

2

MODELOS DA CLASSE STAR ESTRUTURADOS EM ÁRVORE

2.1

Introdução

Apesar de largamente utilizados, modelos lineares para análise de séries temporais são bastante limitados, pois não são capazes de capturar as relações não-lineares que na prática podem existir entre as variáveis. Modelos de regressão linear e auto-regressivos se encaixam nesta descrição, ao serem utilizados para estimar séries nas quais a natureza da relação entre as variáveis não é linear, geram resultados insatisfatórios. Ao realizar uma análise dos resíduos destas estimações, percebe-se que ainda há estrutura presente, que não foi capturada pelo modelo utilizado.

No outro extremo, modelos não-lineares são geralmente de difícil estimação e compreensão. É muito mais intuitivo entender e explicar os modelos descritos anteriormente, do que, por exemplo, dizer que a variável a ser estimada tem uma relação exponencial com seus regressores.

Modelos da classe STAR são lineares por partes, ou seja, existem regiões distintas no espaço das variáveis regressoras em que diferentes modelos lineares melhor se adequam para estimar sua relação com a variável a ser estimada. Ainda no STAR, a transição entre as regiões é regida de forma suave, por meio de uma função de pertinência. Esta flexibilidade permite que relações não-lineares em uma série sejam capturadas por um modelo que é na verdade uma seqüência de combinações lineares entre modelos lineares.

2.2

Formulação Geral dos Modelos STAR

Um modelo STAR se caracteriza pela combinação de dois modelos de regressão através de uma função de transição. Esta função, baseada no

comportamento de uma variável de transição, atribui um peso a cada um dos modelos. Exemplos de funções de transição são a função logística, a exponencial e a degrau.

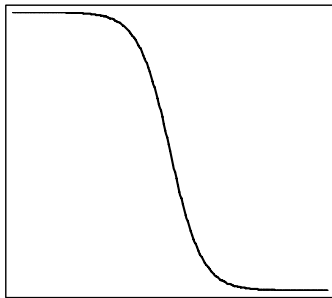


Figura 2.1: Função Logística

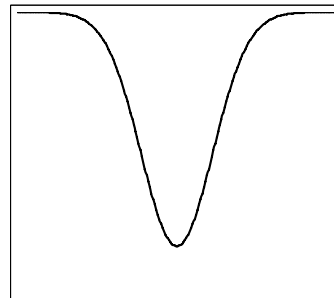


Figura 2.2: Função Exponencial

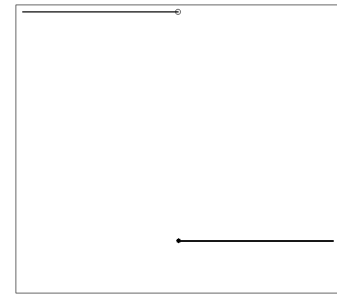


Figura 2.3: Função Degrau

A escolha da função de transição atende a objetivos específicos da modelagem. A logística, por exemplo, é capaz de oferecer uma transição suave entre regimes. Já a exponencial, devido a sua simetria, permite que o mesmo regime esteja ativo para valores altos e baixos da variável e transição. Já com a função degrau, apenas um regime permanece ativo de cada vez.

A variável de transição (s_t) pode ser tanto uma variável regressora, auto-regressiva ou não, uma constante, ou o próprio tempo. No caso da escolha de uma variável regressora, a mesma pode estar defasada no tempo.

O modelo STAR pode ser escrito de forma geral:

$$y_t = (B_1 \cdot X) \cdot F(\Phi; s_t) + (B_2 \cdot X) \cdot (1 - F(\Phi; s_t)) + \varepsilon_t \quad (2-2)$$

Onde

$$\Phi = \begin{bmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_m \end{bmatrix},$$

$$B_i = \begin{bmatrix} \beta_{i,0} \\ \beta_{i,1} \\ \vdots \\ \beta_{i,p} \end{bmatrix}, X = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{bmatrix}, 0 \leq F(\Phi; s_t) \leq 1, E[\varepsilon_t] = 0, E[\varepsilon_t^2] = \sigma^2, i=1,2$$

$$\varepsilon_t \sim N(0, \sigma^2), E[\varepsilon_t \varepsilon_{t+k}] = 0$$

O vetor Φ contém os parâmetros da função de transição. São chamados de parâmetros não-lineares.

