

BOX & JENKINS COM FUNÇÃO DE TRANSFERÊNCIA USANDO O R: UM ESTUDO PARA A SÉRIE DE ENA DO SUBSISTEMA SUL

Pedro Guilherme Costa Ferreira

Instituto Brasileiro de Economia, Fundação Getúlio Vargas
R. Barão de Itambi - 60, Botafogo, Rio de Janeiro, Brasil
pedro.guilherme@fgv.br

Reinaldo Castro Souza

Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro
R. Marquês de São Vicente - 225, Gávea, Rio de Janeiro, Brasil
reinaldo@ele.puc-rio.br

Daiane Marcolino de Mattos

Instituto Brasileiro de Economia, Fundação Getúlio Vargas
R. Barão de Itambi - 60, Botafogo, Rio de Janeiro, Brasil
daiane.mattos@fgv.br

RESUMO

A utilização de variáveis auxiliares nos modelos de Box & Jenkins por meio de funções de transferências pode aperfeiçoar a modelagem e previsão de séries temporais. Este artigo tem como objetivo apresentar tal método utilizando o software R, fato este ainda não abordado na literatura internacional. Além disso, um estudo utilizando as séries de Manchas Solares e SOI (*Southern Oscillation Index*) para modelar a série ENA do subsistema Sul mostrou resultados superiores quando se comparam as métricas do modelo com função de transferência em relação a um modelo ARIMA.

PALAVRAS CHAVE. Box & Jenkins com função de transferência, Energia Natural Afluente, SOI.

Área Principal: EN, EST

ABSTRACT

The use of auxiliary variables in Box & Jenkins time series models through transfer functions can improve the modelling and forecast. This article aims to present this methodology using the R software, which still hasn't been presented in the international literature. Further, a study using Sunspot and SOI (*Southern Oscillation Index*) to model the Southwest subsystem showed ENA time series better results when compared with ARIMA models.

KEYWORDS. Box & Jenkins with Transfer Function, Natural Inflow Energy, SOI.

Main Area: EN, EST

1. Introdução

Os modelos de Box and Jenkins (1970) podem incorporar outras variáveis auxiliares. A forma como essas variáveis auxiliares X_t influenciam a variável resposta Y_t , isto é, como os movimentos dessas variáveis afetam o percurso da variável resposta, é dada por uma função de transferência $f(X_t)$:

$$Y_t = f(X_t) + \varepsilon_t \quad (1)$$

onde ε_t pode ser um ruído branco ou um modelo ARIMA completo.

A função $f(X_t)$ pode agrupar valores passados e/ou presentes de uma ou mais séries temporais (TS), que podem ser do tipo quantitativo ou binário (*dummy*) e essa distinção implica na forma de identificação de $f(X_t)$. No caso de uma TS quantitativa, a forma genérica de uma $f(X_t)$ pode ser denotada pela equação seguinte. Nota-se que para identificar $f(X_t)$, é necessário descobrir os valores de r , s e b e estimar os parâmetros $w_i, i = 0, \dots, s$ e $\delta_j, j = 1, \dots, r$.

$$f(X_t) = \frac{(w_0 + w_1L + w_2L^2 + \dots + w_sL^s)}{(1 - \delta_1L - \delta_2L^2 - \dots - \delta_rL^r)} X_{t-b} \quad (2)$$

Na literatura é possível encontrar artigos sobre o tema: Hanssens (1980), por exemplo, ao modelar vendas de compostos vegetais a partir de gastos com publicidade, mostrou que a metodologia tem um desempenho superior a alguns modelos univariados. Nogales and Conejo (2006) utilizaram o método para prever o preço de eletricidade e mostraram ser uma boa alternativa em relação a modelos ARIMA e Redes Neurais. Kannan and Farook (2013), com foco em aquecimento global, modelaram a temperatura atmosférica através da emissão de CO_2 e concluíram que os resultados são de boa qualidade. McDowell (2002) ensina como pode ser fácil ajustar um modelo de função de transferência usando o software STATA.

Dada a importância da metodologia, este artigo tem como objetivo apresentar tal método utilizando o software R, fato este ainda não abordado na literatura internacional. Para atingir o objetivo, o artigo foi dividido em mais quatro seções além desta. Na seção seguinte, o algoritmo proposto e a metodologia são abordados simultaneamente. Em outras palavras, utilizando as ST disponibilizadas por Box and Jenkins (1970) explicita-se a metodologia proposta pelos autores (e.g. identificar e estimar funções de transferência) utilizando-se o software R. Na seção três aplica-se o algoritmo proposto a um problema real vivido pelo Setor Elétrico Brasileiro (SEB) que é a previsão/simulação de Energia Natural Afluente¹ (ENA). Na seção posterior comparam-se os resultados aferidos pelo modelo de Box & Jenkins com função de transferência e um modelo SARIMA. Por fim, na quinta seção é apresentada a conclusão e trabalhos futuros.

