

4

Metodologia

O método de estimação por mínimos quadrados está fundamentado em algumas premissas, que são necessárias para realizar inferências estatísticas sobre a variável dependente Y . As principais premissas são, [23]:

- (i) Linearidade, isto é, a esperança condicional de Y , $E(Y/X_i)$, deve ser uma função linear nos parâmetros, mas não precisa ser linear nas variáveis explicativas.
- (ii) A esperança condicional dos erros aleatórios ε_i é zero, $E(\varepsilon_i/X_i) = 0$. Isso significa que os fatores não incluídos no modelo e, portanto, agrupados em ε_i , não afetam sistematicamente o valor esperado de Y . Além disso, a ausência de correlação entre ε_i e X_i , indica que X e ε exercem influências separadas e aditivas sobre Y .
- (iii) Homocedasticidade, ou seja, a variância condicional dos erros aleatórios é igual para todas as observações, $Var(\varepsilon_i/X_i) = E(\varepsilon_i^2/X_i) - E^2(\varepsilon_i/X_i) = E(\varepsilon_i^2/X_i) = \sigma^2$.
- (iv) Ausência de correlação serial nos erros, dados dois valores quaisquer de X , X_i e X_j ($i \neq j$), a correlação entre ε_i e ε_j é zero, $Cov(\varepsilon_i, \varepsilon_j/X_{i,j}) = E(\varepsilon_i \cdot \varepsilon_j/X_{i,j}) - E(\varepsilon_i/X_{i,j}) \cdot E(\varepsilon_j/X_{i,j}) = E(\varepsilon_i \cdot \varepsilon_j/X_{i,j}) = 0$. Isso significa que, dado o valor de X , os desvios de quaisquer dois valores de Y em relação à sua média não apresentam padrões sistemáticos de comportamento.
- (v) Os valores de X não devem ser uma constante, portanto a $Var(X)$ deve ser um número positivo finito.
- (vi) Ausência de multicolinearidade perfeita, ou seja, não há dependência linear perfeita entre as variáveis explicativas.

Dadas as premissas, os estimadores de mínimos quadrados apresentam características desejáveis como linearidade, não tendenciosidade, variância mínima e consistência, que podem ser resumidas pela classe de melhor estimador linear não tendencioso.

Para realizar inferências estatísticas acerca dos parâmetros estimados pelo método de mínimos quadrados, através de testes de hipóteses, é necessário relacionar os estimadores (variáveis aleatórias) a uma distribuição de probabilidade. Por isso, inclui-se a premissa de normalidade dos erros, $\varepsilon_i \sim N(0, \sigma^2)$. Dada a premissa de que ε_i segue a distribuição Normal, os estimadores de mínimos quadrados ordinários, que são funções lineares de ε_i , também apresentam distribuição Normal. Esses estimadores, agora, pertencem à classe de melhor estimador não tendencioso, sejam lineares ou não.

Todas as premissas citadas são consideradas hipóteses simplificadoras que facilitam o desdobramento da teoria, porém na prática nem sempre replicam a realidade. Por isso foram desenvolvidos métodos alternativos, como o método de mínimos quadrados ponderados, que contorna a violação da premissa de homocedasticidade, que será o foco da metodologia deste trabalho.

A utilização do método tradicional de mínimos quadrados ordinários sem considerar o efeito heterocedástico nos erros, não altera as propriedades de não tendenciosidade e de consistência do estimador, porém anulam sua eficiência (variância mínima). Desta forma, as inferências e os testes de hipóteses passam a não ter validade.

4.1

Mínimos Quadrados Ponderados

A presença de variâncias desiguais é uma das violações mais comuns, em que a matriz de covariância condicional dos erros \mathbf{C} , não é da forma $I\sigma^2$, e sim uma matriz diagonal com elementos desiguais σ_i^2 . Existem casos em que a matriz apresenta valores não nulos fora da diagonal, indicando erros correlacionados. Para ambos os casos, os estimadores de mínimos quadrados ponderados são eficientes, e pertencem à classe de melhor estimador linear não tendencioso, [14], [39].

A idéia é aplicar transformações apropriadas nas variáveis Y e X , de forma que ao estimar os parâmetros do modelo por mínimos quadrados ordinários $\hat{\beta} = (X^T X)^{-1} X^T Y$, produzam um novo vetor de erros, u , com variância constante $Var(u/X) = I\sigma^2$. Considere o modelo com erro heterocedástico:

$$Y = X\beta + \varepsilon \quad (4.1)$$

Onde:

$$E(\varepsilon/X) = 0, \quad Var(\varepsilon/X) = H(X)\sigma^2 = W^{-1}\sigma^2 \quad e \quad \varepsilon \sim N(0, W^{-1}\sigma^2) \quad (4.2)$$

Sendo $H(X)$ uma função das variáveis explicativas, que determinam a heterocedasticidade. Essa função compõe uma matriz simétrica ($n \times n$) e, para facilitar a interpretação de cálculos adiante, será denominada W^{-1} . Através da fatoração de Cholesky, é possível escrever a matriz W como função de uma matriz triangular superior P , de forma que $W = P^T P$ ou $W^{-1} = P^{-1} P^{-T}$.

Multiplicando ambos os lados da Eq.(4.1) por P , obtêm-se as variáveis transformadas:

$$\begin{aligned} PY &= PX\beta + P\varepsilon \quad ou \\ Z &= Q\beta + u \end{aligned} \quad (4.3)$$

Dessa forma o novo erro é $u = P\varepsilon$ com variância constante:

$$\begin{aligned} Var(u/X) &= E(uu^T/X) \\ &= E(P\varepsilon\varepsilon^T P^T/X) \\ &= P \cdot E(\varepsilon\varepsilon^T/X) \cdot P^T \\ &= P P^{-1} P^{-T} P^T \cdot \sigma^2 \\ &= I\sigma^2 \end{aligned} \quad (4.4)$$

Portanto, o estimador dos parâmetros por mínimos quadrados ponderados é:

$$\begin{aligned} \hat{\beta} &= (X^T P^T P X)^{-1} (X^T P^T P Y) \\ &= (X^T W X)^{-1} (X^T W Y) \end{aligned} \quad (4.5)$$

Os resíduos que devem ser analisados são estimados através da equação:

$$\hat{u} = P(Y - \hat{Y}) \quad (4.6)$$

Onde: $\hat{Y} = X\hat{\beta}$, então:

$$\hat{u} = P(I - X(X^T W X)^{-1} X^T W) Y \quad (4.7)$$

4.1.1

Recursividade

Existem várias formulações para $H(X)$, que variam de acordo com a especificação do modelo e com o comportamento das variáveis. A formulação mais utilizada assume que os erros são não-correlacionados, porém com variâncias desiguais:

$$\text{Var}(\varepsilon/X) = E(\varepsilon\varepsilon^T/X) = \begin{cases} \sigma_i^2, & i = j \\ 0, & i \neq j \end{cases} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \quad (4.8)$$

