

**Modelos auto-regressivos
com limiares**

Sérgio Ricardo Martins

DISSERTAÇÃO APRESENTADA AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE
MESTRE EM ESTATÍSTICA

Área de Concentração: Estatística

Orientadora: Profª. Dra. Clélia Maria de Castro Toloi

São Paulo, março de 2003.

Modelos Auto-Regressivos com Limiares

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Sérgio Ricardo Martins e aprovada pela comissão julgadora.

São Paulo, 21 de março de 2003.

Banca Examinadora:

Professora Dra. Clélia Maria de Castro Toloí (Orientadora) – IME – USP

Professor Dr. Pedro Alberto Morettin – IME – USP

Professora Dra. Silvia Regina da Costa Lopes - UFRGS

Para minha tia Elisabeth e minha avó Anália que são
os motivos da minha existência.

Agradecimentos

A Deus por sempre me dar força, sabedoria e coragem.

A minha mãe Elisabete por ter me propiciado a vida.

A minha tia Maristela por sempre ter acreditado em mim.

Sou grato em especial à minha grande amiga Adriana Bruscato por tudo o que ela representa na minha vida pessoal, profissional e acadêmica.

Agradeço às minhas eternas fontes de inspiração: Lara Fidelis e Ricardo Chiquetto, sem vocês à minha volta a vida não faria sentido.

Sou grato ainda às amigas Delhi Salinas e Airlane Alencar pela confiança na minha capacidade.

Agradeço também à minha orientadora Clélia que, além da amizade, me fez abraçar a Estatística, ainda na graduação, quando me propiciou assistir a um brilhante curso de Séries Temporais. A você dedico este trabalho e todo meu eterno carinho.

Resumo

O modelo auto-regressivo com limiares é um dos possíveis modelos de séries temporais não lineares encontrados na literatura. Ele foi inicialmente proposto por Tong e discutido em detalhes por Tong e Lim. Este trabalho tem como objetivo apresentar as metodologias, para o caso univariado, propostas por Tong e Tsay e, no caso multivariado, proposta por Tsay. Nas aplicações utilizaremos dados epidemiológicos e de poluição na cidade de São Paulo.

Abstract

The threshold autoregressive model is one of the nonlinear time series models available in the literature. The purpose of this work is to present the methodologies, due to Tong and Tsay for both: univariate and multivariate cases. These methodologies are applied to air pollution/mortality and temperature data from São Paulo, Brazil.

Conteúdo

Capítulo 1 - Introdução.....	1
Capítulo 2 - Modelo univariado com limiares (" <i>threshold</i> ").....	3
2.1. Introdução.....	3
2.2. Procedimento de análise proposto por Tong.....	5
2.3. Procedimento de análise proposto por Tsay.....	15
Capítulo 3 - Modelo multivariado com limiares (" <i>threshold</i> ").....	22
3.1. Introdução.....	22
3.2. Teste para verificar a necessidade de um modelo com limiares.....	24
3.3. Identificação do modelo.....	28
3.4. Estimação do modelo.....	30
Capítulo 4 - Aplicações.....	32
4.1. Introdução.....	32
4.2. Modelos univariados com limiares.....	33
4.3. Modelo multivariado com limiares.....	65
4.4 Considerações finais.....	68
Referências Bibliográficas.....	70
Apêndice A - Dados.....	73
Apêndice B - Programas.....	80

Capítulo 1

Introdução

Neste trabalho iremos apresentar alguns procedimentos para ajustar modelos lineares por regimes. Esse tipo de metodologia foi enunciada primeiramente em Tong (1977) e discutida em detalhes por Tong e Lim (1980) e Tong (1983).

O uso de variáveis e valores limiares nos permite modelar uma série temporal não-linear em modelos lineares por regimes. Entre os diversos modelos com limiares, vamos descrever o *Modelo Auto-Regressivo com Limiares Auto-Controlados* (SETAR), sendo este, um dos possíveis modelos de Séries Temporais não-lineares encontrado na literatura.

Nos últimos anos muito se tem estudado sobre os modelos não lineares, principalmente com o progresso da área computacional facilitando a análise desses modelos. Contudo, o modelo SETAR não tem sido muito utilizado em aplicações, principalmente devido à dificuldade de identificação da variável limiar, da estimação dos limiares a ela associados e, também, devido a não existência de um procedimento simples de modelagem.

Em Tong (1983), encontramos um número grande de exemplos de diversas áreas, em que a noção de limiares é claramente aplicável, como: engenharia naval, oceanografia, economia, biologia entre outras.

O maior destaque deste tipo de modelagem é o fato dela captar fenômenos como: ciclos limite, dependência entre amplitude e frequência e fenômenos de salto, que não são captados quando utilizamos modelos lineares. Mais detalhes sobre estes fenômenos podem ser encontrados, por exemplo, em Priestley (1988), Tong e Lim (1980), Tong (1983) e Ventura (2000).

O Capítulo 2 é dedicado à análise da modelagem proposta por Tong (1983) e Tsay (1989), ou seja, apresentamos a metodologia descrita por cada autor para modelarmos um processo SETAR. No Capítulo 3 descrevemos a metodologia proposta por Tsay (1998) para ajustarmos modelos “*threshold*” multivariados. Finalizamos a dissertação com aplicações destes modelos para séries epidemiológicas e de poluição ambiental, na cidade de São Paulo, no período de 18/04/1996 – 31/12/1997 (totalizando 623 observações), no Capítulo 4. Ainda neste capítulo, os modelos univariados não-lineares serão comparados aos modelos lineares propostos em Box *et al.* (1994).

No Apêndice A apresentamos as séries de dados reais utilizadas nas aplicações e no Apêndice B deixamos à disposição dos leitores, os programas utilizados para a identificação preliminar dos modelos estudados. Vale ressaltar que todos foram escritos em linguagem de programação S.

Capítulo 2

Modelo univariado com limiares (“*threshold*”)

2.1 Introdução

Uma série temporal X_t será um processo $SETAR(k; p_1, \dots, p_k; d)$ se satisfizer

$$X_t = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} X_{t-i} + a_t^{(j)}, \quad r_{j-1} \leq X_{t-d} < r_j \quad (2.1)$$

em que : $j = 1, 2, \dots, k$;

d – parâmetro de defasagem, inteiro e positivo;

k – número de regimes separados por $k-1$ limiares r_j ;

p_j – ordem auto-regressiva (AR) no j -ésimo regime;

X_{t-d} – variável limiar e

$\{a_t^{(j)}\}$ – seqüências de variáveis aleatórias independentes e normalmente

distribuídas com média zero e variância $\sigma_{a_j}^2$.

Vale observar que:

- (i) a ordem AR pode ser diferente para cada regime;
- (ii) o modelo *SETAR* torna-se um modelo auto-regressivo não-linear não-homogêneo se e somente se as variâncias dos ruídos $\sigma_j^2 = Var(a_t^{(j)})$ forem diferentes para diferentes regimes;
- (iii) o modelo *SETAR* reduz-se a um modelo de mudança de nível aleatório se os termos constantes, $\phi_0^{(j)}$, forem diferentes para os diferentes regimes; e
- (iv) um modelo *SETAR*(1, p) é simplesmente um modelo AR(p).

As características (ii) e (iii) são de interesse especial em várias aplicações e são ligadas a valores discrepantes e modelos de mudança de nível em análise de séries temporais lineares.

Segundo Priestley (1988) o modelo (2.1) é capaz de captar certos fenômenos que os modelos da classe ARIMA, por exemplo, não conseguem. Tais fenômenos são:

- (i) *ciclos limite* (oscilações contínuas de mesma freqüência) – este é um típico comportamento de séries que apresentam periodicidades; o ciclo limite pode ser estável ou instável. Segundo Tong (1983), este comportamento pode ser detectado, quando existir, observando o comportamento gráfico de X_t em função de X_{t-1} . Nas Figuras 2.1 e 2.2 estão representados os ciclos limites estáveis e instáveis, respectivamente.

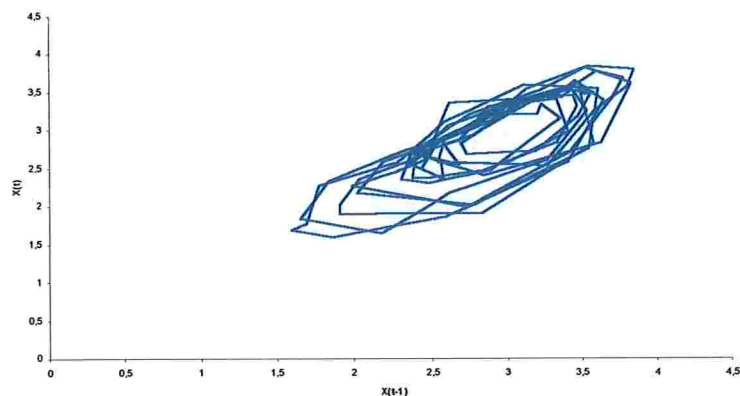


Figura 2.1 – Ciclo limite estável.

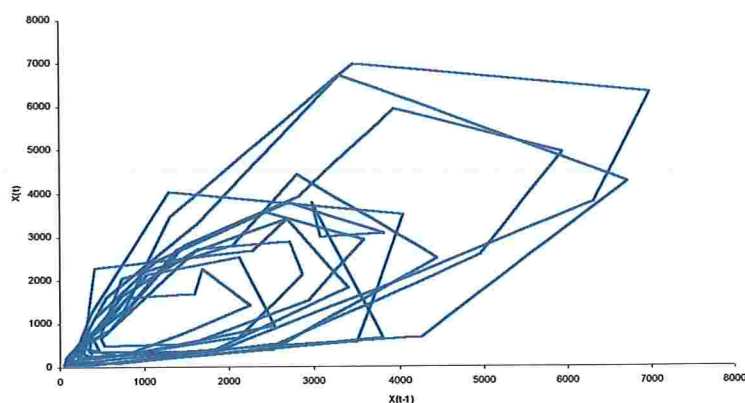


Figura 2.2 – Ciclo limite instável.

(ii) *dependência entre amplitude e freqüência* – em alguns sistemas não-lineares a amplitude do processo é dependente da freqüência, por exemplo, a amplitude da série aumenta quando a freqüência aumenta. Este comportamento é chamado de *dependência entre amplitude e freqüência*.

(iii) *fenômenos de salto* – a amplitude de saída de um sistema não-linear, diferentemente de um sistema linear, pode ter “saltos” em diferentes freqüências dependendo se a freqüência de entrada, de amplitude constante, é monotonicamente crescente ou decrescente. Este comportamento é chamado *fenômeno de salto*.

Mais detalhes sobre estes fenômenos podem ser encontrados, por exemplo, em Priestley (1988), Tong e Lim (1980), Tong (1983), Tong (1990) e Ventura (2000).

2.2 Procedimento de análise proposto por Tong

2.2.1 Métodos gráficos para análise inicial dos dados (identificação)

Numerosos métodos gráficos são utilizados na construção de modelos em séries temporais. Por exemplo, gráfico da série, função de autocorrelação (FAC) e função de autocorrelação parcial (FACP) amostrais, função de densidade espectral amostral, histogramas, gráfico da série tomando diferenças e fazendo transformações, etc. Nesta

sub-seção descreveremos apenas algumas ferramentas gráficas no domínio do tempo. Mais detalhes, destas e de outras possíveis técnicas gráficas veja, por exemplo, Tong (1983).

(a) **Gráfico das Observações.**

Pode nos mostrar a existência, ou não, de fenômenos periódicos. Os gráficos sugeridos são:

- (i) Gráfico da série X_t em função de t e
- (ii) Diagramas de dispersão de $(X_t, X_{t-j}), j = 1, \dots, p$.

(b) **Histogramas univariado de X_t e bivariado para $(X_t, X_{t-j}), j = 1, \dots, p$.**

A análise desses gráficos fornece indicações sobre a simetria e modalidade da distribuição de probabilidades de $\{X_t\}$. Segundo Tong, a existência de modas e antimodas podem ser bastante informativas para a identificação do(s) parâmetro(s) limiar(es).

(c) **Estimativas não-paramétricas da função de regressão de X_t em $X_{t-j}, j = 1, \dots, p$.**

Seja $\{X_t\}$ uma série temporal estacionária com $\gamma_0 = E(X_t^2) < \infty$ e função de autocorrelação $\rho_j, j = 0, \pm 1, \dots$. Estimativas da função de regressão não-paramétrica de X_t em $X_{t-j}, j = 1, 2, \dots, p$, pode nos fornecer informações sobre a não-linearidade de uma série temporal. Denote por

$$m_j(x) = E(X_t | X_{t-j} = x). \quad (2.2)$$

a função de regressão não-paramétrica.

A variância de $m_j(x)$ é dada por

$$E\{X_t - E(X_t | X_{t-j})\}^2 = \gamma_0 - \text{Var}\{E(X_t | X_{t-j})\}. \quad (2.3)$$

A melhor aproximação linear para $E(X_t | X_{t-j})$, no sentido de média quadrática mínima, é dada por $\mu_x + \rho_j(X_{t-j} - \mu_x)$, onde $\mu_x = E(X_t)$. Além disso, a variância de $X_t - [\mu_x + \rho_j(X_{t-j} - \mu_x)]$ dada por

$$\gamma_0 - \gamma_0 \rho_j^2, \quad (2.4)$$

que se torna um limite superior de (2.3), isto é,

$$\gamma_0 - \text{Var}\{E(X_t | X_{t-j})\} \leq \gamma_0 - \gamma_0 \rho_j^2. \quad (2.5)$$

Assim,

$$0 \leq \gamma_0 \rho_j^2 / \text{Var}\{E(X_t | X_{t-j})\} \leq 1. \quad (2.6)$$

Tong denomina a quantidade

$$\mathcal{L} = \inf_{j \in \mathbb{Z}} [\gamma_0 \rho_j^2 / \text{Var}\{E(X_t | X_{t-j})\}], \quad (2.7)$$

de índice de linearidade de segunda ordem.

Se $\text{Var}\{E(X_t | X_{t-j})\} = 0$, para algum j , este índice será definido como sendo 1. É claro que, $0 \leq \mathcal{L} \leq 1$, bastando para tal observar (2.6). A quantidade \mathcal{L} pode ser aproximadamente interpretada como uma medida prática de quão próximo o mecanismo gerador de $\{X_t\}$ é linear. O modelo será linear quando $\mathcal{L} = 1$ se supusermos $\{X_t\}$ gaussiana e um modelo AR(1) com uma função de auto-regressão diferenciável. Mais detalhes sobre este coeficiente de linearidade podem ser encontrados, por exemplo, em Tong (1990).

Do exposto acima, o índice de linearidade em (2.7) sugere que estimativas dos $m_j(x)$ podem ser úteis na identificação do parâmetro de defasagem, d .

Descreveremos, a partir de agora, uma metodologia para encontrar estimativas do índice de linearidade.

Dadas as observações $\{x_1, x_2, \dots, x_N\}$ uma estimativa central de (2.2), para $j = 1, 2, \dots, p$, ($p \ll N$), pode ser dada por

$$\hat{m}_{j,N}(x) = \frac{\sum_{t=1}^{N-j} x_t \delta_N(x - x_{t+j})}{\sum_{t=1}^{N-j} \delta_N(x - x_{t+j})}, \quad (2.8)$$

e

$$\hat{m}_{-j,N}(x) = \frac{\sum_{t=j+1}^N x_t \delta_N(x - x_{t-j})}{\sum_{t=j+1}^N \delta_N(x - x_{t-j})}, \quad (2.9)$$

sendo $\delta_N(z)$ uma seqüência de funções não negativas, de área total unitária, denominada função suavizadora ou janela. Quando $N \rightarrow \infty$, $\delta_N(z)$ converge para uma função delta de Dirac. Tal seqüência pode ser gerada pela introdução de funções fixadas $k(z) \geq 0$, $\int k(z) dz = 1$ e um B_N tal que a família é $\{B_N^{-1}k(z/B_N)\}$, isto é, $\delta_N(z) = B_N^{-1}k(z/B_N)$, com $B_N \rightarrow 0$ quando $N \rightarrow \infty$. Além disso, necessitamos que $\int zk(z) dz = 0$.

Uma forma para $k(z)$, por exemplo, é dada por:

$$k(z) = \begin{cases} 1 - |z|, & \text{para } |z| \leq 1 \\ 0, & \text{caso contrário.} \end{cases} \quad (2.10)$$

Outras formas específicas para $k(z)$ também podem ser utilizadas, para tanto veja, por exemplo, Parzen (1958).

Watson (1964), demonstra que $\hat{m}_j(x)$ dado por (2.8) – (2.9) é um estimador consistente de $m_j(x)$.

2.2.2 Método de estimação dos parâmetros

O procedimento de estimação dos parâmetros de um modelo AR(p) pode ser facilmente estendido para o modelo $SETAR(k; p_1, \dots, p_k; d)$, uma vez que o mesmo consiste de k regimes lineares. Introduziremos o procedimento para o ajuste de um modelo $SETAR(2; p_1, p_2; d)$.

Sejam d e r fixados e $l = \max(p_1, p_2, d)$. As observações (X_{t+1}, \dots, X_N) podem ser divididas em dois conjuntos, seguindo a regra:

X_t pertence ao primeiro conjunto se, e somente se, $X_{t-d} \leq r$, e

X_t pertence ao segundo conjunto se, e somente se, $X_{t-d} > r$.

Denotamos $R_1 = \{x_{j_1}^{(1)}, x_{j_2}^{(1)}, \dots, x_{j_{n_1}}^{(1)}\}$ e $R_2 = \{x_{j_1}^{(2)}, x_{j_2}^{(2)}, \dots, x_{j_{n_2}}^{(2)}\}$, ($n_1 + n_2 = N - l$),

as observações do primeiro e segundo conjunto, respectivamente. Para cada conjunto de dados temos um modelo da forma

$$X_i = A_i \theta_i + e_i, \quad (2.11)$$

com

$$X_i = (x_{j_1}^{(i)}, \dots, x_{j_{n_i}}^{(i)})',$$

$$e_i = (e_{j_1}^{(i)}, \dots, e_{j_{n_i}}^{(i)})',$$

$$\theta_i = (a_0^{(i)}, \dots, a_{p_i}^{(i)}) \text{ e}$$

$$A_i = \begin{bmatrix} 1 & x_{j_1-1}^{(i)} & x_{j_1-2}^{(i)} & \cdots & x_{j_1-p_i}^{(i)} \\ 1 & x_{j_2-1}^{(i)} & x_{j_2-2}^{(i)} & \cdots & x_{j_2-p_i}^{(i)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{j_{n_i}-1}^{(i)} & x_{j_{n_i}-2}^{(i)} & \cdots & x_{j_{n_i}-p_i}^{(i)} \end{bmatrix}$$

para $j = 1, 2$.

Denotando os estimadores de mínimos quadrados por $\hat{\theta}_i$, ($i = 1, 2$), temos que $\hat{e}_i = x_i - A_i \hat{\theta}_i$ e $\|\hat{e}_i\|^2 / n_i$ é a estimativa (aproximada) de máxima verossimilhança da variância, σ_i^2 , para o i -ésimo conjunto. Vale notar que, ao contrário do caso linear, diferentes conjuntos de variáveis regressoras são envolvidas nos A_i , ($i = 1, 2$), para diferentes regimes.

A seguir apresentaremos um procedimento sistemático para a obtenção das estimativas dos parâmetros $d, r, p_1, a_0^{(1)}, a_1^{(1)}, \dots, a_{p_1}^{(1)}, \sigma_1^2, p_2, a_0^{(2)}, a_1^{(2)}, \dots, a_{p_2}^{(2)}, \sigma_2^2$. Consideraremos $\{1, 2, \dots, T\}$ e $\{\tau_1, \tau_2, \dots, \tau_s\}$ os potenciais candidatos para a estimação de d e r , respectivamente.

Primeiro Passo:

a) Fixar $d = d_0$ e $r = r_0$. Denotar por L a máxima ordem auto-regressiva de cada regime; a escolha de L é subjetiva, mas usualmente ligada ao tamanho da amostra ($L = N^\alpha$, $\alpha < 1/2$). Vale ressaltar que a máxima ordem de um regime não necessariamente é a mesma do outro. Aqui trabalharemos com ordens iguais para facilitar a descrição do procedimento. Seja $N_{d_0} = \max(d_0, L)$; normalmente escolhemos $T \leq L$.

b) Arrumar as observações $(x_{N_{d_0}+1}, \dots, x_N)$ em dois conjuntos conforme descrito pela equação (2.11). Estimar, utilizando mínimos quadrados, θ_1 e θ_2 para cada escolha de p_1 e p_2 ; podemos utilizar, por exemplo, o Critério de Informação de Akaike (AIC) para determinarmos a ordem auto-regressiva de cada um dos regimes, isto é, escolher o valor \hat{p}_i tal que

$$AIC(\hat{p}_i) = \min_{0 \leq p_i \leq L} [n_i \ln\{\|\hat{e}_i(p_i)\|^2 / n_i\} + 2(p_i + 1)], \quad i = 1, 2, \quad (2.12)$$

em que $\hat{e}(p_i)$ são os resíduos do ajuste, $x_i - A_i \hat{\theta}_i$, com $\theta_i = (a_0^{(i)}, \dots, a_{p_i}^{(i)})$.

Para $d = d_0$ e $r = r_0$, o MAICE (*Minimum Akaike Criterium Estimation*) de p_i , ($i = 1, 2$), é o valor \hat{p}_i que satisfaz a equação anterior.

Vale observar que $n_1 + n_2 = N - N_{d_0}$. Logo,

$$AIC(d_0, r_0) = AIC(\hat{p}_1) + AIC(\hat{p}_2). \quad (2.13)$$

Segundo Passo:

Fixar $d = d_0$, variar r em $\{\tau_1, \tau_2, \dots, \tau_s\}$ e minimizar o $AIC(d_0, r)$ neste conjunto, ou seja, condicionando $d = d_0$, obter o valor \hat{r} de τ_i é tal que

$$AIC(d_0, \hat{r}) = \min_{r \in \{\tau_1, \tau_2, \dots, \tau_s\}} \{AIC(d_0, r)\}. \quad (2.14)$$

Este será MAICE de r , e os \hat{p}_1, \hat{p}_2 correspondentes ao \hat{r} serão o MAICE de p_1 e p_2 .

Terceiro Passo:

Escolher em $\{1, 2, \dots, T\}$ o MAICE de d . Escolhas de diferentes d alteram $N - N_d$, então, pode-se normalizar o $AIC(d, \hat{r})$ dividindo-o por $N - N_d$. Para cada escolha de d , o AIC normalizado é dado por

$$NAIC(d) = AIC(d, \hat{r}) / (N - N_d). \quad (2.15)$$

O MAICE de d será denotado por \hat{d} e

$$\text{NAIC}(\hat{d}) = \min_{d \in \{1, 2, \dots, T\}} \{\text{NAIC}(d)\}. \quad (2.16)$$

Ao final do terceiro passo, nós completamos a identificação do mínimo AIC, que nos fornece o MAICE de todos os parâmetros listados a partir do primeiro passo.

