

2 Modelos Não Lineares

2.1. Introdução

Nos últimos anos, muitos modelos não-lineares para a análise de séries temporais têm sido propostos. Na econometria clássica, os modelos de múltiplos regimes lineares, tais como o TAR/SETAR, proposto por Tong (1978), e o STAR, proposto por Chan and Tong (1986), Granger & Teräsvirta (1993) e Teräsvirta (1994), que é uma generalização dos dois primeiros, têm recebido muita atenção. Já na Teoria de Aprendizado de Máquina, as redes neurais artificiais (RNA), modelos de particionamento recursivo, especialmente as árvores de classificação e regressão (CART, Breiman et al (1984)), e métodos de regressão não paramétrica são bastante utilizados.

2.2. Modelo STAR

O modelo STAR (Smooth Transition AutoRegression) pode ser visto como um modelo linear autoregressivo, onde os seus coeficientes são determinados pela posição do vetor de variáveis explanatórias dentro do denominado espaço de transição.

Para aplicações práticas, o principal problema do uso deste modelo é descrever a relação entre o espaço de transição e os coeficientes. Por outro lado, o modelo STAR carrega importantes propriedades dos modelos lineares e ferramentas estatísticas para especificação, estimação e testes de diagnósticos.

2.2.1. Formulação Matemática

A formulação matemática destes modelos para uma série temporal univariada observada nos instantes $t = 1-p, 1-(p-1), \dots, -1, 0, 1, \dots, T-1, T$ é dada por:

$$y_t = \beta'_1 z_t (1 - G(x_t; \gamma, c)) + \beta'_2 z_t (G(x_t; \gamma, c)) + \varepsilon_t \quad (2.1)$$

onde

$$z_t = (1, y_{t-1}, y_{t-2}, \dots, y_{t-p})',$$

$$\beta_i = (\beta_{i,0}, \beta_{i,1}, \dots, \beta_{i,p})' ; i = 1, 2$$

e

$$\gamma > 0$$

O erro aleatório ε_t segue uma distribuição condicional normal com média $E[\varepsilon_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}] = 0$ e variância $\text{VAR}[\varepsilon_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}] = \sigma^2$.

A formulação acima descreve um modelo com dois regimes cuja transição entre eles é governada por uma função suave, não-linear, contínua e limitada, G , que assume valores no intervalo $(0,1)$. G é usualmente chamada de função de transição. De acordo com a equação (2.1), o modelo STAR descreve a evolução de uma série temporal como a combinação de dois modelos autoregressivos de ordem p .

A variável de transição x_t pode ser um autoregressor ($x_t = y_{t-d}$), variáveis exógenas, funções de variáveis endógenas ($x_t = h(z_t)$), ou até mesmo uma tendência linear temporal ($x_t = t$).

Um caso particular, o modelo TAR (Threshold AutoRegression), pode ser obtido quando a função G é uma função indicadora do tipo:

$$G(.) = \begin{cases} 1, & x_t \leq c \\ 0, & x_t > c \end{cases} \quad (2.2)$$

Neste caso, o limiar entre os dois regimes é abrupto e determinado por c , o parâmetro de limiar ou locação.

Uma das grandes vantagens na utilização dos modelos de transição suave é a possibilidade de especificar a função de transição de forma a evitar o problema da

busca por um limiar “rígido” entre os regimes. A escolha mais comum para a função de transição é a função logística:

$$G(x_t; \gamma, c) = \frac{1}{1 + e^{-\gamma(x_t - c)}} \quad (2.3)$$

O modelo com esta função de transição é denominado modelo LSTAR (Logistic Smooth Transition AutoRegression).

A função logística possui dois parâmetros, γ e c . O primeiro é o responsável pelo grau de suavidade da função de transição, e o segundo representa o limiar entre os dois regimes. Na situação em que $x_t = c$, a observação pertence a ambos os regimes com igual grau de pertinência.

A Figura (2.1) mostra o gráfico da função logística com o parâmetro de localização $c = 0$ e o parâmetro γ com valores no conjunto $\{1; 2,5; 5; 10; 50\}$, representando diferentes níveis de suavidade para a função logística.