2.3

O Modelo LSTAR

Este modelo é um caso particular da classe STAR com a mudança de regime ditada pela função logística, que é descrita a seguir:

$$f(c; \gamma; y_{t-1}) = \frac{e^{-\gamma(y_{t-1}-c)}}{1 + e^{-\gamma(y_{t-1}-c)}}, \text{ onde } y_{t-1} \text{ é a variável de transição} \quad (2-2)$$

escolhida

Esta função é largamente utilizada para reger a transição entre regimes [2].

Os modelos associados são de ordem auto-regressiva. O modelo é descrito pela equação a seguir:

$$y_t = (B_1 \cdot X) \cdot f(c; \gamma; y_{t-1}) + (B_2 \cdot X) \cdot (1 - f(c; \gamma; y_{t-1})) + \varepsilon_t \quad (2-3)$$

$$\text{Onde } B_i = \begin{bmatrix} \beta_{i,0} \\ \beta_{i,1} \\ \vdots \\ \beta_{i,p} \end{bmatrix}, X = \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix}, \gamma > 0, E[\varepsilon_t] = 0, E[\varepsilon_t^2] = \sigma^2, \varepsilon_t \sim N(0, \sigma^2),$$

$$E[\varepsilon_t \varepsilon_{t+k}] = 0$$

A variável de transição, um dos parâmetros da função logística, que no caso é y_{t-1} , atribui um grau de pertinência a cada um dos modelos. O parâmetro c serve como parâmetro de localização da função. Se for mal escolhido, um dos modelos provavelmente predominará sobre o outro na estimação. Quando a variável de transição equivale ao parâmetro de localização, ambos os modelos recebem 50 por cento de pertinência. Já o γ é um parâmetro de escala da função. Um γ muito baixo irá gerar uma transição muito suave entre os modelos, isto é, no limite, seja qual for o valor da variável de transição, ambos os modelos sempre terão um percentual de pertinência significativo. Se o γ for alto, a transição entre regimes

passa a ser abrupta, e quando um regime estiver ativo, o outro terá um percentual de pertinência muito baixo.

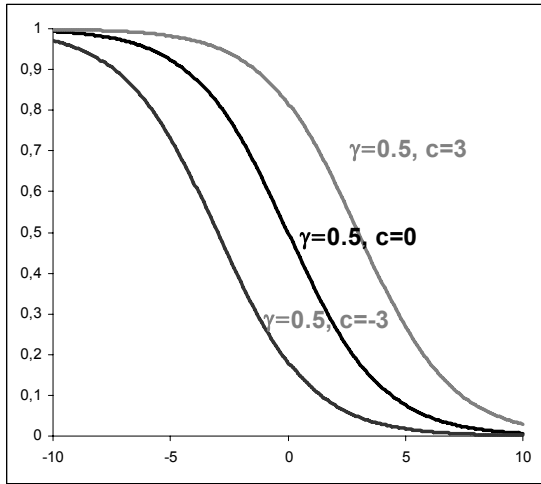


Figura 2.4: Gama fixo com valor baixo, c variando

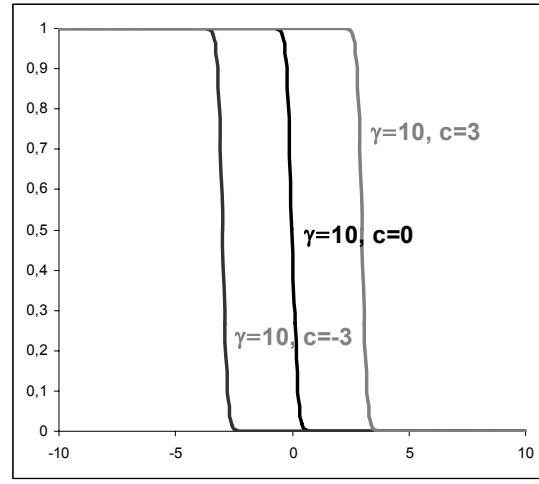


Figura 2.5: Gama fixo com valor alto, c variando

O vetor estimado \bar{Y} , é calculado da seguinte forma:

