

### 3 Modelo STAR-Tree

#### 3.1. Introdução

O modelo STAR-Tree é uma combinação dos modelos STAR (Smooth Transition AutoRegression) e CART (Classification and Regression Tree). A modelagem do STAR-Tree se baseia no conceito de múltiplos regimes definidos por uma árvore binária. Ou seja, é especificado um modelo não linear paramétrico através de uma árvore de decisão.

No STAR-Tree as partições abruptas dos modelos estruturados em árvores são substituídas por partições suaves no espaço de transição e realizadas através de uma árvore de decisão, como no algoritmo CART, desta forma, os coeficientes são determinados através da combinação de diferentes modelos auto-regressivos. Em outras palavras, o STAR-Tree é um modelo da classe STAR estruturado através de uma árvore de decisão que pode utilizar variáveis exógenas no conjunto daquelas que ajudam a compor o espaço de transição.

As principais vantagens do modelo STAR-Tree vêm desta combinação dos modelos STAR e CART, isto é, modelos com transições suaves podem ser especificados através de métodos estatísticos, assim cria-se um ambiente para implementar testes estatísticos de hipóteses para a especificação do modelo. Além disso, a árvore de decisão final pode ser interpretada facilmente, e como no CART, lida como um conjunto de sentenças lógicas.

Para ilustrar essa idéia substituímos a função indicadora  $I(\cdot)$  da equação (2.7) do exemplo da Figura (2.3) pela função logística definida como:

$$G(x_t; \gamma, c) = \frac{1}{1 + e^{-\gamma(x_t - c)}} \quad (3.1)$$

Desta forma obtemos a seguinte equação:

$$y_t = \beta_1 G(x_{2t}; \gamma_0, c_0) + \{\beta_5 G(x_{1t}; \gamma_2, c_2) + \beta_6 [1 - G(x_{1t}; \gamma_2, c_2)]\} \\ * [1 - G(x_{2t}; \gamma_0, c_0)] \quad (3.2)$$

Temos agora o parâmetro adicional  $\gamma_j$ , que controla a suavidade da função logística. Essa é a grande diferença em relação à abordagem CART, pois a divisão do nó não gera dois subconjuntos exclusivos de observações, mas dois conjuntos nebulosos (Zadeh, 1965), aos quais todas as observações pertencem, porém com diferentes graus de pertinência.

É importante notar que o modelo de árvore de regressão é um caso especial da especificação de transição suave quando o parâmetro  $\gamma_j$  vai para o infinito. No caso desse parâmetro se aproximar de zero, temos a situação mais nebulosa na qual não há ganho em se produzir divisões nos nós. O parâmetro  $c_j$  é chamado de parâmetro de localização.

### 3.2. Formulação Matemática

Considera-se a sequência observada de valores reais da variável dependente  $\{y_t\}_{t=1}^T$  gerada pela função  $f(\cdot)$  desconhecida da seguinte forma:

$$y_t = f(x_t) + \varepsilon_t$$

tal que,

$$E[\varepsilon_t] = 0 \text{ e } E[\varepsilon_t^2] = \sigma^2$$

Seja  $(z_t \subseteq x_t)$ , tal que  $z_t \in \mathbb{R}^p$ ,  $p < q$ , e  $\tilde{z}_t = (1, z_t)'$ , o modelo paramétrico definido pela função  $H_{\mathbb{J}\mathbb{T}}(x_t; \psi): \mathbb{R}^{q+1} \rightarrow \mathbb{R}$  indexada pelo vetor de parâmetros  $\psi$  é chamado de STAR-Tree (Smooth Transition AutoRegression Tree) se

$$H_{\mathbb{J}\mathbb{T}}(x_t; \psi) = \sum_{i \in \mathbb{T}} \beta_i' \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) \quad (3.3)$$

onde

$$B_{\mathbb{J}i}(x_t; \theta_i) = \prod_{j \in \mathbb{J}} G(x_{s_j, t}; \gamma_j, c_j)^{\frac{n_{i,j}(1+n_{i,j})}{2}} \left[ 1 - G(x_{s_j, t}; \gamma_j, c_j) \right]^{(1-n_{i,j})(1+n_{i,j})} \quad (3.4)$$

e

$$n_{i,j} \begin{cases} -1, & \text{se o caminho para a folha } i \text{ não inclui o nó ancestral } j \\ & (B_{\mathbb{J}i} = 1); \\ 0, & \text{se o caminho para a folha } i \text{ inclui o filho direito do nó ancestral } j \\ & (B_{\mathbb{J}i} = 1 - G); \\ 1, & \text{se o caminho para a folha } i \text{ inclui o filho esquerdo do nó ancestral } j \\ & (B_{\mathbb{J}i} = G). \end{cases} \quad (3.5)$$