2.2.3 Análise dos Resíduos

Após termos identificado o modelo SETAR, precisamos testar a adequação dos resíduos, isto é, verificar se eles:

- (i) são aproximadamente não correlacionados (ruído branco);
- (ii) têm distribuição normal (Ghaddar e Tong, 1981).

Citamos aqui dois procedimentos. Ver Tong (1983) para demais metodologias.

a) Teste de Ruído Branco

Sejam $\{\hat{\epsilon}_t; t = 1, 2, \dots, N\}$ os resíduos resultantes do ajuste do modelo da série temporal observada $\{X_t; t = 1, 2, \dots, N\}$, amostrada de um processo estocástico estacionário. Seja $\hat{\rho}_k$ a função de autocorrelação amostral de *lag* k , $k = 0, 1, \dots$, de $\{\hat{\epsilon}_t\}$, dada por:

$$\hat{\rho}_k = \frac{\sum_{i=k+1}^N (\hat{\epsilon}_i - \bar{e})(\hat{\epsilon}_{i-k} - \bar{e})}{\sum_{i=1}^N (\hat{\epsilon}_i - \bar{e})^2}, \quad (2.17)$$

em que

$$\bar{e} = \frac{\sum_{i=1}^N \hat{\epsilon}_i}{N}. \quad (2.18)$$

Um teste bem popular é baseado no exame de $\hat{\rho}_k$, $k = 1, 2, \dots, L$ ($L \ll N$). Sob a hipótese nula de ruído branco, é bem conhecido que $\sqrt{N}\hat{\rho}_k$, para $k = 1, 2, \dots, L$ são assintoticamente independentes e normais com média zero e variância 1 (Priestley, 1981). Então, para aceitarmos a hipótese de ruído branco, esperamos que não mais que 5% das autocorrelações amostrais estejam além da banda de confiança $\pm 1,96 / \sqrt{N}$ (Jenkins e Watts, 1968).

Ljung e Box (1978) sugeriram um teste para as autocorrelações dos resíduos estimados, que, apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos. Se o modelo for apropriado, a estatística:

$$Q(K) = N(N+2) \sum_{k=1}^K \frac{\hat{r}_k^2}{(N-k)}, \quad (2.19)$$

em que

K – número de defasagens consideradas; e

\hat{r}_k - k-ésima autocorrelação residual

tem distribuição χ^2 com $K-p-q$ graus de liberdade. A hipótese de ruído branco para os resíduos é rejeitada para valores grandes de Q . Segundo Morettin e Tolo (1987), em geral basta tomar as 20 ou 25 primeiras autocorrelações residuais.

Para o caso de modelos univariados com limiares, Tong (1983) considera razoável normalizar os $\hat{\epsilon}_i$ por $\|\hat{\epsilon}_i\|^2 / n_i$, $i = 1, 2, \dots, K$, referindo-se a eles como resíduos normalizados.

a) Teste para verificar a normalidade

Suponha que os $\hat{\epsilon}_t$ sejam aproximadamente ruído branco. Podemos então proceder a investigação da normalidade. Um histograma contém bastante informação e é um bom indicador gráfico. Coeficientes amostrais de assimetria e curtose também são bons indicadores. O gráfico normal de probabilidade também pode ser utilizado.

O estudo da assimetria de $\{X_t\}$ é de particular interesse, desde que relacionada à irreversibilidade temporal e é esperado que o modelo SETAR, quando escolhido adequadamente, capte esta particularidade.

Os métodos informais, propostos anteriormente, podem ser complementados com um teste formal para verificar a normalidade. Sugerimos, a seguir, um procedimento desenvolvido por Lin e Mudhoekar (1980) para testar a normalidade dos erros. Inicialmente calculamos,

$$g_i = \left\{ \frac{1}{N} \left[\sum_{j \neq i}^N \hat{e}_j^2 - \frac{1}{N-1} \left(\sum_{j \neq i}^N \hat{e}_j \right)^2 \right] \right\}^{1/3}, \quad i = 1, 2, \dots, N \quad (2.20)$$

e, em seguida,

$$R = \frac{\sum_{i=1}^N (\hat{e}_i - \bar{e})(g_i - \bar{g})}{\left\{ \sum_{i=1}^N (\hat{e}_i - \bar{e})^2 \sum_{i=1}^N (g_i - \bar{g})^2 \right\}^{1/2}}. \quad (2.21)$$

Demonstra-se que a estatística de teste

$$T^* = 2^{-1} \left(\frac{N}{3} \right)^{1/2} \ln \left(\frac{1+R}{1-R} \right), \quad (2.22)$$

segue uma distribuição assintoticamente normal com média zero e variância 1, sob a hipótese nula de normalidade dos e_t . A estatística (2.22) é baseada em uma propriedade da distribuição normal: independência da média amostral e da variância amostral.

2.3 Procedimento de análise proposto por Tsay

2.3.1 Teste para não-linearidade “*threshold*”

Este teste foi construído baseando-se no teste *portmanteau* de não-linearidade de Petrucci e Davies (1986), que também é baseado numa auto-regressão arrumada e nos resíduos preditivos.

a) Auto-regressão arrumada e resíduos preditivos

Vamos escrever um modelo AR(p) com n observações, da seguinte forma

$$X_t = (1, X_{t-1}, \dots, X_{t-p})\beta + a_t, \text{ para } t = p+1, \dots, n,$$

em que β é o vetor de $(p+1)$ coeficientes e a_t é o ruído.

Define-se como auto-regressão arrumada à auto-regressão com as observações re-arranjadas, de acordo com um particular regressor. Para o modelo dado por (2.1), a auto-regressão arrumada torna-se útil se for construída de acordo com a variável limiar.

Apenas para exemplificar, vamos considerar o caso em que $k = 2$, ou seja, a situação de um valor limiar, r_1 , não trivial. Assim, para um dado modelo *SETAR*(2; $p_1, p_2; d$) com n observações, a variável limiar, X_{t-d} , pode assumir valores $\{X_h, \dots, X_{n-d}\}$, em que $h = \max\{1, p+1-d\}$. Seja π_i o indexador do tempo da i -ésima menor observação de $\{X_h, \dots, X_{n-d}\}$. Podemos reescrever o modelo (2.1) da seguinte forma

$$X_{\pi_i+d} = \begin{cases} \phi_0^{(1)} + \sum_{v=1}^{p_1} \phi_v^{(1)} X_{\pi_i+d-v} + a_{\pi_i+d}^{(1)}, & \text{se } i \leq s \\ \phi_0^{(2)} + \sum_{v=1}^{p_2} \phi_v^{(2)} X_{\pi_i+d-v} + a_{\pi_i+d}^{(2)}, & \text{se } i > s \end{cases} \quad (2.23)$$

onde s satisfaz $X_{\pi_s} < r_1 < X_{\pi_{s+1}}$.

Assim (2.23) é uma auto-regressão arrumada com os s primeiros casos no primeiro regime e o restante no segundo. A auto-regressão arrumada nos fornece um meio pelo qual as observações pontuais serão agrupadas de forma que todas as observações de um mesmo grupo seguem o mesmo modelo AR linear.

Considerando o modelo (2.23), se o valor r_1 for conhecido, então estimativas consistentes dos parâmetros podem ser facilmente obtidas.

Como o valor limiar é desconhecido, segundo Tsay (1989), podemos prosseguir seqüencialmente:

Para o modelo (2.23), seja $\hat{\beta}_m = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)'$ o vetor de estimadores de mínimos quadrados baseado nas m primeiras observações (Tsay (1989), sugere utilizar $m = (n/10) + p$ para inicializar a recursão), P_m a matriz inversa associada a $X'X$ e $x_{m+1} = (1, x_{t-1}, \dots, x_{t-p})'$ o vetor de regressores das próximas observações a entrar na auto-regressão, isto é $X_{d+\pi_{m+1}}$.

Logo, o estimador de mínimos quadrados recursivo pode ser facilmente calculado por

$$\hat{\beta}_{i+1} = \hat{\beta}_i + K_{i+1} [X_{\pi_{i+1}+d} - x'_{i+1} \hat{\beta}_i],$$

$$D_{i+1} = 1 + x'_{i+1} P_i x_{i+1},$$

$$K_{i+1} = \frac{P_i x_{i+1}}{D_{i+1}},$$

$$P_{i+1} = \left(I - P_i \frac{x_{i+1} x'_{i+1}}{D_{i+1}} \right) P_i,$$

$i = m, \dots, n-d$, respectivamente.

Para mais detalhes sobre esta recursão e suas propriedades, veja Ertel e Fowlkes (1976), Goodwin e Payne (1977) e Tsay (1989).

Os resíduos preditivos e os preditivos padronizados ficam dados por

$$\hat{a}_{\pi_{m+1}+d} = X_{\pi_{m+1}+d} - x'_{m+1} \hat{\beta}_m \quad (2.24)$$

e

$$\hat{\varepsilon}_{\pi_{m+1+d}} = \hat{a}_{\pi_{m+1+d}} / \sqrt{D_{m+1}}. \quad (2.25)$$

Para mais detalhes sobre a consistência destes estimadores veja, por exemplo, Tsay (1989).

b) Teste de linearidade

Este teste baseia-se ainda nos testes de não-linearidade propostos por Keenan (1985) e Tsay (1986).

Considere as hipóteses

H_0 : X_t é linear

H_A : X_t segue um modelo com limiares. (2.26)

Para p e d fixos, o número de observações na auto-regressão é $n-d-h+1$, com $h = \max\{1, p+1-d\}$. Assumindo que a auto-regressão recursiva inicia com m observações, temos $n-d-m-h+1$ resíduos preditivos.

Ajustando a regressão

$$\hat{\varepsilon}_{\pi_i+d} = w_0 + \sum_{v=1}^p w_v X_{\pi_i+d-v} + \varepsilon_{\pi_i+d}, \quad i = m+1, \dots, n-d-h+1 \quad (2.27)$$

por mínimos quadrados, testamos (2.26) utilizando a estatística

$$F^*(p, d) = \frac{(\sum \hat{\varepsilon}_t^2 - \sum \hat{\varepsilon}_t^2) / (p+1)}{\sum \hat{\varepsilon}_t^2 / (n-d-m-p-h)}. \quad (2.28)$$

em que $\hat{\varepsilon}_t$ são os resíduos do ajuste do modelo em (2.27).

O argumento (p, d) de F^* é usado para mostrar a dependência de p e d no cálculo da estatística.

Resultado: Suponha que X_t é um processo linear estacionário AR de ordem p , isto é, X_t segue o modelo (2.1) para $k = 1$. Então, para n grande a estatística (2.28) segue aproximadamente uma distribuição *F-Snedecor* com $(p+1)$ e $(n-d-m-p-h)$ graus de liberdade. Ainda, $(p+1)F^*(p, d)$ é assintoticamente uma variável aleatória Qui-quadrado com $(p+1)$ graus de liberdade.

Para mais detalhes sobre este teste, ver Tsay (1989).

2.3.2 Especificação da variável limiar

A maior dificuldade em se ajustar um modelo SETAR está na especificação da defasagem da variável limiar, que é a mais importante na natureza dos modelos não-lineares com limiares. Tong e Lim (1980), por exemplo, usam o critério de informação de Akaike, AIC, para selecionar d , após escolherem todos os outros parâmetros. Tsay (1989), propõe um procedimento diferente que seleciona d , antes de localizar os valores limiares, r_i . O método proposto foi motivado pela performance da estatística (2.28) na análise de conjuntos de dados reais.

a) Seleção do parâmetro de defasagem d

Assumindo conhecida a ordem p do modelo auto-regressivo, para um certo processo SETAR, selecionaremos uma estimativa para o parâmetro de defasagem, denotado por d_p , utilizando a estatística

$$F^*(p, d_p) = \max_{v \in S} \{F^*(p, v)\}, \quad (2.29)$$

em que, $F^*(p, v)$ é dada por (2.28)

O índice p significa que a estimativa para d depende de p e S é um conjunto de inteiros positivos pré-especificados, isto é, uma coleção de possíveis valores de d . Este conjunto S de possíveis limiares de defasagem pode ser $\{1, \dots, p\}$ para um dado valor de p . Pode-se incluir defasagens sazonais quando elas existirem no processo. Para simplificar, assuma que todas as estatísticas $F^*(p, \nu)$ de (2.29) tenham o mesmo número de graus de liberdade. Isto pode ser obtido por uma seleção correta do ponto de partida m da recursão. Quando os graus de liberdade forem diferentes, uma maneira é calcular o nível descritivo associado à estatística F e selecionar d_p baseando-se no menor valor obtido. É interessante tentar vários valores para d nos dados a serem analisados, tais como utilizar o máximo e o segundo maior valor de d de (2.28) em (2.29).

b) Localização dos limiares

Assumindo, por exemplo, $k = 2$ e o verdadeiro valor de r_1 satisfazendo $X_{\pi_s} < r_1 < X_{\pi_{s+1}}$, notamos que qualquer valor no intervalo $[X_{\pi_s}, X_{\pi_{s+1}})$ é bom para ser uma estimativa de r_1 , pois, todos eles darão o mesmo resultado de ajuste para especificar o modelo SETAR. Em geral, um procedimento é fornecer estimativas intervalares de cada valor limiar (r_i) ou utilizar percentis amostrais para estimativas pontuais. Tsay (1989) utiliza a segunda opção, adotando a aproximação de Tong e Lim (1980) para considerar os percentis empíricos como candidatos a valores limiar.

Os métodos propostos por Tsay (1989) para localizar os limiares utilizam diagramas de dispersão de várias estatísticas versus a variável limiar especificada. Embora os gráficos não sejam testes formais, eles fornecem informações úteis na localização do limiar.

Os gráficos utilizados são:

- (i) diagrama de dispersão dos resíduos padronizados de (2.25) ou resíduos ordinários de (2.24) versus X_{t-d_p} ;
- (ii) diagrama de dispersão de estimativas recursivas padronizadas de cada coeficiente auto-regressivo versus X_{t-d_p} .

Na montagem da auto-regressão arrumada, o modelo SETAR consiste de vários modelos que mudam a cada ocorrência do valor limiar, r_i . Logo, o resíduo preditivo é obtido utilizando estimadores viesados a partir dos valores limiar. Sendo assim, um diagrama de dispersão dos resíduos padronizados versus a variável limiar, pode revelar a localização do valor limiar de um modelo SETAR. Por outro lado, para uma série linear o gráfico é aleatório, exceto no começo da recursão.

Quanto ao uso de diagramas de dispersão de razões t recursivas de um certo coeficiente AR versus a variável limiar, podemos relacionar o seguinte:

- (a) mostram a significância de um particular coeficiente AR;
- (b) quando o coeficiente é significativo as razões t , gradual e suavemente, convergem para um valor fixado no decorrer da recursão;
- (c) o gráfico da razão t do termo constante, $\hat{\phi}_0$, é importante porque significa mudanças de nível;
- (d) a utilidade do gráfico não é limitada ao caso de um único valor limiar e,
- (e) a técnica pode ser utilizada iterativamente. Por exemplo, diminuir o número de observações no primeiro regime para detectar o segundo limiar.

Mais detalhes sobre o comportamento das razões t , veja Tsay (1989).

2.3.3 Modelando um processo SETAR

Assim como Tong (1983), Tsay (1989) sugere um procedimento para a construção do modelo SETAR:

Passo 1 - Selecione a ordem p e o conjunto das possíveis defasagens S .

Passo 2 - Ajuste auto-regressões arrumadas para um dado p e todos os elementos de S e faça o teste de não-linearidade (2.28). Se a não-linearidade do processo for detectada, selecione o parâmetro de defasagem d_p pelo método descrito na seção (2.3.2).

Passo 3 – Para um dado p e d_p , localize o(s) valor(es) limiar(es), r_i , usando os diagramas de dispersão sugeridos na seção (2.2.1).

Passo 4 - Refine a ordem AR e o(s) valor(es) limiar(es), se necessário, em cada regime utilizando técnicas de auto-regressão linear.

Para o Passo 1, a ordem p pode ser selecionada considerando a função de autocorrelação parcial (FACP) de X_t ou algum critério de informação, por exemplo, o AIC. Tsay (1989) opta pela FACP, pois:

- (a) a FACP sugere um valor de p razoável;
- (b) o critério de informação pode ser enganoso quando o processo é de fato não-linear e,
- (c) a ordem p pode ser refinada, se desejada, no Passo 4.

No Passo 3, razões t de vários coeficientes AR podem ser examinadas tão exhaustivamente quanto forem os coeficientes AR significantes. Diagramas de dispersão de coeficiente AR não-significantes são, em geral, não informativos.

O modelo refinado no Passo 4 pode utilizar critérios de informação, tal como AIC, devido a natureza linear do modelo SETAR. Detalhes do uso de AIC na modelagem de um SETAR podem ser encontrados na seção (2.2.2), em Tong e Lim (1980) e Tong (1983).

Capítulo 3

Modelo multivariado com limiares (“threshold”)

3.1 Introdução

Considere uma série de tempo ν -dimensional $\mathbf{y}_t = (y_{1t}, \dots, y_{\nu t})'$ e outra ν -dimensional, de variáveis exógenas, $\mathbf{x}_t = (x_{1t}, \dots, x_{\nu t})'$. Sejam $-\infty < r_0 < r_1 < \dots < r_{k-1} < r_k < \infty$ os limiares. Então \mathbf{y}_t segue um modelo “*threshold*” multivariado com variável limiar z_t e defasagem d se satisfizer

$$\mathbf{y}_t = \mathbf{c}_j + \sum_{i=1}^p \phi_i^{(j)} \mathbf{y}_{t-i} + \sum_{i=1}^q \beta_i^{(j)} \mathbf{x}_{t-i} + \varepsilon_t^{(j)}, \quad \text{se } r_{j-1} < z_{t-d} < r_j, \quad j = 1, 2, \dots, k \quad (3.1)$$

em que, k é o número de regimes, \mathbf{c}_j é um vetor de constantes e p e q são números inteiros não negativos.

As inovações satisfazem a relação $\varepsilon_t^{(j)} = \sum_j^{1/2} a_t$, onde $\sum_j^{1/2}$ são matrizes simétricas positiva definida e $\{a_t\}$ é uma seqüência de vetores aleatórios serialmente não correlacionados, com média zero e matriz de covariâncias I (matriz identidade). Assumiremos estacionariedade e distribuição de probabilidade contínua para a variável limiar z_t . O modelo (3.1) possui k regimes e é um modelo linear por partes no espaço limiar z_{t-d} , mas é não linear no tempo quando $k > 1$. Iremos supor que a variável limiar z_t seja conhecida, porém a defasagem d , o número de regimes k e os limiares r_i serão desconhecidos.

Os objetivos desse capítulo são:

(i) propor um teste estatístico para detectar a necessidade de utilizar um modelo “*threshold*”, contra a alternativa de um modelo linear; ou seja, $k = 1$ versus $k > 1$. Esse teste é uma generalização do teste encontrado em Tsay (1989), para o caso univariado, e possui uma distribuição assintótica Qui-quadrado.

(ii) considerar um procedimento para a construção de um modelo multivariado com limiares que inclua a estimação da defasagem, d , e dos limiares, r_i . Utilizaremos o método de mínimos quadrados condicionais na estimação e o critério de informação de Akaike (AIC) na seleção do modelo. O procedimento é mais simples de ser aplicado do que aquele proposto em Tsay (1989), que utiliza diagramas de dispersão que exigem interpretações subjetivas.

Este capítulo será organizado da seguinte forma: na seção 3.2, consideraremos uma análise de regressão re-arranjada que transforma o modelo (3.1) num problema de mudança de ponto. Em seguida, utilizaremos o método de mínimos quadrados recursivos, a fim de obter resíduos preditivos da regressão re-arranjada, que serão padronizados para serem utilizados na construção de um teste estatístico a ser proposto. Na seção 3.3, proporemos um procedimento para a construção do modelo. Na seção 3.4, dedicada à estimação do modelo, falaremos do procedimento de mínimos quadrados condicionais. O critério AIC é utilizado a fim de selecionar os limiares, assim como os outros parâmetros.

3.2 Teste para verificar a necessidade de um modelo com limiares

Considere as hipóteses

$$\begin{aligned} H_0: & \mathbf{y}_t \text{ é linear} \\ H_A: & \mathbf{y}_t \text{ segue um modelo com limiares.} \end{aligned} \tag{3.2}$$

O teste (3.2) tem chamado muito a atenção nos últimos anos, em parte devido a dificuldade de se encontrar os limiares r_i , que são indefinidos sob a hipótese nula, quando o teste da razão de verossimilhanças é usado; ver Chan e Tong (1990) e Hansen (1996a). A maioria dos testes baseados na verossimilhança considera o caso univariado com dois regimes, isto é, um único limiar, e utilizam simulações para obtenção de valores críticos. Por outro lado, Petruccelli e Davies (1986) e Tsay (1989) transformaram o teste num problema de detecção de mudança de ponto, utilizando o conceito da auto-regressão re-arranjada e utilizaram os resíduos preditivos na construção do teste estatístico, que não envolve parâmetros indefinidos. Mais especificamente, estes testes utilizam a variável limiar na construção da regressão re-arranjada, porém não dependem de outras características do modelo alternativo. Eles são simples e possuem distribuições familiares. Por exemplo, o teste estatístico de Tsay (1989) possui uma distribuição assintótica F-Snedecor. Algumas simulações mostraram que esta última aproximação tem um poder razoável quando o tamanho da amostra é moderado ou grande. Veja Balke e Fomby (1997), para mais detalhes.

Nesta seção, estudaremos uma generalização proposta por Tsay (1998), para o teste estatístico encontrado em Tsay (1989), para o modelo multivariado (3.1). Um estudo sobre o desempenho da estatística de teste para amostras finitas pode ser visto em Tsay (1998), por exemplo. Várias razões justificam tal extensão:

- (i) o teste é bem simples e possui uma distribuição assintótica Qui-quadrado. Esta é uma característica importante quando o tamanho da amostra é grande, tal como na análise de dados de alta frequência em finanças, em que o tamanho da amostra pode facilmente exceder a 5000 observações;

- (ii) o teste é amplamente aplicável, incluindo sistemas co-integrados;
- (iii) o teste se mostra poderoso na constatação da não-linearidade “*threshold*”.