$$\bar{Y} = \bar{B}_1^T \cdot Z + \bar{B}_2^T \cdot (U - Z) \quad \text{onde} \quad (2-4)$$

$$\bar{B}_i = \begin{bmatrix} \bar{\beta}_{i,0} \\ \bar{\beta}_{i,1} \\ \vdots \\ \bar{\beta}_{i,p} \end{bmatrix}, \quad i=1,2, \quad U = \begin{bmatrix} 1 & \cdots & 1 & 1 \\ 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cdots & 1 & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 \cdot f(c; \gamma; y_{t-(n-p)}) & \cdots & 1 \cdot f(c; \gamma; y_{t-2}) & 1 \cdot f(c; \gamma; y_{t-1}) \\ y_{t-(n-p)} \cdot f(c; \gamma; y_{t-(n-p)}) & \cdots & y_{t-2} \cdot f(c; \gamma; y_{t-2}) & y_{t-1} \cdot f(c; \gamma; y_{t-1}) \\ \vdots & \vdots & \vdots & \vdots \\ y_{t-n} \cdot f(c; \gamma; y_{t-(n-p)}) & \cdots & y_{t-(p+1)} \cdot f(c; \gamma; y_{t-1}) & y_{t-p} \cdot f(c; \gamma; y_{t-1}) \end{bmatrix}$$

Os vetores \bar{B}_i contêm os estimadores dos parâmetros auto-regressivos do modelo. Um modelo LSTAR(p) possui um total de $2p+2$ parâmetros, sendo parte deles lineares (os dos vetores \bar{B}_i), parte não-lineares (c e γ).

As estacionariedades de ambos os modelos auto-regressivos por si só garantem a estacionariedade global do modelo LSTAR, conforme[7]. Porém, sob

determinadas condições, ainda que um dos modelos possua raiz unitária, o LSTAR pode ser globalmente estacionário.