Define-se  $\mathbb{J}_i$  como o subconjunto de  $\mathbb{J}$  contendo os índices dos nós geradores que formam o caminho para a folha (nó terminal)  $i$ . Logo  $\theta_i$  é o vetor contendo todos os parâmetros  $(\gamma_k, c_k)$  tal que  $k \in \mathbb{J}_i, i \in \mathbb{T}$ .

$G(\cdot)$  é a função logística definida na equação (3.1). As funções  $B_{\mathbb{J}i}, 0 < B_{\mathbb{J}i} < 1$ , são conhecidas como funções de pertinência. Note que  $\sum_{j \in \mathbb{J}} B_{\mathbb{J}i}(x_t; \theta_j) = 1, \forall x_t \in \mathbb{R}^{q+1}$ .

O motivo de  $z_t$  ser um subconjunto de  $x_t$  (conjunto de variáveis de transição) é evitar regressores não estacionários nos modelos lineares locais, uma vez que podemos ter séries não estacionárias como variáveis de transição. Por exemplo,  $x_t$  pode conter uma tendência linear, que é uma variável de transição interessante quando existe uma possível quebra estrutural nas séries.

### 3.3. Construção do Modelo

O processo de construção do modelo STAR-Tree é uma adaptação do modelo proposto por Dijk, Terasvirta e Franses (2002). Como mencionado no Capítulo (2), o ciclo de modelagem envolve três passos: especificação, estimação e avaliação do modelo.

A especificação do modelo consiste na escolha de variáveis relevantes, a seleção do nó a ser dividido, se for o caso, e a seleção da variável de transição.

As candidatas a possíveis variáveis de transição podem ser variáveis exógenas ou defasagens da variável dependente. Quando o número de variáveis é muito grande, é importante selecionar as variáveis relevantes. Para isso pode-se usar de conhecimento prévio sobre a série ou critérios de informação (AIC e BIC), que podem ser obtidos com uma aproximação linear da série. Essa pré-seleção reduz o tempo computacional, porém pode causar efeitos adversos na especificação do modelo final.

A seleção do nó a ser dividido e da variável de transição será realizada através de uma sequência de testes de Multiplicador de Lagrange (ML), seguindo a idéia apresentada em Luukkonen, Saikkone, and Teräsvirta (1988). Desta forma estamos usando inferência estatística para determinar o crescimento da árvore.

A estimação dos parâmetros do modelo será feita por mínimos quadrados não-lineares (MQNL), que é equivalente à Quase-Máxima Verossimilhança (QML), como discutido no Capítulo (2).

O próximo passo é a avaliação do modelo estimado. Em geral, modelos estruturados por árvores são avaliados pelo seu desempenho (habilidade preditiva) em dados fora da amostra (*out-of-sample*).

Seguindo o princípio “específico para geral”, começamos o ciclo de modelagem a partir do nó raiz (profundidade 0) e os passos gerais são:

1. Seleção das variáveis relevantes.
2. Especificação do modelo através da seleção, na profundidade  $d$ , do nó a ser dividido, se não for o nó raiz, e da variável de transição. Para isso usa-se o teste ML.
3. Estimação dos parâmetros.
4. Avaliação do modelo estimado, testando a necessidade de mudar o nó a ser dividido, mudar a variável de transição, ou remover a divisão do nó.
5. Usar o modelo final para fazer previsão ou para o uso descritivo.

O ciclo começa no nó raiz, testando a hipótese nula de um modelo constante global contra o mais simples modelo STAR-Tree, com 2 nós terminais. Para isso, precisamos estimar os parâmetros.

### 3.4. Estimação dos Parâmetros

Considerando o modelo STAR-Tree na equação (3.3), temos o vetor de parâmetros  $\psi$ , tal que  $\psi = (\beta_1, \dots, \beta_K, \theta'_1, \dots, \theta'_K)'$ , onde  $\beta_i$  é o vetor dos parâmetros lineares, e  $\theta'_i$  é o vetor dos parâmetros não-lineares  $(\gamma_j, c_j)$  do nó terminal  $i$  gerado pelo nó  $j$ .