2. Algoritmo proposto e metodologia: uma abordagem utilizando as ST disponibilizadas por Box & Jenkins (1970)

Para identificar e estimar uma função de transferência (FT), será utilizado um exemplo clássico extraído de Box and Jenkins (1970) que se passa no contexto de uma investigação de otimização adaptativa em que um aquecedor a gás foi utilizado na combinação de ar e metano para formar uma mistura de gases contendo CO_2 (dióxido de carbono). A alimentação de ar foi mantida constante, mas a taxa de alimentação de metano poderia ser variada de qualquer maneira desejada. Após a combinação, a concentração de CO_2 resultante nos gases de exaustão foi medida. A finalidade do exemplo é encontrar a forma como a variável Taxa de Alimentação de Metano (X_t) se relaciona com Concentração de CO_2 (Y_t). Cada variável representa uma série temporal de 296 observações.

¹Em linhas gerais, denomina-se ENA (Energia Natural Afluente) a quantidade de eletricidade que pode ser gerada pelo parque hidrelétrico com a água que chega às centrais hidrelétricas. Esta energia é estimada assumindo que o nível dos reservatórios esteja num patamar médio de 65% de sua capacidade total e uma política de operação. Vale lembrar que este valor pode mudar de acordo com a política de operação (Terry et al., 1986).

2.1. Pacotes e dados necessários

Durante a modelagem de FT serão utilizados alguns pacotes do R. São eles: **utils**, **stats**, **graphics**, **forecast** e **TSA**. Os três primeiros são pacotes básicos do R, assim, não é necessária a instalação deles. Os outros pacotes, no entanto, precisam ser instalados. Pode-se usar a função `install.packages()` para tal finalidade.

```
> install.packages("forecast")
> install.packages("TSA")
```

O download dos dados do exemplo pode ser feito diretamente pelo R se possuir conexão com internet. O local onde ficará o arquivo baixado é o diretório de trabalho, que pode ser indicado pela função `getwd()`. Execute as linhas de comando abaixo para efetuar o download do arquivo, que está nomeado de `gas.csv`. Acesse o diretório de trabalho para conferir se o download foi executado conforme desejado.

```
> download.file(url = "http://datasets.connectmv.com/file/gas-furnace.csv",
+              destfile = "./gas.csv")
```

Para ler no R o arquivo baixado anteriormente utiliza-se a função `read.csv()`. Além disso, é preciso transformar esse conjunto de dados em um objeto de séries temporais fazendo uso da função `ts()`.

```
> dados <- read.csv("gas.csv")
> dados.ts <- ts(dados)
```

As colunas do objeto `dados.ts` referem-se à variável independente X_t e à variável dependente Y_t , nessa ordem. O gráfico das duas variáveis é exposto a seguir. Nota-se que as séries não têm um comportamento estacionário² em todo o espaço de tempo, com variações não constantes e uma leve tendência de queda para X_t , e crescimento para Y_t .

```
> plot(dados.ts, main = "Input Gas Rate (X) & CO2 (Y)")
```

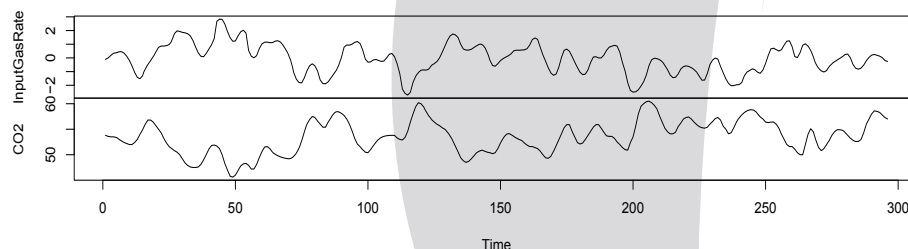


Figura 1: Input Gas Rate (X) & CO2 (Y)

²Para confirmar essa afirmação o mais correto seria fazer um teste de raiz unitária, contudo como o objetivo do artigo é apenas explorar a metodologia Box & Jenkins com função de transferência, não explorou-se tal assunto. Para explorar melhor os testes de raiz unitária ver: Dickey and Fuller (1979); Phillips and Perron (1988).

2.2. Construindo funções de transferência

Esses são os seguintes passos para obter um modelo de BJ com função de transferência:

1. Calcular a função de correlação cruzada entre Y_t e X_t ;
2. Identificar r , s e b ;
3. Estimar o modelo BJ com TF;
4. Verificar se o modelo é adequado.