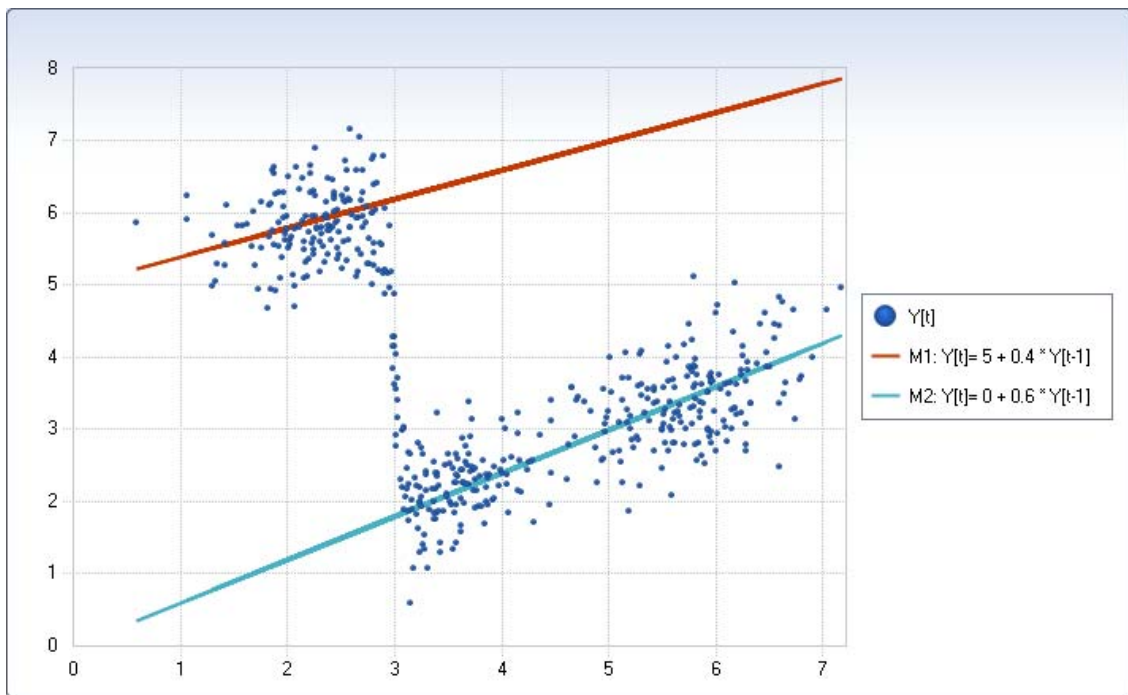


Figura 2.6: Exemplo de regiões de dados distintas com modelos específicos de estimação

2.4

Árvores de Decisão

Árvores de decisão são estruturas bastante utilizadas em *data mining* e aprendizado de máquina. Elas particionam os dados em diferentes sub-grupos ou regiões, que são tratados por modelos específicos localizados nas folhas destas estruturas. São chamadas de árvores de classificação quando a variável dependente é categórica, e árvores de regressão quando a mesma é contínua.

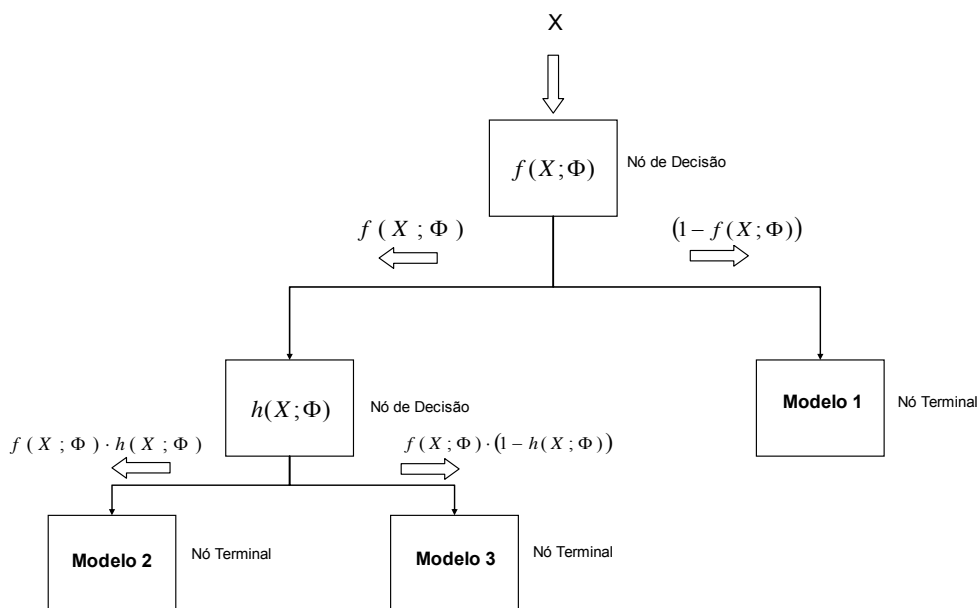


Figura 2.7: Exemplo de árvore de decisão

A Figura 2.7 mostra um exemplo deste tipo de estrutura. O dado de entrada é submetido a sucessivas avaliações e ao fim, um ou mais modelos são combinados para gerar um resultado final. Uma árvore de decisão real pode vir a ser bem maior do que a do exemplo. Seu ciclo de construção começa com um único nó de decisão e só decide-se por expandi-la caso a divisão de nós gere um ganho considerado suficiente. Quando não há mais ganho, pára-se de expandir a árvore em uma determinada ramificação.