Para a estimação dos parâmetros, leva-se em conta a hipótese de que a sequência  $\{\varepsilon_t\}_{t=1}^T$  é formada por variáveis aleatórias normalmente distribuídas

com média zero e variância constante  $\sigma^2 < \infty$ , desta forma, podemos usar o princípio da máxima verossimilhança, que sob hipótese de normalidade, é equivalente a mínimos quadrados não lineares. No caso de erros não gaussianos, podemos usar o estimador de quase-máxima verossimilhança (QMV).

O estimador de quase-máxima verossimilhança, que é equivalente ao de mínimos quadrados não lineares, do vetor de parâmetros do modelo é dado por:

$$\hat{\psi} = \underset{\psi \in \Psi}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^T q_t(\psi) = \underset{\psi \in \Psi}{\operatorname{argmin}} \left\{ \frac{1}{T} \sum_{t=1}^T [y_t - H_{\mathbb{J}\mathbb{T}}(x_t; \psi)]^2 \right\} \quad (3.6)$$

Condicionado ao conhecimento dos parâmetros  $\theta_i$  na equação (3.3),  $i = 1, \dots, K$ , o modelo da equação (3.3) é simplesmente uma regressão linear e o vetor de parâmetros  $\beta = (\beta_1, \dots, \beta_K)'$  pode ser estimado por mínimos quadrados ordinários (MQO) como

$$\hat{\beta} = [Z(\theta)'Z(\theta)]^{-1}Z(\theta)'y, \quad (3.7)$$

onde

$$y = (y_1, \dots, y_T)',$$

$$\theta = (\theta_1', \dots, \theta_k')'$$

e

$$Z(\theta) = \begin{bmatrix} \tilde{z}_1 B_1(x_1; \theta_1) & \cdots & \tilde{z}_1 B_K(x_1; \theta_K) \\ \vdots & \ddots & \vdots \\ \tilde{z}_T B_1(x_T; \theta_1) & \cdots & \tilde{z}_T B_K(x_T; \theta_K) \end{bmatrix}$$

Esse procedimento é chamado de mínimos quadrados concentrados.

Os parâmetros do vetor  $\theta_i$  são estimados condicionalmente a  $\beta$  através de mínimos quadrados não-lineares que faz uso de procedimentos de otimização não linear. No nosso caso usamos o algoritmo de Levenberg-Marquadt, completando a estimação da iteração.

Como o algoritmo de mínimos quadrados não lineares (MQNL) é muito sensível aos valores iniciais, é feita uma busca de valores iniciais por grid (grade de valores), que busca a maximização da função de log-verossimilhança concentrada.

### 3.5. Algoritmo de Crescimento

Considere um modelo STAR-Tree com  $K$  folhas, e queremos testar se o nó terminal  $i^* \in \mathbb{T}$  deve ou não ser dividido. Escrevemos o modelo como:

$$y_t = \sum_{i \in \mathbb{T} - \{i^*\}} \beta'_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \beta'_{2i^*+1} \tilde{z}_t B_{\mathbb{J}2i^*+1}(x_t; \theta_{2i^*+1}) + \beta'_{2i^*+2} \tilde{z}_t B_{\mathbb{J}2i^*+2}(x_t; \theta_{2i^*+2}) + \varepsilon_t, \quad (3.8)$$

onde

$$B_{\mathbb{J}2i^*+1}(x_t; \theta_{2i^*+1}) = B_{\mathbb{J}i^*}(x_t; \theta_{i^*})G(x_{i^*t}; \gamma_{i^*}, c_{i^*})$$

$$B_{\mathbb{J}2i^*+2}(x_t; \theta_{2i^*+2}) = B_{\mathbb{J}i^*}(x_t; \theta_{i^*})[1 - G(x_{i^*t}; \gamma_{i^*}, c_{i^*})]$$

A equação (3.8) pode ser escrita da forma:

$$y_t = \sum_{i \in \mathbb{T} - \{i^*\}} \beta'_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \phi' \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) + \lambda' \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*})G(x_{i^*t}; \gamma_{i^*}, c_{i^*}) + \varepsilon_t \quad (3.9)$$

onde  $\phi = \beta_{2i^*+2}$  e  $\beta_{2i^*+1} - \beta_{2i^*+2}$ .