Define-se W , como a matriz diagonal de pesos w_i , onde $w_i = \frac{1}{\sigma_i^2}$:

$$W = \begin{bmatrix} w_1 & 0 & 0 & \dots & 0 \\ 0 & w_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & w_n \end{bmatrix} \quad (4.9)$$

Na prática é muito difícil obter informações específicas sobre a estrutura da matriz W e as estimativas para cada σ_i^2 . Uma solução para esse problema é utilizar o algoritmo de mínimos quadrados em estágios ou recursivo. Em um primeiro estágio, supõe-se $W=I$. Embora seja uma suposição errônea, o intuito é examinar os resíduos gerados, identificar algum padrão de comportamento e formular uma função $H(X)$. Essa função irá estimar σ_i^2 , por meio das variáveis X consideradas relevantes, e esses valores invertidos irão fornecer \hat{w}_i . Neste contexto, a matriz W pode ser escrita em função da matriz diagonal P que recebe $\hat{w}_i^{1/2}$. Esse procedimento pode ser executado em dois estágios, ou de forma iterativa obedecendo algum critério de parada.

4.1.2

Estimação e Inferência

A matriz de covariância do estimador é dada por:

$$\text{Var}(\hat{\beta}) = \sigma^2 \cdot (X^T W X)^{-1} \quad (4.10)$$

A precisão ou erro padrão do estimador pode ser obtido aplicando a raiz quadrada na Eq.(4.10). Outras medidas da qualidade do ajuste são: SSR (*Regression Sum of Squares*), que mede a variação do valor esperado de Y , $E(Y)$, ou a variação de Y que pode ser explicada pela variação de X através do modelo; SSE (*Error Sum of Squares*), que mede a variação dos erros ou a variação de Y não explicada pelo modelo; SSTO (*Total Sum of Squares*), que mede a variação total de Y , e é a soma de SSR e SSE.

$$\begin{aligned} SSR &= (Q\hat{\beta})^T (Q\hat{\beta}) = (PX\hat{\beta})^T (PX\hat{\beta}) \\ &= \hat{\beta}^T X^T P^T P X \hat{\beta} \\ &= \hat{\beta}^T X^T W X \hat{\beta} \\ &= Y^T W X (X^T W X)^{-1} X^T W Y \end{aligned} \quad (4.11)$$

$$SSE = u^T u = \varepsilon^T P^T P \varepsilon = \varepsilon^T W \varepsilon = (Y - X\beta)^T W (Y - X\beta) \quad (4.12)$$

$$SSTO = Z^T Z = (PY)^T P Y = Y^T W Y \quad (4.13)$$

O fator de correção $n\bar{Z}^2$ representa a variação de Y que pode ser explicada através do modelo mais simples $z_t = \beta_0 + u_t$, que contém apenas uma constante e o erro. Seja Z a transformação da variável Y (Eq.(4.3)), e, n o número de observações. Quando o modelo não contém o intercepto, β_0 , deve-se subtrair esse termo das Eq.(4.11) e Eq.(4.13), obtendo o SSTO corrigido. Embora não muito usual, o fator de correção também pode ser escrito sob a forma matricial $Z^T \mathbb{1}\mathbb{1}^T Z/n$.

A partir dessas medidas, calcula-se um indicador muito utilizado para avaliar o ajuste do modelo. O coeficiente de determinação, R^2 , mede a proporção da variação total de Y explicada pelo modelo, $0 \leq R^2 \leq 1$:

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} = \frac{SSR}{SSTO} \text{ ou} \quad (4.14)$$

$$R^2 = 1 - \frac{SSE}{SSTO} \quad (4.15)$$

Uma propriedade do coeficiente de determinação, R^2 , é ser uma função não decrescente do número de variáveis explicativas do modelo. A inclusão de uma variável X , não reduz o valor de R^2 . Observando a Eq.(4.15), o denominador SSTO não depende do número de variáveis X , no entanto, o numerador SSE tende a diminuir com o aumento de variáveis X , ou pelo menos não aumenta. Em vista disso, para comparar dois modelos desenvolvidos para a mesma variável Y , mas com número diferente de variáveis X , criou-se uma medida ajustada do coeficiente de determinação que considera o número de parâmetros do modelo, k .

$$R_{aj}^2 = 1 - (1 - R^2) \left(\frac{n-1}{n-k} \right) \quad (4.16)$$

O erro padrão do estimador dos parâmetros na Eq.(4.10) depende de σ^2 , que pode ser estimado pelo MSE (*Error Mean Square*):

$$\hat{\sigma}^2 = \frac{u^T u}{n-k} = \frac{SSE}{n-k} = MSE \quad (4.17)$$

Dada a premissa de normalidade dos erros, a significância dos coeficientes pode ser testada através do teste de hipótese, formulado com a hipótese nula $H_0: \beta_j = 0$ e hipótese alternativa $H_1: \beta_j \neq 0, j = 1, 2, \dots, p$. A estatística de teste é:

$$T = \frac{\hat{\beta}_j}{\sqrt{MSE \cdot c_{jj}}} \sim t_{n-k, 1-\alpha/2} \quad (4.18)$$

Onde c_{jj} é o elemento, linha j e coluna j , da matriz $(X^T W X)^{-1}$, e $t_{n-k, 1-\alpha/2}$ é o quantil da distribuição t-Student tabelado, tal que $P(T < t) = \alpha/2$, também conhecido como valor crítico. Portanto, rejeita-se H_0 se $|T| > t_{n-k, 1-\alpha/2}$. Com isso, o intervalo de confiança de nível $(1-\alpha)100\%$ para β_j é:

$$\hat{\beta}_j \pm t_{n-k, 1-\alpha/2} \sqrt{\text{MSE} \cdot c_{jj}} \quad (4.19)$$

Para testar a significância simultânea dos coeficientes angulares, usa-se o teste de hipótese formulado com a hipótese nula $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ e hipótese alternativa $H_1: \text{pelo menos um } \beta_j \neq 0, j = 1, \dots, k$. Esse teste está relacionado à especificação das variáveis explicativas no modelo, e avalia se as variáveis X , em conjunto, têm efeito sobre a variável Y . A estatística de teste é:

$$F = \frac{SSR/gl}{SSE/gl} = \frac{\frac{SSR}{k-1}}{\frac{SSE}{n-k}} = \frac{MSR}{MSE} \sim f_{k-1, n-k, 1-\alpha} \quad (4.20)$$

Onde, MSR (*Regression Mean Square*) é a razão entre SSR e seus graus de liberdade (gl), e $f_{k-1, n-k, 1-\alpha}$ é o valor tabelado, ou valor crítico, da distribuição F-Snedecor. Portanto, rejeita-se H_0 se $F > f_{k-1, n-k, 1-\alpha}$.