Existem vários algoritmos de crescimento de árvores de decisão, como por exemplo, o *CART (Classification and Regression Trees)*. Neste algoritmo, a definição de ganho após uma divisão de nós é uma ponderação entre a soma dos erros quadráticos da estimação com a complexidade do modelo. Seu princípio é de que reduções pequenas no erro não se justificam, se o aumento da quantidade de parâmetros no modelo for significativo. Neste caso, seria melhor ficar com o modelo mais simples, que geraria um resultado similar com um custo computacional menor. Sendo assim, os nós considerados desnecessários são podados da árvore.

A equação de um modelo em árvore é obtida de forma recursiva. O modelo mostrado na Figura 2.7 se escreve da seguinte forma:

$$y_i = f(X; \Phi) \cdot [h(X; \Phi') \cdot M_2 + (1 - h(X; \Phi')) \cdot M_3] + (1 - f(X; \Phi)) \cdot M_1 + \varepsilon_i \quad (2-5)$$

A formulação geral de um modelo estruturado em árvore se dá pela equação:

$$y_i = F(X) + \varepsilon_i \quad (2-6)$$

Onde

$$F(X) = \sum_{i=1}^K M_i \cdot H_i(X; \Phi_i) \quad (2-7)$$

$$0 \leq H_i(X; \Phi_i) \leq 1$$

Os M_i são os diferentes modelos obtidos nos nós terminais. Normalmente são modelos simples que são aplicados às partições de dados impostas pelos nós de decisão. X é o vetor de variáveis explicativas, que serve como variável de divisão ou de transição.

As funções H_i são produtórias das funções intermediárias de partição dos dados encontradas no caminho entre a folha onde se encontra o modelo M_i e o nó raiz. Podem ser tanto contínuas como funções indicadoras, dependendo do tipo de árvore utilizada. Se forem indicadoras, temos regiões de dados bem definidas. Se forem contínuas, atribuem diferentes graus de pertinência aos dados. Os vetores contêm os parâmetros de construção das funções H_i . Exemplos são fatores de escala e de localização.

Existem, portanto, uma série de parâmetros a serem estimados para cada configuração de uma árvore. A Figura 2.8 ilustra o ciclo de estimação:

