{T+1} | I_T] \end{aligned} \quad (91)$$

Pode ser escrita como:

$$\begin{aligned} E[X_{T+2}|I_T] &= \phi_1(\phi_1 X_T - \theta_1 u_T) \\ &= \phi_1^2 X_T - \phi_1 \theta_1 u_T \end{aligned} \quad (92)$$

O erro de previsão dois passos à frente é dado por:

$$\begin{aligned}
v_{T+2} &= X_{T+2} - E[X_{T+2}|I_T] \\
&= \phi_1 X_{T+1} + u_{T+2} - \theta_1 u_{T+1} - \phi_1 E[X_{T+1}|I_T] \\
&= u_{T+2} - \theta_1 u_{T+1} + \phi_1 (X_{T+1} - E[X_{T+1}|I_T]) \\
&= u_{T+2} - \theta_1 u_{T+1} + \phi_1 v_{T+1} \\
&= u_{T+2} + (\phi_1 - \theta_1) u_{T+1}
\end{aligned} \tag{93}$$

A variância do erro de previsão é dada por:

$$\begin{aligned}
Var(v_{T+2}) &= Var(u_{T+2} - \theta_1 u_{T+1} + \phi_1 v_{T+1}) \\
&= Var(u_{T+2}) + \theta_1^2 Var(u_{T+1}) + \phi_1^2 Var(v_{T+1}) - 2\theta_1 \phi_1 Cov(u_{T+1}, v_{T+1}) \\
&= \sigma^2(1 + (\phi_1 - \theta_1)^2)
\end{aligned} \tag{94}$$

A função de previsão para o modelo  $ARMA(1, 1)$  será dada por:

$$X_{T+l|T} = \begin{cases} \phi_1 X_T - \theta_1 u_T & \text{para } l = 1 \\ \phi_1 X_{T+l-1|T} & \text{para } l \geq 2 \end{cases} \tag{95}$$

O erro de previsão  $l$  passos à frente é dado por:

$$v_{T+l} = \begin{cases} u_{T+1} & \text{para } l = 1 \\ u_{T+l} - \theta_1 u_{T+l-1} + \phi_1 v_{T+l-1} & \text{para } l \geq 2 \end{cases} \tag{96}$$

A variância do erro de previsão  $l$  passos à frente será dada por:

$$Var(v_{T+l}) = \begin{cases} u_{T+1} & \text{para } l = 1 \\ \sigma^2 + \sigma^2 \theta_1^2 + \sigma^2 \phi_1^2 Var(v_{T+l-1}) - 2\sigma^2 \theta_1 \phi_1 & \text{para } l \geq 2 \end{cases} \tag{97}$$

## 6.4 Previsão de um $ARMA(p, q)$

Considere um modelo  $ARMA(p, q)$  dado por:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q} \quad \text{para } t = 1, \dots, T \tag{98}$$

e desejamos fazer previsão  $l$  passos a frente.

As previsões serão calculadas através do operador esperança condicional na informação disponível, isto é  $E[X_{T+l} | I_T] \triangleq X_{T+l|T}$  para  $l = 1, 2, \dots$ . Temos então

$$X_{T+l|T} = \phi_1 X_{T+l-1|T} + \dots + \phi_p X_{T+l-p|T} + u_{T+l|T} + \theta_1 u_{T+l-1|T} + \dots + \theta_q u_{T+l-q|T} \quad (99)$$

onde

$$X_{T+l|T} = X_{T+l} \text{ se } l < 0 \quad (100)$$

e

$$u_{T+l|T} = \begin{cases} 0 & \text{se } l > 0 \\ u_{T+l} & \text{se } l \leq 0 \end{cases} \quad (101)$$

Como estamos assumindo que o processo é Gaussiano temos que:

$$X_{T+l} \sim N(X_{T+l|T} : MSE(X_{T+l|T})) \quad (102)$$

onde

$$MSE(X_{T+l|T}) = \sigma^2(1 + \psi_1^2 + \dots + \psi_l^2) \quad (103)$$

e  $\psi'$ s são os coeficientes da representação  $MA(\infty)$ .