3.2.1 Regressão re-arranjada

Dadas as observações $\{y_t, x_t, z_t\}$, $t = 1, \dots, n$, o objetivo é detectar a não-linearidade “*threshold*” de y_t , supondo que p , q e d sejam conhecidos. Para esta finalidade, utilizaremos o método de mínimos quadrados e reescreveremos (3.1) como um modelo de regressão re-arranjada,

$$y_t = X_t' \Phi + \varepsilon_t, \quad t = h+1, \dots, n \quad (3.3)$$

em que $h = \max(p, q, d)$, $X_t = (1, y_{t-1}, \dots, y_{t-p}, x_{t-1}, \dots, x_{t-q})'$ é um regressor $(p+q+1)$ -dimensional e Φ - é a matriz de parâmetros.

Sob a hipótese nula de (3.2), os estimadores de mínimos quadrados de (3.3) são consistentes. Por outro lado, os estimadores serão viesados, sob a hipótese alternativa.

A variável limiar, z_{t-d} , que foi utilizada para construir o modelo (3.3), assume valores em $S = \{z_{h+1-d}, \dots, z_{n-d}\}$. Considere as estatísticas de ordem de S e denote o i -ésimo menor elemento de S por $z_{(i)}$. Além disso, seja $t(i)$ o índice temporal de $z_{(i)}$. Assim, a regressão re-arranjada baseada na ordem crescente da variável limiar z_{t-d} fica dada por

$$y_{t(i)+d} = X_{t(i)+d}' \Phi + \varepsilon_{t(i)+d}, \quad i = 1, \dots, n-h. \quad (3.4)$$

É importante notar que em (3.4) a dinâmica da série y_t não é modificada, ou seja, a variável independente de y_t é X_t , para todo t . O que mudou foi a seqüência na qual as informações entraram para o início da regressão, isto é, a ordem de entrada é por linha, se colocarmos a regressão na forma matricial. Uma característica importante da regressão re-arranjada é que ela efetivamente transforma um modelo com limiares em

um problema de mudança de ponto, uma vez que a regressão é re-arranjada de acordo com a ordem crescente da variável limiar z_{t-d} .

3.2.2 Teste Estatístico

Há várias maneiras de se detectar a mudança de regime do modelo em (3.4). Aqui utilizaremos os resíduos preditivos e o método recursivo de mínimos quadrados. A idéia é simples: se y_t for linear, então os estimadores dos parâmetros, por mínimos quadrados recursivos, da regressão re-arranjada (3.4), serão consistentes, de forma que os resíduos preditivos serão aproximadamente ruídos brancos e serão não correlacionados com o regressor $X_{t(i)+d}$. Por outro lado, se y_t seguir um modelo com limiares, então, os resíduos preditivos não serão ruídos brancos, já que os estimadores de mínimos quadrados da regressão serão viesados. Neste caso, os resíduos preditivos serão correlacionados com o regressor $X_{t(i)+d}$.

Denote por $\hat{\phi}_m$, a matriz de estimativas de mínimos quadrados da matriz de parâmetros da equação (3.4), com $i = 1, \dots, m$; ou seja, as estimativas da regressão re-arranjada utilizando observações associadas com os m menores valores de z_{t-d} . Sejam

$$\hat{e}_{t(m+1)+d} = y_{t(m+1)+d} - \hat{\phi}'_m X_{t(m+1)+d}, \quad (3.5)$$

e

$$\hat{\eta}_{t(m+1)+d} = \frac{\hat{e}_{t(m+1)+d}}{(1 + X'_{t(m+1)+d} V_m X_{t(m+1)+d})^{1/2}}, \quad (3.6)$$

os resíduos preditivos e preditivos padronizados, respectivamente, da regressão (3.4), com

$$V_m = \left[\sum_{i=1}^m X_{t(i)+d} X'_{t(i)+d} \right]^{-1}.$$

As quantidades (3.5) e (3.6) podem ser obtidas de maneira eficaz pelo algoritmo recursivo de mínimos quadrados.

Considere a regressão

$$\hat{\eta}'_{t(l)+d} = \mathbf{X}'_{t(l)+d} \Psi + \mathbf{w}'_{t(l)+d}, \quad l = m_0 + 1, \dots, n-h, \quad (3.7)$$

em que, m_0 denota o valor de partida na estimação da regressão recursiva. O problema em questão aqui é testar as hipóteses

$$\begin{aligned} H_0: \Psi &= 0 \\ H_A: \Psi &\neq 0, \end{aligned} \quad (3.8)$$

no modelo proposto em (3.7).

A estatística de teste é dada por

$$C(d) = [n - h - m_0 - (v_p + v_q + 1)] \cdot \{\ln[\det(\mathbf{S}_0)] - \ln[\det(\mathbf{S}_1)]\}, \quad (3.9)$$

onde a defasagem, d , indica que o teste depende da variável limiar z_{t-d} , $\det(S_i)$ denota o determinante da i -ésima matriz S , dada por

$$\mathbf{S}_0 = \frac{1}{n-h-m_0} \sum_{l=m_0+1}^{n-h} \hat{\eta}_{t(l)+d} \hat{\eta}'_{t(l)+d},$$

e

$$\mathbf{S}_1 = \frac{1}{n-h-m_0} \sum_{l=m_0+1}^{n-h} \hat{w}_{t(l)+d} \hat{w}'_{t(l)+d},$$

com \hat{w}_t denotando o resíduo do modelo ajustado em (3.7).

Sob a hipótese nula de que \mathbf{y}_t é linear e sob algumas condições de regularidade, $C(d)$ é uma variável aleatória com distribuição assintótica Qui-quadrado com $\nu(\rho\nu + \nu q + 1)$ graus de liberdade. Para informações sobre a derivação deste teste, ver Tsay (1998).

3.3 Identificação do modelo

Identificar um modelo “*threshold*” multivariado, adequado para um determinado conjunto de dados, envolve a escolha de muitos parâmetros. Exceto para a identificação da variável limiar z_t , o problema mais difícil pode ser a especificação do número de regimes, ou seja, a identificação de k . A complexidade computacional e os dados podem restringir k a um valor pequeno, tal como 2 ou 3. Nesta parte, suporemos que z_t e k sejam dados. Quando k é desconhecido, supomos que ele seja 2 ou 3 e utilizamos algum critério estatístico para fazer a seleção. De maneira informal, pode-se dividir os dados em subgrupos de acordo com os “percentuais empíricos” de z_{t-d} e utilizar o teste estatístico (3.9) para detectar qualquer mudança no modelo dentro de cada subgrupo. Esta análise pode fornecer uma estimativa preliminar de k e algumas possíveis localizações dos limiares.

Quando z_t e k são dados, podemos utilizar, por exemplo, o AIC para selecionar o modelo, supondo que $0 \leq p \leq p_0$; $0 \leq q \leq q_0$; $1 \leq d \leq d_0$. Em alguns casos, utilizando os resultados do teste em (3.9), para diferentes defasagens, selecionamos de uma maneira mais refinada o valor de d , o que resulta em uma simplificação futura. Isto está baseado na idéia de que o teste é mais poderoso quando o valor de d é especificado corretamente. Dados p , q , d e k , o AIC de um modelo “*threshold*” multivariado em (3.1) é definido por

$$\text{AIC}(p, q, d, k) = \sum_{j=1}^k \{2 \cdot \ln[L_j(p, q, d, k)] + 2\nu(\nu p + \nu q + 1)\}, \quad (3.10)$$

em que, $L_j(p, q, d, k)$ é a função de verossimilhança do regime j utilizada para avaliar os estimadores de máxima verossimilhança de \mathbf{c}_j , $\phi_i^{(j)}$ e $\beta_i^{(j)}$. Se as inovações forem normais multivariadas, então o AIC se reduz a:

$$AIC(p, q, d, k) = \sum_{j=1}^k [n_j \cdot \ln(|\hat{\Sigma}_j|) + 2v(vp + vq + 1)], \quad (3.11)$$

com

$$\hat{\Sigma}_j = \frac{1}{n_j} \sum_t^* \hat{\varepsilon}_t^{(j)} \hat{\varepsilon}_t^{(j)'}, \quad (3.12)$$

em que,

n_j – é o número de observações do j -ésimo regime;

\sum_t^* - é a somatória das observações no j -ésimo regime, e

$\hat{\varepsilon}_t^{(j)}$ - são os resíduos do modelo ajustado no j -ésimo regime.

O AIC tem sido usado na literatura para selecionar modelos auto-regressivos com limiares; ver, por exemplo, Tong (1983 e 1990), Tsay (1989) e as referências lá contidas. Quando p , q e k são fixados, o AIC é assintoticamente equivalente a selecionar o modelo que tenha a menor variância residual generalizada utilizando o método dos mínimos quadrados condicionais.

3.4 Estimação do modelo

Nesta seção faremos considerações sobre o método de estimação do modelo proposto em (3.1). Para tal, iremos supor que p , q e k têm valores conhecidos e que a variável limiar, z_t , seja dada. Entretanto a defasagem, d , e os limiares, r_i , serão considerados desconhecidos, ou seja, deverão ser estimados a partir de algum critério. Tsay (1998), generaliza os resultados encontrados por Chan (1993) e Hansen (1996b) para o caso univariado do modelo (3.1). Para facilidade de apresentação, consideraremos o caso em que $k = 2$ e escreveremos o modelo na forma

$$\mathbf{y}_t = \begin{cases} \mathbf{X}_t' \Phi_1 + \sum_1^{1/2} \mathbf{a}_t, & \text{se } z_{t-d} \leq r_1 \\ \mathbf{X}_t' \Phi_2 + \sum_2^{1/2} \mathbf{a}_t, & \text{se } z_{t-d} > r_1 \end{cases}, \quad (3.13)$$

em que, $\mathbf{a}_t = (a_{1t}, \dots, a_{kt})'$.

Assumiremos que:

- i. z_{t-d} é estacionária, contínua com função densidade de probabilidade positiva, $f(r)$, em um subconjunto delimitado da linha real, digamos $R_0 \subset R$, e
- ii. $d \in \{1, \dots, d_0\}$, onde d_0 é um inteiro positivo fixado.

Os parâmetros do modelo (3.13) são $(\Phi_1, \Phi_2, \Sigma_1, \Sigma_2, r_1, d)$ e suas estimativas de mínimos quadrados condicionais podem ser obtidas em duas etapas:

Etapa 1: inicialmente, para d e r_1 fixados, o modelo (3.13) se reduz a duas regressões lineares multivariadas separadas, para as quais as estimativas de mínimos quadrados de Φ_i e Σ_i , $i = 1, 2$, são facilmente obtidas. Os estimadores são dados por

$$\hat{\Phi}_i(r_1, d) = \left(\sum_{t,(i)} \mathbf{X}_t \mathbf{X}_t' \right)^{-1} \left(\sum_{t,(i)} \mathbf{X}_t \mathbf{y}_t' \right), \quad (3.14)$$

e

$$\hat{\Sigma}_i(r_1, d) = \frac{\sum_{t,(i)} (y_t - X_t' \hat{\Phi}_i^*) (y_t - X_t' \hat{\Phi}_i^*)'}{n_i - k}, \quad (3.15)$$

em que, $\sum_{t,(i)}$ denota a somatória das observações no i -ésimo regime, $\hat{\Phi}_i^* = \hat{\Phi}_i(r_1, d)$, n_i é o tamanho da amostra do i -ésimo regime e v é a dimensão de X_t , satisfazendo à condição $v < n_i$, para $i = 1, 2$. Denote a soma de quadrados dos resíduos por

$$\mathbf{S}(r_1, d) = \mathbf{S}_1(r_1, d) + \mathbf{S}_2(r_1, d), \quad (3.16)$$

em que, $\mathbf{S}_i(r_1, d)$, denota o traço de $(n_i - k)\hat{\Sigma}_i(r_1, d)$.

Etapa 2: as estimativas de mínimos quadrados de r_1 e d podem ser obtidas por:

$$(\hat{r}_1, \hat{d}) = \underset{r_1, d}{\operatorname{argmin}} \mathbf{S}(r_1, d), \quad (3.17)$$

em que $1 \leq d \leq d_0$ e $r_1 \in R_0$.

Os estimadores de mínimos quadrados dos parâmetros das equações (3.14) e (3.15) são,

$$\hat{\Phi}_i = \hat{\Phi}_i(\hat{r}_1, \hat{d}), \quad (3.18)$$

e

$$\hat{\Sigma}_i = \hat{\Sigma}_i(\hat{r}_1, \hat{d}), \quad (3.19)$$

respectivamente.

Chan (1993) e Hansen (1996b) demonstram que as propriedades assintóticas dos estimadores por mínimos quadrados são válidas, quando utilizamos o modelo (3.13).

Capítulo 4

Aplicações

4.1 Introdução

Neste capítulo serão aplicadas as técnicas de modelagem propostas nos Capítulos 2 e 3 às séries epidemiológicas e de poluição ambiental, na cidade de São Paulo, medidas diariamente, no período de 18/04/1996 – 31/12/1997 (totalizando 623 observações), ver Apêndice A. Os programas necessários para testar não-linearidade *threshold* tanto na caso univariado quanto no caso multivariado foram desenvolvidos em linguagem S e podem ser encontrados no Apêndice B.

As séries utilizadas neste capítulo são descritas a seguir:

- (a) Número de óbitos por causas respiratórias para idosos maiores de 65 anos (Y_t) fornecido pela Companhia de Tecnologia de Saneamento Ambiental (CETESB);

- (b) Média da concentração do agente poluidor PM_{10} (X_t), material particulado, isto é, poeira composta de partículas menores do que 10 micra, calculada por meio das concentrações fornecidas por 13 estações da CETESB espalhadas pela cidade de São Paulo, este agente foi medido em $\mu g/m^3$;
- (c) Temperatura mínima, medida em $^{\circ}C$, (Z_t) fornecida pelas estações da CETESB situadas no Parque Dom Pedro e no Parque do Estado na Água Funda.

De acordo com a definição da CETESB para padrão primário de qualidade do ar: “...concentrações de poluentes que ultrapassadas poderão afetar a saúde da população, podendo ser entendidos como níveis máximo toleráveis ...”, utilizam-se os seguintes valores de padrões de qualidade do ar para o Estado de São Paulo (decreto estadual no. 8468 de 08/09/76):

Pm_{10} - $240 \mu g/m^3$
 SO_2 - $365 \mu g/m^3$
 CO - $40000 \mu g/m^3$ ou 35 ppm
 O_3 - $160 \mu g/m^3$

Segundo a orientação da CETESB, esses limites não devem ser excedidos mais de uma vez ao ano.

4.2 Modelos univariados com limiares

4.2.1 Análise da série número de óbitos por causas respiratórias para idosos maiores de 65 anos (Y_t)

Na Figura 4.1 apresentamos a série Y_t e podemos notar que a mesma apresenta um crescimento no inverno, ou seja, nos meses de junho a agosto e decai nos meses subsequentes. Sendo assim, notamos que sua média não é constante, logo, esta série pode não ser considerada estacionária de segunda ordem.

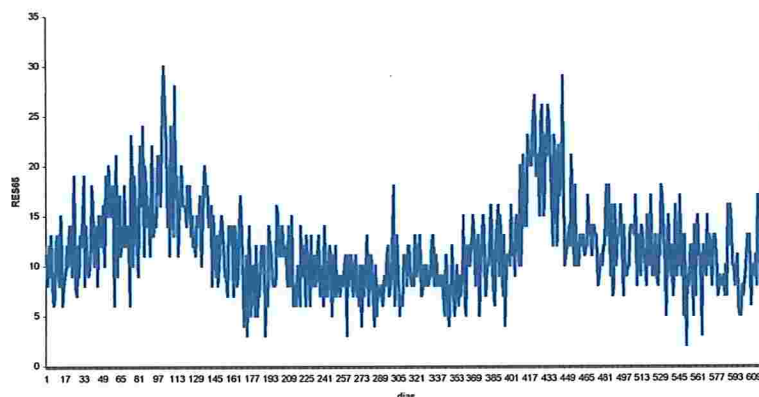


Figura 4.1 – Valores observados da série número de óbitos por causas respiratórias para idosos maiores de 65 anos.

ADF Test Statistic	-0.482494	1% Critical Value*	-2.5690
		5% Critical Value	-1.9399
		10% Critical Value	-1.6159

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(RES65)
 Method: Least Squares
 Date: 12/18/02 Time: 22:39
 Sample(adjusted): 9 623
 Included observations: 615 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES65(-1)	-0.005655	0.011720	-0.482494	0.6296
D(RES65(-1))	-0.774250	0.041675	-18.57829	0.0000
D(RES65(-2))	-0.619189	0.051121	-12.11221	0.0000
D(RES65(-3))	-0.443501	0.055657	-7.968415	0.0000
D(RES65(-4))	-0.411825	0.056183	-7.330062	0.0000
D(RES65(-5))	-0.304027	0.055743	-5.454105	0.0000
D(RES65(-6))	-0.244480	0.051011	-4.792724	0.0000
D(RES65(-7))	-0.122489	0.040635	-3.014399	0.0027
R-squared	0.390102	Mean dependent var		0.013008
Adjusted R-squared	0.383069	S.D. dependent var		4.556181
S.E. of regression	3.578652	Akaike info criterion		5.400772
Sum squared resid	7773.696	Schwarz criterion		5.458289
Log likelihood	-1652.737	F-statistic		55.46407
Durbin-Watson stat	2.008305	Prob(F-statistic)		0.000000

Quadro 4.1 – Saída do *software* Eviews para testar a presença de raiz unitária na série número de óbitos por causas respiratórias para idosos maiores de 65 anos.

Para verificarmos se a série estudada apresenta ou não raiz unitária realizamos um teste de Dickey – Fuller (Dickey e Fuller, 1979), cujos resultados encontram-se no Quadro 4.1 e, analisando os mesmos, podemos verificar que ao compararmos o valor da estatística de teste, $\hat{\tau} = -0,483$, com o valor crítico, a 5%, $\tau = -1,939$ verificamos que não existem evidências para rejeição da hipótese nula (a série Y_t não é estacionária), ou seja, a série em questão apresenta raiz unitária. Para solucionarmos o problema da ausência de estacionariedade foi aplicada à série uma diferença de primeira ordem.

A série transformada (ΔY_t) está apresentada na Figura 4.2 e podemos verificar que a transformação estabilizou a média da série. Lembramos que este cuidado se faz necessário pois para modelarmos uma série utilizando os procedimentos propostos por Tsay (1989), Tong (1983) e Box et al. (1994) as séries analisadas devem ser estacionárias.

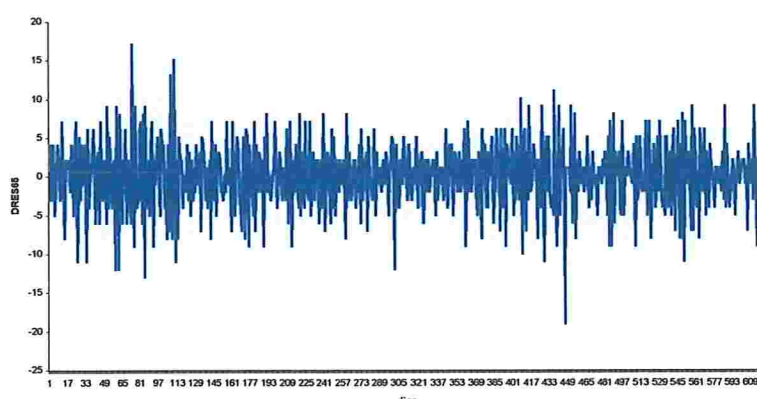


Figura 4.2 – Valores observados da série número de óbitos por causas respiratórias para idosos maiores de 65 anos, após tomarmos a primeira diferença.

Ao observarmos conjuntamente as Figuras 4.3 e 4.4, que contêm as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) amostrais das séries Y_t e ΔY_t , respectivamente, notamos que a transformação foi adequada e parece ter tornado a série estacionária.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.****	.****	1	0.500	0.500	156.64	0.000
.****	.**	2	0.476	0.301	298.60	0.000
.****	.**	3	0.480	0.239	443.15	0.000
.***	.*	4	0.412	0.088	550.16	0.000
.***	.*	5	0.441	0.147	672.74	0.000
.***	.*	6	0.415	0.083	781.27	0.000
.***	.*	7	0.433	0.123	899.94	0.000
.***	.*	8	0.438	0.103	1021.2	0.000
.***	.	9	0.398	0.033	1121.6	0.000
.***	.	10	0.405	0.051	1225.8	0.000
.***	.*	11	0.454	0.140	1356.8	0.000
.***	*	12	0.352	-0.060	1435.6	0.000
.***	.	13	0.372	0.013	1523.8	0.000
.***	.*	14	0.409	0.084	1630.8	0.000
.***	.	15	0.374	0.028	1720.4	0.000
.***	.	16	0.361	-0.013	1804.1	0.000
.***	.	17	0.333	-0.029	1875.3	0.000
.***	.	18	0.347	0.015	1952.6	0.000

Figura 4.3 – FAC e FACP da série número de óbitos por causas respiratórias para idosos maiores de 65 anos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.****	.****	1	-0.478	-0.478	142.58	0.000
.	.***	2	-0.029	-0.333	143.10	0.000
. *	*	3	0.073	-0.153	146.41	0.000
*	**	4	-0.095	-0.195	152.08	0.000
.	*	5	0.055	-0.120	153.98	0.000
.	*	6	-0.044	-0.151	155.23	0.000
.	*	7	0.015	-0.124	155.36	0.000
.	.	8	0.045	-0.049	156.63	0.000
.	*	9	-0.049	-0.067	158.17	0.000
.	*	10	-0.041	-0.155	159.26	0.000
. *	.	11	0.153	0.048	174.06	0.000
*	.	12	-0.122	-0.022	183.58	0.000
.	*	13	-0.017	-0.091	183.77	0.000
. *	.	14	0.072	-0.035	187.09	0.000
.	.	15	-0.023	0.006	187.42	0.000
.	.	16	0.017	0.024	187.60	0.000
.	.	17	-0.044	-0.021	188.83	0.000
.	.	18	0.025	-0.010	189.22	0.000

Figura 4.4 – FAC e FACP da primeira diferença da série número de óbitos por causas respiratórias para idosos maiores de 65 anos.

Verificando a FAC e a FACP da Figura 4.4 utilizaremos uma ordem auto-regressiva, inicialmente, igual a 7 para verificarmos a presença de não-linearidade *threshold* nesta série, segundo o procedimento proposto por Tsay (1989) e descrito na seção 2.3.1. Os resultados deste teste encontram-se na Tabela 4.1.