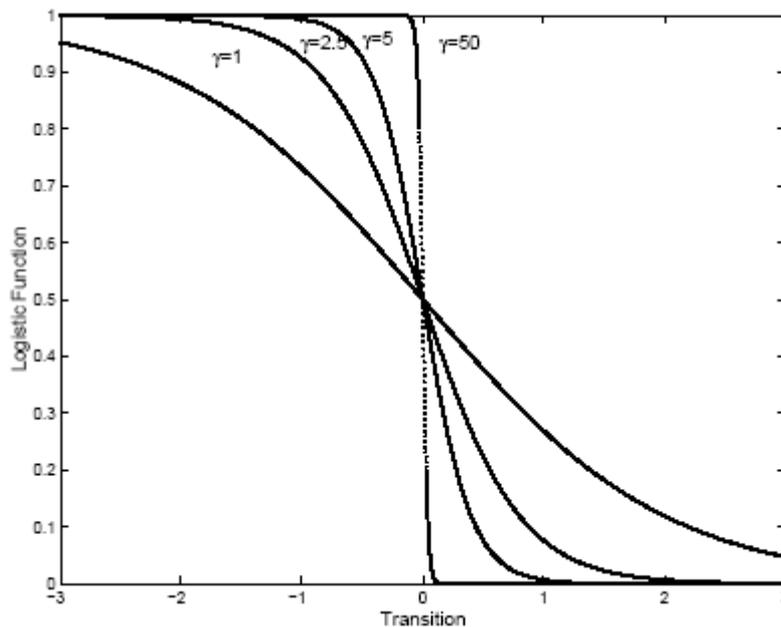


Figura 2.1: Função logística com parâmetros fixos

É importante notar que quando γ vai para infinito, a função logística se aproxima de uma função do tipo degrau e a transição de um regime para o outro se torna quase instantânea, ou seja, transição abrupta. Neste caso temos um

modelo TAR. No caso de $x_t = y_{t-d}$, temos um SETAR (Self-Exciting Threshold Autoregression).

Quando γ tende para zero, a função logística se torna igual a uma constante (0,5) e o LSTAR se reduz a um modelo linear. Ou seja, não existe nenhuma distinção entre os regimes.

Alguns autores classificam os parâmetros do modelo STAR, equação (2.1), em lineares e não lineares. Os parâmetros lineares são os coeficientes dos modelos autoregressivos ligados aos regimes, e os não lineares são os parâmetros da função de transição.

O modelo STAR com dois regimes pode ser visto como uma média ponderada de dois modelos AR, com pesos $G(x_t; \gamma, c)$ e $(1 - G(x_t; \gamma, c))$ determinados pela função de transição. Para o mesmo valor de γ , a distância entre o valor de x_t e c determina o grau de pertinência dos regimes do modelo.

Pela suposição de normalidade e independência dos erros, os estimadores de mínimos quadrados não-lineares e máxima verossimilhança são assintoticamente normais se satisfeitas algumas condições de regularidade. A partição em parâmetros lineares e não-lineares é particularmente interessante para estimação via maximização da função de log-verossimilhança concentrada.

2.2.2. Construção do Modelo

A modelagem STAR parte de um modelo simples e aumenta sua complexidade de acordo com os resultados dos testes aplicados. O processo de construção destes modelos segue um ciclo, proposto em van Dijk, Terasvirta e Franses (2002). Os passos estão a seguir:

1. Especificação de um modelo auto-regressivo de ordem p .

A escolha leva em consideração critérios de informação como o AIC e o BIC, e é importante que os resíduos do modelo sejam aproximadamente um ruído branco.

2. Teste da hipótese de linearidade contra uma alternativa da família STAR.

O teste usado é o Multiplicador de Lagrange, onde a hipótese nula é a de linearidade. Em caso de rejeição da hipótese nula, é feita a seleção da variável de transição (x_t) e da forma da função de transição.

3. Estimação dos parâmetros do modelo STAR selecionado.

A estimação dos parâmetros lineares é feita por Mínimos Quadrados Ordinários (MQO), e a dos parâmetros não-lineares é feita por mínimos quadrados não-lineares (MQNL), que sob a normalidade dos erros, é equivalente à Máxima Verossimilhança (MV).

4. Realização do diagnóstico do modelo.

Os testes para a validação do modelo testam a correlação dos resíduos.