Podemos então obter os intervalos de confiança de previsão que serão dados por:

$$X_{T+l|T} \pm t_{n-p-q}(\alpha/2) * s(1 + \psi_1^2 + \dots + \psi_l^2)^{1/2} \quad (104)$$

## 6.5 Previsão para um $ARIMA(0, 1, 0)$ com constante

Considere uma passeio aleatório com *drift*, isto é,

$$X_t = X_{t-1} + \mu + u_t \quad \text{para } t = 1, \dots, T \quad (105)$$

onde  $u_t \sim NI(0, \sigma^2)$ .

Observe que a média não condicional de  $X_t$  é dada por



$$\begin{aligned}
E(X_t) &= E(X_0 + \mu t + \sum_{i=0}^{t-1} u_{t-i}) \\
&= X_0 + \mu t
\end{aligned} \tag{106}$$

que cresce linearmente, diferente do caso estacionário em que a média é constante.

Mas por outro lado a previsão um passo à frente, que é a média condicional no passado, é dada por

$$E(X_{T+1} | X_T) = X_T + \mu \tag{107}$$

e o erro de previsão um passo à frente é dado por

$$\begin{aligned}
v_{T+1} &= X_{T+1} - E[X_{T+1} | I_T] \\
&= X_T + \mu + u_{T+1} - (X_T + \mu) \\
&= u_{T+1}
\end{aligned} \tag{108}$$

Observe que a variância não condicional de (105) é dada por

$$Var(X_t) = t\sigma^2 \tag{109}$$

e a variância condicional é dada por

$$Var(X_{T+1} | X_T) = \sigma^2 \tag{110}$$

e portanto a variância do erro de previsão um passo à frente será dada por

$$Var(v_{T+1}) = Var(u_{T+1}) = \sigma^2 \tag{111}$$

que é igual ao caso estacionário.

Mas a previsão dois passos à frente será dada por

$$\begin{aligned}
E(X_{T+2} | X_T) &= E(X_{T+1} + \mu | X_T) \\
&= E(X_{T+1} | X_T) + \mu \\
&= X_T + 2\mu
\end{aligned} \tag{112}$$

e o erro de previsão dois passos à frente será dada por

$$\begin{aligned}
v_{T+2} &= X_{T+2} - E[X_{T+2}|I_T] \\
&= X_{T+1} + \mu + u_{T+2} - (E(X_{T+1} | X_T) + \mu) \\
&= v_{T+1} + u_{T+2} \\
&= u_{T+1} + u_{T+2}
\end{aligned} \tag{113}$$

e a variância do erro de previsão dois passos à frente será dada por

$$Var(v_{T+2}) = 2\sigma^2 \tag{114}$$

Observe que diferente do caso estacionário a função de previsão será dada pelo último valor da série observado mais duas vezes a constante e a variância do erro de previsão dois passos à frente é o dobro da variância do erro um passo à frente.

Temos então que a previsão  $l$  passos à frente será dada por

$$\begin{aligned}
E(X_{T+l}|X_T) &= E(X_{T+l-1} + \mu|X_T) \\
&= E(X_{T+l-1}|X_T) + \mu \\
&= X_T + (l-1)\mu + \mu \\
&= X_T + l\mu
\end{aligned} \tag{115}$$

que cresce linearmente

O erro de previsão será dado por

$$\begin{aligned}
v_{T+l} &= X_{T+l} - E[X_{T+l}|I_T] \\
&= X_{T+l-1} + \mu + u_{T+l} - (E(X_{T+l-1} | X_T) + \mu) \\
&= u_{T+l} + u_{T+l-1} + \dots + u_{T+1}
\end{aligned} \tag{116}$$

e a variância do erro de previsão será dada por

$$Var(v_{T+l}) = l\sigma^2 \tag{117}$$

## 6.6 Previsão para Transformações de Variáveis

Em algumas situações transformações dos dados são utilizadas e queremos fazer previsões para as séries transformadas, por exemplo se  $X_t$  é a série original queremos prever  $Y_t = g(X_t)$ , isto é dada as previsões  $X_{T+l|T}$  queremos obter

$$Y_{T+l|T} = g(X_{T+l|T}) \quad (118)$$

Se a função  $g(\cdot)$  for linear as previsões da série original são facilmente obtidas de (118) uma vez que previsões são invariante a transformações lineares. Por outro lado se a transformação for contínua e admitir uma inversa, temos que a previsão da série original é obtida através da expressão a seguir