Para testar a significância estatística da partição do nó, uma hipótese nula conveniente seria  $H_0: \gamma_{i^*} = 0$  contra a hipótese alternativa  $H_a: \gamma_{i^*} > 0$ . Olhando para a equação (3.9), percebe-se que hipótese nula equivalente seria  $H'_0: \lambda = 0$ . Contudo, na equação (3.9) fica claro que os parâmetros  $\lambda$  e  $c_{i^*}$  podem assumir diferentes valores sem, no entanto, alterar a função de quase-verossimilhança, gerando um problema de identificação, discutido em Davies (1977) e Davies (1987).

A solução adotada para resolver esse problema, proposta por Luukkonen, Saikkone, and Teräsvirta (1988), é aproximar a função logística  $G(\cdot)$  da equação (3.1) por uma expansão de Taylor de terceira ordem em torno de  $\gamma_{i^*} = 0$ . Isso nos dá a seguinte equação:

$$G(x_t; \gamma, c) = \frac{1}{2} + 2\gamma(x_t - c) + \frac{3}{2}\gamma^2(x_t - c)^2 + \frac{1}{2}\gamma^3(x_t - c)^3 + R_3(x_t; \gamma, c) \quad (3.10)$$

onde  $R_3(x_t; \gamma, c)$  é o resto.

Substituindo a equação (3.10) na equação (3.9), temos

$$y_t = \sum_{i \in \mathbb{T} - \{i^*\}} \beta'_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) + \alpha'_0 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) + \alpha'_1 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t} + \alpha'_2 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t}^2 + \alpha'_3 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t}^3 + e_t \quad (3.11)$$

onde  $e_t = \varepsilon_t + \lambda' \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) + R_3(x_{i^*t}; \gamma_{i^*}, c_{i^*})$ . Os parâmetros  $\alpha_k, k = 0, \dots, 3$  são funções dos parâmetros originais do modelo,  $(\phi, \lambda, \gamma_{i^*}, c_{i^*})$ .

Desta maneira, a hipótese nula pode ser construída da seguinte forma:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0 \quad (3.12)$$

Note que sob  $H_0$ , o resto da expansão de Taylor some, e temos que  $e_t = \varepsilon_t$ , logo, as propriedades dos resíduos permanecem inalteradas sob a hipótese nula e a inferência assintótica pode ser usada.

A função de log-verossimilhança para a  $t$ -ésima observação assume a forma:

$$l_t = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left\{ y_t - \sum_{i \in \mathbb{T} - \{i^*\}} \beta'_i \tilde{z}_t B_{\mathbb{J}i}(x_t; \theta_i) - \alpha'_0 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) - \alpha'_1 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t} - \alpha'_2 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t}^2 - \alpha'_3 \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \theta_{i^*}) x_{i^*t}^3 \right\}^2 \quad (3.13)$$

Derivando a equação (3.13) sob a hipótese nula, e assumindo que  $E(|x_{1t}|^\delta) < \infty, l = 1, \dots, q$ , para algum  $\delta > 6$ , e que  $\varepsilon_t \sim \text{NID}(0, \sigma^2)$ , temos o seguinte resultado para o teste ML:

$$ML = \frac{1}{\hat{\sigma}^2} \sum_{t=1}^T \hat{u}_t \hat{v}'_t \left\{ \sum_{t=1}^T \hat{v}_t \hat{v}'_t - \sum_{t=1}^T \hat{v}_t \hat{h}'_t \left( \sum_{t=1}^T \hat{h}_t \hat{h}'_t \right)^{-1} \sum_{t=1}^T \hat{h}_t \hat{v}'_t \right\}^{-1} \sum_{t=1}^T \hat{v}_t \hat{u}_t \quad (3.14)$$

onde

$$\hat{u}_t = y_t - H_{\mathbb{J}\mathbb{T}}(x_t; \hat{\psi}),$$

$$\hat{h}_t = \left. \frac{\partial H_{\mathbb{J}\mathbb{T}}(x_t; \psi)}{\partial \psi} \right|_{\psi=\hat{\psi}}$$

$$= \left[ \tilde{z}_t B_{\mathbb{J}i_1}(x_t; \psi), \dots, \tilde{z}_t B_{\mathbb{J}i_K}(x_t; \psi), \beta'_{i_1} \tilde{z}_t \left. \frac{\partial B_{\mathbb{J}i_1}(x_t; \theta_{i_1})}{\partial \theta_{i_1}} \right|_{\theta_{i_1}=\hat{\theta}_{i_1}}, \dots, \beta'_{i_K} \tilde{z}_t \left. \frac{\partial B_{\mathbb{J}i_K}(x_t; \theta_{i_K})}{\partial \theta_{i_K}} \right|_{\theta_{i_K}=\hat{\theta}_{i_K}} \right]'$$