2.2.1. Calcular a função de correlação cruzada entre Y e X

Como visto na seção 1, para identificar funções de transferência, inicialmente, basta estimar valores para r , s e b . A identificação desses valores é feita calculando a função de correlação cruzada (CCF) entre Y_t e as variáveis auxiliares (apenas uma neste exemplo). A CCF entre as séries temporais Y_t e X_t mostra as correlações entre elas para diferentes defasagens no tempo. É definida como:

$$ccf(k) = \frac{c_{xy}(k)}{s_x s_y}, \quad k = 0, \pm 1, \pm 2, \dots \quad (3)$$

onde:

s_x e s_y representam, respectivamente, os desvios-padrão de X_t e Y_t .

$c_{xy}(k)$ representa a covariância entre as duas variáveis no lag k :

$$c_{xy}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), & k = 0, 1, 2, \dots \\ \frac{1}{n} \sum_{t=1}^{n+k} (y_t - \bar{y})(x_{t-k} - \bar{x}), & k = 0, -1, -2, \dots \end{cases}$$

Para $k > 0$, a CCF mostra o relacionamento entre X no tempo t e Y no tempo futuro $t + k$. Para valores negativos de k , tem-se o relacionamento entre X no tempo t e Y no tempo passado $t - k$.

É importante saber que a CCF é afetada pela autocorrelação de X_t e Y_t , e se as TS não forem estacionárias, o resultado da CCF não refletirá realmente o grau de associação entre elas (Hamilton, 1994; Phillips and Perron, 1988). O método de pré-branqueamento³ é sugerido por BJ para corrigir esse problema.

O pré-branqueamento permite eliminar a estrutura de tendência (determinística ou estocástica) presente numa série temporal. O método consiste nas seguintes etapas:

- a. Ajustar um modelo ARIMA para a série independente X_t ;
- b. Filtrar Y_t pelo modelo encontrado em (a), isto é, o modelo de Y_t é o mesmo modelo de X_t (com os mesmos parâmetros estimados);
- c. Salvar os resíduos dos dois modelos;
- d. Calcular a CCF entre os resíduos obtidos em (c).

A seguir discute-se cada etapa separadamente aplicando-as às ST disponíveis por BJ.

a. Ajustar um modelo ARIMA para a série independente X_t : Para identificar as ordens do modelo ARIMA para X_t , usa-se a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF)⁴.

³Existe também o pré-branqueamento duplo, em que é ajustado para cada variável seu próprio modelo ARIMA. No entanto, por tornar as duas séries um ruído branco, a correlação entre elas pode ser puramente devida ao acaso.

⁴Um método mais robusto que auxilia na identificação de modelos ARIMA são os critérios de informação, como por exemplo AIC (Akaike, 1973) e BIC (Schwarz, 1978). Os mesmos serão utilizados na seção ??, na qual é apresentado um exemplo mais atual.

```
> acf(dados.ts[, "InputGasRate"], lag.max = 36)
> pacf(dados.ts[, "InputGasRate"], lag.max = 36)
```

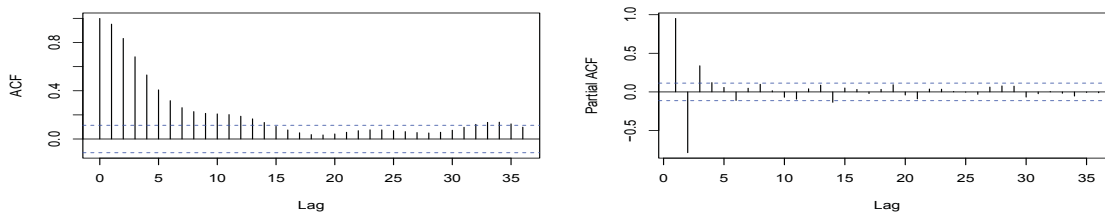


Figura 2: ACF and PACF: Input Gas Rate (X)

A queda exponencial da ACF e o corte brusco da PACF no lag 3 (figura 2) sugerem um modelo ARIMA(3,0,0). Ao ajustar o modelo ARIMA (função `Arima` do pacote **forecast**), a constante não foi significativa, sendo esta excluída do modelo. A seguir são apresentadas as rotinas para a estimação do modelo.

```
> library("forecast")
> modelo_x <- Arima(dados.ts[, "InputGasRate"], order = c(3, 0, 0),
+                 include.mean = F)
```

b. Filtrar Y_t pelo modelo encontrado em (a): Para filtrar Y_t por meio do modelo de X_t , também utiliza-se a função `Arima`, no entanto, é necessário acrescentar o argumento `model` indicando o modelo já estimado anteriormente.

```
> modelo_y <- Arima(dados.ts[, "CO2"], model = modelo_x)
```

c. Salvar os resíduos dos dois modelos: Os resíduos podem ser obtidos pela função `resid()` e serão salvos em novos objetos: `alpha` e `beta` para os modelos de X e Y , respectivamente.

```
> alpha <- resid(modelo_x)
> beta <- resid(modelo_y)
```