$$X_{T+l|T} = g^{-1}(Y_{T+l|T}) \quad (119)$$

Um caso particular de (119) é quanto  $g(\cdot) = \ln(\cdot)$  e neste caso temos que a previsão para a série original é dada por

$$X_{T+l|T} = \exp(Y_{T+l|T}) \quad (120)$$

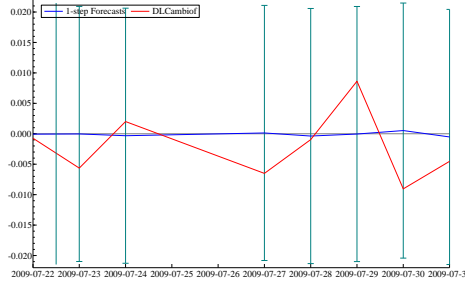
Mas no caso de distribuição normal a previsão ótima para a série original não é dada por (120) mas sim por

$$X_{T+l|T} = \exp(Y_{T+l|T} + 0.5Var(v_{T+l})) \quad (121)$$

## 6.7 Previsão para a série *DLDolar*

O modelo  $AR(1)$  para a série *DLDolar* foi reestimado sem as últimas 7 observações. Foram feitas as previsões e o gráfico abaixo compara as previsões, com os seus respectivos intervalos de confiança de previsão, com os valores observados para estas últimas observações.

Figura 19: Previsões para as últimas 7 observações para *DLDolar*



Observe que os valores observados estão dentro dos Intervalos de Confiança de Previsão, indicando que o modelo é capaz de prever bem. Observe também que as bandas de previsão são muito largas fazendo com que os valores previstos tenham muita incerteza associados a eles.

## 6.8 Exercícios

1. Para a série Log(Ibx) faça previsões para as 20 últimas observações usando o modelo que voce determinou na seção anterior, reestimando o modelo para a amostra sem as últimas 20 observações.
2. Refaça o exercício anterior para a série Log(SP500).
3. Refaça o exercício anterior para a série Log(Ibv).
4. Refaça o exercício anterior para a série ipca mas usando 10 observações para o período de previsão.

## 7 Avaliação de Previsões

Existem várias medidas que são usadas para avaliar previsões. Suponha que a amostra de previsão é para  $l = T + 1, \dots, T + h$  e denotando por  $X_t$  a série observada e por  $X_{t|T}$  os valores previstos podemos definir as seguinte medidas de avaliação de previsão

1. Raíz do Erro Médio Quadrático de Previsão = REMQP =  $\sqrt{\sum_{t=T+1}^{T+h} (X_{t|T} - X_t)^2 / h}$
2. Erro Médio Absoluto = EMA =  $\sum_{t=T+1}^{T+h} |X_{t|T} - X_t| / h$
3. Erro Médio Absoluto em Percentagem = EMAP =  $\sum_{t=T+1}^{T+h} |\frac{X_{t|T} - X_t}{X_t}| / h$
4. Índice de Desigualdade de Theil = IDT =  $\frac{\sqrt{\sum_{t=T+1}^{T+h} (X_{t|T} - X_t)^2 / h}}{\sqrt{\sum_{t=T+1}^{T+h} X_{t|T}^2 / h} + \sqrt{\sum_{t=T+1}^{T+h} X_t^2 / h}}$

Observe que as duas primeiras medidas são dependentes da escala da variável observada e devem ser usadas como uma medida relativa para a mesma variável entre diferentes modelos. As duas últimas medidas não dependem de escala e o coeficiente de Theil é um número entre 0 e 1 onde 0 representa um ajuste perfeito

Observe que o EMQ pode ser decomposto da seguinte forma

$$\sum_{t=T+1}^{T+h} (X_{t|T} - X_t)^2 / h = \left( \sum_{t=T+1}^{T+h} (X_{t|T}/h) - \bar{X} \right)^2 + (s_{X_{t|T}} - s_{X_t})^2 + 2(1-\rho)s_{X_{t|T}}s_{X_t} \quad (122)$$

e se dividirmos ambos os lados pelo lado esquerdo temos três proporções que correspondem a:

$$\begin{aligned} 1. \text{ Proporção do Viés} &= \frac{\sum_{t=T+1}^{T+h} (X_{t|T}/h - \bar{X})^2}{\sum_{t=T+1}^{T+h} (X_{t|T} - X_t)^2 / h} \\ 2. \text{ Proporção da Variância} &= \frac{(s_{X_{t|T}} - s_{X_t})^2}{\sum_{t=T+1}^{T+h} (X_{t|T} - X_t)^2 / h} \\ 3. \text{ Proporção da Covariância} &= \frac{2(1-\rho)s_{X_{t|T}}s_{X_t}}{\sum_{t=T+1}^{T+h} (X_{t|T} - X_t)^2 / h} \end{aligned}$$

## 8 Testes para Comparação de Previsões

A abordagem mais comum é classificar as previsões de acordo com a função de perda associada, tipicamente o EQPM (Erro Quadrático Médio de Previsão) ou o EMAP (Erro Médio Absoluto de Previsão).