e

$$\hat{v}_t = [\tilde{z}_t B_{\mathbb{J}i^*}(x_t; \hat{\theta}_{i^*}) x_{i^*t}, \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \hat{\theta}_{i^*}) x_{i^*t}^2, \tilde{z}_t B_{\mathbb{J}i^*}(x_t; \hat{\theta}_{i^*}) x_{i^*t}^3]'$$

seguem uma distribuição  $\chi^2$  assintótica com  $m = 3(p + 1)$  graus de liberdade.

A partir da suposição da normalidade dos erros, a matriz de informação é um bloco diagonal, e assim, podemos assumir que a variância do erro é constante.

O teste ML também pode ser aplicado na sua versão do teste F, seguindo os seguintes passos:

1. Estimar o modelo da equação (3.3) com  $K$  regimes. Para isso regredimos os resíduos  $\hat{u}_t$  em  $\hat{h}_t$  e calculamos a soma dos novos resíduos quadráticos  $SSR_0 = \sum_{t=1}^T \tilde{u}_t^2$ . Os novos resíduos  $\tilde{u}_t$  serão ortogonais à  $\hat{h}_t$ .
2. Fazer a regressão de  $\tilde{u}_t$  em  $\hat{h}_t$  e  $\hat{v}_t$ . Calcular a soma dos resíduos quadráticos obtidos a partir desta regressão,  $SSR_1 = \sum_{t=1}^T \hat{v}_t^2$ .
3. Calcular a estatística  $\chi^2$

$$ML_{\chi^2} = T \frac{SSR_0 - SSR_1}{SSR_0}, \quad (3.15)$$

ou a versão  $F$  do teste

$$ML_F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(T - n - m)}, \quad (3.16)$$

onde  $n = (q + 2)h + p + 1$ .

Sob  $H_0$ , o teste  $ML_{\chi^2}$  segue uma distribuição  $\chi^2$  assintótica com  $m$  graus de liberdade, e  $ML_F$  segue aproximadamente uma distribuição  $F$  com  $m$  e  $T - n - m$  graus de liberdade, onde  $T$  é o número de observações.

### 3.6. Ciclo de Modelagem – Sequência de Testes

Uma vez definidos a estimação do modelo e o teste ML para o crescimento da árvore, o algoritmo da modelagem se torna simples.

Uma sequência de  $n$  testes de hipótese do tipo ML, como descrito na Seção (2.5), é executada. Neste momento, decidir dividir um nó de maneira errônea pode ser muito prejudicial à complexidade da árvore, já que esta é função do número de nós terminais. Deste modo, temos que tentar controlar o erro do tipo 1 (mais divisões de nós são consideradas significantes do que deveriam ser, ou seja, a árvore é superestimada). Para isso adotamos um procedimento de diminuir o nível de significância do teste conforme a árvore cresce.

Para o  $n$ -ésimo teste na sequência, se ele é aplicado em um nó na  $d$ -ésima profundidade, o nível de significância do teste é

$$\alpha(d, n) = \frac{\alpha}{n^d}, \quad (3.17)$$

onde  $\alpha$  é o nível de significância do primeiro teste.

Forçando o teste a ser mais rigoroso nas maiores profundidades, diminuimos a importância do uso de técnicas de podagem a posteriori, citadas no Capítulo (2).

Para cada divisão, é feita a estimação dos parâmetros, e desta forma o modelo final é estimado.

### 3.7. Previsão

A flexibilidade da metodologia STAR-Tree permite que sejam criados três tipos de previsão diferentes:

1. Combinação de Regimes (RC):

Aplicação direta do modelo estimado. Utiliza-se a equação resultante da estimação para obter a previsão um passo a frente. Assim, a previsão é obtida a partir da soma ponderada dos modelos lineares locais, na qual os pesos são as pertinências das observações em cada regime.

2. Máxima Pertinência (MM):

Observa-se em qual regime a observação apresenta maior pertinência e aplica-se o modelo local correspondente.

3. Combinação Adaptativa de Regimes (ARC):

Metodologia similar à RC, porém os parâmetros lineares são re-estimados a cada passo da previsão, utilizando as últimas 252 observações, que correspondem a um ano de observação.