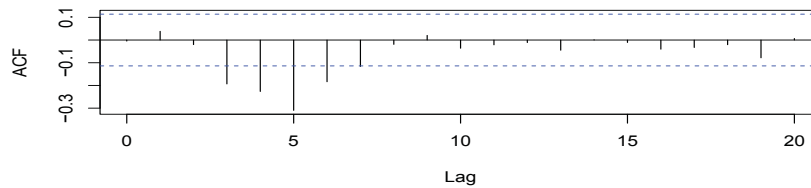
d. Calcular a CCF entre os resíduos obtidos em (c): Feito o pré-branqueamento, é possível calcular a CCF (figura 3) entre os resíduos utilizando a função `ccf()` do pacote **stats**. A CCF nesse gráfico mostra o relacionamento entre Y_t e os lags defasados de X_t a partir dos coeficientes de correlação. Nota-se que não há correlação significativa entre Y_t e X_t no tempo presente ($t = 0$) e a primeira correlação significativa é dada para $t = 3$, ou seja, entre Y no tempo presente e X defasada em 3 lags.

```
> ccf(beta, alpha, xlim = c(0, 20))
```

2.2.2. Identificar r , s e b

- b : refere-se ao primeiro lag significativo. Representa a primeira defasagem de X que entrará no modelo. Neste caso, $b = 3$.
- s : número de lags crescentes depois de b . Representa as próximas defasagens de X que entrarão no modelo. Logo, $s = 2$.
- r : por haver queda exponencial⁵ após os lags crescentes, $r = 1$.

⁵Queda exponencial na CCF indica parâmetros no denominador (r), enquanto picos indicam parâmetros no numerador (s). Veja um exemplo em (Box and Jenkins, 1970, fig. 10.6).


 Figura 3: CCF: Input Gas Rate (X) and CO₂ (Y)

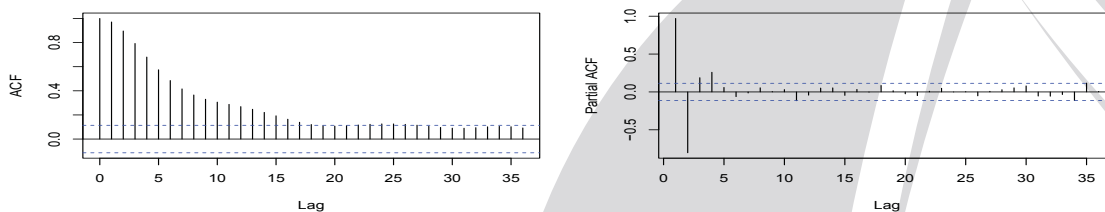
Portanto, o modelo contém X_{t-3} , X_{t-4} e X_{t-5} e a $f(X_t)$ é definida como

$$f(X_t) = \frac{(w_0 + w_1L + w_2L^2)}{(1 - \delta_1L)} X_{t-3}$$

2.2.3. Estimar o modelo BJ com FT

Uma vez identificada a forma de $f(X_t)$, inicia-se a etapa de estimação do modelo com função de transferência. Inicialmente, é preciso identificar a ordem do modelo ARIMA para a série Y_t , assim como foi feito para X_t .

```
> acf(dados.ts[, "CO2"], lag.max = 36, main = "ACF: CO2 (Y)")
> pacf(dados.ts[, "CO2"], lag.max = 36, main = "PACF: CO2 (Y)")
```


 Figura 4: ACF and PACF: CO₂ (Y)

A queda exponencial da ACF e o corte brusco da PACF no lag 2 (figura 4) sugerem um modelo ARIMA(2,0,0). Estima-se o modelo com função de transferência utilizando a função `arimax()` do pacote **TSA**.

- `x`: série dependente (Y_t);
- `order`: ordem do modelo ARIMA para Y_t ;
- `xtransf`: série independente (X_t) já defasada no lag b ;
- `transfer`: valores de r e s na forma `list(c(r, s))`.

Para defasar a variável X_t em três lags, pode-se usar a função `lag()` do pacote **stats**.

```
> x_novo <- lag(dados.ts[, "InputGasRate"], k = -3)
```

Como três valores de X_t foram "perdidos" para estimar o modelo, é preciso cortar os três primeiros valores de Y_t para os dois conjuntos de dados terem o mesmo tamanho.


```
> dados_novo <- cbind(x_novo, dados.ts[, "CO2"])
> dados_novo <- na.omit(dados_novo)
> colnames(dados_novo) <- c("InputGasRate", "CO2")
```

Com os dados na forma correta, estima-se um modelo para a variável dependente CO₂.

```
> modelo_ft <- arimax(x = dados_novo[, "CO2"], order = c(2, 0, 0),
+                   xtransf = dados_novo[, "InputGasRate"],
+                   transfer = list(c(1, 2)))
```

O modelo pode ser representado pela seguinte equação:

$$Y_t = 53.4 + \frac{(-0.5310 - 0.3801L - 0.5180L^2)}{(1 - 0.5490L)} X_{t-3} + \frac{1}{1 - 1.5272L + 0.6288L^2} e_t$$