Essas comparações são determinísticas; avalia-se se um EQPM é maior que os outros, mas não se a diferença entre eles é estatisticamente significativa.

### 8.1 Testes para Comparação de Previsão - Morgan-Granger-Newbold

O primeiro teste é devido a Granger and Newbold (1986), também conhecido como o teste de Morgan-Granger-Newbold, e requer que os erros de

previsão tenham média zero, sejam normalmente distribuídos e não correlacionados.

Se  $\mathbf{e}_1$  e  $\mathbf{e}_2$  indicam os erros de previsão de dois modelos concorrentes, o teste é baseado nas variáveis auxiliares:

$$u_{1,T+J} = e_{1,T+J} - e_{2,T+J} \quad (123)$$

$$u_{2,T+J} = e_{1,T+J} + e_{2,T+J} \quad (124)$$

Observe que

$$E(\mathbf{u}'_1 \mathbf{u}_2) = MSFE_1 - MSFE_2 \quad (125)$$

onde  $\mathbf{u}'_j = (u_{j,T+1}, \dots, u_{j,T+H})$  e  $j = 1, 2$ .

A hipótese de interesse é se  $\mathbf{u}_1$  e  $\mathbf{u}_2$  são não correlacionados ou não.

A estatística de teste é dada por:

$$\frac{r}{\sqrt{(H-1)^{-1}(1-r^2)}} \sim t_{H-1} \quad (126)$$

onde  $t_{H-1}$  é a distribuição t-Student com  $H-1$  graus de liberdade,  $H$  é o tamanho da amostra de avaliação e  $r$  é definido por:  $r = \frac{\sum_{i=1}^H u_{1,T+i} u_{2,T+i}}{\sqrt{\sum_{i=1}^H u_{1,T+i}^2 \sum_{i=1}^H u_{2,T+i}^2}}$ .

## 8.2 Testes para Comparação de Previsão - Diebold & Mariano

O segundo teste é devido a Diebold and Mariano (1995), com a estatística de teste dada por

$$DM = \frac{\sum_{j=1}^H d_j / H}{\sigma_d} = H^{1/2} \frac{\bar{d}}{\sigma_d} \quad (127)$$

onde

$$d_j = g(e_{1j}) - g(e_{2j}) \quad (128)$$

e  $g$  é uma função de perda de interesse, por exemplo, perda quadrática  $g(e) = e^2$  ou absoluta  $g(e) = |e|$ ,  $e_1$  e  $e_2$  são os erros de duas previsões concorrentes e  $\sigma_d^2$  é a variância de  $\bar{d}$ .

Esta variância é estimada por:

$$\hat{\sigma}_d^2 = \left( \gamma_0 + 2 \sum_{i=1}^{h-1} \gamma_i \right) \quad \text{com} \quad \gamma_k = \frac{1}{H} \sum_{t=k+1}^H (d_t - \bar{d})(d_{t-k} - \bar{d}) \quad (129)$$

onde  $h$  é o horizonte de previsão.

A hipótese nula é:

$$H_0 : E(d) = 0 \quad (130)$$

e sob a hipótese nula,  $DM$  tem uma distribuição assintótica normal padrão.

Uma versão modificada da estatística  $DM$  foi proposta por Harvey et al. (1998) e é dada por:

$$HLN = \left( \frac{H + 1 - 2h + H^{-1}h(h-1)}{HH} \right) DM \quad (131)$$

e ela é distribuída como uma  $t$  de Student com  $H1$  graus de liberdade.

Quando os modelos são aninhados, a distribuição da estatística de teste torna-se não padrão e é uma função de Movimentos Brownianos.

Uma solução é usar a estimação rolante (amostra crescente) em vez da estimação recursiva, veja Giacomini and White (2006).