Tabela 4.1 - Teste para verificação de não-linearidade *threshold* na série ΔY_t .

d	F-estimada	nível descritivo
1	0,8002	0,6026
2	1,1576	0,3229
3	1,9284	0,0536
4	0,2605	0,9781
5	0,8757	0,5367
6	1,3784	0,2030
7	0,3758	0,9334

Pelos resultados da Tabela 4.1, constatamos que a série ΔY_t , segundo o procedimento proposto por Tsay, segue um modelo auto-regressivo com um único regime, uma vez que não foi detectada a presença de não-linearidade *threshold* ao nível de 5% de significância.

Verificaremos agora, segundo a metodologia proposta por Tong na seção 2.2.1 (c), se conseguiremos detectar a presença de não-linearidade *threshold* na série ΔY_t .

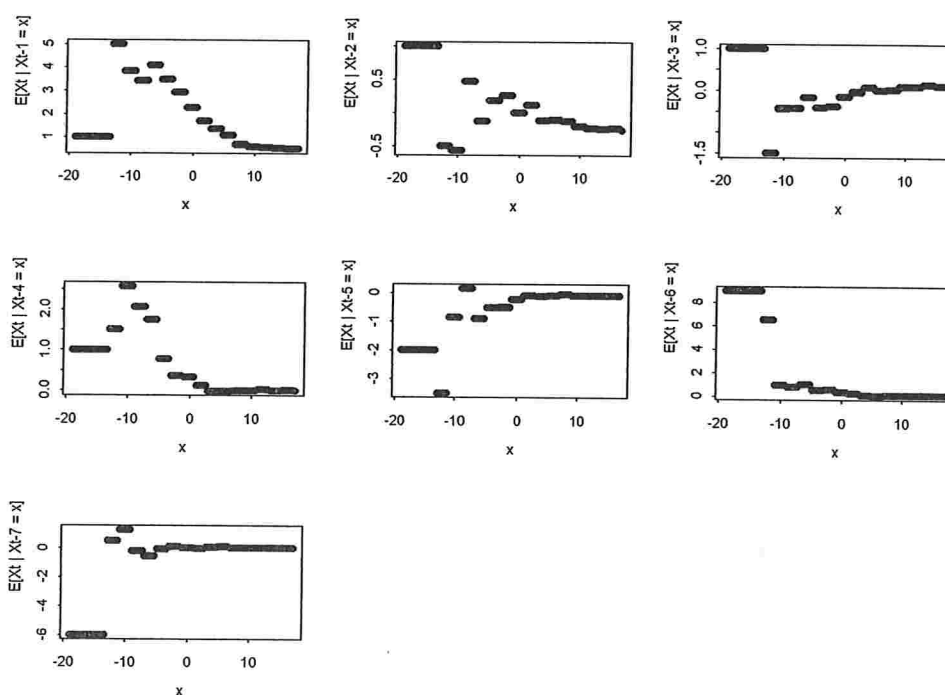


Figura 4.5 – Estimativas não-paramétricas, dadas por (2.8), da função de regressão de ΔY_t em ΔY_{tj} , $j = 1, \dots, 7$.

Tabela 4.2 – Estimativas do índice de linearidade de segunda ordem proposto por Tong, dadas por (2.7), para a série ΔY_t .

d	ξ
1	0,98
2	0,92
3	0,97
4	0,88
5	0,94
6	0,91
7	0,92

Ao observarmos a Figura 4.5, desconsiderando o início da recursão e calculando os índices de linearidade, dado por (2.7), dispostos na Tabela 4.2, não evidenciamos a presença de não-linearidade *threshold*, uma vez que todas as estimativas deste índice ficaram muito próximas de uma unidade, assim, um modelo linear também pode ser adotado para a série ΔY_t , segundo a metodologia proposta por Tong (1983).

Assim a série ΔY_t pode ser ajustada por um modelo linear, já que nenhuma das técnicas propostas pelos dois autores estudados foi capaz de evidenciar a necessidade de um modelo *threshold*. Conseqüentemente ajustaremos um modelo ARMA para esta série, seguindo a metodologia proposta em Box et al. (1994). O Quadro 4.2 apresenta os resultados da estimação do modelo para a série ΔY_t .

Dependent Variable: DRES65
Method: Least Squares
Date: 12/18/02 Time: 22:46
Sample(adjusted): 2 623
Included observations: 622 after adjusting endpoints
Convergence achieved after 5 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.826331	0.023124	-35.73536	0.0000
R-squared	0.392328	Mean dependent var		0.006431
Adjusted R-squared	0.392328	S.D. dependent var		4.544639
S.E. of regression	3.542698	Akaike info criterion		5.369260
Sum squared resid	7793.990	Schwarz criterion		5.376387
Log likelihood	-1668.840	Durbin-Watson stat		1.953694
Inverted MA Roots	.83			

Quadro 4.2 – Saída do *software* Eviews com a estimação do modelo para a série ΔY_t .

Analisando o Quadro 4.2, percebemos que a série ΔY_t pode ser ajustada por um modelo de médias móveis de primeira ordem.

O modelo linear ajustado para a série ΔY_t encontra-se descrito na equação a seguir:

$$\Delta Y_t = a_t - 0,826a_{t-1}. \quad (4.1)$$

Transformando a equação (4.1), de forma a obtermos o modelo para a série original (Y_t), teremos:

$$Y_t = Y_{t-1} + a_t - 0,826a_{t-1}. \quad (4.2)$$

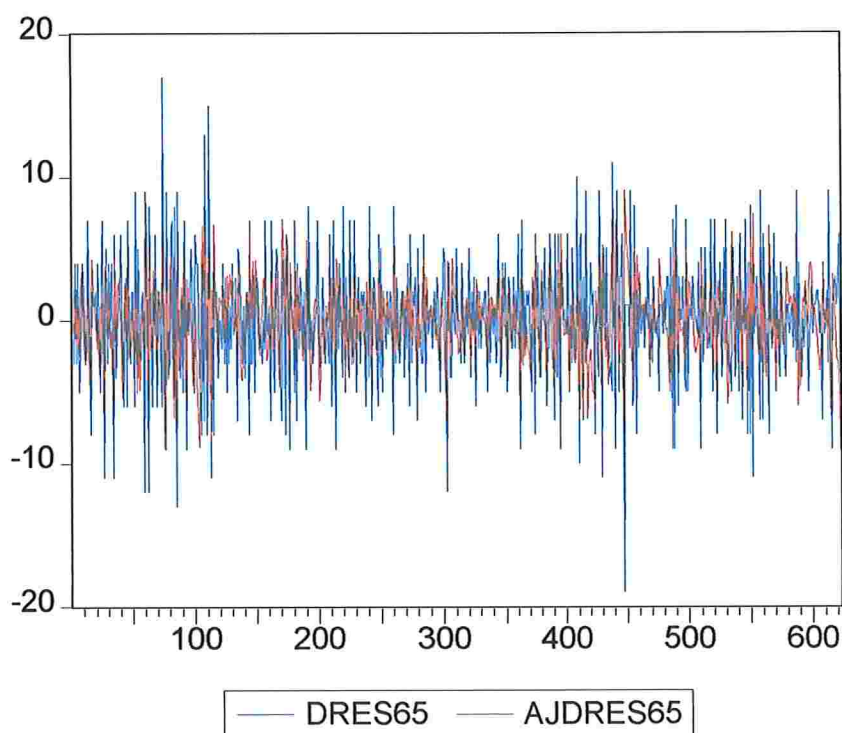


Figura 4.6 - Gráficos da diferença de primeira ordem da série Y_t e da série ajustada pelo modelo MA(1).

Ao analisarmos a equação (4.1), notamos que o presente da série Y_t depende do seu valor observado no dia anterior mais uma combinação linear de erros no presente com o dia anterior. Este modelo apresentou um AIC de 5,369 com uma variância estimada de 12,55. Na Figura 4.6 encontramos o gráfico da série observada e o da série ajustada pelo modelo proposto em (4.2), podemos notar que o modelo estimado conseguiu captar bem a evolução da série observada.

A Figura 4.7 mostra que o modelo adotado está bem ajustado, uma vez que não verificamos a presença de correlação serial entre os resíduos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	0.023	0.023	0.3178	
. .	. .	2	-0.002	-0.003	0.3213	0.571
. .	. .	3	0.031	0.031	0.9210	0.631
* .	* .	4	-0.081	-0.082	4.9956	0.172
. .	. .	5	-0.005	-0.001	5.0143	0.286
. .	. .	6	-0.040	-0.042	6.0177	0.305
. .	. .	7	0.012	0.020	6.1112	0.411
. .	. .	8	0.034	0.027	6.8609	0.444
. .	. .	9	-0.032	-0.031	7.4919	0.485
. .	. .	10	0.001	-0.005	7.4926	0.586
. *	. *	11	0.116	0.118	16.092	0.097
* .	* .	12	-0.070	-0.073	19.229	0.057
. .	. .	13	-0.015	-0.014	19.376	0.080
. *	. *	14	0.074	0.071	22.877	0.043
. .	. .	15	0.022	0.038	23.184	0.057
. .	. .	16	0.016	0.004	23.344	0.077
. .	. .	17	-0.024	-0.020	23.707	0.096
. .	. .	18	0.023	0.025	24.059	0.118

Figura 4.7 – FAC e FACP dos resíduos do ajuste do modelo (4.1).

4.2.2 Análise da série média da concentração do agente poluidor PM_{10} (X_t), material particulado

A série X_t é apresentada na Figura 4.8; podemos perceber que esta série pode ser não estacionária, uma vez que sua média e variância não se mostram constantes ao longo do tempo.

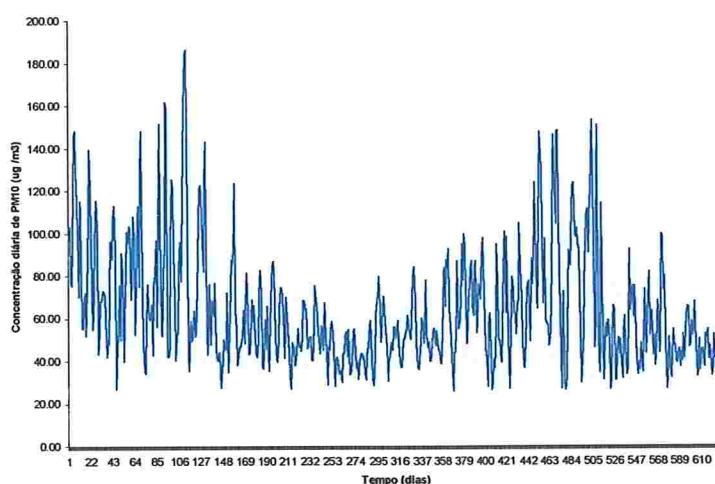


Figura 4.8 – Valores observados da série média da concentração do agente poluidor PM₁₀, material particulado.

ADF Test Statistic	-1.651465	1% Critical Value*	-2.5690
		5% Critical Value	-1.9400
		10% Critical Value	-1.6159

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(PM10)

Method: Least Squares

Included observations: 612 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PM10(-1)	-0.018786	0.011376	-1.651465	0.0992
D(PM10(-1))	-0.020402	0.041287	-0.494142	0.6214
D(PM10(-2))	-0.402218	0.040962	-9.819182	0.0000
D(PM10(-3))	-0.239917	0.043507	-5.514396	0.0000
D(PM10(-4))	-0.147946	0.044556	-3.320462	0.0010
D(PM10(-5))	-0.191050	0.044554	-4.288041	0.0000
D(PM10(-6))	-0.134629	0.044336	-3.036521	0.0025
D(PM10(-7))	-0.019860	0.044187	-0.449467	0.6533
D(PM10(-8))	-0.163349	0.042949	-3.803340	0.0002
D(PM10(-9))	-0.099503	0.040021	-2.486265	0.0132
D(PM10(-10))	-0.107583	0.040216	-2.675113	0.0077
R-squared	0.219843	Mean dependent var	-0.047761	
Adjusted R-squared	0.206862	S.D. dependent var	20.98003	
S.E. of regression	18.68445	Akaike info criterion	8.711071	
Sum squared resid	209814.4	Schwarz criterion	8.790457	
Log likelihood	-2654.588	F-statistic	16.93579	
Durbin-Watson stat	2.031585	Prob(F-statistic)	0.000000	

Quadro 4.3 – Saída do *software* Eviews para testar a presença de raiz unitária na série média da concentração do agente poluidor PM₁₀, material particulado.

Para verificarmos se a série estudada apresenta ou não raiz unitária realizamos um teste de Dickey – Fuller (Dickey e Fuller, 1979), cujos resultados encontram-se no Quadro 4.3 e, analisando os mesmos, podemos verificar que ao compararmos o valor da estatística de teste, $\hat{\tau} = -1,651$, com o valor crítico, a 5%, $\tau = -1,94$ observamos que não existem evidências para rejeição da hipótese nula (a série X_t não é estacionária), ou seja, a série em questão apresenta raiz unitária. Para solucionarmos o problema da ausência de estacionariedade foi aplicada à série uma diferença de primeira ordem. A série transformada (ΔX_t) está apresentada na Figura 4.9.

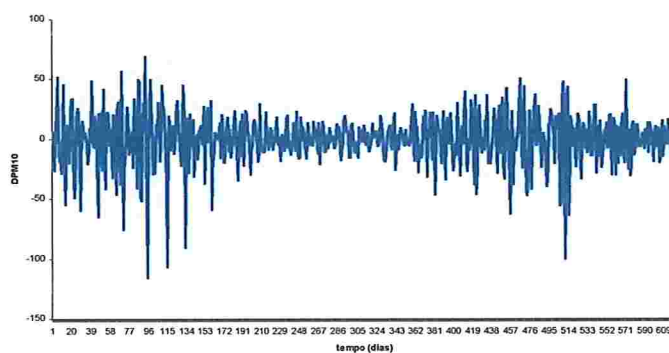


Figura 4.9 – Valores observados da diferença da série média da concentração do agente poluidor PM_{10} , material particulado.

Analisando conjuntamente as Figuras 4.10 e 4.11, que contêm as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) amostrais das séries X_t e ΔX_t respectivamente, notamos que a transformação parece ter sido adequada. Ainda, analisando a FAC e a FACP da Figura 4.11 utilizaremos uma ordem auto-regressiva, inicialmente, igual a 13 para verificarmos a presença de não-linearidade *threshold* nesta série, segundo o procedimento proposto por Tsay (1989) e descrito na seção 2.3.1. Os resultados deste teste encontram-se na Tabela 4.3.

Fazendo o teste para a verificação de não-linearidade *threshold* (Tsay, 1989), verificamos, segundo os resultados da Tabela 4.3, que para $d = 9$ chegamos a um nível descritivo de 0,00000482, o que nos leva a concluir que a série ΔX_t apresenta mais de um regime.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *****	*****	1	0.716	0.716	321.25	0.000
. ***	** .	2	0.396	-0.240	419.72	0.000
. **	. **	3	0.273	0.203	466.52	0.000
. **	. .	4	0.260	0.057	508.95	0.000
. **	. .	5	0.239	0.016	545.02	0.000
. **	. *	6	0.224	0.080	576.56	0.000
. *	. .	7	0.196	-0.014	600.86	0.000
. *	. *	8	0.106	-0.114	607.92	0.000
.	9	0.050	0.053	609.47	0.000
. * .	. .	10	0.072	0.047	612.77	0.000
. * .	. .	11	0.126	0.058	622.81	0.000
. ** .	. *	12	0.199	0.152	648.04	0.000
. ** .	. *	13	0.290	0.153	701.60	0.000
. *** .	. *	14	0.338	0.075	774.67	0.000
. ** .	. .	15	0.299	-0.003	831.87	0.000
. ** .	. .	16	0.244	0.016	869.95	0.000
. ** .	. *	17	0.250	0.068	910.16	0.000
. ** .	. *	18	0.298	0.081	967.38	0.000

Figura 4.10 – FAC e FACP da série média da concentração do agente poluidor PM₁₀, material particulado.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.	1	0.065	0.065	2.6361	0.104
. *** .	*** .	2	-0.347	-0.353	78.020	0.000
. ** .	* .	3	-0.195	-0.163	101.97	0.000
. . .	* .	4	0.011	-0.106	102.05	0.000
. . .	* .	5	-0.014	-0.162	102.17	0.000
.	6	0.020	-0.054	102.42	0.000
. . *	. .	7	0.115	0.043	110.80	0.000
. * .	* .	8	-0.060	-0.125	113.04	0.000
. * .	* .	9	-0.136	-0.102	124.70	0.000
. . .	* .	10	-0.055	-0.110	126.59	0.000
. . .	** .	11	-0.040	-0.197	127.61	0.000
. . .	* .	12	-0.030	-0.181	128.18	0.000
. . *	* .	13	0.082	-0.098	132.43	0.000
. . *	. .	14	0.155	-0.025	147.68	0.000
.	15	0.028	-0.036	148.19	0.000
. . *	* .	16	-0.111	-0.094	156.14	0.000
. . *	* .	17	-0.072	-0.098	159.43	0.000
. . *	. .	18	0.129	0.063	170.05	0.000

Figura 4.11 – FAC e FACP da primeira diferença da série média da concentração do agente poluidor PM₁₀, material particulado.

Tabela 4.3 - Teste para verificação de não-linearidade *threshold* na diferença de primeira ordem da série média da concentração do agente poluidor PM₁₀, material particulado.

d	F-estimada	nível descritivo
1	2,12394	0,00967
2	2,55486	0,00147
3	0,90871	0,54923
4	2,33312	0,00395
5	1,76663	0,04052
6	2,19166	0,00726
7	2,75132	0,00060
8	2,53502	0,00162
9	3,75179	4,82.10⁻⁶
10	2,17963	0,00765
11	1,87949	0,02617
12	2,37497	0,00329
13	1,76004	0,0415

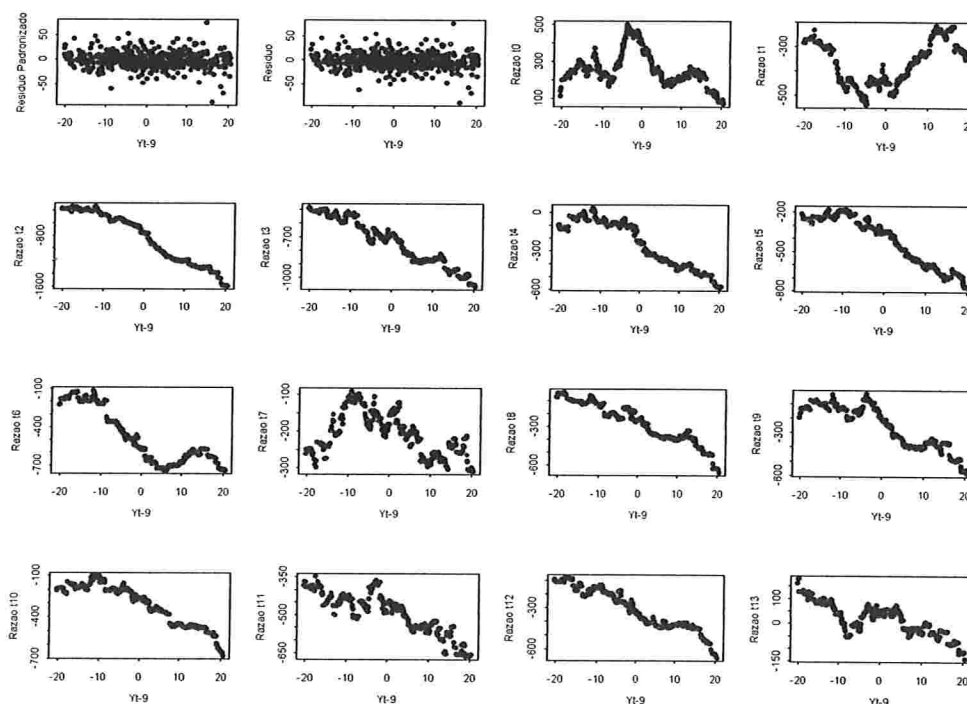


Figura 4.12 – Gráfico dos resíduos padronizados, resíduos e razões t dos coeficientes ajustados para a série ΔX_t , segundo a metodologia proposta por Tsay (1989).

Analisando os gráficos apresentados na Figura 4.12, concluímos que, de acordo com as razões t do sétimo coeficiente, por exemplo, os possíveis valores para o limiar são $r_1 = \{-8,90; -8,85; -8,80\}$, pois próximos a estes valores há uma mudança no comportamento do valor estimado.

Ao ajustarmos os modelos para cada um destes três possíveis valores de limiares, optamos pelo modelo SETAR (2; 7, 13; 9) com $r_1 = -8,85$ já que o mesmo apresentou o menor valor de AIC = 8,762 e uma variância estimada igual a 370,02.

O modelo final para a série ΔX_t é dado por

$$\Delta X_t = \begin{cases} 0,2748 \Delta X_{t-7} + a_t^{(1)}, & \Delta X_{t-9} \leq -8,85 \\ (0,0735) \\ -0,0851 \Delta X_{t-1} - 0,2221 \Delta X_{t-2} - 0,1469 \Delta X_{t-3} - 0,1002 \Delta X_{t-5} - \\ (0,0460) \quad (0,0447) \quad (0,0472) \quad (0,0456) \\ -0,1954 \Delta X_{t-6} + 0,1656 \Delta X_{t-10} + 0,1331 \Delta X_{t-13} + a_t^{(2)}, & -8,85 < \Delta X_{t-9}. \\ (0,0448) \quad (0,0444) \quad (0,0444) \end{cases} \quad (4.3)$$

- Observações:** (i) entre parênteses encontram-se os erros padrões associados a cada uma das estimativas dos parâmetros
(ii) no primeiro regime utilizamos 112 observações na estimação, enquanto que no segundo, 451.

Isolando a variável original, X_t , em (4.3) teremos

$$X_t = \begin{cases} X_{t-1} + 0,2748 X_{t-7} - 0,2748 X_{t-8} + a_t^{(1)}, & X_{t-9} - X_{t-10} \leq -8,85 \\ 0,9149 X_{t-1} - 0,1370 X_{t-2} + 0,0752 X_{t-3} + 0,1469 X_{t-4} - 0,1002 X_{t-5} - \\ -0,0952 X_{t-6} + 0,1954 X_{t-7} + 0,1656 X_{t-10} - 0,1656 X_{t-11} + 0,1331 X_{t-13} - \\ -0,1331 X_{t-14} + a_t^{(2)}, & X_{t-9} - X_{t-10} > -8,85. \end{cases} \quad (4.4)$$

Podemos observar que o modelo (4.4), para a série original, apresenta dois regimes sendo que o valor presente da série X_t dependerá de seus valores a 1, 7 e 8 dias passados se a diferença entre o nono e o décimo dia for inferior a $-8,85 \mu\text{g}/\text{m}^3$. Caso a diferença da série a nove e dez dias no passado seja superior $-8,85 \mu\text{g}/\text{m}^3$ o valor presente da série dependerá de seus valores passados a 1, 2, 3, 4, 5, 6, 7, 10, 11, 13 e 14 dias. Vale ressaltar que a maior influência no valor presente da série, em ambos os regimes, se dá no valor da série observada no dia anterior.