5. Re-especificação do modelo de acordo com os resultados do diagnóstico.

6. Utilização do modelo com propósitos descritivos ou de previsão.

Uma importante extensão do modelo STAR é o MRSTAR (Multiple Regime Smooth Transition AutoRegression) que permite a criação de mais de dois regimes. A representação de um MRSTAR de 4 regimes pode ser mostrada da seguinte forma:

$$\begin{aligned}
 y_t = & [\beta'_1 z_t (1 - G_1(x_{1t}; \gamma_1, c_1)) + \beta'_2 z_t (G_1(x_{1t}; \gamma_1, c_1))] * (1 - G_2(x_{2t}; \gamma_2, c_2)) \\
 & + [\beta'_3 z_t (1 - G_1(x_{1t}; \gamma_1, c_1)) + \beta'_4 z_t (G_1(x_{1t}; \gamma_1, c_1))] \\
 & * (G_2(x_{2t}; \gamma_2, c_2)) + \varepsilon_t
 \end{aligned}
 \tag{2.4}$$

Assumindo as variáveis x_{1t} e x_{2t} conhecidas, nota-se que os regimes na equação (2.4) são ponderados por uma composição de funções logísticas (G_1 e G_2) que somam 1. Estas composições podem ser vista como funções de pertinência.

Uma descrição mais detalhada do MRSTAR pode ser vista em van Dijk and Franses (1999).

2.3. Modelo CART

Os modelos estruturados em árvores são modelos não paramétricos, cuja filosofia é utilizar modelos simples para sub-amostras dos dados, dividindo o problema em partes. Frequentemente esses modelos são estimados através de algoritmos recursivos de particionamento.

Proposto por Breiman, Friedman, Olshen e Stone (1984), o algoritmo CART (Classification And Regression Trees) é usado para prever variáveis dependentes contínuas (regressão) e categóricas (classificação) através do particionamento recursivo do espaço de variáveis de transição.

A metodologia CART tem como principal atrativo a interpretabilidade proporcionada pela estrutura de árvore de decisão obtida no modelo final que, também pode ser lido, como um conjunto de sentenças lógicas a respeito das variáveis explanatórias. Esta é uma característica importante, pois permite ao analista acessar prontamente o que o modelo está fazendo e tomar decisões com mais clareza.

O modelo estimado é representado por um gráfico com o formato de uma árvore binária de decisão, com $N \in \mathbb{N}$ nós ancestrais, ou nós de divisão, e $K \in \mathbb{N}$ nós terminais, também chamados de folhas. A árvore cresce a partir do nó raiz em direção aos nós terminais.

Para ajudar na parametrização dos modelos a serem utilizados nesta dissertação, adota-se um padrão de numeração dos nós da árvore, conforme a Figura (2.2). Atribui-se o valor 0 para a raiz (profundidade 0), e na seqüência, da esquerda para a direita, números inteiros de forma crescente. Quando o nó não é gerado, salta-se o seu número correspondente e continua-se a numeração no próximo nó gerado.