### 8.3 Testes para Comparação de Previsão - Conjunto de modelos com confiança (*model confidence set*) (MCS)

O conjunto de modelos com confiança(MCS) é uma técnica de seleção de modelos de previsão desenvolvida por Hansen et al. (2011)

Consiste em um algoritmo que classifica um conjunto de previsões de um conjunto de modelos  $M^*$  que contém o conjunto dos melhores modelos selecionados a partir de um conjunto inicial de modelos.

O conjunto  $M^0$  é o conjunto que contém os melhores modelos definidos a partir de um critério de qualidade preditiva.

O conjunto que contém os melhores modelos é definido por  $M^* = \{i \in M^0 :: E(d_\tau^{i,j}) \leq 0 \text{ para todo } j \in M^0\}$

Seja  $M^\dagger$  o conjunto complementar, i.e.  $M^\dagger = \{i \in M^0 :: E(d_\tau^{i,j}) > 0 \text{ para todo } j \in M^0\}$  no qual  $g(e_\tau^i)$  é alguma função perda e

$$d_{\tau}^{i,j} = g(e_{\tau}^i) - g(e_{\tau}^j) \quad (132)$$

$$e_{\tau}^i = \tilde{y}_{t+\tau+h}^{t+\tau-i} - y_{t+\tau+h} \quad (133)$$

MCS seleciona o modelo usando um teste de equivalência,  $\delta_M$ , e uma regra de eliminação,  $\rho_M$ .

A regra de equivalência é aplicada ao conjunto  $M = M^0$ .

Se a regra de equivalência for rejeitada em um nível de confiança selecionado, então há, com alta probabilidade, um grupo de modelos ruins em termos de poder preditivo que devem ser eliminados do conjunto de bons modelos.

Neste caso, uma regra de eliminação,  $\rho_M$ , é usada para remover modelos com baixo poder preditivo do conjunto de bons modelos.

Feito isso, usamos novamente a regra de equivalência.

O procedimento é repetido até que a hipótese preditiva de equivalência no conjunto analisado, não seja rejeitada.

O conjunto de modelos da última etapa ( $\widetilde{M}_F$ ) é selecionado e deve conter os melhores modelos em um determinado nível de significância.

A hipótese nula do teste de equivalência é dada por:

$$H_M^0 : E(d_{\tau}^{i,j}) = 0 \text{ para todo } i, j \in M \quad (134)$$

onde  $M \subset M^0$ .

A hipótese alternativa é dada por:

$$H_M^1 : E(d_{\tau}^{i,j}) \neq 0 \text{ para todo } i, j \in M \quad (135)$$

Um ponto importante a enfatizar é que podem existir modelos melhores fora do conjunto inicial de modelos "candidatos"  $M^0$ .

O objetivo é classificar um conjunto particular de modelos para obter  $M^*$

A hipótese nula pode ser testada a partir das seguintes estatísticas

$$T_{R,M} = \max_{i,j \in M} |t_{ij}| \quad (136)$$

onde  $t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\text{Var}(\bar{d}_{ij})}}$  para algum  $i$  e  $j$ .

A estatística de teste dada por (136) tem uma distribuição estatística não padrão, mas pode ser simulada usando técnicas de bootstrap.

A regra de eliminação é dada por:



$$\rho_M = \arg \max_{i \in M} \left\{ \sup_{j \in M} (t_{ij}) \right\} \quad (137)$$

O algoritmo MCS possui os seguintes passos:

- (i) Inicia o procedimento configurando o conjunto inicial de modelo a ser analisado  $M = M^0$ ;
  - (ii) Testa  $H_M^0$  usando  $\delta_M$  e um nível de significância  $\alpha$ ;
  - (iii) Se  $H_M^0$  não for rejeitado, o procedimento termina e o conjunto final é  $\widehat{M}_{1-\alpha}^* = M$ , caso contrário, utilizamos a regra de eliminação  $\rho_M$  para remover modelos do conjunto  $M$  e voltar ao passo (i).
- Os autores sugerem que o MCS possui as seguintes propriedades estatísticas:

- (i)  $\lim_{n \rightarrow \infty} (M^* \subset \widehat{M}_{1-\alpha}^*) > 1 - \alpha$  e;
- (ii)  $\lim_{n \rightarrow \infty} (i^\dagger \in \widehat{M}_{1-\alpha}^*) = 0$  para todo  $i^\dagger \in M^\dagger$