Lembrando que o valor esperado de uma observação pontual de Y , $E(Y_j)$, pode ser definido como:

$$E(Y_j) = \hat{y}_j = x_j^T \hat{\beta}, \quad x_j^T = \begin{bmatrix} 1 \\ x_{j1} \\ \vdots \\ x_{jk} \end{bmatrix} \quad (4.21)$$

Assim, o intervalo de confiança de nível $(1-\alpha)100\%$ para $E(Y_j)$ é:

$$\hat{y}_j \pm t_{n-k, 1-\alpha/2} \sqrt{\text{MSE} \cdot q_j^T (X^T W X)^{-1} q_j} \quad (4.22)$$

Onde, q_j é a linha j da matriz Q de variáveis explicativas transformadas. Na Eq.(4.3), $Q=PX$.

E o intervalo de confiança de nível $(1-\alpha)100\%$ para valores previstos de Y , a partir de novas observações de X , é:

$$\hat{y}_j \pm t_{n-k-1, 1-\alpha/2} \sqrt{MSE \cdot [1 + q_j^T (X^T W X)^{-1} q_j]} \quad (4.23)$$

4.2

Testes de Especificação

O desenvolvimento de um modelo a partir de dados do tipo série temporal apresenta algumas particularidades, e outras premissas são necessárias à estimação. Algumas propriedades que devem ser observadas são: estacionariedade das séries, estrutura de autocorrelação, efeitos espúrios, entre outras. A seguir, serão apresentados os fundamentos dos principais testes estatísticos utilizados no processo de modelagem.

4.2.1

Estacionariedade

Uma série temporal é considerada estacionária no sentido estrito, quando todos os momentos de sua distribuição de probabilidade não se alteram ao longo do tempo. Na maior parte das situações práticas, é suficiente que a série seja fracamente estacionária, isto é, quando os dois primeiros momentos (média e variância) são constantes ao longo do tempo. Contudo, se a série for fracamente estacionária e com distribuição Normal, também é estritamente estacionária, pois todos os momentos da distribuição Normal podem ser definidos por sua média e variância.

A estacionariedade da série é uma premissa fundamental, pois garante que os resultados do estudo do seu comportamento possam ser generalizados para outros períodos de tempo. As séries que não apresentam essa característica devem passar por transformações em seus dados originais. A transformação mais comum consiste em tomar diferenças sucessivas da série original, até obter uma série estacionária. A primeira diferença de y_t é definida por:

$$\Delta y_t = y_t - y_{t-1} \quad (4.24)$$

A segunda diferença é:

$$\begin{aligned}\Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) \\ \Delta^2 y_t &= \Delta y_t - \Delta y_{t-1} \\ \Delta^2 y_t &= y_t - 2y_{t-1} + y_{t-2}\end{aligned}\tag{4.25}$$

De modo geral, a n -ésima diferença de y_t é $\Delta^n y_t = \Delta(\Delta^{n-1} y_t)$.

4.2.1.1

Teste de Dickey-Fuller Aumentado

Se uma série deve ser diferenciada d vezes antes de tornar-se estacionária, então ela contém d raízes unitárias. Os testes de raízes unitárias são capazes de detectar se a série foi suficientemente diferenciada para se tornar estacionária. Considere o modelo:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad -1 \leq \rho \leq 1\tag{4.26}$$

Onde, ε_t é o termo que representa o erro e segue um processo de ruído branco. Se o modelo possui raiz unitária ($\rho = 1$), então pode ser classificado como passeio aleatório sem deslocamento, que é um processo estocástico não-estacionário. A idéia central dos testes de raiz unitária é verificar se ρ estimado é estatisticamente igual a um, através da regressão de Y_t contra o seu valor defasado. O modelo pode ser reescrito conforme Eq.(4.27):

$$\begin{aligned}y_t - y_{t-1} &= \rho y_{t-1} - y_{t-1} + \varepsilon_t \\ \Delta y_t &= (\rho - 1)y_{t-1} + \varepsilon_t \\ \Delta y_t &= \delta y_{t-1} + \varepsilon_t\end{aligned}\tag{4.27}$$

Na prática, o teste verifica se $H_0: \delta = 0$ com a mesma interpretação. Porém, sob a hipótese nula, a estatística de teste da significância do coeficiente não segue a distribuição t usual. Dickey e Fuller demonstraram que o valor estimado do coeficiente segue a estatística τ (*tau*), cujos valores críticos são tabelados. O teste de Dickey-Fuller ainda possui outras vertentes de acordo com a natureza do

processo, com valores críticos distintos para cada um dos casos. Para o caso de um passeio aleatório com deslocamento, tem-se:

$$\Delta y_t = \beta_1 + \delta y_{t-1} + \varepsilon_t \quad (4.28)$$

E para um passeio aleatório com deslocamento e tendência estocástica (t):

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \varepsilon_t \quad (4.29)$$

O caso mais simples do teste é utilizar um modelo AR(1) e estimar o coeficiente δ por mínimos quadrados. Porém, o pressuposto de que ε_t é ruído branco e, portanto não-correlacionado, nem sempre é válido. Para esses casos utiliza-se o teste Dickey-Fuller Aumentado (ADF), que tem a mesma distribuição assintótica de Dickey-Fuller. O teste ADF consiste em estimar o seguinte modelo:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-i} + \varepsilon_t \quad (4.30)$$

Onde, ε_t é um termo de erro de ruído branco puro, e os termos de diferenças defasadas $\Delta y_{t-1} = (y_{t-1} - y_{t-2}), \Delta y_{t-2} = (y_{t-2} - y_{t-3}), etc$, devem ser incluídos m vezes para que o erro não apresente correlação serial. O valor de m é determinado empiricamente, através da significância de $\hat{\alpha}_i, i=1, \dots, m$, ou através dos critérios de AIC e BIC. A estatística de teste τ é a razão do coeficiente estimado $\hat{\delta}$ e seu desvio-padrão. Rejeita-se H_0 , se $|\tau| > C$, e neste caso conclui-se que a série é estacionária.