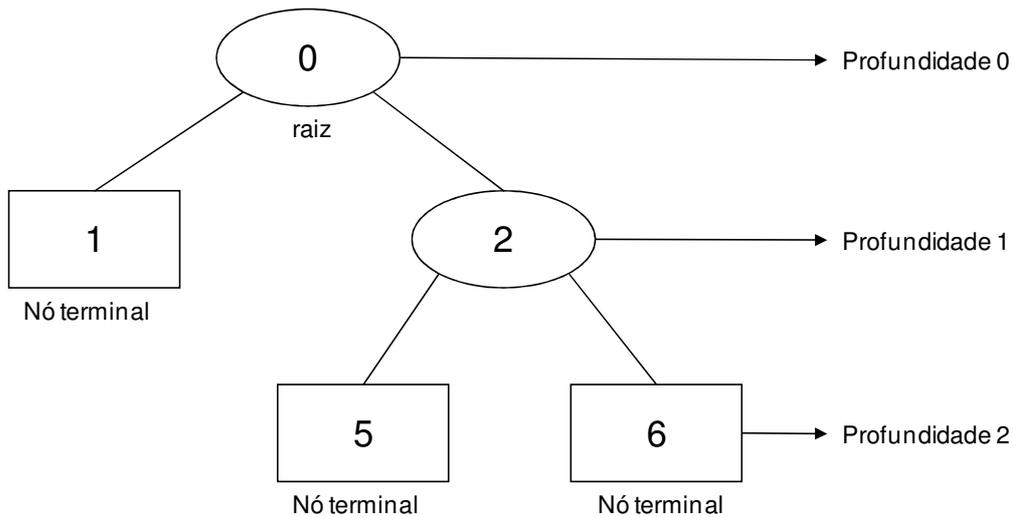


Figura 2.2: Numeração de uma árvore de decisão

2.3.1. Formulação Matemática

Considere $x_t = (x_{1t}, \dots, x_{qt})' \in \mathbb{X} \subseteq \mathbb{R}^q$ um vetor que contém q variáveis explicativas para uma resposta univariada contínua $y_t \in \mathbb{R}$, $t = 1, \dots, T$. O vetor x_t pode conter lags de y_t , no caso de séries temporais, como também variáveis exógenas. A relação entre y_t e x_t segue um modelo de regressão da forma

$$y_t = f(x_t) + \varepsilon_t, \quad (2.5)$$

onde a forma funcional $f(\cdot)$ é desconhecida e não há suposições sobre a distribuição do termo aleatório ε_t .

Para representar matematicamente um modelo complexo de árvore de regressão, introduzimos a seguinte notação. Como mencionado anteriormente, o nó raiz está na posição 0 e um nó ancestral na posição j gera um filho do lado esquerdo na posição $2j+1$ e um filho do lado direito na posição $2j+2$. Cada nó ancestral tem uma variável de transição associada $x_{s_j t} \in x_t$, onde $s_j \in \mathbb{S} = \{1, 2, \dots, q\}$. Seja \mathbb{J} e \mathbb{T} os conjuntos dos índices dos nós ancestrais e dos nós terminais, respectivamente. Desta forma, a arquitetura completa da árvore pode ser determinada por \mathbb{J} e \mathbb{T} .

Seguindo Lewis and Stevens (1991) um modelo estruturado por árvore com K folhas é um modelo de particionamento recursivo que aproxima $f(\cdot)$ por uma

função geral não-linear $H(x_t; \psi)$ de x_t e definida pelo vetor de parâmetros $\psi \in \mathbb{R}^r$, onde r é o número total de parâmetros.

Usualmente, $H(\cdot)$ é uma função constante definida por K sub-regiões $k_i(\theta), i = 1, \dots, K$, de algum domínio $K \subset \mathbb{R}^q$.

$$f(x_t) \approx H(x_t; \psi) = \sum_{i=1}^K \beta_i I_i(x_t; \theta_i) \tag{2.6}$$

onde

$$I_i(x_t; \theta_i) = \begin{cases} 1, & \text{se } x_t \in k_i(\theta_i) \\ 0, & \text{cc.} \end{cases}$$

e

$$\psi = (\beta_1, \dots, \beta_K, \theta'_1, \dots, \theta'_K)'$$

A Figura (2.3) é um exemplo de um modelo gerado por uma árvore de regressão que explica a relação entre a variável resposta y e um conjunto de duas variáveis explicativas x_1 e x_2 ($q = 2$). Define-se $c_j, j = 0, 1, \dots, N$, como o valor limite da partição k_i que determinará a inclusão da observação na região.