## 8.4 Testes de Comparação de Previsões - Classificando os modelos: valores-p para MCS

A regra de eliminação,  $\rho_M$  define uma sequência de conjuntos aleatórios,  $M_0 = M_1 \supset M_2 \supset \dots \supset M_{m_0}$ , onde  $M_i = \rho_i, \dots, \rho_{m_0}$  e  $m_0$  é o número de elementos em  $M_0$ ,  $\rho_{m_0}$  é o primeiro elemento a ser eliminado,  $\rho_{m_1}$  é o segundo a ser eliminado, e assim por diante.

No final, apenas um modelo sobrevive.

Configure o valor-p do modelo final para um.

É armazenado nos valores-p dos modelos deletados se eles forem maiores do que o valor-p do modelo previamente eliminado.

Se o valor-p da remoção atual for armazenado. Os valores-p do MCS são importantes pois facilitam para o analista avaliar um dado conjunto  $\widehat{M}_{1-\alpha}^*$ .

## 9 A Combinação de Previsões

Quando modelos alternativos estão disponíveis, podemos combiná-los, construindo uma previsão conjunta.

Vamos supor que duas previsões  $\hat{y}_1$  e  $\hat{y}_2$  estejam disponíveis para a mesma série alvo  $y$ .

Os erros de previsão associados são  $e_1$  e  $e_2$ .

Queremos construir a previsão combinada linearmente, isto é:

$$\hat{y}_c = \alpha \hat{y}_1 + (1 - \alpha) \hat{y}_2 \quad (138)$$

os pesos podem ser escolhidos de forma a minimizar o *MSFE* de  $\hat{y}_c$ .

De (138) temos:

$$\begin{aligned} e_c &= y - \hat{y}_c \\ &= \alpha y - \alpha \hat{y}_1 + (1 - \alpha) y - (1 - \alpha) \hat{y}_2 \\ &= \alpha e_1 + (1 - \alpha) e_2 \end{aligned} \quad (139)$$

onde a variância de (139) (*MSFE*) é dada por:

$$\sigma_{e_c}^2 = \alpha^2 \sigma_{e_1}^2 + (1 - \alpha)^2 \sigma_{e_2}^2 + 2\alpha(1 - \alpha)\rho\sigma_{e_1}\sigma_{e_2} \quad (140)$$

onde  $\rho$  é o coeficiente de correlação entre  $e_1$  e  $e_2$ .

O peso ótimo, os minimizadores de (140) são (veja Bates and Granger (1969))

$$\alpha^* = \frac{\sigma_{e_2}^2 - \rho\sigma_{e_1}\sigma_{e_2}}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho\sigma_{e_1}\sigma_{e_2}} \quad (141)$$

o que resulta em:

$$\sigma_{e_c, optimal}^2 = \frac{\sigma_{e_1}^2 \sigma_{e_2}^2 (1 - \rho^2)}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho\sigma_{e_1}\sigma_{e_2}} \quad (142)$$

Podemos mostrar que:

$$\begin{aligned}
\sigma_{e_{c,optimal}}^2 - \sigma_{e_1}^2 &= \frac{\sigma_{e_1}^2 \sigma_{e_2}^2 (1 - \rho^2)}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho\sigma_{e_1}\sigma_{e_2}} - \sigma_{e_1}^2 \\
&= \frac{\sigma_{e_1}^2 \sigma_{e_2}^2 (1 - \rho^2) - \sigma_{e_1}^2 (\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho\sigma_{e_1}\sigma_{e_2})}{\sigma_{e_1}^2 + \sigma_{e_2}^2 - 2\rho\sigma_{e_1}\sigma_{e_2}} \\
&= \frac{\sigma_{e_1}^2 (\sigma_{e_2}^2 - \rho^2 \sigma_{e_2}^2 - \sigma_{e_1}^2 - \sigma_{e_2}^2 + 2\rho\sigma_{e_1}\sigma_{e_2})}{(\sigma_{e_1} - \rho\sigma_{e_2})^2 + \sigma_{e_2}^2 (1 - \rho^2)} \\
&= \frac{\sigma_{e_1}^2 (-\rho^2 \sigma_{e_2}^2 - \sigma_{e_1}^2 + 2\rho\sigma_{e_1}\sigma_{e_2})}{(\sigma_{e_1} - \rho\sigma_{e_2})^2 + \sigma_{e_2}^2 (1 - \rho^2)} \\
&= \frac{-\sigma_{e_1}^2 (\sigma_{e_1} - \rho\sigma_{e_2})^2}{(\sigma_{e_1} - \rho\sigma_{e_2})^2 + \sigma_{e_2}^2 (1 - \rho^2)} \leq 0 \\
\Rightarrow \sigma_{e_{c,optimal}}^2 &\leq \sigma_{e_1}^2 \tag{143}
\end{aligned}$$