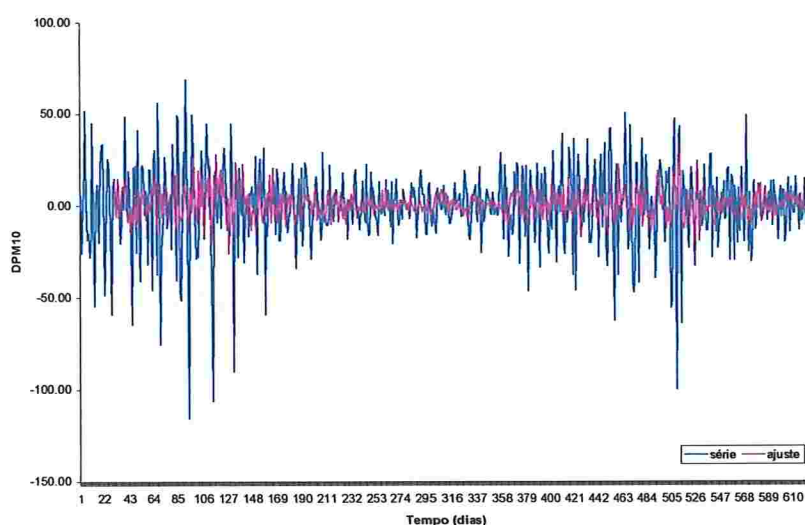


Figura 4.13 - Gráficos da diferença de primeira ordem da série média da concentração do agente poluidor PM_{10} , material particulado, e da série ajustada pelo modelo SETAR (2; 7, 13; 9).

A Figura 4.13 nos mostra a série original e a ajustada, segundo a metodologia proposta por Tsay (1989). Podemos notar que o modelo consegue acompanhar de maneira satisfatória a evolução da série temporal estudada.

Já a Figura 4.14 nos mostra a FAC e FACP dos resíduos do ajuste deste modelo, e podemos perceber pelas mesmas que não existe correlação serial entre os resíduos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	-0.037	-0.037	0.7978	0.372
. .	. .	2	-0.032	-0.033	1.3836	0.501
. .	. .	3	-0.028	-0.031	1.8513	0.604
. .	. .	4	-0.007	-0.010	1.8763	0.759
* .	* .	5	-0.094	-0.097	7.0515	0.217
. .	. .	6	-0.007	-0.016	7.0805	0.313
. . *	. . *	7	0.097	0.090	12.550	0.084
* .	* .	8	-0.071	-0.071	15.474	0.051
. .	. .	9	-0.028	-0.031	15.940	0.068
* .	* .	10	-0.062	-0.073	18.182	0.052
. .	. .	11	0.008	-0.005	18.216	0.077
* .	* .	12	-0.067	-0.058	20.823	0.053
. .	. .	13	0.033	0.012	21.497	0.090
. .	. .	14	0.006	-0.011	23.129	0.186

Figura 4.14 – FAC e FACP dos resíduos do ajuste do modelo (4.3).

Verificaremos agora, segundo a metodologia proposta por Tong na seção 2.2.1 (c), se conseguiremos detectar a presença de não-linearidade *threshold* na série ΔX_t .

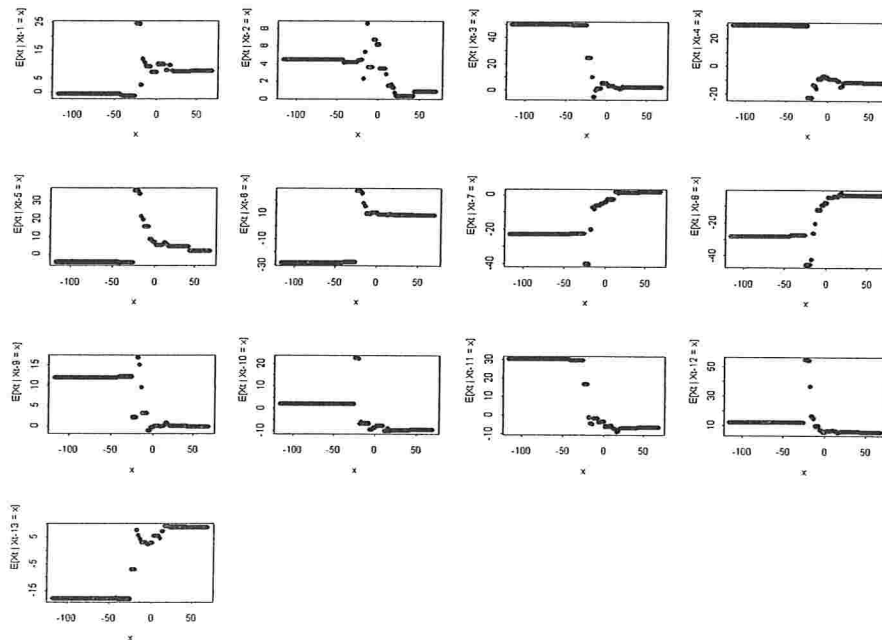


Figura 4.15 – Estimativas não-paramétricas, dadas por (2.8), da função de regressão de ΔX_t em ΔX_{t-j} , $j = 1, \dots, 13$.

Tabela 4.4 – Estimativas do índice de linearidade de segunda ordem proposto por Tong, dadas por (2.7), para a série ΔX_t .

d	$\hat{\epsilon}$
1	0,12721
2	0,16431
3	0,00157
4	$6,40 \cdot 10^{-6}$
5	0,00074
6	$2,58 \cdot 10^{-5}$
7	0,00225
8	0,00031
9	0,00726
10	0,00344
11	0,00165
12	0,00109
13	0,00115

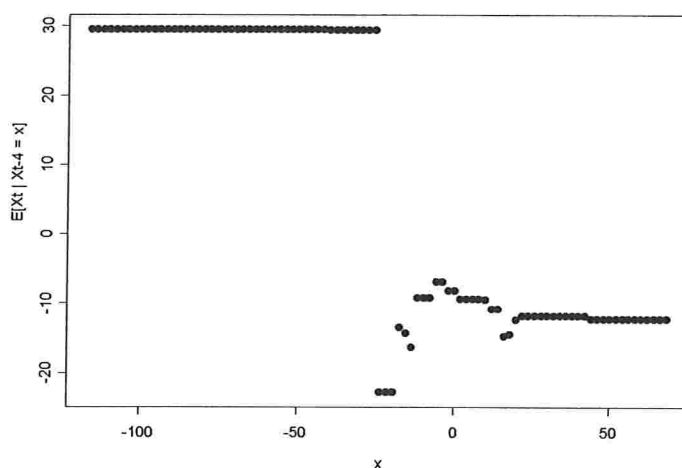


Figura 4.16 – Estimativas não-paramétricas, dadas por (2.8), da função de regressão de ΔX_t em ΔX_{t-4} .

Ao observarmos a Figura 4.15, percebemos quebras de comportamento em alguns gráficos de estimativas não paramétricas da função de regressão de ΔX_t . Calculando o índice de não-linearidade, dado por (2.7), para $j = 1, \dots, 13$, disposto na Tabela 4.4, notamos que para $d = 4$ o valor estimado deste índice foi de $0,00000064$, o

que evidencia a não-linearidade da série em estudo, ainda, verificamos que a série ΔX_t apresenta dois regimes e, observando, a seguir, detalhadamente a Figura 4.16 das estimativas da regressão não-paramétrica de ΔX_t em função de ΔX_{t-4} concluímos que os possíveis valores para o limiar são $r_1 = \{-16,88; -16,68; -16,54\}$.

Ao ajustarmos os modelos para cada um destes três possíveis valores de limiares, optamos pelo modelo SETAR (2; 7, 12; 4) com $r_1 = -16,68$ já que o mesmo apresentou o menor valor de AIC = 8,860 com uma variância estimada de 418,34.

O modelo final para a série ΔX_t é dado por

$$\Delta X_t = \begin{cases} 0,2304 \Delta X_{t-1} - 0,2773 \Delta X_{t-2} - 0,2508 \Delta X_{t-7} + a_t^{(1)}, & \Delta X_{t-4} \leq -16,68 \\ (0,1093) & (0,1065) & (0,1051) \\ -0,2640 \Delta X_{t-2} - 0,1376 \Delta X_{t-3} - 0,1554 \Delta X_{t-4} - \\ (0,0455) & (0,0450) & (0,0469) \\ -0,1333 \Delta X_{t-8} - 0,1272 \Delta X_{t-12} + a_t^{(2)}, & -16,68 < \Delta X_{t-4}. \end{cases} \quad (4.5)$$

Observações: (i) entre parênteses encontram-se os erros padrões associados a cada uma das estimativas dos parâmetros
(ii) no primeiro regime utilizamos 97 observações na estimação, enquanto que no segundo, 451.

Isolando a variável original, X_t , em (4.5) teremos

$$X_t = \begin{cases} 1,2304 X_{t-1} - 0,5077 X_{t-2} + 0,2773 X_{t-3} - \\ - 0,2508 X_{t-7} + 0,2508 X_{t-8} + a_t^{(1)}, & X_{t-4} - X_{t-5} \leq -16,68 \\ X_{t-1} - 0,2640 X_{t-2} + 0,11264 X_{t-3} - 0,0178 X_{t-4} + 0,1554 X_{t-5} - 0,1333 X_{t-8} + 0,1333 X_{t-9} - \\ - 0,1272 X_{t-12} + 0,1272 X_{t-13} + a_t^{(2)}, & X_{t-4} - X_{t-5} > -16,68. \end{cases} \quad (4.6)$$

Podemos observar que o modelo (4.6), para a série original, apresenta dois regimes sendo que o valor presente da série X_t dependerá de seus valores a 1, 2, 3, 7 e 8 dias passados se a diferença entre o quarto e o quinto dia for inferior a $-16,68 \mu\text{g}/\text{m}^3$. Caso a diferença da série a quatro e cinco dias no passado seja superior $-16,68 \mu\text{g}/\text{m}^3$ o valor presente da série dependerá de seus valores passados a 1, 2, 3, 4, 5, 8, 9, 12 e 13 dias. Vale ressaltar que a maior influência no valor presente da série, em ambos os regimes, se dá no valor da série observada no dia anterior.

A Figura 4.17 nos mostra a série original e a ajustada, segundo a metodologia proposta por Tong (1983). Podemos perceber que o modelo estimado consegue acompanhar bem a evolução da série observada.

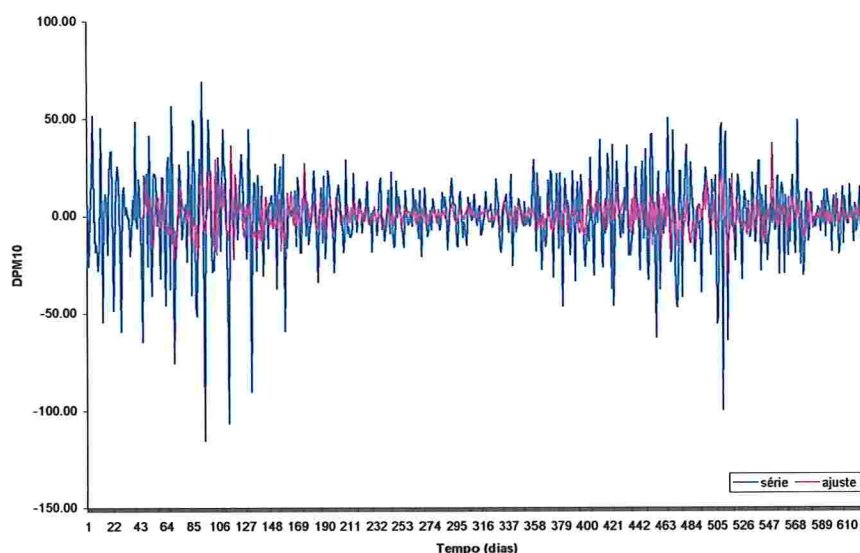


Figura 4.17 - Gráficos das séries ΔX_t e da ajustada pelo modelo SETAR (2; 7, 12; 4).

Já a Figura 4.18 nos mostra a FAC e FACP dos resíduos do ajuste deste modelo, podemos perceber pelas mesmas que não existe correlação serial entre os resíduos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
.	1 -0.055	-0.055	1.7046	0.192
.	2 -0.005	-0.008	1.7212	0.423
* . .	* . .	3 -0.090	-0.091	6.4058	0.093
.	4 -0.017	-0.027	6.5667	0.161
.	5 -0.018	-0.022	6.7463	0.240
* . .	* . .	6 -0.077	-0.089	10.122	0.120
.	7 0.024	0.010	10.456	0.164
* . .	* . .	8 -0.074	-0.080	13.600	0.093
.	9 0.017	-0.009	13.760	0.131
.	10 -0.005	-0.009	13.777	0.183
.	11 -0.010	-0.030	13.834	0.242
.	12 -0.011	-0.023	13.901	0.307
.	13 0.033	0.028	14.541	0.337
. * . .	. * . .	14 0.086	0.075	18.907	0.169
.	15 0.025	0.036	19.273	0.202
.	16 0.007	0.012	19.302	0.253
.	17 -0.034	-0.017	19.972	0.276
. * . .	. * . .	18 0.089	0.100	24.670	0.134

Figura 4.18 – FAC e FACP dos resíduos do ajuste do modelo (4.5).

Dependent Variable: DPM10
Method: Least Squares
Sample(adjusted): 20 623
Included observations: 604 after adjusting endpoints
Convergence achieved after 8 iterations
Backcast: 11 19

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DPM10(-8)	-0.097912	0.040717	-2.404680	0.0165
DPM10(-14)	0.129803	0.039771	3.263751	0.0012
DPM10(-18)	0.170778	0.039905	4.279575	0.0000
MA(1)	-0.099681	0.039184	-2.543931	0.0112
MA(2)	-0.456785	0.035882	-12.73004	0.0000
MA(3)	-0.271426	0.040159	-6.758763	0.0000
MA(9)	-0.107260	0.034102	-3.145232	0.0017
R-squared	0.290172	Mean dependent var	-0.054007	
Adjusted R-squared	0.283038	S.D. dependent var	20.87914	
S.E. of regression	17.67913	Akaike info criterion	8.594168	
Sum squared resid	186593.3	Schwarz criterion	8.645203	
Log likelihood	-2588.439	F-statistic	40.67471	
Durbin-Watson stat	2.018565	Prob(F-statistic)	0.000000	
Inverted MA Roots	.98	.62 -.40i	.62+.40i	.12+.71i
	.12 -.71i	-.43+.66i	-.43 -.66i	-.75 -.28i
	-.75+.28i			

Quadro 4.4 – Saída do *software* Eviews com a estimação do modelo para a série ΔX_t .

Finalmente, como critério de comparação, ajustamos um modelo ARMA (18, 9) incompleto à série ΔX_t , sendo que o AIC deste ajuste vale 8,594 e a variância estimada é igual a 312,55. O Quadro 4.4 apresenta os resultados da estimação do modelo para a série ΔX_t .

O modelo linear final ajustado para a série ΔX_t é dado por

$$\begin{aligned} \Delta X_t = & - 0,0979\Delta X_{t-8} - 0,1298\Delta X_{t-14} + 0,1708\Delta X_{t-18} - \\ & (0,0407) \quad (0,0397) \quad (0,0399) \\ & - a_t - 0,0997a_{t-1} - 0,4567a_{t-2} - 0,2714a_{t-3} - 0,1072a_{t-9} \end{aligned} \quad (4.7)$$

Observação: entre parênteses encontram-se os erros padrões associados a cada uma das estimativas dos parâmetros.

Isolando a variável original, X_t , em (4.7) teremos

$$\begin{aligned} X_t = & X_{t-1} - 0,0979X_{t-8} + 0,0979X_{t-9} - 0,1298X_{t-14} + 0,1298X_{t-15} + \\ & + 0,1708X_{t-18} - 0,1708X_{t-19} - a_t - 0,0997a_{t-1} - \\ & - 0,4567a_{t-2} - 0,2714a_{t-3} - 0,1072a_{t-9} \end{aligned} \quad (4.8)$$

Podemos observar para o modelo (4.8) que o valor presente da série X_t dependerá de seus valores a 1, 8, 9, 14, 15, 18 e 19 dias passados, além de uma combinação linear de erros no presente e passado a 1, 2, 3 e 9 dias.

A Figura 4.19 nos mostra a série original e a ajustada, segundo a metodologia proposta por Box et al. (1994). Podemos perceber que o modelo estimado consegue acompanhar bem a evolução da série observada.

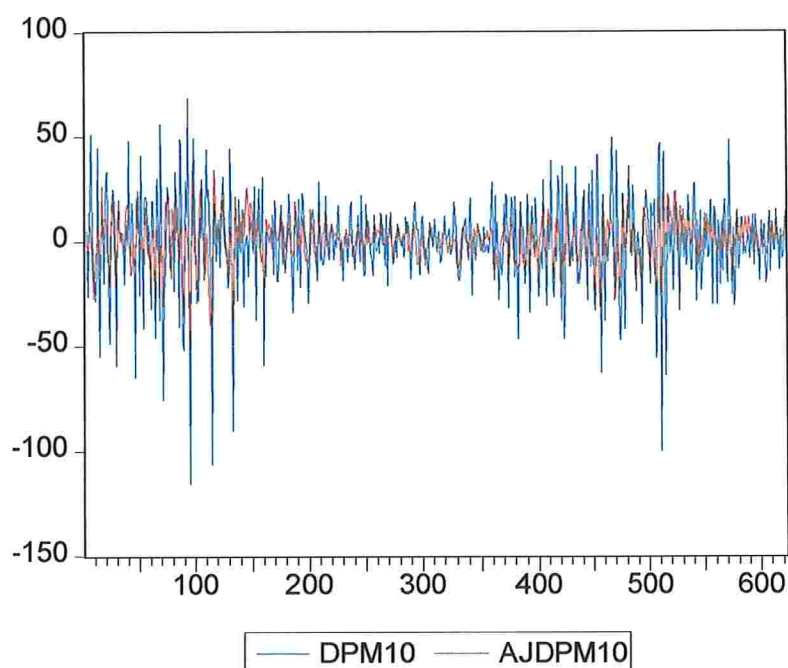


Figura 4.19 - Gráficos das séries ΔX_t e da ajustada pelo modelo ARMA (18, 9).

Já a Figura 4.20 nos mostra a FAC e FACP dos resíduos do ajuste deste modelo e podemos perceber pelas mesmas que não existe correlação serial entre os resíduos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
. .	. .	1 -0.012	-0.012	0.0884	
. .	. .	2 -0.030	-0.030	0.6256	
. .	. .	3 -0.021	-0.022	0.8882	
. .	. .	4 0.001	0.000	0.8888	
. .	. .	5 -0.012	-0.013	0.9723	0.324
. .	. .	6 -0.049	-0.050	2.4371	0.296
. *	. *	7 0.073	0.071	5.6791	0.128
. .	. .	8 -0.006	-0.008	5.7028	0.222
. .	. .	9 -0.049	-0.048	7.2034	0.206
* .	* .	10 -0.060	-0.059	9.4338	0.151
. .	. .	11 -0.026	-0.033	9.8669	0.196
. .	. .	12 0.027	0.020	10.321	0.243
. .	. *	13 0.063	0.067	12.742	0.175
. .	. .	14 -0.019	-0.025	12.976	0.225
. *	. *	15 0.070	0.070	16.052	0.139
. .	. .	16 0.016	0.020	16.211	0.182
. .	. .	17 -0.042	-0.035	17.334	0.185
. .	. .	18 0.002	0.010	17.336	0.239

Figura 4.20 – FAC e FACP dos resíduos do ajuste do modelo (4.7).

Apresentamos na Tabela 4.5 os valores do critério de Akaike e da variância residual para cada um dos modelos ajustados. Comparando os valores de AIC para cada um dos três modelos ajustados, não notamos uma diferença tão expressiva entre tais valores (variações percentuais inferiores a 3%). Já ao compararmos as variâncias percebemos que o modelo linear foi o que apresentou a menor variabilidade.

Tabela 4.5 – Valores do critério de Akaike e variância residual para cada um dos modelos ajustados.

Modelo	AIC	$\hat{\sigma}_a^2$
SETAR(2; 7, 13; 9)	8,762	370,02
SETAR(2; 7, 12; 4)	8,860	418,34
ARMA(18, 9)	8,594	312,55

4.2.3 Análise da série temperatura mínima (Z_t), medida em °C

A série Z_t está apresentada na Figura 4.21; podemos perceber que esta série parece ser não estacionária. Também verificamos uma aparente tendência crescente, que na realidade é parte de um ciclo com periodicidade de 365 dias.

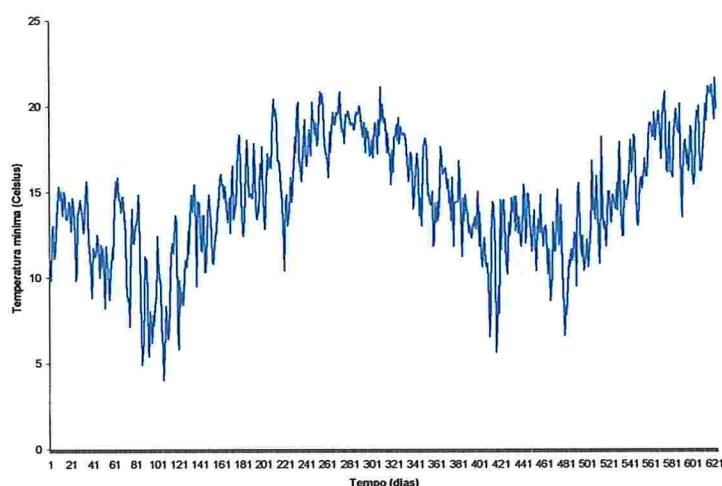


Figura 4.21 – Valores observados da série temperatura mínima, medida em (°C).

Para verificarmos se a série estudada apresenta ou não raiz unitária realizamos um teste de Dickey – Fuller (Dickey e Fuller, 1979), cujos resultados encontram-se no Quadro 4.5 e, analisando os mesmos, podemos verificar que ao compararmos o valor da estatística de teste, $\hat{\tau} = -0,166$, com o valor crítico, a 5%, $\tau = -1,939$ observamos que não existem evidências para rejeição da hipótese nula (a série Z_t não é estacionária), ou seja, a série em questão apresenta raiz unitária. Para solucionarmos o problema da ausência de estacionariedade foi aplicada à série uma diferença de primeira ordem. A série transformada (ΔZ_t) está apresentada na Figura 4.22.