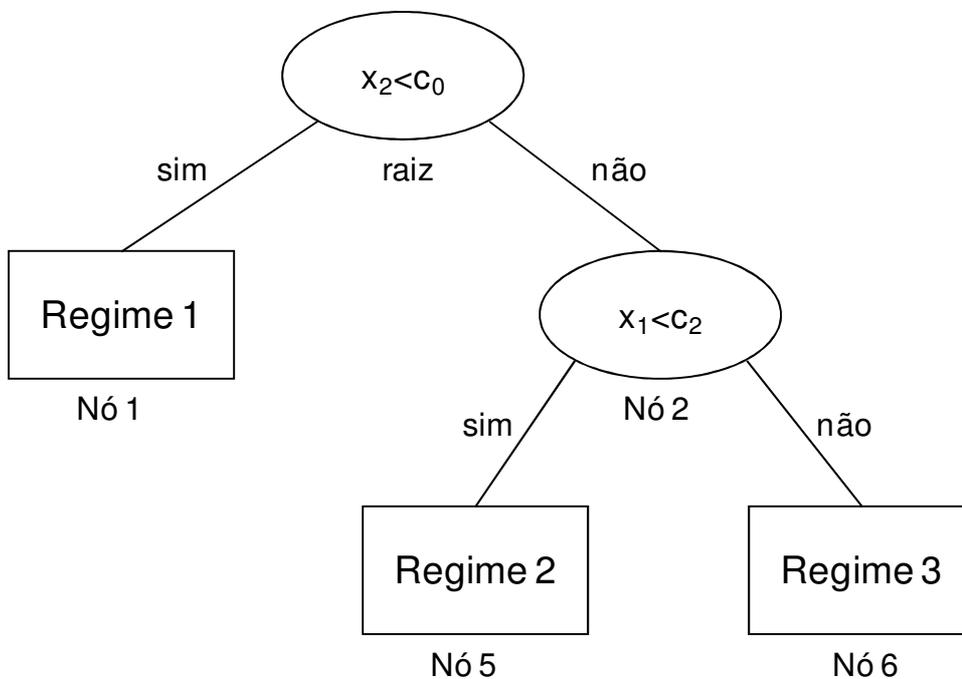


Figura 2.3: Exemplo de modelo de árvore de regressão com 3 regimes

O modelo final para a arquitetura da Figura (2.3) tem a seguinte forma:

$$y_t = \beta_1 I(x_{2t}; c_0) + \{\beta_5 I(x_{1t}; c_2) + \beta_6 [1 - I(x_{1t}; c_2)]\} * [1 - I(x_{2t}; c_0)] \quad (2.7)$$

onde

$$I(x_{s_j t}; c_j) = \begin{cases} 1, & \text{se } x_{s_j t} \leq c_j \\ 0, & \text{cc.} \end{cases}$$

Neste contexto, definimos as sub-regiões $k_i(\theta), i = 1, \dots, K$, da equação (2.6) por hiperplanos ortogonais aos eixos das variáveis de transição. No caso do exemplo da Figura (2.3), temos os seguintes hiperplanos:

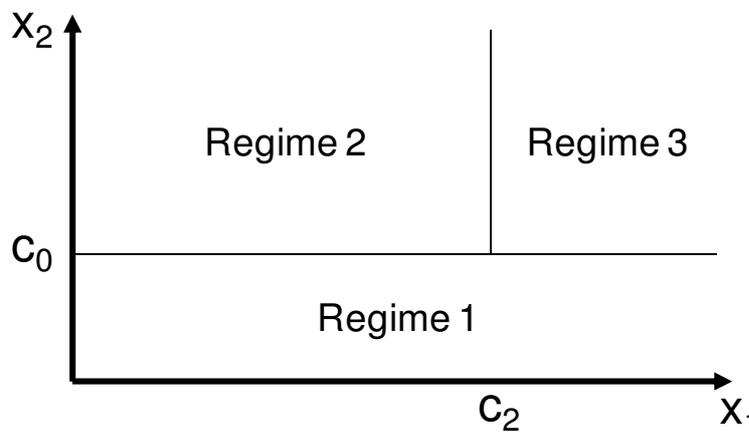


Figura 2.4: Hiperplanos que definem os regimes da árvore

Apesar dos modelos estruturados em árvore serem populares ferramentas para regressões não-paramétricas em áreas como bioestatística, medicina, ecologia, em economia, a sua aplicação é limitada. No entanto, alguns modelos econométricos não-lineares já propostos na literatura são casos especiais de árvores de regressão.

2.3.2. Algoritmo de Crescimento

A arquitetura da árvore é definida a partir de um ciclo iterativo que escolhe um nó a cada passo para ser subdividido e gerar mais 2 nós. A cada iteração, além

do nó a ser dividido, também é especificada uma variável de transição e o limiar desta divisão (c_j). A escolha desta especificação visa minimizar a soma dos erros quadráticos de previsão. Para a raiz da árvore (primeira divisão), a equação a ser minimizada pode ser vista a seguir em (2.8).

$$\text{SEQ}(s_0) = \sum_{t=1}^T \{y_t - [\beta_1 I(x_t; s_0, c_0) + \beta_2 (1 - I(x_t; s_0, c_0))]\}^2 \quad (2.8)$$

Após a especificação, estimam-se os parâmetros dos modelos locais para as observações alocadas dentro dos nós gerados pela divisão. Esse ciclo se repete até que não haja mais ganho em efetuar subdivisões na árvore.

Com o modelo final estimado, é possível realizar cortes de algumas folhas, técnica conhecida como podagem (pruning), a partir de medidas de custo e complexidade, ou da capacidade preditiva do modelo.