4.2.2

Normalidade

A premissa de normalidade dos erros confere maior eficiência aos estimadores e possibilita inferências estatísticas, através da aplicação de testes de hipótese. Porém a justificativa teórica para inclusão desta premissa está relacionada com a interpretação do termo aleatório ε_t e o Teorema Central do Limite (TCL). Uma vez que, ε_t representa a influência combinada de um grande número de variáveis

não incluídas explicitamente no modelo; o TCL demonstra que a distribuição da soma de um grande número de variáveis aleatórias independentes com mesma distribuição, tende a uma Normal na medida em que o número de variáveis aumenta. Uma variante do teorema garante boa aproximação Normal, para um número menor de variáveis ou para variáveis não estritamente independentes. Na prática, o pressuposto de normalidade pode ser verificado através de testes estatísticos aplicados nos resíduos gerados pelo modelo.

4.2.2.1

Teste de Jarque-Bera

O teste de Jarque-Bera é um teste paramétrico e assintótico, e tem o objetivo de verificar se a distribuição de probabilidade de uma variável segue uma distribuição Normal, com base na hipótese nula H_0 : *distribuição Normal* e hipótese alternativa H_1 : *distribuição não é Normal*. A estatística de teste é:

$$JB = n \left[\frac{\hat{S}^2}{\sigma} + \frac{(\hat{K} - 3)^2}{24} \right] \quad (4.31)$$

Onde,

n = tamanho da amostra;

\hat{S} = coeficiente de assimetria amostral;

\hat{K} = coeficiente de curtose amostral.

O teste compara as medidas de forma da distribuição empírica da variável com as medidas características da curva Normal, que apresenta assimetria igual a zero e curtose igual a 3. Dessa forma, o termo $(\hat{K}-3)$ representa o excesso de curtose. A estatística JB , assintoticamente, segue a distribuição χ^2 com dois graus de liberdade, portanto os valores críticos do teste são $[\chi^2_{2,1-\alpha/2}, \chi^2_{2,\alpha/2}]$, e rejeita-se H_0 se JB estiver fora desse intervalo.

4.2.2.2

Teste de Kolmogorov-Smirnov

O teste de normalidade de Kolmogorov- Smirnov é um teste não-paramétrico, que compara a máxima diferença absoluta entre a função de distribuição acumulada da Normal e a função de distribuição acumulada dos dados, com base na hipótese nula H_0 : *distribuição Normal* e hipótese alternativa H_1 : *distribuição não é Normal*. A estatística de teste é:

$$KS = \max (|F(x) - G(x)|) \quad (4.32)$$

Onde,

$F(x)$ = função de distribuição acumulada empírica $P(X \leq x)$;

$G(x)$ = função de distribuição acumulada da Normal $P(Z \leq z)$.

Esta função representa a distância vertical máxima entre os gráficos de $F(x)$ e $G(x)$ sobre a amplitude dos possíveis valores de x . O valor de $F(x)$ é calculado através da frequência relativa acumulada observada e $G(x)$ pode ser encontrado na tabela da distribuição normal padronizada. A hipótese nula é rejeitada com $(1-\alpha)100\%$ de confiança, se KS for maior que o valor crítico tabelado. Para $\alpha=0,05$ e $n > 40$, o valor crítico é dado por $1,36/\sqrt{n}$.

O teste de normalidade de Kolmogorov- Smirnov possui algumas desvantagens, devido ao seu baixo poder¹¹.

4.2.3

Autocorrelação

A premissa de ausência de correlação serial nos erros pressupõe que o erro relacionado a qualquer das observações não é influenciado pelo erro de qualquer outra observação. Na prática, os testes estatísticos são aplicados nos resíduos

¹¹O poder do teste é a probabilidade de rejeitar a hipótese nula quando esta é de fato. Na prática, é importante que se tenham testes com nível de significância próximos do nível de significância nominal e que o poder seja alto, mesmo em situações de amostras pequenas.

gerados pelo modelo, com o objetivo de verificar se têm comportamento puramente aleatório. A presença de correlação serial nos resíduos pode indicar problemas de especificação, pois alguma característica da série não foi captada pelo modelo, sugerindo a inclusão de mais defasagens da variável dependente, ou defasagens adicionais das variáveis exógenas, ou novas variáveis causais.

4.2.3.1

Teste *d* de Durbin-Watson

O teste de Durbin-Watson verifica a correlação serial de primeira ordem. A facilidade de seu cálculo faz com que seja empregado rotineiramente, porém esse teste possui algumas restrições quanto a seu uso. Dentre elas, o modelo não deve conter valores defasados da variável dependente como variável explicativa.

A sua metodologia consiste em definir as variáveis explicativas e calcular a regressão por mínimos quadrados ordinários, com base na hipótese nula $H_0: \rho_1 = 0$ e hipótese alternativa $H_1: \rho_1 \neq 0$. A estatística de teste é:

$$d = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2} \cong 2(1 - \hat{\rho}_1) \quad (4.33)$$

Onde:

\hat{u}_t = resíduo calculado pela regressão;

$\hat{\rho}_1$ = coeficiente de autocorrelação de primeira ordem amostral.

Como $-1 \leq \hat{\rho}_1 \leq 1$, implica em $0 \leq d \leq 4$. Se $d = 2$, $\hat{\rho}_1 \cong 0$, e não há evidência de autocorrelação. Portanto, a autocorrelação entre os resíduos é indicada por valores significativamente diferentes de 2. Os valores críticos para esse teste são tabelados, d_L e d_U , com base no tamanho da amostra e no número de variáveis, e as regras decisão são:

Tabela 4-1– Teste de Durbin Watson: regras de decisão.

d	$[0, d_L[$	$[d_L, d_U[$	$[d_U, 4-d_U[$	$[4-d_U, 4-d_L[$	$[4-d_L, 4[$
Decisão	Rejeitar H_0	Sem decisão	Não rejeitar H_0	Sem decisão	Rejeitar H_0

4.2.3.2

Teste h de Durbin-Watson

Uma modificação do teste d de Durbin-Watson foi proposta com o objetivo de contornar uma de suas limitações e detectar a autocorrelação em modelos autoregressivos. O teste verifica a autocorrelação de primeira ordem para grandes amostras em modelos com valores defasados da variável dependente como variáveis explicativas. As hipóteses do teste são: $H_0: \rho_1 = 0$ e $H_1: \rho_1 \neq 0$. O resultado fundamental assintótico de Durbin é:

$$h = \hat{\rho}_1 \sqrt{\frac{n}{1 - n(\text{Var}(b_1))}} \sim N(0,1) \quad (4.34)$$

Onde:

n = tamanho da amostra;

$\text{Var}(b_1)$ = variância do coeficiente de y_{t-1} pelo ajuste de mínimos quadrados ordinários. O teste não é aplicável se $n(\text{Var}(b_1)) > 1$.