ADF Test Statistic	-0.166211	1% Critical Value*	-2.5690
		5% Critical Value	-1.9399
		10% Critical Value	-1.6159

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(TEMP)

Method: Least Squares

Date: 12/19/02 Time: 10:31

Sample(adjusted): 8 623

Included observations: 616 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TEMP(-1)	-0.000708	0.004262	-0.166211	0.8680
D(TEMP(-1))	-0.194828	0.040615	-4.796987	0.0000
D(TEMP(-2))	-0.233029	0.041096	-5.670309	0.0000
D(TEMP(-3))	-0.281300	0.041700	-6.745856	0.0000
D(TEMP(-4))	-0.152487	0.041670	-3.659397	0.0003
D(TEMP(-5))	-0.116771	0.040923	-2.853439	0.0045
D(TEMP(-6))	-0.079837	0.040421	-1.975137	0.0487
R-squared	0.108828	Mean dependent var	0.011688	
Adjusted R-squared	0.100048	S.D. dependent var	1.651437	
S.E. of regression	1.566649	Akaike info criterion	3.747053	
Sum squared resid	1494.723	Schwarz criterion	3.797317	
Log likelihood	-1147.092	F-statistic	12.39502	
Durbin-Watson stat	2.003103	Prob(F-statistic)	0.000000	

Quadro 4.5 – Saída do *software* Eviews para testar a presença de raiz unitária na série Z_t .

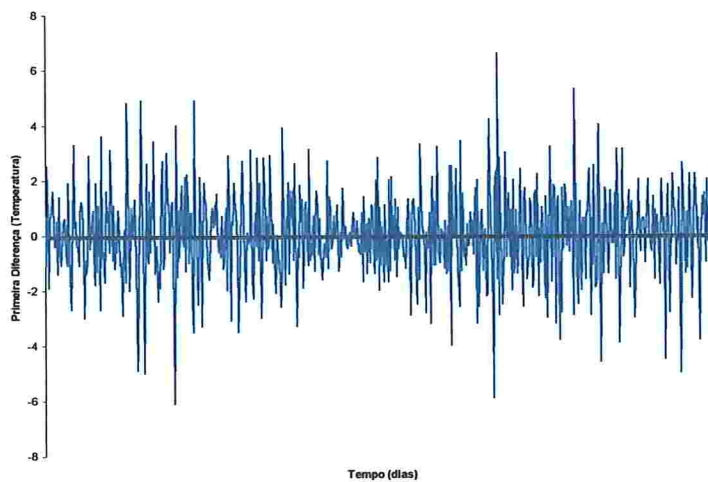


Figura 4.22 – Valores observados da diferença de primeira ordem da série temperatura mínima, medida em (°C).

Analisando conjuntamente as Figuras 4.23 e 4.24, que contêm as funções de autocorrelação (FAC) e autocorrelação parcial (FACP) amostral das séries Z_t e ΔZ_t , respectivamente, notamos que a transformação foi adequada e parece ter tornado a série estacionária.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*****	*****	1	0.882	0.882	486.47	0.000
*****	. .	2	0.785	0.033	872.42	0.000
*****	. *	3	0.724	0.116	1201.4	0.000
*****	. *	4	0.704	0.179	1513.4	0.000
*****	. *	5	0.691	0.079	1813.9	0.000
*****	. *	6	0.679	0.080	2104.5	0.000
*****	. .	7	0.664	0.055	2383.3	0.000
*****	. .	8	0.643	0.011	2644.7	0.000
*****	. .	9	0.626	0.045	2893.2	0.000
*****	. *	10	0.621	0.071	3138.4	0.000
*****	. .	11	0.620	0.046	3382.8	0.000
*****	. *	12	0.626	0.087	3632.5	0.000
*****	. *	13	0.646	0.132	3898.7	0.000
*****	. *	14	0.662	0.080	4178.7	0.000
*****	. .	15	0.652	-0.019	4451.3	0.000
*****	. .	16	0.640	0.036	4714.3	0.000
*****	. .	17	0.635	0.041	4973.5	0.000
*****	. .	18	0.628	-0.004	5227.5	0.000

Figura 4.23 – FAC e FACP da série Z_t .

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
* . .	* . .	1	-0.101	-0.101	6.3768	0.012
* . .	* . .	2	-0.144	-0.156	19.316	0.000
* . .	** . .	3	-0.180	-0.220	39.691	0.000
. . .	* . .	4	-0.032	-0.117	40.329	0.000
. . .	* . .	5	-0.006	-0.103	40.348	0.000
. . .	* . .	6	0.013	-0.080	40.458	0.000
.	7	0.033	-0.031	41.138	0.000
. . .	* . .	8	-0.019	-0.061	41.378	0.000
. . .	* . .	9	-0.046	-0.083	42.714	0.000
. . .	* . .	10	-0.017	-0.063	42.905	0.000
. . .	* . .	11	-0.029	-0.093	43.423	0.000
* . .	* . .	12	-0.058	-0.146	45.595	0.000
. . .	* . .	13	0.023	-0.084	45.924	0.000
. *	14	0.110	0.017	53.679	0.000
.	15	0.009	-0.041	53.732	0.000
. . .	* . .	16	-0.041	-0.060	54.834	0.000
.	17	0.009	0.000	54.891	0.000
.	18	0.018	0.000	55.102	0.000

Figura 4.24 – FAC e FACP da série ΔZ_t .

Ainda, analisando a FAC e a FACP da Figura 4.24 utilizaremos uma ordem auto-regressiva, inicialmente, igual a 13 para verificarmos a presença de não-linearidade *threshold* nesta série.

Os resultados do teste aplicado, de acordo com a metodologia proposta por Tsay (1989), para verificar a presença de não-linearidade *threshold* (ver seção 2.3.1), na série estudada, encontram-se na Tabela 4.6.

Observando os resultados dispostos na Tabela 4.6 verificamos que para $d = 1$ encontramos um nível descritivo de 0,0002, o que nos leva a concluir que a série ΔZ_t apresenta mais de um regime.

Analisando a Figura 4.25, concluímos que, de acordo com os gráficos dos resíduos e com os gráficos das razões t das estimativas dos coeficientes da regressão re-arranjada, a série ΔZ_t apresenta dois regimes e os possíveis valores para o limiar são $r_1 = \{-0,1; 0,0; 0,1\}$.

Tabela 4.6 - Teste para verificação de não-linearidade *threshold* na série ΔZ_t .

d	F-estimada	nível descritivo
1	2,5586	0,0002
2	0,7143	0,7607
3	1,6055	0,0734
4	1,9295	0,0214
5	1,5016	0,1056
6	0,9793	0,4732
7	0,8842	0,5762
8	2,0407	0,0136
9	1,7891	0,0372
10	1,2516	0,2338
11	1,2307	0,2485
12	2,3713	0,0033
13	1,2794	0,2154

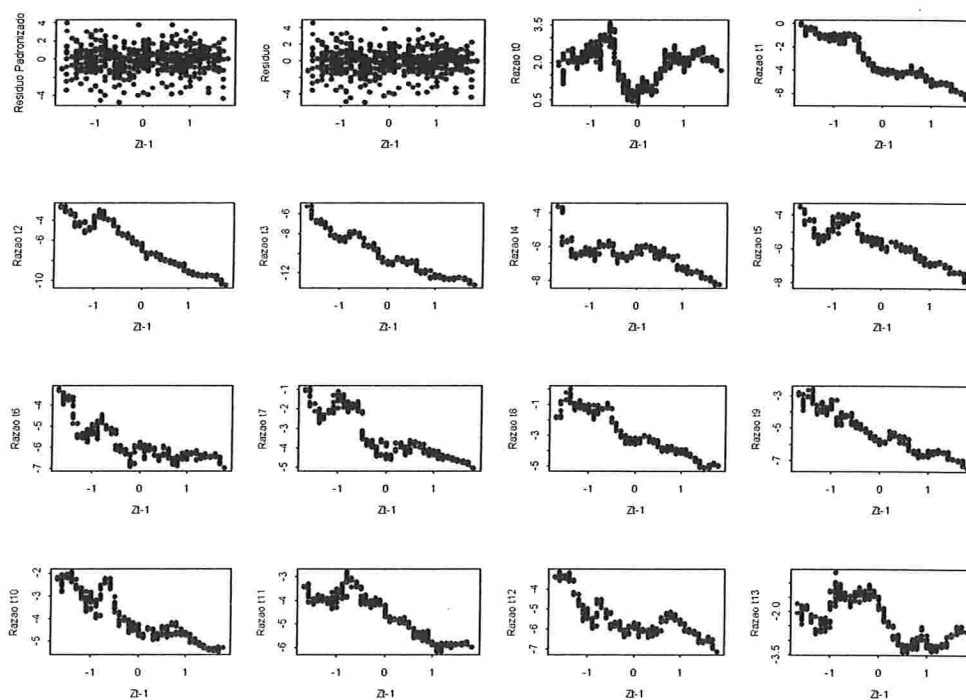


Figura 4.25 – Gráfico dos resíduos padronizados, resíduos e razões t dos coeficientes ajustados para a série ΔZ_t , segundo a metodologia proposta por Tsay (1989).

Trabalhando com a metodologia proposta por Tong (1983) chegamos às mesmas conclusões que aquelas tiradas usando a metodologia proposta por Tsay (1989), ou seja, a regressão não paramétrica nos leva a concluir que para $d = 1$ temos uma estimativa do índice de linearidade igual a 0,09649, veja Tabela 4.7, o que nos leva a concluir que a série ΔZ_t é não-linear e, ao observarmos a Figura 4.26, concluímos que a série apresenta dois regimes e os possíveis valores para o limiar são $r_1 = \{-0,1; 0,0; 0,1\}$.

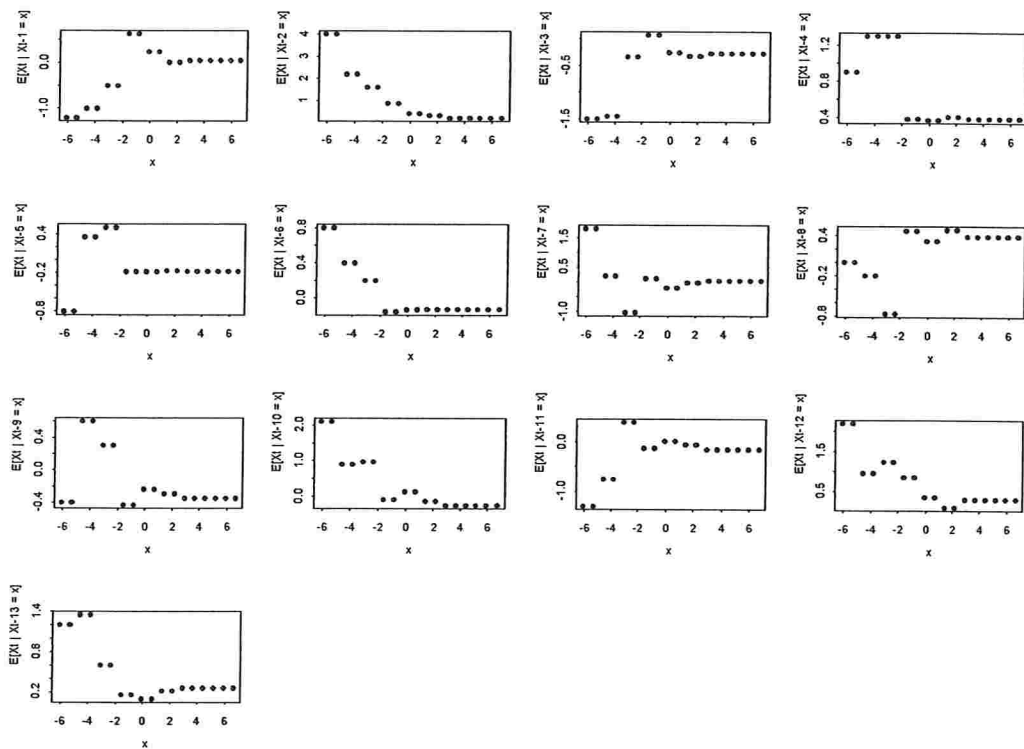


Figura 4.26 – Estimativas não-paramétricas, dadas por (2.8), da função de regressão de ΔZ_t em ΔZ_{t-j} , $j = 1, \dots, 13$.

Tabela 4.7 – Estimativas do índice de linearidade de segunda ordem proposto por Tong, dadas por (2.7), para a série ΔZ_t .

d	£
1	0,09649
2	0,30299
3	0,17789
4	0,10940
5	0,22410
6	0,68954
7	0,64818
8	0,23410
9	0,11229
10	0,26357
11	0,99714
12	0,22360
13	0,37995

Ao ajustarmos os modelos para cada um destes três possíveis valores de limiares, optamos pelo modelo SETAR (2; 3, 2; 1) com $r_1 = -0,1$ já que o mesmo apresentou o menor valor de AIC = 4,29 com uma variância estimada de 2,73.

O modelo final para a série ΔZ_t é dado por

$$\Delta Z_t = \begin{cases} -1,300 + 0,1269 \Delta Z_{t-3} + a_t^{(1)} & , \quad \Delta Z_{t-1} \leq -0,1 \\ (-18,46) \quad (2,20) & \\ 1,2350 - 0,08603 \Delta Z_{t-2} + a_t^{(2)} & , \quad -0,1 < \Delta Z_{t-1} \\ (22,91) \quad (-1,54) & \end{cases} \quad (4.9)$$

Observações: (i) entre parênteses encontram-se os erros padrões associados a cada uma das estimativas dos parâmetros
(ii) no primeiro regime utilizamos 265 observações na estimação, enquanto que no segundo, 282.

Isolando a variável original, Z_t , em (4.9) teremos

$$Z_t = \begin{cases} -1,3000 + Z_{t-1} + 0,1269 Z_{t-3} - 0,1269 Z_{t-4} + a_t^{(1)} & , Z_{t-1} - Z_{t-2} \leq -0,1 \\ 1,2350 + Z_{t-1} - 0,0860 Z_{t-2} + 0,0860 Z_{t-3} + a_t^{(2)} & , Z_{t-1} - Z_{t-2} > -0,1 \end{cases} \quad (4.10)$$

Podemos observar que o modelo (4.10), para a série original, apresenta dois regimes, sendo que o valor presente da série Z_t dependerá de uma constante mais seus valores a 1, 3, e 4 dias passados se a diferença entre o primeiro e o segundo dia for inferior a $-0,1^\circ\text{C}$. Caso a diferença da série a um e dois dias no passado seja superior a $-0,1^\circ\text{C}$ o valor presente da série dependerá de uma constante e de seus valores passados a 1, 2 e 3 dias.

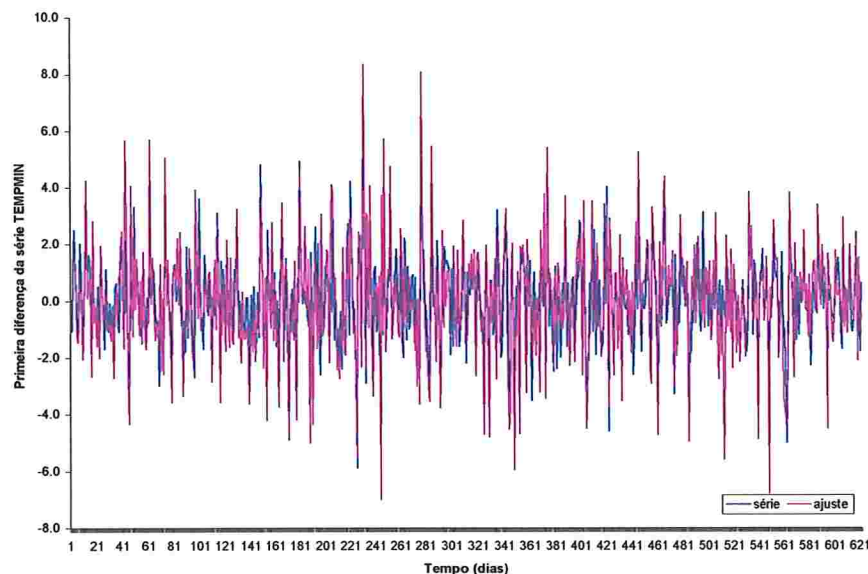


Figura 4.27 - Gráficos das séries ΔZ_t e da ajustada pelo modelo SETAR (2; 3, 2; 1).

A Figura 4.27 nos mostra a série original e a ajustada, segundo as metodologias propostas por Tong (1983) e Tsay (1989), nela percebemos o alto grau de ajuste do modelo proposto, uma vez que ambas as séries estão muito próximas.

Já a Figura 4.28 nos mostra a FAC e FACP dos resíduos do ajuste deste modelo, e podemos perceber pelas mesmas que não existe correlação serial entre os resíduos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.	1	-0.032	-0.032	0.4597	0.498
.	2	-0.021	-0.022	0.6589	0.719
.	3	-0.009	-0.010	0.6925	0.875
.	4	-0.040	-0.042	1.4187	0.841
.	5	-0.012	-0.015	1.4812	0.915
.	6	-0.009	-0.012	1.5172	0.958
.	7	-0.027	-0.030	1.8527	0.968
.	8	-0.016	-0.020	1.9654	0.982
.	9	-0.039	-0.043	2.6506	0.977
.	10	0.002	-0.003	2.6527	0.988
.	11	0.057	0.053	4.1396	0.966
.	12	0.018	0.019	4.2930	0.978
.	13	0.025	0.025	4.5792	0.983
. * .*	. * .*	14	0.074	0.076	7.0371	0.933
.	15	0.037	0.047	7.6483	0.937
.	16	-0.006	0.002	7.6671	0.958
.	17	0.038	0.045	8.3332	0.959
*	18	-0.069	-0.057	10.534	0.913

Figura 4.28 – FAC e FACP dos resíduos do ajuste do modelo (4.9).

Finalmente, como critério de comparação, ajustamos um modelo MA(3) à série ΔZ_t , sendo que o AIC deste ajuste vale 3,70 com uma variância estimada de 2,36. O Quadro 4.6 apresenta os resultados da estimação do modelo para a série ΔZ_t .

O modelo linear final ajustado para a série ΔZ_t é dado por

$$\Delta Z_t = a_t - 0,2414a_{t-1} - 0,2526a_{t-2} - 0,2508a_{t-3} \quad (4.11)$$

(0,0388) (0,0389) (0,0390)

Observação: entre parênteses encontram-se os erros padrões associados a cada uma das estimativas dos parâmetros.

Isolando a variável original, Z_t , em (4.11) teremos

$$Z_t = Z_{t-1} + a_t - 0,2414a_{t-1} - 0,2526a_{t-2} - 0,2508a_{t-3} \quad (4.12)$$

Dependent Variable: DTEMP
 Method: Least Squares
 Date: 12/19/02 Time: 11:13
 Sample(adjusted): 2 623
 Included observations: 622 after adjusting endpoints
 Convergence achieved after 5 iterations
 Backcast: -1 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.241465	0.038881	-6.210425	0.0000
MA(2)	-0.252565	0.038900	-6.492726	0.0000
MA(3)	-0.250874	0.039042	-6.425763	0.0000
R-squared	0.135487	Mean dependent var		0.014309
Adjusted R-squared	0.132694	S.D. dependent var		1.649644
S.E. of regression	1.536301	Akaike info criterion		3.701444
Sum squared resid	1460.977	Schwarz criterion		3.722824
Log likelihood	-1148.149	F-statistic		48.50518
Durbin-Watson stat	1.977094	Prob(F-statistic)		0.000000
Inverted MA Roots	.87	-.31+.44i	-.31 -.44i	

Quadro 4.6 – Saída do *software* Eviews com a estimação do modelo para a série ΔZ_t .

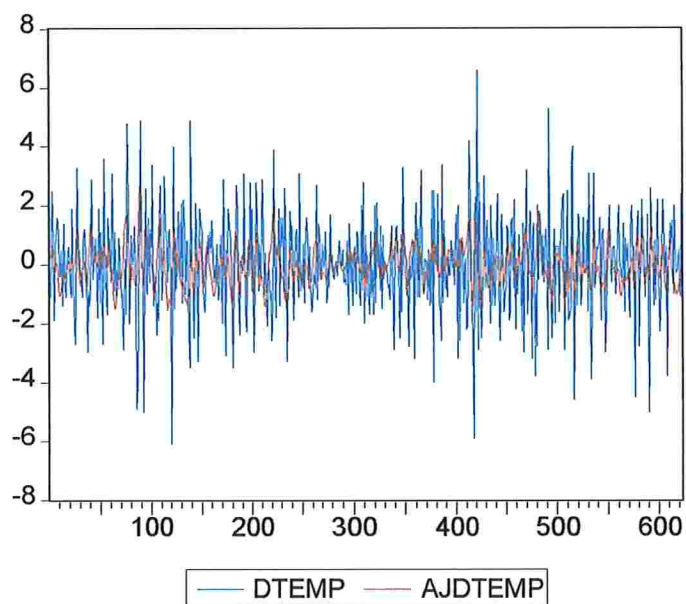


Figura 4.29 - Gráficos das séries ΔZ_t e da ajustada pelo modelo MA(3).

Podemos observar para o modelo (4.12) que o valor presente da série Z_t dependerá de seu valor a 1 dia passado, além de uma combinação linear de erros no presente e passado a 1, 2 e 3 dias.

A Figura 4.29 nos mostra a série original e a ajustada, segundo a metodologia proposta por Box et al. (1994). Podemos perceber que o modelo estimado consegue acompanhar razoavelmente bem a evolução da série observada.

Já a Figura 4.30 nos mostra a FAC e FACP dos resíduos do ajuste deste modelo, e podemos perceber pelas mesmas que não existe correlação serial entre os resíduos.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	0.009	0.009	0.0502	
. .	. .	2	0.023	0.023	0.3767	
. .	. .	3	0.034	0.033	1.0910	
. .	. .	4	-0.023	-0.025	1.4359	0.231
. .	. .	5	-0.007	-0.008	1.4690	0.480
. .	. .	6	-0.007	-0.007	1.4966	0.683
. .	. .	7	0.002	0.004	1.4984	0.827
. .	. .	8	-0.041	-0.041	2.5693	0.766
* .	* .	9	-0.071	-0.070	5.7451	0.452
. .	. .	10	-0.039	-0.036	6.6864	0.462
. .	. .	11	-0.031	-0.025	7.2977	0.505
. .	. .	12	-0.056	-0.052	9.3129	0.409
. .	. .	13	0.020	0.020	9.5578	0.480
. *	. *	14	0.108	0.110	16.953	0.109
. .	. .	15	0.029	0.030	17.494	0.132
. .	. .	16	0.001	-0.009	17.496	0.178
. .	. .	17	0.050	0.037	19.068	0.162
. .	. .	18	0.047	0.044	20.495	0.154

Figura 4.30 – FAC e FACP dos resíduos do ajuste do modelo (4.11).