2.2.4. Verificar se o modelo é adequado

Para avaliar se o modelo ajustado é adequado, as seguintes análises são propostas⁶: ACF dos resíduos e CCF entre os resíduos e a variável auxiliar X_t pré-branqueada. Se houver padrões de correlação, sugere-se que o modelo não esteja bem especificado, e conseqüentemente deve ser modificado.

```
> residuos <- resid(modelo_ft)
> acf(residuos, na.action = na.omit, lag.max = 36)
> ccf(residuos, alpha, na.action = na.omit)
```

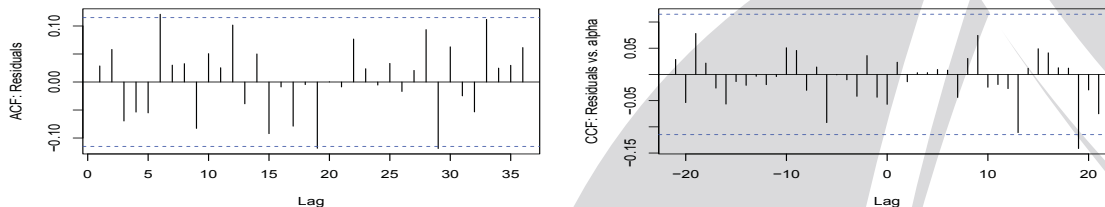


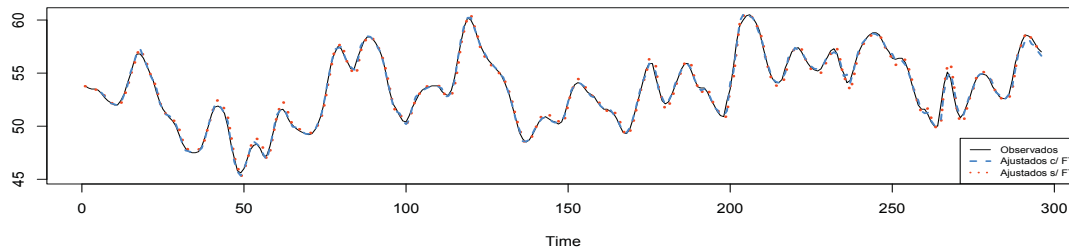
Figura 5: ACF: Resíduos

A partir da análise das funções de correlação (figura 5), pode-se concluir que esse modelo com função de transferência está adequado. O gráfico dos valores observados versus ajustados pelos modelos com e sem função de transferência é exposto na figura 6.

```
> modelo_y <- Arima(dados.ts[, "CO2"], order = c(2, 0, 0), include.mean = T)
> ajustados <- fitted(modelo_y)
> ajustados_ft <- fitted(modelo_ft)
> ts.plot(dados.ts[, "CO2"], ajustados, ajustados_ft, lty = c(1, 3, 2),
+         lwd = c(1, 3, 2), col = c(1, "orangered", "dodgerblue"))
```

A partir da análise dos modelos com e sem FT, nota-se que ambos são adequados para a modelagem de CO₂ dentro da amostra. Os resultados do modelo com função de transferência parecem ser menos erráticos, porém, a diferença não é expressiva entre os dois modelos: o MAPE para o modelo com função de transferência é de 0.3% e para o outro modelo é de 0.5%. Outra forma de verificar qual dos dois modelos é mais eficiente é analisar suas respectivas previsões para fora da amostra. Essa análise será feita na seção ?? com um exemplo mais atual.

⁶Para um melhor diagnóstico dos resíduos deveriam ser executados, por exemplo, os testes de Ljung and Box (1978); Engle (1982); Jarque and Bera (1980). No estudo de caso apresentado na seção três, os mesmos são utilizados.


 Figura 6: TS observada (CO₂) e valores ajustados com e sem FT

3. Estudo de Caso: Energia Natural Afluente (subsistema Sul)

No que tange à geração de energia, o Brasil possui cerca de 90% de seu potencial composto por usinas hidrelétricas. Dadas as incertezas associadas aos fenômenos naturais e ainda sob a influência de fenômenos influentes de longo prazo, como o El Niño, La Niña e as Manchas Solares, a atividade de planejamento e operação energética do sistema hidrelétrico brasileiro torna-se um problema complexo, estocástico e de solução não trivial.

É importante lembrar que para o planejamento do sistema elétrico brasileiro é necessário o estudo e o ajuste das séries de Energia Natural Afluente (ENA) dos 4 subsistemas brasileiros (Sudeste/Centro-Oeste, Sul, Nordeste e Norte) que compõem o Sistema Interligado Nacional (Ferreira et al., 2015a,b).