Na prática, o coeficiente de autocorrelação de primeira ordem $\hat{\rho}_1$ pode ser estimado por $(1 - d/2)$, onde d foi calculado na Eq.(4.33). Assintoticamente $h \sim N(0,1)$, e com um nível de 95% de confiança, tem-se as regras de decisão:

- (i) se $h > 1,96$, rejeita-se a hipótese nula – não há autocorrelação positiva de primeira ordem;
- (ii) se $h < -1,96$, rejeita-se a hipótese nula – não há autocorrelação negativa de primeira ordem;
- (iii) se h está no intervalo $(-1,96; 1,96)$, não rejeita-se a hipótese nula – não há autocorrelação de primeira ordem positiva ou negativa.

A partir do estudo de Durbin, Breusch e Godfrey desenvolveram o teste LM (*Lagrange Multiplier*) que consegue evitar algumas restrições, e é considerado

estatisticamente mais poderoso que o teste h de Durbin-Watson, sendo, portanto preferível ao teste h .

4.2.3.3

Teste de Box-Pierce-Ljung

Box e Pierce (1970) sugeriram um teste para diagnosticar autocorrelações que apresentassem um valor excessivamente elevado, em diferentes defasagens. Uma modificação deste teste foi proposta por Ljung e Box (1978) e, em vez de testar a autocorrelação em cada defasagem, testa a autocorrelação global de um bloco pré-definido. O teste de Box-Pierce-Ljung, também conhecido por Ljung-Box, permite detectar quebras específicas no comportamento aleatório. As hipóteses do teste são: $H_0: \rho_1, \rho_2, \dots, \rho_L = 0$ e $H_1: \text{pelo menos um } \rho_j \neq 0, j = 1, \dots, L$, e a estatística de teste é dada por:

$$Q(L) = n_{ef}(n_{ef} + 2) \sum_{j=1}^L \frac{\hat{\rho}_j^2}{(n_{ef} - j)} \quad (4.35)$$

Onde:

L = número de defasagens das funções de autocorrelação;

n = tamanho da amostra;

n_{ef} = número efetivo de observações = $n - d$;

d = ordem de diferença da série.

A distribuição limite de Q foi derivada a partir do pressuposto que os erros seguem um processo autoregressivo (AR). Assim, a estatística $Q(L)$ segue a distribuição χ^2 com $(L - k)$ graus de liberdade, onde k é o número de parâmetros estimados no modelo AR para os erros. Rejeita-se H_0 para valores elevados de $Q(\cdot)$. Em geral, é suficiente utilizar $L=20$.

Segundo [13], o efeito na distribuição da estatística de teste não é conhecido quando aplicado em erros que não sigam processos puramente autoregressivos (AR), ou que possuam variáveis exógenas.

4.2.3.4

Teste de Breusch-Godfrey

O teste de Breusch-Godfrey também é conhecido como teste LM (*Lagrange Multiplier*) para autocorrelação. Após definir o modelo e estimar seus coeficientes, o teste consiste em efetuar uma regressão do resíduo contra o próprio resíduo defasado no tempo e as variáveis explicativas. As hipóteses do teste são: $H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0$ e $H_1: \text{pelo menos um } \rho_j \neq 0, j = 1, \dots, p$. Suponha que o modelo especificado seja:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad (4.36)$$

Suponha que o termo de erro siga um processo AR(p).

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t \quad (4.37)$$

Após estimar a equação acima, e definir a ordem de defasagem p , obtém-se a estatística de teste, que está baseada na regressão:

$$\hat{u}_t = \alpha_0 + \alpha_1 x_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \dots + \hat{\rho}_p \hat{u}_{t-p} + v_t \quad (4.38)$$

Onde os últimos termos $\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-p}$ são os resíduos estimados pela Eq.(4.37). Por isso, o tamanho efetivo da amostra usada para estimar a Eq.(4.38) é $(n - p)$. Assintoticamente tem-se:

$$(n - p) \cdot R^2 = (n - p) \cdot \frac{\sum_{t=1}^{n-p} (\hat{u}_t - \bar{u})^2}{\sum_{t=1}^{n-p} (u_t - \bar{u})^2} \sim \chi_p^2 \quad (4.39)$$

Esse teste possui algumas vantagens, pois permite a inclusão de valores defasados de Y como variáveis explicativas; pode ser aplicado mesmo que os erros não sigam processos puramente autoregressivos, através da modificação da Eq.(4.37); permite a inclusão de variáveis exógenas e considera autocorrelações de ordens simples ou mais elevadas.

4.2.4

Heterocedasticidade

A natureza do problema estudado pode sugerir um padrão de comportamento na variância do erro relacionado a alguma variável explicativa, porém existem alguns métodos formais de análise dos resíduos para detecção da condição de heterocedasticidade.

4.2.4.1

Teste de White

O teste de White é o mais utilizado, pois não depende da hipótese de normalidade e é de fácil aplicação. Após definir o modelo e estimar seus coeficientes, o teste consiste em efetuar uma regressão dos quadrados dos resíduos contra: as variáveis explicativas X , seus valores ao quadrado e seus produtos cruzados. As hipóteses do teste são: $H_0: Var(\varepsilon_t) = \sigma^2$ e $H_1: Var(\varepsilon_t) = \sigma_t^2$. A estatística de teste está baseada na própria regressão formulada:

$$\hat{u}_t^2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{1j} x_1^2 + \beta_{2j} x_2^2 + \dots + \beta_{ij} x_1 x_2 + \dots + v_t \quad (4.40)$$

Assintoticamente tem-se:

$$n \cdot R^2 = n \cdot \frac{\sum_{t=1}^n (\hat{u}_t - \bar{u})^2}{\sum_{t=1}^n (u_t - \bar{u})^2} \sim \chi_{k-1}^2 \quad (4.41)$$

Onde,

n = tamanho da amostra;

k = número de parâmetros.

O teste de White possui uma limitação quanto a um número elevado de variáveis explicativas no modelo, pois a inclusão de todos os termos e combinações, formulados para o teste, pode consumir rapidamente os graus de liberdade. Uma opção para testar a heterocedasticidade nesses casos é aplicar o teste Koenker-Bassett.

4.2.4.2

Teste de Koenker-Bassett

O teste de Koenker-Bassett, assim como o teste de White, está embasado no quadrado dos resíduos, porém, em vez de realizar uma regressão contra as variáveis explicativas e suas combinações, o teste realiza uma regressão dos quadrados dos resíduos contra os quadrados dos valores estimados de Y . A regressão formulada para o teste é:

$$\hat{u}_t^2 = \beta_0 + \beta_1(\hat{Y}_t)^2 + v_t \quad (4.42)$$

A hipótese nula afirma que os erros são homocedásticos, através do coeficiente β_1 . Logo, $H_0: \beta_1 = 0$ e $H_1: \beta_1 \neq 0$. Os valores críticos do teste podem ser obtidos por meio do habitual teste t para significância de coeficientes.