Apresentamos na Tabela 4.8 os valores do critério de Akaike e da variância residual para cada um dos modelos ajustados.

Tabela 4.8 – Valores do critério de Akaike e variância residual para cada um dos modelos ajustados.

Modelo	AIC	$\hat{\sigma}_a^2$
SETAR(2; 3, 2; 1)	3,29	1,73
MA(3)	3,70	2,36

Comparando os valores de AIC e de variância residual dos dois modelos ajustados, verificamos que o modelo não-linear apresentou resultados um pouco superiores aos resultados obtidos utilizando uma modelagem linear. A qualidade dos ajustes também pode ser verificada ao compararmos as Figuras 4.27 e 4.29, neste caso notamos que o melhor ajuste realmente parece ter sido o do modelo não-linear.

4.3 Modelo multivariado com limiares

Serão utilizadas na aplicação desta seção as três séries trabalhadas anteriormente, sendo que Y_t será utilizada como variável resposta, X_t como variável explicativa e ΔZ_t como variável *threshold*. Vale ressaltar que a única série que será transformada é Z_t , pois, de acordo com a metodologia proposta por Tsay (1998), apenas a variável limiar necessita ser estacionária e, como foi visto anteriormente, no nosso caso a série de temperatura utilizada não é.

Dando continuidade à análise verificaremos se a hipótese nula em (3.2) deve, ou não, ser refutada, já que à partir da rejeição da mesma é que poderemos pensar em ajustar um modelo *threshold* para a série resposta. Para tal utilizamos a metodologia descrita na seção 3.2. Os valores iniciais propostos para as defasagens autorregressivas da variável resposta e explicativa serão, respectivamente, 8 e 4, obtidos a partir da análise da função de autocorrelação parcial das séries Y_t e X_t , que encontram-se nas Figuras 4.3 e 4.10, respectivamente.

De acordo com os resultados observados na Tabela 4.9, não existem evidências para rejeitarmos a hipótese nula em (3.2), logo, não há a necessidade de ajustarmos um modelo *threshold* para a série Y_t , com variável explicativa X_t e variável limiar ΔZ_t , ou seja, um modelo linear múltiplo deve explicar de maneira razoável a variável resposta. O Quadro 4.7 apresenta os resultados da estimação do modelo para a série Y_t .

Tabela 4.9 – Resultados do teste para verificar a necessidade de um modelo com limiares, com $n = 614$, $h = 8$ e $m = 124$ ($m = 5\sqrt{n}$).

d	1	2	3	4
C(d)	12,65	12,88	13,91	11,34
nível descritivo	0,47	0,44	0,38	0,58

Dependent Variable: RES65
Method: Least Squares
Date: 12/19/02 Time: 12:30
Sample(adjusted): 9 623
Included observations: 615 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.537045	0.554668	2.771108	0.0058
RES65(-1)	0.204831	0.039524	5.182404	0.0000
RES65(-2)	0.134918	0.040847	3.302990	0.0010
RES65(-3)	0.166504	0.040426	4.118785	0.0000
RES65(-5)	0.101627	0.040631	2.501241	0.0126
RES65(-7)	0.114132	0.039718	2.873568	0.0042
RES65(-8)	0.103770	0.040400	2.568555	0.0105
PM10	0.009206	0.005499	1.674049	0.0946
R-squared	0.401804	Mean dependent var	11.93008	
Adjusted R-squared	0.394905	S.D. dependent var	4.555484	
S.E. of regression	3.543613	Akaike info criterion	5.381093	
Sum squared resid	7622.215	Schwarz criterion	5.438610	
Log likelihood	-1646.686	F-statistic	58.24527	
Durbin-Watson stat	2.016441	Prob(F-statistic)	0.000000	

Quadro 4.7 – Saída do *software* Eviews com a estimação do modelo para a série Y_t .

O modelo linear final ajustado para a série Y_t é dado por

$$\begin{aligned}
 Y_t = & 1,838 + 0,209Y_{t-1} + 0,141Y_{t-2} + 0,170Y_{t-3} + \\
 & (0,525) \quad (0,039) \quad (0,041) \quad (0,040) \\
 & + 0,103Y_{t-5} + 0,115Y_{t-7} + 0,110Y_{t-8} + 0,009X_t + a_t \quad (4.13) \\
 & (0,041) \quad (0,039) \quad (0,040) \quad (0,005)
 \end{aligned}$$

Observação: entre parênteses encontram-se os erros padrões associados a cada uma das estimativas dos parâmetros.

Podemos observar para o modelo (4.13) que o valor presente da série Y_t dependerá de uma constante e de seus valores passados a 1, 2, 3, 5, 7 e 8 dias e verificamos, também, que a série X_t foi importante para explicar a variável resposta apenas no presente.

A Figura 4.31 nos mostra a série original e a ajustada, segundo a metodologia proposta por Box et al. (1994). Podemos perceber que o modelo estimado consegue acompanhar razoavelmente bem a evolução da série observada.

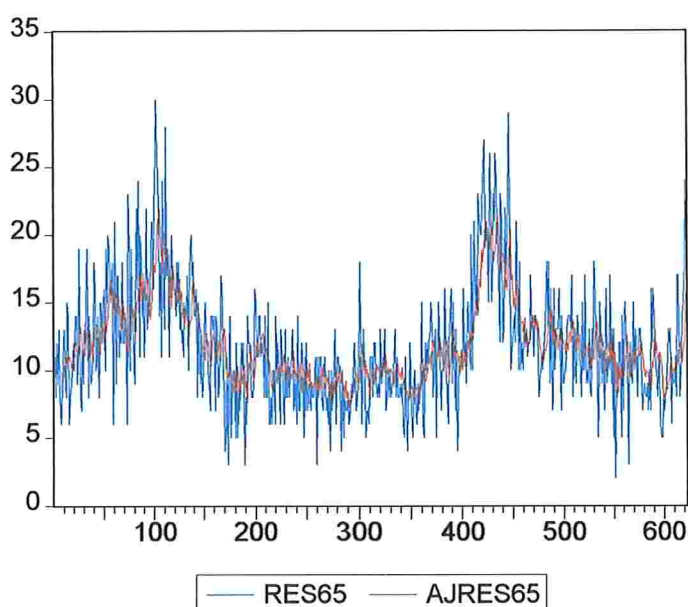


Figura 4.31 - Gráficos da série Y_t e ajustada pelo modelo múltiplo.

Já a Figura 4.32 nos mostra a FAC e FACP dos resíduos do ajuste deste modelo, e podemos perceber pelas mesmas que não existe correlação serial entre os resíduos.

Ainda, apresentamos na Tabela 4.10 os valores do critério de Akaike e da variância residual para o modelo (4.1) e (4.13). Podemos notar que a variância residual dos dois modelos foi a mesma, enquanto que o AIC de ambos ficaram bem próximos (variação percentual inferior a 1%). Logo, o modelo (4.1) parece ser mais adequado uma vez que é mais simples de ser estimado.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	-0.008	-0.008	0.0445	0.833
. .	. .	2	-0.007	-0.007	0.0764	0.963
. .	. .	3	-0.031	-0.031	0.6605	0.882
. .	. .	4	-0.010	-0.011	0.7258	0.948
. .	. .	5	-0.017	-0.017	0.8989	0.970
. .	. .	6	0.013	0.012	1.0104	0.985
. .	. .	7	-0.022	-0.022	1.3027	0.988
. .	* .	8	-0.057	-0.058	3.3027	0.914
. .	. .	9	-0.017	-0.018	3.4823	0.942
. .	. .	10	-0.002	-0.005	3.4857	0.968
. *	. *	11	0.127	0.124	13.626	0.254
* .	* .	12	-0.060	-0.062	15.868	0.197
. .	. .	13	-0.004	-0.006	15.880	0.256
. .	. .	14	0.040	0.039	22.469	0.096

Figura 4.32 – FAC e FACP dos resíduos do ajuste do modelo (4.13).

Tabela 4.10 – Valores do critério de Akaike e variância residual para os modelos ajustados (4.1) e (4.13).

Modelo	AIC	$\hat{\sigma}_a^2$
(4.1)	5,369	12,55
(4.13)	5,381	12,55

4.4 Considerações Finais

Esta dissertação objetivou descrever e aplicar alguns procedimentos para ajustar modelos lineares por regimes, nos casos univariado e multivariado, a algumas séries epidemiológicas e de poluição ambiental, na cidade de São Paulo. A seguir descreveremos algumas das vantagens e desvantagens de cada um dos procedimentos apresentados nos Capítulos 2 e 3 e aplicados às séries de número de óbitos por causas respiratórias para idosos maiores de 65 anos, média da concentração do agente poluidor PM_{10} e temperatura mínima, com os resultados discutidos no Capítulo 4:

- (a) O modelo SETAR não tem sido muito utilizado em aplicações, principalmente devido à dificuldade de identificação da variável limiar, da estimação dos limiares a ela

- associados e, também, devido a não existência de um procedimento simples de modelagem bem como de um *software* específico para tais aplicações;
- (b) Tanto a modelagem proposta em Tong (1983) quanto a descrita em Tsay (1989) apresentam como desvantagem a subjetividade dos métodos aplicados já que a estimação do limiar e do número de regimes são feitas a partir de métodos gráficos;
 - (c) A vantagem do método proposto por Tsay (1989), em relação ao utilizado por Tong (1983), está no fato de mesmo detectar o valor da defasagem via um teste de não-linearidade *threshold*, o que torna a análise um pouco menos subjetiva;
 - (d) Observando os resultados descritos no Capítulo 4, percebemos que a qualidade do ajuste dos modelos utilizando a metodologia linear ficou muito próxima à não-linear, entretanto, no caso do ajuste da série de temperatura mínima, o modelo linear por regimes se mostrou mais adequado (vide Figuras 4.27, 4.29 e Tabela 4.5);
 - (e) Apesar da qualidade dos ajustes lineares e não-lineares estarem muito próximas, a vantagem da metodologia não-linear está no fato de detectar e quantificar as mudanças de regime podendo dar às mesmas interpretações práticas importantes que, no caso linear, são impossíveis de ser encontradas;
 - (f) Quanto ao método multivariado, não foi possível ajustar um modelo linear por regimes para a série de número de óbitos, com o poluente PM_{10} como variável exógena, uma vez que o teste de não-linearidade *threshold* não detectou a existência de mais de um regime neste caso. Futuramente aplicaremos esta metodologia à análise de dados financeiros, nas quais os limiares, r_i , são funções dos custos de transação, taxas de interesses ou riscos econômicos e, neste caso, o volume de informações é bem maior que nos casos de séries epidemiológicas e de poluição ambiental;
 - (g) Finalmente, vale ressaltar que todos os programas utilizados para a identificação preliminar dos modelos estudados no Capítulo 4 foram escritos em linguagem de programação S e que, ainda, não existem, ainda, técnicas inferenciais disponíveis para detectar o número de regimes e o valor dos limiares a eles associados.

Referências Bibliográficas

- [1] BALKE, N. S. and FOMBY, T. B. (1997). Threshold cointegration. *International Economic Review*, **38**, 627-645.
- [2] BILLINGSLEY, P. (1961). The Lindenberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society*, 788-792.
- [3] BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994). *Time Series Analysis, Forecasting and Control*. 3rd edition. Prentice Hall: New Jersey Holden-Day: San Francisco (Revised edition, 1976).
- [4] BRILLINGER, D. R. (1966). An extremal property of the conditional expectation. *Biometrika*, **53**, 594-595.
- [5] CHAN, K. S. and TONG, H. (1990). On likelihood ratio tests for threshold autoregression, *Journal of the Royal Statistical Society, B*, **52**, 469-476.
- [6] CHAN, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, **21**, 520-533.
- [7] DICKEY, D.A. and FULLER, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root *Journal of the American Statistical Association*, **74**, 427-431.

- [8] ERTHEL, J. E. and FOWLKES, E. B. (1976). Some algorithms for linear spline and piecewise multiple linear regression. *Journal of the American Statistical Association*, **71**, 640-648.
- [9] GHADDAR, D. K. and TONG, H. (1981). Data transformation and self-exciting threshold autoregression. *J. Roy. Statist. Soc., C*, **30**, 238-248.
- [10] GOODWIN, G. C, and PAYNE, R. L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press: New York.
- [11] HANSEN, B. E. (1996a). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, **64**, 413-430.
- [12] HANSEN, B. E. (1996b). *Sample Splitting and Threshold Estimation*. Boston College: Dept. of Economics.
- [13] JENKINS, G. M. and WATTS, D. G. (1968). *Spectral Analysis and its Application*. Holden-Day: San Francisco.
- [14] KEENAN, D. M. (1985). A Tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, **72**, 39-44.
- [15] LIN, C. C. and MUDHOEKAR, G. S. (1980). A simple test for normality against asymmetric alternatives. *Biometrika*, **67**, 455-461.
- [16] LJUNG, G. M. and BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297-303.
- [17] MORETTIN, P.A. e TOLOI, C.M.C. (1987). *Previsão de Séries Temporais*. 2ª edição. Atual Editora: São Paulo.
- [18] PARZEN, E. (1958). On asymptotically efficient consistent estimates of the spectral density function of stationary time series. *J. Roy. Statist. Soc., B*, **20**, 303-322.
- [19] PETRUCCELLI, J. and DAVIES, N. (1986). A Portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika*, **73**, 687-694.
- [20] PRIESTLEY, M. B. (1981). *Spectral analysis and Time Series, Vols I and II*. Academic Press: London.
- [21] PRIESTLEY, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis*. Academic Press: London.
- [22] ROSENBLATT, M. (1969). Conditional probability density and regression estimators. *Multivariate Analysis*, **2**, 25-31.

- [23] SÁFADI, T. (1997). Análise Bayesiana de Alguns Modelos de Séries Temporais. Tese de Doutorado, *Instituto de Matemática e Estatística - USP*, São Paulo.
- [24] TONG, H. (1977). Discussion of a paper by A. J. Lawrence and N.T. Kottegodac. *J. Roy. Statist. Soc., A*, **140**, 34-35.
- [25] TONG, H. and LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc., B*, **42**, 245-292.
- [26] TONG, H. (1982). An Index of Non-Linearity in Time Series Analysis, Tech. Rep. No. 155, *Depts. of Maths (Stats.)*, UMIST.
- [27] TONG, H. (1983). *Threshold Models in Non-linear Time Series Analysis (Lecture Notes in Statistics)*, Vol 21. Springer-Verlag: New York.
- [28] TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford, U.K.: Oxford University Press.
- [29] TSAY, R. S. (1986). Nonlinearity tests for time series. *Biometrika*, **73**, 461-466.
- [30] TSAY, R. S. (1989). Testing and modeling threshold autoregressive processes. *J. of the American Statistical Association*, **84**, 231-240.
- [31] TSAY, R. S. (1998). Testing and modeling multivariate threshold models. *J. of the American Statistical Association*, **93**, 1188-1202.
- [32] VENTURA, A. M. (2000). Alguns Testes de Linearidade para Séries Temporais nos Domínios do Tempo e da Frequência. Dissertação de Mestrado, *Instituto de Matemática e Estatística - USP*, São Paulo.
- [33] WATSON, G. S. (1964). Smooth regression analysis. *Sankya, series A*, **26**, 359-372.
- [34] YULE, G. U. (1927). On method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philos. Trans. Roy. Soc., London, series A*, **226**, 267-298.

APÊNDICE A

DADOS

Tabela A.1 – Dados sobre o número de óbitos diários, na cidade de São Paulo, por causas respiratórias para idosos maiores de 65 anos, no período de 18/04/1996 a 31/12/1997. (*)

11	16	15	14	7	11	14	11
8	10	22	18	14	11	7	10
12	11	16	15	13	14	7	7
9	14	24	18	8	14	10	4
13	8	11	13	9	11	12	10
8	15	20	15	14	12	11	5
6	15	18	12	17	11	5	8
7	12	14	13	15	8	11	7
11	14	16	11	11	14	7	7
13	16	11	15	4	8	12	8
10	10	15	14	6	8	7	6
8	19	22	14	11	15	9	7
15	15	13	17	3	6	9	9
14	20	15	10	9	6	7	8
6	19	14	15	14	8	9	11
8	15	16	19	5	6	8	12
9	16	21	20	9	10	8	7
10	18	21	17	10	10	10	8
12	6	16	18	10	6	11	9
10	15	22	14	5	14	3	14
14	21	26	14	12	9	11	18
14	9	30	16	5	12	8	6
9	17	26	8	9	10	11	9
12	17	22	15	7	6	11	13
19	11	14	12	10	13	8	9
8	13	18	9	12	11	7	5
7	12	11	13	10	9	9	7
12	18	24	8	12	6	11	7
9	12	16	10	3	13	7	6
13	12	13	13	8	8	8	11
12	14	28	15	6	8	6	11
15	12	17	13	14	11	10	11
19	6	19	12	12	8	4	8
8	23	11	9	12	10	10	12
14	19	16	9	9	11	7	12
10	10	18	7	8	13	9	9
9	19	20	14	8	7	10	8
10	15	16	14	9	8	13	9
12	12	16	13	16	10	6	8
18	9	16	14	15	6	11	13

(*) ler por colunas

Tabela A.1 – Dados sobre o número de óbitos diários, na cidade de São Paulo, por causas respiratórias para idosos maiores de 65 anos, no período de 18/04/1996 a 31/12/1997. (continuação)

9	6	16	13	12	17	15	8
11	5	11	22	15	9	11	9
10	12	12	17	18	10	11	11
13	10	9	20	13	9	3	13
7	12	14	23	18	13	9	12
8	10	15	29	9	9	12	13
10	12	12	10	16	8	10	6
8	13	10	11	7	13	9	10
10	15	20	12	8	11	15	9
8	13	10	13	16	18	10	9
9	8	15	14	10	17	13	11
9	11	21	12	13	13	12	8
11	13	14	21	13	10	8	17
13	5	17	15	13	5	12	17
8	11	14	10	16	10	13	8
11	11	23	18	12	15	13	9
8	15	21	10	7	11	11	11
8	11	20	12	14	9	7	14
9	7	20	10	9	12	8	15
8	10	24	13	9	7	9	18
9	10	25	12	10	10	8	24
6	11	27	12	10	16	8	21
5	16	19	13	13	9	9	15
11	10	21	11	14	11	7	
7	8	19	12	14	10	10	
4	6	15	12	14	17	7	
8	12	24	17	13	9	16	
8	14	26	13	17	13	12	
12	16	15	14	8	5	16	
7	9	18	11	13	13	14	
5	15	23	14	10	2	10	
8	11	21	14	9	9	10	
10	7	26	13	14	8	8	
8	13	24	13	13	9	10	
6	4	20	8	11	12	11	
9	8	17	9	10	12	6	
7	11	12	10	8	5	5	
9	11	23	11	15	14	5	
9	11	21	10	10	7	8	
15	10	12	12	10	13	7	

(*) ler por colunas

Tabela A.2 – Dados diários sobre a média da concentração do agente poluidor PM₁₀, material particulado, na cidade de São Paulo, no período de 18/04/1996 a 31/12/1997. (*)

97,70	94,45	42,77	63,55	54,57	39,31	47,06	40,88
102,94	87,85	75,99	51,55	37,89	49,67	43,41	42,23
76,63	106,53	80,41	53,92	42,69	65,12	56,35	34,80
75,33	112,70	96,60	74,03	46,64	74,16	54,30	30,97
94,65	91,83	55,94	105,36	46,37	74,42	44,80	42,83
145,95	26,76	105,03	121,28	49,81	71,06	67,07	46,90
148,17	47,34	151,42	122,39	50,99	52,21	59,43	56,24
129,41	53,93	104,93	117,84	63,78	41,30	43,27	58,77
114,21	75,42	53,18	95,94	48,08	69,95	28,62	40,90
85,78	49,58	51,89	81,81	67,79	60,73	46,39	31,63
70,05	90,66	81,05	126,32	81,08	54,76	45,12	28,42
115,03	80,88	93,02	143,06	62,15	43,81	54,68	42,49
110,31	39,64	161,67	52,64	43,01	33,09	58,38	61,36
55,53	60,02	158,25	43,16	44,36	26,76	52,89	68,11
54,99	81,72	42,58	60,24	51,02	48,54	44,25	73,78
66,15	100,61	41,89	75,74	68,74	46,36	27,88	79,42
71,90	95,53	46,36	47,60	58,30	46,10	40,68	64,55
51,91	103,32	95,72	68,10	65,76	37,51	41,57	48,67
73,44	101,06	125,23	66,47	50,93	40,95	36,84	58,02
106,20	68,93	120,81	61,08	44,37	49,36	33,63	70,04
139,41	88,46	91,79	76,48	41,52	55,11	35,32	69,54
120,63	107,60	68,66	45,49	49,18	46,36	29,74	58,01
71,82	98,40	40,42	40,02	72,06	47,46	43,34	54,69
54,83	52,30	52,08	40,77	82,58	44,63	43,79	45,95
71,13	76,22	53,99	44,05	71,66	50,75	51,86	30,48
96,27	106,38	83,86	27,39	37,53	68,23	53,46	39,60
115,17	112,78	95,64	33,85	36,18	67,83	41,77	41,73
102,80	74,97	77,50	39,42	44,86	63,10	54,35	48,33
43,39	131,10	93,54	49,81	59,21	64,27	33,30	45,82
53,43	147,97	137,87	45,94	45,30	45,70	35,84	44,98
68,01	72,35	158,79	45,44	65,59	48,18	39,79	55,55
68,39	48,17	183,88	71,94	43,61	51,19	53,63	55,66
72,90	35,33	186,25	34,68	35,14	50,99	54,67	50,61
72,33	34,04	149,88	44,85	58,20	41,00	43,89	53,90
70,70	60,19	43,33	62,01	77,88	40,14	36,53	58,55
50,23	75,85	35,18	87,34	84,66	56,18	39,02	48,40
41,71	63,60	53,65	84,90	86,51	75,32	31,41	39,66
49,81	59,37	58,68	91,98	81,40	68,19	39,82	36,59
47,85	60,59	49,37	123,27	52,10	66,19	43,21	37,51
95,93	66,48	52,74	64,08	43,76	53,11	43,24	49,67

(*) ler por colunas

Tabela A.2 – Dados diários sobre a média da concentração do agente poluidor PM₁₀, material particulado, na cidade de São Paulo, no período de 18/04/1996 a 31/12/1997. (continuação)