A base de dados utilizada está disponibilizada publicamente na página do Operador Nacional do Sistema Elétrico, ONS⁷. Os dados utilizados são mensais e o horizonte de estudo inicia-se em janeiro de 1932 e termina em dezembro de 2008, totalizando 77 anos de dados e um total de 924 observações. A unidade de medida da ENA é MWmed. As covariáveis presentes são:

- SOI: Southern Oscillation Index, ou Índice de Oscilação Sul (SOI), é uma medida das flutuações de grande escala na pressão do ar que ocorrem entre o oeste e leste do Pacífico tropical (i.e., o estado da Oscilação do Sul) durante os episódios de ocorrência dos fenômenos meteorológicos El Niño e La Niña.
- Manchas Solares (MS)

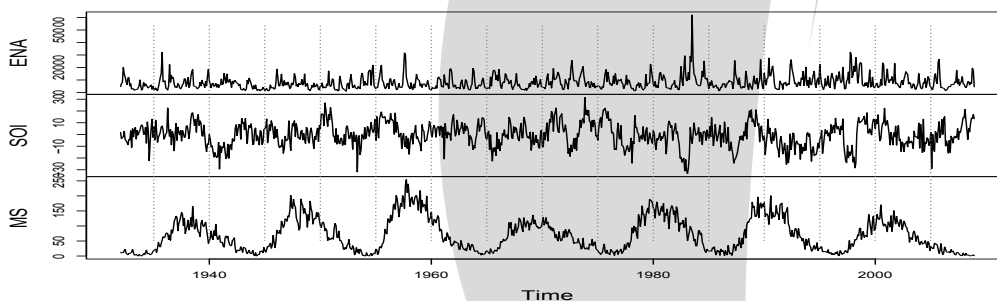


Figura 7: ENA, SOI and MS

Como se pode observar na figura 7, as séries parecem ser estacionárias, apesar de existirem alguns pontos discrepantes. Uma análise do teste de raiz unitária Augmented Dickey-Fuller

⁷Visite www.ons.org.br.

(ADF) permite concluir que não há indícios de raiz unitária em nenhuma das três séries temporais (tabela 1). Dessa forma, não é necessário diferenciá-las.

TS	Equação	Lag	Estatística τ	Valor Crítico
ENA	Trend	15	-7.3913	-3.41
SOI	Trend	13	-7.6717	-3.41
MS	Drift	20	-6.0147	-2.86

Tabela 1: Augmented Dickey-Fuller Test

Serão seguidos os mesmos passos vistos na seção 2 para ajustar um modelo de BJ com FT. Como existe mais de uma covariável, o procedimento de pré-branqueamento acontecerá duas vezes. É importante ressaltar que com o objetivo de estabilizar a variância, foi aplicada a transformação log em ENA.

3.1. Pré-branqueamento

O modelo ARIMA identificado para as séries independentes SOI e MS são, respectivamente, um ARIMA(3,0,0) sem constante e ARIMA(3,0,0) com constante. Os coeficientes dos modelos foram significativos e, embora de acordo com a figura 8 existam autocorrelações significativas em alguns lags, acredita-se que isso ocorra pela quantidade de dados observados, tornando o intervalo de confiança mais estreito. Dessa forma, considerou-se ambos os modelos adequados. Em seguida, filtrou-se a série ENA pelos dois modelos anteriores e os resíduos foram armazenados.

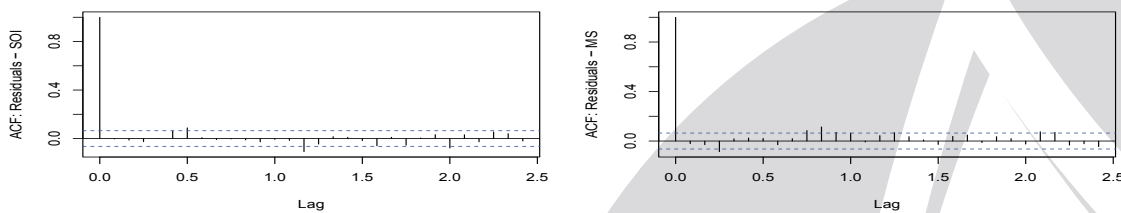


Figura 8: ACF: Resíduos modelo SOI e MS

3.2. Identificar r, s e b

A CCF entre as variáveis pré-branqueadas não resultou em lags significativos que indicassem que valores passados ou presentes das covariáveis teriam relação com valores presentes da ENA como se pode ver na figura 9. Dessa forma, preferiu-se então tentar, exaustivamente, ajustar um modelo com FT que minimizasse o critério de informação de Schwarz (BIC) a partir de diversas combinações entre as duas covariáveis defasadas do lag 1 até o lag 12, e que todas fossem significativas no modelo.