4.2.5

Outliers

Outliers são considerados observações extremas ou discrepantes, que podem ser provenientes de algum erro de medição ou de algum efeito adverso. A presença de *outliers* na estimação e inferência de parâmetros, em qualquer análise de dados, pode comprometer os resultados e levar a conclusões falsas ou a uma estimação enganosa. Em geral, para modelagem de dados em *cross-section*¹², recomenda-se descartar essas observações, a não ser que, exista uma evidência direta que estas observações extremas representem uma circunstância que deva ser considerada. Porém, quando se analisam dados tipo série temporal, o mesmo procedimento não pode ser aplicado. Além das razões óbvias de discretização da série no tempo, a observação de uma variável no tempo t está correlacionada com outras observações da série. Os *outliers* em séries provocam efeitos como mudança, (abrupta ou suave) no nível da série, e até alterações de sua tendência.

¹²Uma ou mais variáveis coletadas no mesmo ponto do tempo t .

Os *outliers* podem ser classificados de acordo com os diferentes tipos de efeito que provocam no processo gerador da série. Em geral, os dois efeitos mais considerados são o aditivo e o de inovação, [42]. Considere uma série temporal estacionária $\{Z_t\}$, sem *outliers*, gerada por um modelo ARMA¹³(p, q):

$$\begin{aligned} \phi(B)z_t &= \theta(B)e_t \\ \phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q, \quad B^k z_t = z_{t-k} \end{aligned} \quad (4.43)$$

Onde $e_t \sim iid(0, \sigma^2)$. Logo, $z_t = \frac{\theta(B)}{\phi(B)} e_t$ ou $z_t = \frac{1}{\pi(B)} e_t$.

O *outlier* aditivo (AO) pode ser considerado uma grande discrepância, que incide apenas na t -ésima observação. Num instante $t=T$, o valor observado y_t possui um termo aditivo, δ_{AO} , que representa a magnitude do *outlier*, este termo é igual a zero para $t \neq T$. AO pode ser definido como:

$$y_t = \frac{\theta(B)}{\phi(B)} e_t + \delta_{AO} d_T, \quad d_T = \begin{cases} 1, & t = T \\ 0, & \text{cc} \end{cases} \quad (4.44)$$

O *outlier* de inovação (IO) pode ser considerado um choque extraordinário na série de inovações $\{e_t\}$ no instante $t=T$. Em outras palavras, IO é o resultado da aleatoriedade natural do processo. Este tipo de *outlier* incide nas observações subsequentes, através da memória do modelo dada pelas defasagens que compõe o termo $\theta(B)/\phi(B)$. IO pode ser definido como:

$$y_t = \frac{\theta(B)}{\phi(B)} (e_t + \delta_{IO} d_T) \quad (4.45)$$

Onde δ_{IO} representa a magnitude do *outlier* IO.

A ocorrência de AO, na maioria das vezes, indica um erro no registro, e torna-se necessária uma ação no sentido de ajustar o instrumento de medição. No entanto, se IO ocorre, nenhum ajuste da operação de medição é necessário.

¹³Os resultados obtidos nesta seção podem ser generalizados para o processo ARIMA(p, d, q), modificando o valor de $\pi(B)$, segundo a expressão: $\phi(B)\Delta^d y_t = \theta(B)e_t$.

4.2.5.1

Análise de Intervenção

O AO pode ser corrigido com intervenção do tipo *pulse*, que considera uma variável *dummy* com valor igual a um no momento da ocorrência do *outlier* e zero caso contrário, $t \neq T$. O modelo proposto para AO é:

$$Y_t = Z_t + \delta_{AO} d_T, \quad d_T = \begin{cases} 1, & t = T \\ 0, & \text{cc} \end{cases} \quad (4.46)$$

O IO pode ser corrigido com intervenção do tipo *step* na série de inovações, com valor igual a zero antes do momento da ocorrência do *outlier*, e igual a um posteriormente a ele, $t \geq T$. O modelo proposto para IO é:

$$Y_t = Z_t + \delta_{IO} \frac{1}{\pi(B)} d_T, \quad d_T = \begin{cases} 0, & t < T \\ 1, & t \geq T \end{cases} \quad (4.47)$$

Embora o AO afete a série das observações apenas no período $t=T$, o impacto na série de resíduos ultrapassa um período. Em um processo $AR(p)$, por exemplo, o IO irá afetar apenas o resíduo da observação em $t=T$. Por outro lado, o AO detectado na observação $y_{t=T}$, afetará os p resíduos subsequentes. Este efeito pode enviesar a estimação dos coeficientes e sua precisão, além disso, tem consequências para posteriores análises dos resíduos. Em geral, o IO tem efeito menos prejudicial. Portanto, quando a intervenção for realizada diretamente nos resíduos, as técnicas se invertem: os resíduos terão o efeito AO eliminado através da intervenção *step*, ou o efeito IO eliminado através da intervenção *pulse*.

4.2.5.2

Detecção de Outliers

Além da análise gráfica da série, a ocorrência de *outliers* na série pode ser avaliada através de um teste baseado na estatística de razão de verossimilhança. Este teste, não só avalia a presença de *outlier*, como determina o tipo de *outlier*, o período de ocorrência e a estimativa do impacto ou magnitude do *outlier* identificado.

As Eq.(4.46) e Eq.(4.47) podem ser generalizadas:

$$Y_t = Z_t + \delta_i \cdot v_i(B) \cdot d_T \quad (4.48)$$

Onde,

$i = AO, IO$;

δ_i = impacto inicial do outlier;

$v_i(B)$ = estrutura dinâmica do efeito do outlier = $\begin{cases} 1, & i - AO \\ 1/\pi(B), & i - IO \end{cases}$

Considere a expressão geral dos resíduos da série com *outliers* $\{Y_t\}$:

$$\begin{aligned} \phi(B)y_t &= \theta(B)a_t \\ a_t &= \pi(B) \cdot y_t \end{aligned} \quad (4.49)$$

Substituindo a Eq.(4.48) na Eq.(4.49), a expressão dos resíduos pode ser avaliada em cada *outlier*:

$$a_t = \delta_i[\pi(B) \cdot v_i(B) \cdot d_T] + e_t \quad (4.50)$$

Definindo $x_t = \pi(B) \cdot v_i(B) \cdot d_T$, a expressão anterior fica:

$$a_t = \delta_i x_t + e_t \quad (4.51)$$

Através da equação linear (Eq.(4.51)), o estimador por mínimos quadrados para o impacto do *outlier* é:

$$\hat{\delta}_i = \frac{\sum_{t=1}^n a_t x_t}{\sum_{t=1}^n x_t^2} \quad (4.52)$$