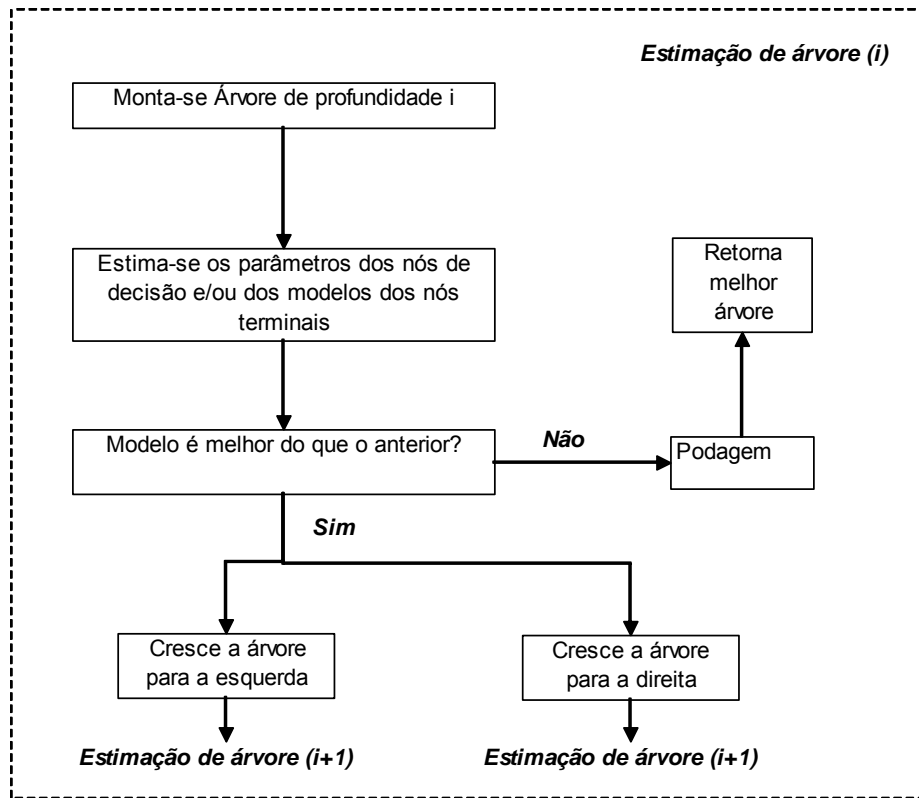


Figura 2.8: Ciclo de estimação de uma árvore de decisão

O modelo resultante $F(X)$ em (2-7) é uma função linear por partes definida em K subregiões no R_m , onde m tem a mesma dimensionalidade do vetor X . Os modelos de previsão contidos nos nós terminais são lineares, porém, a cada entrada de um novo vetor X , o modelo resultante $F(X)$ consegue capturar movimentos não lineares, já que a cada instante de tempo torna-se uma nova ponderação de modelos lineares.

O somatório dos erros quadráticos de uma árvore é obtido da seguinte forma:

$$SSE_{Arv(i)} = \sum_{t=1}^T (y_t - F_{Arv(i)}(X_t))^2 \quad (2-8)$$

Na metodologia CART, para se determinar que uma árvore seja superior a outra, utiliza-se um critério que pondera maximização da diminuição do erro com aumento de complexidade.

2.5

O Modelo STAR-Tree

Seguindo a abordagem das árvores de decisão, o modelo LSTAR pode ser generalizado para uma classe de modelo chamada MRSTAR, ou Multiple Regime Smooth Transition Auto-Regression, para que possa particionar os dados em mais do que dois regimes.

O exemplo em (2-5) pode ser reescrito no modelo MRSTAR da seguinte forma:

$$y_t = f(y_{t-1}; c_1; \gamma_1) \cdot [f(y_{t-1}; c_2; \gamma_2) \cdot (B_2 X) + (1 - f(y_{t-1}; c_2; \gamma_2)) \cdot (B_3 X)] + (1 - f(y_{t-1}; c_1; \gamma_1)) \cdot (B_1 X) + \varepsilon_t \quad (2-10)$$

$$\text{Onde } B_i = \begin{bmatrix} \beta_{i,0} \\ \beta_{i,1} \\ \vdots \\ \beta_{i,p} \end{bmatrix}, X = \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix}$$

Os modelos contidos nos nós de uma árvore STAR são AR de ordem p . Neste trabalho, a função de pertinência escolhida para o modelo é a função logística, já descrita anteriormente. Ainda, a transição é univariada e governada pela variável y_{t-1} .

A formulação de (2-7) pode ser reescrita como:

$$F(X) = \sum_{i=1}^K [\beta_{i,0} \quad \beta_{i,1} \quad \cdots \quad \beta_{i,p}] \cdot \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \cdot H_i(y_{t-1}; \Phi_i) \quad (2-11)$$

$$0 \leq H_i(X; \Phi_i) \leq 1$$

H_i é o produtório das funções de pertinência dos nós de decisão superiores ao nó terminal em que um determinado modelo AR se encontra. Um modelo STAR-Tree possui $2 \cdot N_d + p \cdot N_t$ parâmetros a serem estimados, aonde N_d é o número de nós de decisão (aonde os respectivos c e γ s precisam ser estimados), e N_t são os nós terminais, aonde p parâmetros de cada AR devem ser estimados.

A estimação dos parâmetros não-lineares da função logística é uma tarefa complexa. Os próximos capítulos irão abordar com mais profundidade algumas metodologias disponíveis para estimar tanto os parâmetros lineares e não-lineares de uma árvore, como o crescimento da própria árvore.