\*\*\*\*\*

onde a igualdade vale se  $\rho = \frac{\sigma_{e_1}}{\sigma_{e_2}}$ , o que implica que  $\hat{y}1$  ou  $\hat{y}2$  são previsões ótimas.

Se os erros de previsão forem não correlacionados,  $\rho = 0$ , (141) se reduz a:

$$\alpha^* = \frac{\sigma_{e_2}^2}{\sigma_{e_1}^2 + \sigma_{e_2}^2} \tag{144}$$

que são pesos comumente usados em aplicações empíricas mesmo com erros correlacionados.

Na prática,  $\alpha$  não é conhecido e deve ser estimado.

Uma estimativa de  $\alpha$  é obtida pela regressão, sobre a amostra de estimação:

$$y = \alpha \hat{y}1 + (1 - \alpha) \hat{y}2 + e \tag{145}$$

ou

$$e_2 = \alpha(\hat{y}1 - \hat{y}2) + e \tag{146}$$

## 10 Abrangente em Previsão

Abrangente em Previsão, introduzido por Chong and Hendry (1986), refere-se à questão de se a previsão de um passo de um modelo pode explicar os erros de previsão feitos por outro (que, portanto, não é eficiente de forma forte, para uma aplicação recente para Orçamento americano veja Ericsson and Martinez (2019) e para previsão da inflação brasileira veja Vicente and Valls Pereira (2022)).

Pode-se usar a regressão (146) e testar se  $\alpha = 0$ .

Se  $\alpha \neq 0$ , a diferença entre  $\hat{y}_1$  e  $\hat{y}_2$  pode explicar parcialmente  $e_2$  e, portanto, o segundo modelo não pode abranger completamente o primeiro na previsão.

Um teste mais direto pode ser baseado na regressão:

$$e_1 = \delta \hat{y}_2 + u \quad (147)$$

e requer que  $\delta = 0$  para que o segundo modelo não abranja completamente o primeiro na previsão.

## 11 Apêndice I - Critérios de Informação

Critério de Informação de Akaike, veja Akaike (1973), é definido por:

$$AIC = -\frac{2}{T} \ln(\text{verossimilhança}) + \frac{2}{T} \times (k) \quad (148)$$

onde a função de verossimilhança é avaliada no estimador de máxima verossimilhança e  $k$  é o número de parâmetros.

Critério de Informação de Schwarz (BIC), veja Schwarz (1978), é definido por:

$$BIC = -\frac{2}{T} \ln(\text{verossimilhança}) + \frac{\ln(T)}{T} \times (k) \quad (149)$$

$BIC$  tente a escolher a ordem correta para um modelo  $AR$ , enquanto que  $AIC$  tendem a super-parametrizar o modelo.

## 12 Apêndice II - Inferência para Estimadores de Máxima Verossimilhança

Se o tamanho da amostra  $T$  é suficientemente grande temos que a distribuição do estimador de máxima verossimilhança  $\hat{\theta}$  pode ser aproximado por uma distribuição normal, isto é:

$$\hat{\theta} \approx N(\theta_o : T^{-1}\mathbf{I}^{-1}) \quad (150)$$

onde  $\theta_o$  é o verdadeiro vetor de parâmetros e  $\mathbf{I}$  é a matriz de informação que pode ser estimada de dois jeitos.