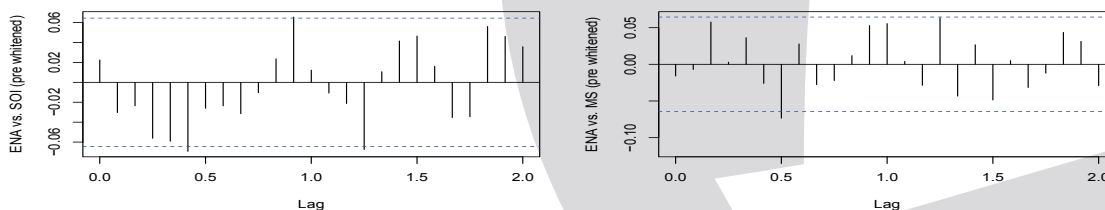


Figura 9: CCF: ENA vs. SOI e MS (pré-branqueadas)

Para defasar as variáveis no lag i , $i = 1, \dots, 12$, foram utilizados os seguintes códigos:

```
> SOI_i <- lag(SOI, k = -i)
> MS_i <- lag(MS, k = -i)
```

3.3. Estimar o modelo BJ com FT

O modelo ARIMA encontrado para ENA (log transformada), utilizando a ACF e PACF foi um ARIMA(1,0,0)(0,0,1). A tabela 2 mostra três modelos ajustados com FT e o modelo BJ sem FT. O primeiro modelo foi obtido considerando minimizar o critério de informação BIC e a significância dos parâmetros. O segundo e o terceiro modelo, foram obtidos apenas separando as covariáveis do primeiro modelo. Nota-se que o modelo sem covariáveis (modelo 4) possui o menor critério de informação, embora não haja mudanças expressivas quando se compara aos outros três modelos.

Modelo	Covariáveis	BIC
1	SOI(-7) MS(-12)	1389.66
2	SOI(-7)	1387.11
3	MS(-12)	1387.58
4	Sem covariáveis	1384.82

Tabela 2: Critérios de informação dos modelos BJ com FT

É importante informar que a função `arimax` ainda não permite realizar previsões por meio de funções já implementadas no R. Nesse caso, foi utilizada a função `Arima`, com o argumento `xreg` para especificar covariáveis, para o ajuste dos modelos 1, 2 e 3. Esta forma de ajuste considera nulo o termo r^8 da $f(X_t)$ - equação 2. Por isso, a função de transferência passa a ser definida da seguinte forma (no caso de apenas uma variável):

$$f(X_t) = (w_0 + w_1L + \dots + w_sL^s)X_t \quad (4)$$

Assim, os três modelos são definidos como:

$$\begin{aligned} \text{Modelo 1: } Y_t &= 8.6773 - 0.0046SOI(-7) + 0.0014MS(-12) + \frac{(1 - 0.1714L^{12})e_t}{1 - 0.5972L} \\ \text{Modelo 2: } Y_t &= 8.7762 - 0.0045SOI(-7) + \frac{(1 - 0.1667L^{12})e_t}{1 - 0.5991L} \\ \text{Modelo 3: } Y_t &= 8.6810 - 0.0014MS(-12) + \frac{(1 - 0.1698L^{12})e_t}{1 - 0.6017L} \\ \text{Modelo 4: } (1 - 0.6037L)Y_t &= (1 - 0.1651L^{12})e_t \end{aligned}$$

3.4. Verificar se o modelo está adequado

Além de testar se há alguma estrutura de correlação entre os resíduos de cada modelo e as covariáveis pré-branqueadas, outros testes estatísticos podem ser usados para confirmar se o modelo está adequado. Foram aplicados os testes de Ljung-Box, Jarque-Bera e ARCH para verificar, respectivamente, autocorrelação, normalidade e heterocedasticidade nos resíduos. O teste de Ljung-Box para autocorrelação foi aplicado no lag 24 para os resíduos dos quatro modelos, e não há indícios de autocorrelação residual, considerando 95% de confiança; O teste de Jarque-Bera foi aplicado aos resíduos de cada modelo, e mesmo com a remoção de outliers, os dados ainda não foram considerados aproximadamente normais ao nível de confiança de 95%. No entanto, essa característica já não era esperada uma vez que a distribuição da ENA não é simétrica; Não há indícios de heterocedasticidade nos resíduos dos quatro modelos ao considerar 95% de confiança, de acordo com o teste Arch.

Assim, os modelos parecem similares quanto a adequação, uma vez que o resultado dos testes são semelhantes. Para verificar qual modelo se destaca dos outros, será feita uma análise fora

⁸É importante lembrar que o termo r já havia sido identificado como zero ao analisar-se a CCF

da amostra. Aquele modelo que possuir menor erro de previsão será definido como superior em relação aos outros.