50,58	65,32	46,01	73,09	55,69	49,92	51,60	58,28
52,92	87,10	40,83	76,18	91,81	26,59	61,69	51,66
55,34	83,19	27,58	77,47	86,09	31,23	63,13	59,87
61,20	92,51	57,06	49,07	84,43	51,68	42,50	67,47
55,06	64,57	62,07	60,57	111,23	65,75	51,66	47,41
51,70	55,98	56,92	88,02	121,73	63,89	37,53	32,59
49,88	50,36	26,07	75,88	123,31	30,82	47,74	35,63
68,64	34,51	30,14	90,04	99,50	31,64	67,96	49,99
78,97	25,52	40,35	123,75	97,83	43,71	63,31	35,43
84,01	43,59	36,70	102,71	102,02	50,60	43,65	44,78
67,74	43,61	55,65	69,53	95,11	42,96	50,57	45,59
48,78	66,26	94,48	64,48	92,10	50,02	99,13	41,56
37,09	86,66	76,79	105,81	52,44	38,52	98,23	37,00
35,88	54,76	50,12	147,41	29,45	31,63	72,94	52,24
40,90	56,45	49,44	141,66	40,44	53,43	79,68	50,73
49,06	59,22	40,00	129,64	52,88	60,81	48,87	54,31
59,91	80,28	39,09	66,74	77,11	47,04	26,88	40,71
57,02	94,72	70,49	74,96	95,22	33,32	37,92	46,51
48,17	77,62	93,60	97,18	107,66	36,52	50,93	38,00
56,92	99,10	100,32	59,06	110,96	64,08	37,67	32,87
77,74	94,35	62,30	57,23	90,64	91,94	32,05	37,89
51,94	47,82	98,21	56,50	108,55	63,48	43,32	52,08
46,00	49,20	51,82	46,96	120,76	73,38	54,37	40,82
48,74	67,93	26,65	51,33	132,79	74,87	47,94	
39,83	75,64	39,22	53,97	152,98	60,32	46,67	
39,62	82,64	66,67	82,36	97,19	74,83	40,36	
48,09	86,51	78,97	132,24	45,95	51,41	38,90	
52,97	66,24	73,34	145,89	59,14	39,62	42,87	
54,86	61,25	61,89	121,79	103,65	33,38	45,47	
52,24	83,83	61,95	103,95	150,49	39,52	37,08	
46,97	86,79	52,48	147,26	50,48	35,56	43,87	
53,79	52,83	66,56	148,15	34,41	48,20	39,69	
48,84	58,89	68,87	106,49	70,96	40,56	52,26	
45,58	75,81	104,53	59,16	113,72	34,54	40,95	
44,42	72,28	84,06	26,88	49,66	53,41	53,83	
38,73	68,51	68,88	49,60	31,08	73,36	61,69	
38,48	89,02	48,42	72,23	49,36	43,22	65,24	
52,69	97,53	37,82	30,16	57,11	53,00	64,92	
81,09	83,65	36,31	26,22	50,92	69,32	46,22	
83,68	57,40	48,63	32,48	58,82	81,59	48,57	

(*) ler por colunas

Tabela A.3 – Dados diários sobre a temperatura mínima da cidade de São Paulo, medida em °C, no período de 18/04/1996 a 31/12/1997. (*)

10,9	11,7	13,0	5,8	16,0	14,6	17,3	19,4
9,8	11,2	13,5	9,8	15,4	12,8	16,5	18,9
12,3	11,2	14,8	8,3	15,4	12,8	17,4	18,9
13,0	11,6	13,9	9,2	14,6	15,6	18,6	19,0
11,1	12,5	9,0	8,4	15,3	17,2	18,2	18,6
11,7	11,8	4,9	9,2	14,1	16,3	17,1	18,8
12,6	10,0	5,4	11,0	13,2	17,0	20,2	19,6
14,2	11,9	6,3	11,0	13,6	16,5	19,3	19,5
15,3	10,6	11,2	10,6	14,6	16,4	19,0	19,5
14,9	11,7	11,1	12,7	12,6	19,3	18,3	20,0
14,5	10,9	11,0	11,4	15,5	20,4	19,0	19,4
15,0	8,2	6,0	13,6	16,5	19,4	17,7	18,9
13,6	11,8	5,4	14,8	13,4	19,8	18,2	18,2
15,0	11,0	8,0	13,8	13,9	18,9	19,8	19,0
14,6	10,4	7,2	14,6	14,3	16,8	20,8	17,3
13,5	8,7	6,2	15,4	16,2	16,8	20,1	18,7
13,4	10,3	7,8	13,0	17,5	16,5	20,7	18,1
13,8	10,8	7,2	9,5	18,3	15,3	20,0	18,5
14,4	11,7	8,4	14,4	18,1	14,4	18,4	17,1
13,8	11,1	9,0	14,4	14,6	11,8	17,4	17,4
12,7	14,2	12,4	14,2	12,7	10,4	17,1	18,0
14,6	15,6	11,0	11,7	12,4	14,3	16,6	17,0
14,0	14,7	10,0	11,5	13,5	14,8	15,8	18,1
12,5	15,8	9,6	13,6	16,2	13,0	18,5	19,0
9,8	14,6	7,2	13,6	18,0	13,6	17,3	18,6
10,4	14,4	5,9	10,3	17,1	13,9	18,7	17,2
13,7	13,8	4,0	10,4	14,7	15,8	19,6	19,2
14,0	14,7	5,6	12,3	14,8	14,4	19,1	18,3
14,5	14,7	8,3	13,7	14,9	16,0	18,9	21,1
14,1	13,6	7,1	14,8	14,6	16,6	19,6	19,1
13,7	12,6	6,4	13,9	14,7	18,2	19,5	20,1
12,6	9,7	7,0	12,6	17,8	17,3	19,7	19,0
13,4	8,6	10,0	11,0	16,2	19,9	20,8	19,3
14,6	8,8	11,3	10,8	13,9	20,2	19,5	19,0
15,6	7,1	12,0	11,4	13,4	16,9	18,5	17,3
15,1	11,9	11,4	12,4	14,2	16,4	18,6	18,4
12,1	14,0	12,4	12,7	13,8	15,6	17,8	17,3
11,2	12,0	13,6	13,8	16,6	17,4	19,5	17,1
9,8	12,0	13,1	14,2	15,7	17,9	19,4	15,4
8,8	13,0	7,0	15,7	17,6	19,2	19,7	17,4

(*) ler por colunas

Tabela A.3 – Dados diários sobre a temperatura mínima da cidade de São Paulo, medida em °C, no período de 18/04/1996 a 31/12/1997. (continuação)

16,1	13,2	11,8	13,0	8,3	15,0	19,0	16,2
18,2	14,4	13,8	15,4	7,8	15,0	18,8	15,4
18,2	13,3	11,2	14,6	9,6	13,6	18,2	16,0
18,9	14,4	10,6	12,0	11,0	13,2	18,2	18,2
18,0	17,6	11,4	13,2	10,6	14,8	19,6	19,7
19,3	17,0	12,3	14,9	11,6	14,5	18,0	19,4
17,8	16,0	11,0	14,8	11,0	14,3	18,9	20,0
18,8	16,2	10,6	13,2	11,4	14,0	19,0	16,2
18,3	16,3	10,8	12,2	12,6	15,5	19,7	16,2
18,4	16,4	8,6	11,5	12,4	13,9	18,7	16,5
18,4	15,2	6,5	12,8	9,5	17,0	16,9	17,9
18,3	15,7	8,2	13,9	14,8	17,9	18,1	18,6
17,5	14,3	12,4	12,0	15,5	14,0	20,1	20,1
16,8	14,6	14,4	10,4	14,3	12,4	20,2	19,1
15,6	13,3	14,1	11,8	11,8	12,5	20,8	21,1
16,0	15,8	11,5	13,4	11,2	15,6	16,3	21,1
17,3	11,8	5,6	12,6	12,4	14,8	16,1	20,8
16,9	14,3	7,2	14,8	10,4	14,5	17,2	20,7
14,0	14,4	9,5	12,6	11,0	15,1	19,0	21,2
14,4	14,4	7,9	12,8	11,6	15,6	16,2	20,3
15,6	14,4	14,5	11,8	12,2	16,4	16,1	19,2
16,9	16,8	11,6	12,8	10,6	18,0	15,8	21,6
17,2	15,4	14,4	13,0	11,3	16,1	18,0	19,8
16,1	13,8	14,5	12,5	12,2	17,1	19,0	
13,6	11,2	12,0	10,2	14,4	18,3	19,8	
14,6	14,6	10,9	11,6	16,8	18,1	18,7	
13,0	13,3	10,2	8,6	13,9	15,1	18,6	
16,3	14,8	13,2	9,3	14,5	13,4	18,4	
17,6	14,0	12,4	10,0	13,4	13,0	20,1	
18,0	13,7	13,1	13,2	15,9	13,4	15,1	
18,1	12,6	14,6	11,5	14,0	15,4	13,5	
17,6	13,0	13,8	11,6	12,3	15,8	16,1	
14,8	12,6	14,7	13,4	10,8	15,2	17,7	
14,9	12,3	12,7	15,1	14,2	15,7	18,0	
14,4	13,1	14,7	11,9	18,2	16,3	16,6	
14,2	13,0	13,3	12,8	13,6	16,9	17,4	
15,0	13,5	12,6	14,2	12,8	15,9	16,2	
11,8	12,8	13,5	10,4	13,2	15,9	16,6	
12,2	13,3	12,1	9,4	11,8	16,9	18,8	
14,3	15,0	11,8	6,6	13,5	18,9	18,5	

(*) ler por colunas

APÊNDICE B

PROGRAMAS

B.1 – Programa referente às estimativas não-paramétricas da função de regressão de X_t em X_{t+j} , proposta em Tong (1990).

```
xt<-# Colocar a série a ser analisada neste objeto#
p<-# Colocar a ordem do processo auto-regressivo neste objeto#
N<-length(xt)
min1<-min(xt)
minimo<-round(min1,1)
max1<-max(xt)
maximo<-ceiling(max1)
h<-1.5
x<-seq(minimo,maximo,h)
n<-length(x)
BN<-1/sqrt(N)
mjx<-matrix(0,n,p)
soma1<-0
soma2<-0
for (j in 1:p){
soma1<-0
soma2<-0
for (i in 1:n){
y<-j+1
for (l in y:N){
Z<-(x[i]-xt[l-j])/BN
if (abs(Z) <= 1)
K<-1-abs(Z)
else (K<-0)
delta<-K/BN

soma1<-soma1 + xt[l]*delta
soma2<-soma2 + delta

mjx[i,j]<-soma1/soma2
```

B.2 – Programa referente ao teste F de linearidade *threshold* proposto em Tsay (1989), caso univariado.

```
remove(ls())

#Introduzir os dados#

Y<- #colocar neste objeto a série a ser analisada#

n<-length(Y)

p<- #colocar neste objeto a ordem auto-regressiva da série#

d<- #variar de 1 até p o valor deste objeto para que, em cada caso,
      o programa calcule o nível descritivo do teste#

p1<-p+1
valor<-p+1-d
h<-max(1,valor)
n1<-n-d
n2<-n-d-h+1
Y<-matrix(Y,1,n)
s<-seq(n)
s<-matrix(s,1,n)
Y1<-Y[,h:n1]
Y1<-matrix(Y1,1,n2)
s1<-s[,h:n1]
s1<-matrix(s1,1,n2)
indord<-order(Y1,s1)
s2<-s1[c(indord)]
Y2<-Y1[c(indord)]
s3<-s2+d
Y3<-Y[c(s3)]
Y3<-matrix(Y3,n2,1)
s4<-matrix(,n2,p)
```



```

for (i in 1:n2) {
for (j in 1:p){
s4[i,j]<-s3[i]-j
}
}

```

```

Y4<-Y[c(s4)]
Y4<-matrix(Y4,n2,p)
y41<- matrix(, n2, 1)

```

```

for (i in 1:n2){
y41[i]<-1
}

```

```

matrizX<-append(y41, Y4)
matrizX<- matrix(matrizX, n2, p1)

```

```

m<-(n/10)+p
m<-trunc(m)

```

```

Ym<-Y3[1:m,]

```

```

matrizXm<- matrizX[1:m,]

```

#foram criados oito objetos de beta0 a beta7, pois, a ordem auto-regressiva utilizada neste caso foi 7, sendo assim, se a ordem do processo a ser testado for diferente crie mais (ou menos) objetos#

```

beta0<-matrizXm[,1] #este objeto contém uma coluna de 1's#
beta1<-matrizXm[,2]
beta2<-matrizXm[,3]
beta3<-matrizXm[,4]
beta4<-matrizXm[,5]
beta5<-matrizXm[,6]
beta6<-matrizXm[,7]
beta7<-matrizXm[,8]

```

```
#Regressão com as m primeiras observações#
```

```
mprimeiros<-lm(Ym ~ beta0+beta1+beta2+beta3+beta4+beta5+beta6+beta7-1)
```

```
betam<-coef(mprimeiros)
```

```
betam<-matrix(betam, 1,p1)
```

```
Pm<-t(matrizXm)%*%(matrizXm)
```

```
Pm<-solve(Pm)
```

```
n3<-m+1
```

```
n4<-n-d-h-m+1
```

```
n5<-n4+1
```

```
BETA<-matrix(,n5,p1)
```

```
BETA[1,]<-betam
```

```
D<-matrix(,n4, 1)
```

```
A<-matrix(,n4, 1)
```

```
E<-matrix(,n4, 1)
```

```
K<-matrix(,n4,p1)
```

```
Identidade<-diag(p1)
```

```
Variancias<-matrix(,n5,p1)
```

```
for (h in 1:p1){
```

```
  Variancias[1,h]<-Pm[h,h]
```

```
}
```

```
#Regressao atualizações (passo m+ 1)#
```

```
for (k in n3:n2) {
```

```
  k1<-k-m
```

```
  k2<-k-(m-1)
```

```
  xlinha<-t(matrizX[k,])
```

```
  D[k1]<-1+xlinha%*%Pm%*%matrizX[k,]
```

```
  inversoD<-1/D[k1]
```

```

km1<-Pm%%matrizX[k,]
K[k1,]<-inversoD*km1

BETA[k2,]<- BETA[k1,] + K[k1,]*(Y3[k]-(xlinha%%BETA[k1,]))

a<-matrizX[k,]%%xlinha
a1<-a*inversoD

Pm<-(Identidade - Pm%%a1)%%Pm

for (j in 1:p1){
  Variancias[k2,j]<-Pm[j,j]
}

A[k1] <- Y3[k] - (xlinha%%BETA[k1,])
E[k1]<-A[k1]*sqrt(D[k1])*inversoD #no objeto E encontram-se os resíduos padronizados#

}

#Regressão - teste F#

comeco<-n3
fim<-n2

Y5m<-matrizX[comeco:fim,]

beta.zero<-Y5m[, 1]
beta.um<-Y5m[,2]
beta.dois<-Y5m[,3]
beta.tres<-Y5m[,4]
beta.quatro<-Y5m[,5]
beta.cinco<-Y5m[,6]
beta.seis<-Y5m[,7]
beta.sete<-Y5m[,8]

```

```
fit.f<-lm(E ~ beta.zero+beta.um+beta.dois+beta.tres+beta.quatro+beta.cinco+beta.seis+beta.sete-1)
```

```
erro1<-0  
erro2<-0  
for (i in 1:n4)  
{  
erro1<-erro1+E[i,]^2  
}
```

```
erro2<-resid(fit.f)  
erro2<-matrix(erro2, n4, 1)
```

```
erro3<-0
```

```
for (i in 1:n4)  
{  
erro3<-erro3+(erro2[i,]^2)  
}
```

```
p1  
p2<-n-d-m-p-h  
p2  
d  
numerador<-(erro1-erro3)/(p1)  
denominador<-erro3/(p2)  
estatf<-numerador/denominador  
estatf  
niveldescritivo<-1-pf(estatf, p1, p2)  
niveldescritivo
```

B.3 – Programa referente ao teste de linearidade *threshold* proposto em Tsay (1998), caso multivariado.

```
remove(ls())

Y<- #colocar neste objeto as séries de entrada#

attach(Y)

Y<-cbind(res65, pm10, temp) #no caso estamos trabalhando com estas três séries#

Y<-matrix(Y,nrow(Y),ncol(Y))

cont<-nrow(Y)
n<-nrow(Y)
Ya<-matrix(res65, cont, 1) #recebe a série resposta#
Yb<-matrix(pm10, cont, 1) #recebe a variável explicativa#
Yc<-matrix(temp, cont, 1) #recebe a variável limiar#

p<- #colocar a ordem auto-regressiva da série resposta#
p2<-p+1

q<- #colocar a ordem auto-regressiva da série utilizada como variável explicativa#
q2<-q+1

d<- #variar este valor de 1 até o mínimo entre p e q para calcular, em cada caso, o valor da estatística
      C(d)#

h<-max(p, q, d)

z.inf<-h+1-d
z.sup<-n-d
s<-n-h
colunas<-p2+q2+1
dimens.X<-p+q+1

matrizresposta<-matrix(NA, cont, p)
```

```

for (i in 1:p){
matrizresposta[(i+1):cont, i]<-Ya[1:(cont-i),]
}

matrizresp<-append(Ya, matrizresposta)
matrizresp<- matrix(matrizresp, cont, p2)

matrizexplicativa<-matrix(NA, cont, q)

for (i in 1:q){
matrizexplicativa[(i+1):cont, i]<-Yb[1:(cont-i),]
}

matrizexplicativa<- matrix(matrizexplicativa, cont, q)

matrizexplic<-append(Yb, matrizexplicativa)
matrizexplic<-matrix(matrizexplic, cont, q2)

mat.completa<-append(matrizresp, matrizexplic)
mat.completa<-append(mat.completa, Yc)
mat.completa<-matrix(mat.completa, cont, columnas)

matriz.limiar<-mat.completa[z.inf:z.sup,]
matriz.ordenada <- matriz.limiar[order(matriz.limiar[,columnas]), 1:ncol(matriz.limiar)]

Y3<-matriz.ordenada[,1]
Y3<-matrix(Y3, s, 1)

matriz1<-matrix(1, s, 1)
matrizX1<-matriz.ordenada[,2:p2]
matrizX2<-matriz.ordenada[(p2+2):(columnas-1)]
matrizX3<-append(matriz1, matrizX1)
matrizX4<-append(matrizX3, matrizX2)
matrizX<-matrix(matrizX4, s, dimens.X)

```

```

m<-5*sqrt(n)
m<-trunc(m)

Ym<-Y3[1:m,]

matrizXm<- matrizX[1:m,]

#aqui foram criados objetos de beta0 a beta12, pois, supusemos p = 8 e q = 4#

beta0<-matrizXm[,1]
beta1<-matrizXm[,2]
beta2<-matrizXm[,3]
beta3<-matrizXm[,4]
beta4<-matrizXm[,5]
beta5<-matrizXm[,6]
beta6<-matrizXm[,7]
beta7<-matrizXm[,8]
beta8<-matrizXm[,9]
beta9<-matrizXm[,10]
beta10<-matrizXm[,11]
beta11<-matrizXm[,12]
beta12<-matrizXm[,13]

n3<-m+1
n4<-n-d-h-m+1
n5<-n4+1
p1<-dimens.X
n2<-n-d-h+1

#Regressao m primeiros#

mprimeiros<-lm(Ym
beta0+beta1+beta2+beta3+beta4+beta5+beta6+beta7+beta8+beta9+beta10+beta11+beta12-1,
na.action=na.omit)

```

```

betam<-coef(mprimeiros)
betam<-matrix(betam,1,p1)

Pm<-t(matrizXm)%*(matrizXm)
Pm<-solve(Pm)

BETA<-matrix(,n5,p1)
BETA[1,]<-betam
D<-matrix(,n4,1)
A<-matrix(,n4,1)
E<-matrix(,n4,1)
K<-matrix(,n4,p1)
Identidade<-diag(p1)
Variancias<-matrix(,n5,p1)

for (h in 1:p1){
Variancias[1,h]<-Pm[h,h]
}

#Regressao atualizacoes#

for (k in n3:n2) {

k1<-k-m
k2<-k-(m-1)

xlinha<-t(matrizX[k,])

D[k1]<-1+xlinha%*Pm%*matrizX[k,]
inversoD<-1/D[k1]

km1<-Pm%*matrizX[k,]
K[k1,]<-inversoD*km1

BETA[k2,]<- BETA[k1,] + K[k1,]*(Y3[k,]-(xlinha%*BETA[k1,]))

```



```

a<-matrizX[k,]%*%xlinha
a1<-inversoD*a

Pm<-(Identidade - Pm%*%a1)%*%Pm
for (j in 1:p1){
  Variancias[k2,j]<-Pm[j,j]
}

A[k1] <- Y3[k] - (xlinha%*%BETA[k1,])
E[k1]<-A[k1]*sqrt(D[k1])*inversoD

}

```

```
#Regressao_teste C(d)#
```

```
valor<- #inserir o número de passos da recursão#
```

```

neta<-E[1:valor]
neta<-matrix(neta,valor,1)
matrizXneta<-matrizX[1:valor,]

```

```
#aqui foram criados objetos de beta.zero a beta.doze, pois, supusemos p = 8 e q = 4#
```

```

beta.zero<-matrizXneta[,1]
beta.um<-matrizXneta[,2]
beta.dois<-matrizXneta[,3]
beta.tres<-matrizXneta[,4]
beta.quatro<-matrizXneta[,5]
beta.cinco<-matrizXneta[,6]
beta.seis<-matrizXneta[,7]
beta.sete<-matrizXneta[,8]
beta.oito<-matrizXneta[,9]
beta.nove<-matrizXneta[,10]
beta.dez<-matrizXneta[,11]
beta.onze<-matrizXneta[,12]
beta.doze<-matrizXneta[,13]

```

```
fit.qui<-lm(neta ~ beta.zero + beta.um + beta.dois + beta.três + beta.quatro + beta.cinco + beta.seis +  
beta.sete + beta.oito + beta.nove + beta.dez + beta.onze + beta.doze - 1, na.action=na.omit)
```

```
erro1<-0  
erro2<-0  
for (i in 1:valor)  
{  
erro1<-erro1+E[i,]^2  
}
```

```
erro2<-resid(fit.qui)  
erro2<-matrix(erro2, valor, 1)
```

```
erro3<-0
```

```
for (i in 1:valor)  
{  
erro3<-erro3+(erro2[i,]^2)  
}
```

```
S0<-erro1/valor  
S1<-erro3/valor
```

```
Cd<-(valor-(p-q+1))*(log(S0)-log(S1))  
Cd
```

```
gl1<-(p+q+1)
```

```
niveldescritivo<-1-pchisq(Cd, gl1)  
niveldescritivo
```