A primeira forma de estimar é através da matriz de derivadas segundas, isto é,

$$\hat{\mathbf{I}}_{2D} = -T^{-1} \frac{\partial^2 \log l(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}} \quad (151)$$

onde

$$\log l(\theta) = \sum_{t=1}^T \log l_t(\theta) \quad (152)$$

Em geral a matriz de derivadas segundas é calculada numericamente. Substituindo (151) em (150) temos que a matriz de variância-covariância de  $\hat{\theta}$  pode ser aproximada por:

$$E[(\hat{\theta} - \theta_o)(\hat{\theta} - \theta_o)'] = \left[ -\frac{\partial^2 \log l(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\hat{\theta}} \right]^{-1} \quad (153)$$

Uma segunda forma de estimar a matriz de informação é através da estimativa de *outer-product* (produto externo) que é dada por:

$$\hat{\mathbf{I}}_{OP} = T^{-1} \sum_{t=1}^T \left[ q_t(\hat{\theta}) \right] \left[ q_t(\hat{\theta}) \right]' \quad (154)$$

onde  $q_t(\hat{\theta})$  denota o escore da  $t$ -ésima observação avaliado no estimador de máxima verossimilhança, isto é:

$$q_t(\hat{\theta}) = \frac{\partial \log l(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}}$$

e neste caso a matriz de variância-covariância de  $\hat{\theta}$  pode ser aproximada por:

$$E[(\hat{\theta} - \theta_o)(\hat{\theta} - \theta_o)'] = \left[ \sum_{t=1}^T \left[ q_t(\hat{\theta}) \right] \left[ q_t(\hat{\theta}) \right]' \right]^{-1} \quad (155)$$

Os estimadores de máxima verossimilhança geram estimadores com boas propriedades quando as inovações são não i.i.d. e não Gaussianas, os desvios padrões calculados usando (151) e (154) não serão válidos. Uma aproximação para a matriz de variância e covariância no caso de má especificação da distribuição probabilística é dada por:

$$E[(\hat{\theta} - \theta_o)(\hat{\theta} - \theta_o)'] \approx T^{-1} \left\{ \hat{\mathbf{I}}_{2D} \hat{\mathbf{I}}_{OP}^{-1} \hat{\mathbf{I}}_{2D} \right\}^{-1} \quad (156)$$

que é matriz de variância e covariância do estimador de quasi-verossimilhança.

## 13 Referências

### Referências

- H. Akaike. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60(2):255–265, 1973. doi: 10.1093/biomet/60.2.255.
- J. M. Bates and C. W. J. Granger. The combination of forecasts. *Operational Research Quarterly*, 20(4):451–468, 1969.
- G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.
- R. L. Brown, J. Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):149–192, 1975. ISSN 00359246. URL <http://www.jstor.org/stable/2984889>.
- Y. Y. Chong and D. Hendry. Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, 53(4):671–690, 1986. URL <https://EconPapers.repec.org/RePEc:oup:restud:v:53:y:1986:i:4:p:671-690>.

- F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.
- N. R. Ericsson and A. B. Martinez. Evaluating government budget forecasts. *The Palgrave Handbook of Government Budget Forecasting*, pages 37–69, 2019.
- R. F. Galbraith and J. I. Galbraith. On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of Applied Probability*, 11(1):63–71, mar 1974. doi: 10.2307/3212583.
- R. Giacomini and H. White. Tests of conditional predictive ability. *Econometrica*, 74:1545–1578, 2006.
- C. W. J. Granger and P. Newbold. *Forecasting Economic Time Series*. Academic Press, 2 edition, 1986.
- P. R. Hansen, A. Lunde, and J. M. Nason. The model confidence set. *Econometrica*, 2(79):453–497, 2011.
- A. C. Harvey. *The Econometric Analysis of Time Series - 2nd Edition (London School of Economics Handbooks in Economics)*. The MIT Press. ISBN 9780262081894.
- D. S. Harvey, S. J. Leybourne, and P. Newbold. Tests for forecast encompassing. *Journal of Business and Economic Statistics*, 16:254–259, 1998.
- C. M. Jarque and A. K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3): 255–259, jan 1980. doi: 10.1016/0165-1765(80)90024-5.
- G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, mar 1978. doi: 10.1214/aos/1176344136.
- P. L. Valls Pereira. Exact likelihood function for a regression model with MA(1) errors. *Economics Letters*, 24(2):145–149, jan 1987. doi: 10.1016/0165-1765(87)90241-2.
- H. d. O. R. Vicente and P. L. Valls Pereira. Forecasting inflation using online daily prices: A midas approach for brazil. *SSRN Electronic Journal*, 2022. ISSN 1556-5068. doi: 10.2139/ssrn.4044997.