3.5. Desempenho dos modelos fora da amostra

A previsão de modelos BJ no R é estimada pela função `forecast` do pacote **forecast**⁹. Foram estimados 6 passos à frente (jan/2009 a jun/2009) para cada um dos quatro modelos, com intervalo de confiança de 95% para a predição, utilizando a forma dos códigos abaixo. Em modelos com covariáveis, é necessário especificar seus valores futuros h-passos à frente para que a previsão h-passos à frente da variável alvo seja estimada. Logo, precisa-se de 6 valores futuros para as covariáveis. Esses valores estão disponíveis pois as variáveis foram defasadas em pelo menos 7 lags. Se não estivessem disponíveis, esses valores deveriam ser estimados.

```
> library("forecast")
> previsaol <- forecast(modelo_ft1, h = 6, level = 0.95,
+                       xreg = dados_prev[,c("SOI_7", "MS_12")])
```

	Observado	Modelo 1	Modelo 2	Modelo 3	Modelo 4
jan/09	5702	4816.438	5013.87	4960.932	5155.946
feb/09	5018	4902.927	5204.699	5002.253	5299.537
mar/09	4492	4898.023	5198.68	5117.163	5419.337
abr/09	1865	5126.876	5516.195	5486.839	5885.795
may/09	2635	5575.077	6026.881	5915.264	6377.187
jun/09	3877	5695.744	6165.191	6135.452	6620.883

Tabela 3: Previsões pontuais 6 passos à frente para ENA do Subsistema Sul

Os modelos para o estudo de caso ENA do subsistema Sul do Brasil se mostraram similares em relação aos valores ajustados dentro da amostra. No entanto, quando se comparam as previsões, os resultados vistos nas seções anteriores permitem concluir que o modelo de Box & Jenkins com função de transferência com as duas covariáveis SOI e MS se mostra superior ao modelo ARIMA sem covariáveis. Há uma redução significativa do erro absoluto médio percentual (MAPE) conforme observado na tabela 4.

Modelos	MAPE (%)
1	60.04
2	69.17
3	67.36
4	77.37

Tabela 4: Erro Absoluto Médio Percentual (MAPE)

4. Conclusão e Trabalhos Futuros

O trabalho cumpre seu objetivo de mostrar como se deve aplicar a metodologia Box & Jenkins com função de transferência utilizando o software R. Conforme observado ao longo do trabalho, essa não é uma tarefa trivial e foi preciso utilizar uma gama de pacotes (eg. **TSA**, **forecast**) para atingir tal objetivo.

Mais dois casos relevantes devem ser destacados: o primeiro é a aplicação do método de BJ com FT a um problema real, que é a previsão de Energia Natural Afluente na região Sul do Brasil, tarefa esta de difícil solução. Segundo, ressalta-se o ineditismo do trabalho ao abordar

⁹Mais informações sobre a utilidade do pacote ver Hyndman and Khandakar (2008).

a metodologia de BJ com FT utilizando o software R e permitindo que novos usuários possam replicar os estudos apresentados e inserir novas covariáveis ao modelo proposto pelos autores.

Trabalhos futuros irão na direção de tentar melhorar a previsão da série temporal ENA do Sul, uma vez que, mesmo utilizando função de transferência no modelo tenha acrescentado melhorias, considera-se o erro de previsão ainda bastante alto. Nesse sentido, a inserção de novas variáveis no modelo e/ou a definição de um novo método é um caminho a ser seguido. Além disso, estimar as séries temporais de SOI e MS, e atualizar o modelo proposto com novas variáveis são tarefas que se fazem necessárias.

Referências

- H. Akaike. Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, 1973.
- G. E. P. Box and G. M. Jenkins. *Time Series Analysis forecasting and control*. Holden Day, San Francisco, 1970.
- D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 1979.
- R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50, 1982.
- P. G. C. Ferreira, F. L. C. Oliveira, and R. C. Souza. The stochastic effects on the brazilian electrical sector. *Energy Economics*, 49, 2015a.
- P. G. C. Ferreira, R. C. Souza, and A. L. M. Marcato. The par(p) interconfigurations model used by the brazilian electric sector. *International Journal of Electrical Power & Energy Systems*, 73, 2015b.
- J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- D. M. Hanssens. Bivariate time-series analysis of the relationship between advertising and sales. *Applied Economics*, 12, 1980.
- R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27, 2008. URL <http://www.jstatsoft.org/v27/i03/>.
- C. M. Jarque and A. K. Bera. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters*, 1980.
- K.S. Kannan and A. J. Farook. Transfer function modeling for global warming. *International Journal of Scientific Research*, 2, 2013.
- G. M. Ljung and G. E. P. Box. On a measure of a lack of fit in time series models. *Biometrika*, 1978.
- A. McDowell. From the help desk: Transfer functions. *The Stata Journal*, 2002.
- F. J. Nogales and A. J. Conejo. Electricity price forecasting through transfer function models. *Journal of the Operational Research Society*, 57, 2006.
- P. C. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 1988.
- G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6, 1978.
- L. A. Terry, M. V. Pereira, T. A. Neto, L. F. Silva, and P. R. Sales. Coordinating the energy generation of the brazilian national hydrotherma electrical generating system. *Interfaces*, 16, 1986.