E a variância do estimador é:

$$Var(\hat{\delta}_i) = \frac{\sigma_e^2}{\sum_{t=1}^n x_t^2} \quad (4.53)$$

Substituindo x_t nas Eq.(4.52) e Eq.(4.53), têm-se os estimadores para o impacto de cada *outlier*:

$$\hat{\delta}_{AO} = \frac{a_T - \sum_{j=1}^{n-T} \pi_j a_{T+j}}{\sum_{j=0}^{n-T} \pi_j^2} = \varphi_{AO}^2 \pi(F) a_T, \quad \text{Var}(\hat{\delta}_{AO}) = \varphi_{AO}^2 \sigma_e^2 \quad (4.54)$$

$$\hat{\delta}_{IO} = a_T, \quad \text{Var}(\hat{\delta}_{IO}) = \sigma_e^2 \quad (4.55)$$

Onde:

$$\varphi_{AO}^2 = (1 + \pi_1^2 + \pi_2^2 + \dots + \pi_{n-T}^2)^{-1};$$

$$\pi(F) = (1 - \pi_1 F - \pi_2 F^2 - \dots - \pi_{n-T} F^{n-T});$$

$$F^k a_T = a_{T+k}.$$

O efeito de um IO no momento T pode ser estimado pelo resíduo desse período a_T , como era esperado, porém o melhor estimador do efeito AO é uma combinação linear de a_T, a_{T+1}, \dots , ponderados de acordo com π do modelo ARMA(p, q).

Finalmente, as estatísticas de teste são dadas pelas razões:

$$\lambda_{AO,T} = \frac{\hat{\delta}_{AO}}{\varphi_{AO} \sigma_e} \quad e \quad \lambda_{IO,T} = \frac{\hat{\delta}_{IO}}{\sigma_e} \quad (4.56)$$

As hipóteses do teste são: $H_0: \delta_{AO} = \delta_{IO} = 0$, ou seja, não existem outliers na série, contra $H_A: \delta_{AO} \neq 0$ e $H_I: \delta_{IO} \neq 0$. Sob H_0 , ambas as estatísticas $\lambda_{AO,T}$ e $\lambda_{IO,T}$ seguem distribuição assintótica $N(0, 1)$.

O teste consiste em um procedimento iterativo. Primeiro, estima-se a série $\{Y_t\}$ pressupondo que não existam *outliers*, e sob esta condição, $a_t = e_t$. Os resíduos e a variância residual serão utilizados como estimativas dos parâmetros das estatísticas de teste. Em seguida, as estatísticas de teste são calculadas para todos os períodos de tempo e obtém-se o máximo absoluto:

$$\begin{aligned} \hat{\lambda}_{AO,max} &= \max \{ |\hat{\lambda}_{AO,t}| \mid 1 \leq t \leq n \} \\ \hat{\lambda}_{IO,max} &= \max \{ |\hat{\lambda}_{IO,t}| \mid 1 \leq t \leq n \} \end{aligned} \quad (4.57)$$

Determina-se $\hat{\lambda} = \max \{\hat{\lambda}_{AO,max}, \hat{\lambda}_{IO,max}\}$ e compara-se com o valor crítico C . Se $\hat{\lambda} < C$, não existem outliers. Porém, se $\hat{\lambda} \geq C$, é possível determinar o tipo, o período de ocorrência e o impacto estimado do *outlier*.

Os valores críticos, obtidos por simulação, foram selecionados por vários autores que sugerem $C = 3$ para alta sensibilidade, $C = 3,5$ para média sensibilidade e $C = 4$ para baixa sensibilidade.

Após identificar o *outlier*, aplica-se a intervenção na série de resíduos de acordo com o tipo encontrado, AO ou IO. Define-se o novo resíduo \hat{e}_T :

$$\begin{aligned} \hat{e}_t &= \hat{a}_t - \pi(B)\hat{\delta}_{AO}d_t, & t \geq T \\ \hat{e}_T &= \hat{a}_T - \hat{\delta}_{IO}, & t = T \end{aligned} \quad (4.58)$$

A partir da nova série de resíduos, livre do efeito do *outlier* em $t=T$, calcula-se a nova variância residual, que será a estimativa de σ_e^2 . As etapas anteriores são repetidas até que nenhum *outlier* seja identificado. As estatísticas de testes são recalculadas, sempre com base na estimativa inicial dos parâmetros, mas alterando os resíduos \hat{e}_t e $\hat{\sigma}_e^2$.

4.2.6

Critérios para Seleção de Modelos

O desenvolvimento do modelo envolve a decisão de incluir ou excluir variáveis, de acordo com seu poder explicativo e considerando o princípio da parcimônia. Alguns critérios de comparação entre modelos auxiliam essa decisão, os principais são AIC (*An Information Criterion*) ou Critério de Akaike, e BIC (*Bayesian Information Criterion*) ou Critério de Schwarz.

4.2.6.1

Critério de Akaike

Este critério consiste em aplicar uma penalidade ao acréscimo de novas variáveis ao modelo, e tem o objetivo de comparar modelos com diferentes estruturas.

Também tem sido empregado para determinar a extensão da defasagem em um modelo autoregressivo. AIC é formalmente definido como a função log-verossimilhança negativa avaliada nos parâmetros estimados, e penalizada pelo número de parâmetros estimados. O menor valor de AIC indica o melhor modelo. AIC é definido por:

$$AIC = -2 \ln(\hat{V}_k) + 2k \quad (4.59)$$

Onde:

k = número de parâmetros efetivamente estimados;

\hat{V}_k = função de máxima verossimilhança do modelo avaliada no máximo.

4.2.6.2

Critério de Schwarz

Este critério está intimamente relacionado com o critério de Akaike, e consiste em aplicar uma penalidade ainda maior ao acréscimo de novas variáveis ao modelo. O menor valor de BIC indica o melhor modelo. BIC é definido por:

$$BIC = -2 \ln(V_k) + k \cdot \ln(n) \quad (4.60)$$

Onde,

n = tamanho da série;

k = número de parâmetros.

4.2.7

Medidas de Aderência do Modelo

Após definir o modelo e realizar o diagnóstico dos resíduos, a próxima etapa consiste em avaliar o desempenho do modelo. Em geral, reserva-se um conjunto de observações mais recentes da série para testar a capacidade preditiva do modelo. Esse conjunto de dados não é utilizado na fase do ajuste dos modelos, e é denominado período *out-of-sample* ou período de validação. Então, o modelo final é aplicado no período de validação de acordo com seu propósito e objetivo, que pode ser previsão, simulação, entre outros. Espera-se que os resultados obtidos

estejam próximos aos valores reais observados. Assim, algumas medidas de aderência são utilizadas como ferramenta para mensurar a proximidade entre o valor, previsto ou simulado (\hat{y}_t), e o valor real (y_t). As principais medidas são: RMSE (*Root Mean Square Error*), MAE (*Mean Absolute Error*), MAPE (*Mean Absolute Percent Error*), MPE (*Mean Percent Error*), SDPE (*Standard Deviation Percentage Error*), definidas como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_t - y_t)^2} \quad (4.61)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_t - y_t| \quad (4.62)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right|, y_t \neq 0 \quad (4.63)$$

$$MPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_t - y_t}{y_t} \right), y_t \neq 0 \quad (4.64)$$

$$SDPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_t - y_t}{y_t} - MPE \right)^2}, y_t \neq 0 \quad (4.65)$$

4.3

Estimação Robusta para Outliers

A partir da detecção de *outliers* (desenvolvido na Seção 4.2.5.2), o próximo passo é eliminar o efeito dos *outliers* da série contaminada.

A estimação dos parâmetros por mínimos quadrados envolve equações para cada \hat{y}_t ajustado. No contexto de série temporal, uma observação suspeita, y_T , será incluída nas $p+1$ equações consecutivas. Por isso, desenvolveu-se um procedimento iterativo que elimina as equações contaminadas pelo *outlier* e, obtém uma estimativa robusta de y_T .

Se a série $\{Y_t\}$ segue um processo $AR(p)$, este pode ser representado por:

$$Y = X\phi + a_t$$

$$\hat{Y} = X\hat{\phi} \quad (4.66)$$

Onde X é a matriz formada pelo vetor Y defasado p vezes, resultando em $(n-p)$ equações:

$$\begin{bmatrix} \hat{y}_{p+1} \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} y_p & \cdots & y_1 \\ y_{p-1} & & y_2 \\ \vdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{bmatrix}$$

$$(n-p) \times 1 \quad (n-p) \times p \quad p \times 1$$

O vetor dos valores ajustados \hat{Y} pode ser reescrito como:

$$\hat{\phi} = (X^T X)^{-1} (X^T Y)$$

$$\hat{Y} = X\hat{\phi} = X[(X^T X)^{-1} X^T] Y = HY \quad (4.67)$$

Onde H é a matriz de projeção. Logo, o vetor de resíduos, R , pode ser escrito como:

$$R = (I - H)Y \quad (4.68)$$

Sabendo que existe um *outlier* em $t=T$. A matriz de variáveis explicativas X , e os vetores Y e R , podem ser decompostos em três partes:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \begin{matrix} (T-p) \times p \\ k \times p \\ (n-T-k) \times p \end{matrix}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \begin{matrix} (T-p) \times 1 \\ k \times 1 \\ (n-T-k) \times 1 \end{matrix}$$

$$R = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} \begin{matrix} (T-p) \times 1 \\ k \times 1 \\ (n-T-k) \times 1 \end{matrix}$$

Onde, k é o número de equações que serão eliminadas. O vetor de resíduos, ainda pode ser escrito sob a forma mais detalhada:

$$R = \begin{bmatrix} I - H_{11} & -H_{12} & -H_{13} \\ -H_{21} & I - H_{22} & -H_{23} \\ -H_{31} & -H_{32} & I - H_{33} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

Onde, $H_{ij} = X_i[(X^T X)^{-1} X_j^T]$ e $i, j = 1, 2, 3$. Após a eliminação das k equações, o novo estimador é:

$$\hat{\varphi}^* = (X_1^T X_1 + X_3^T X_3)^{-1} (X_1^T Y_1 + X_3^T Y_3) \quad (4.69)$$

4.3.1

Procedimento Iterativo

Seja $\{Z_t\}$ uma série temporal estacionária sem *outliers*, gerada por um modelo ARMA (p, q) , é sabido que esse processo pode ser aproximado por um AR $(p+q)$. Lembrando que os coeficientes π de um processo ARMA são obtidos pela equação $\pi(B) = \frac{\phi(B)}{\theta(B)}$, e por causa da invertibilidade $\theta(B)^{-1}$, estes coeficientes decaem e se aproximam de zero para alguma defasagem p^* , onde $p^* = p + q$.

As etapas do procedimento incluem: estimar a série $\{Y_t\}$ por um processo $AR(p^*)$ de ordem suficientemente elevada, e obter as primeiras estimativas $\hat{\varphi}$ pelo método de mínimos quadrados. Com base nos resultados do teste da razão de verossimilhança, é possível identificar o conjunto de *outliers*, o tipo e onde estão localizados na série. Seja $t=T$ a posição do *outlier*, o ajuste da série será:

- (i) Tipo AO: elimina-se $(T - p^*)$ equações até T e obtém-se $\hat{\varphi}^*$. Substitui-se a T -ésima observação pelo seu valor esperado condicional a todas as outras observações $E(y_T/y_t, t \neq T)$:

$$y_t^* = y_t, \quad t \neq T$$

$$y_T^* = \sum_{j=1}^{p^*} \hat{\eta}_j (y_{T+j} + y_{T-j}), \quad t = T \quad (4.70)$$

$$\text{Onde, } \hat{\eta}_j = \frac{\hat{\varphi}_j^* - \sum_{i=1}^{p^*-j} \hat{\varphi}_i^* \hat{\varphi}_{i+j}^*}{1 + \sum_{i=1}^{p^*} \hat{\varphi}_i^{*2}}$$

- (ii) Tipo IO: elimina-se a T -ésima equação e obtém-se $\hat{\Phi}^*$. Substitui-se a T -ésima observação por:

$$\begin{aligned} y_t^* &= y_t, & t < T \\ y_T^* &= y_T - \hat{a}_T, & t = T \\ y_t^* &= y_t - \hat{\psi}_{t-T}^* \hat{a}_T, & t > T \end{aligned} \quad (4.71)$$

Onde, \hat{a}_t é a série de resíduos gerada pela estimativa de $\hat{\Phi}^*$ e $\hat{\psi}^*$ é o coeficiente do polinômio de grau infinito na definição $1 - \psi_1 B - \psi_2 B^2 - \dots = (1 - \pi_1 B - \dots - \pi_{p^*} B^{p^*})^{-1}$.

A aplicação deste procedimento resulta em uma série livre dos efeitos dos *outliers* tipo AO e/ou IO, pronta para ser utilizada em metodologias de modelagem.