

## 2 Modelagem

Este capítulo apresenta o arcabouço econométrico para a análise das séries de volatilidade. A discussão do modelo estruturado em árvore com transições suaves é baseada em da Rosa, Veiga e Medeiros (2003) e Medeiros, da Rosa e Veiga (2005), onde detalhes e provas podem ser encontrados.

### 2.1 Uma Breve Introdução às árvores de Regressão

Seja  $\mathbf{x}_t = (x_{1t}, \dots, x_{qt})' \in \mathbb{X} \subseteq \mathbb{R}^q$  um vetor contendo  $q$  variáveis explicativas (covariadas ou preditoras) para uma resposta contínua univariada  $RV_t \in \mathbb{R}$ ,  $t = 1, \dots, T$ . Suponha que a relação entre  $RV_t$  e  $\mathbf{x}_t$  segue um modelo de regressão da forma

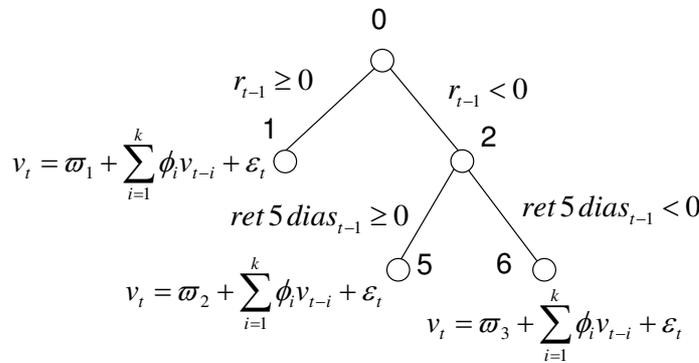
$$RV_t = f(\mathbf{x}_t) + \varepsilon_t, \quad (2-1)$$

onde a função  $f(\cdot)$  é desconhecida e não há hipóteses sobre a distribuição do termo de erro aleatório  $\varepsilon_t$  –a princípio. Uma árvore de regressão é um modelo não-paramétrico baseado numa partição recursiva do espaço  $\mathbb{X}$  que aproxima a função  $f(\cdot)$  como uma soma de modelos locais, cada qual determinado por  $K \in \mathbb{N}$  regiões (partições) diferentes de  $\mathbb{X}$ . Usualmente, o modelo é representado por um gráfico com o formato de uma árvore de decisão binária germinando  $N \in \mathbb{N}$  nós divisores e  $K \in \mathbb{N}$  nós terminais (também chamados de folhas), crescendo de um nó original até os terminais. Geralmente, as partições são definidas por um conjunto de hiperplanos ortogonais ao eixo de uma dada variável preditora, designada *variável de transição*. A referência mais importante em modelos de regressão em árvore é a abordagem de Classificação e Árvores de Regressão (*Classification and Regression Trees*, CART) proposta por Breiman, Friedman, Olshen e Stone (1984), contexto onde os modelos locais são apenas constantes.

Com o objetivo de expressar matematicamente o modelo de regressão em árvore, a seguinte notação é introduzida. O nó original está na posição 0 e o nó divisor  $j$  gera dois novos nós nas posições  $2j + 1$  e  $2j + 2$ , respectivamente. Cada nó divisor está associado com uma variável de transição  $x_{s_j t} \in \mathbf{x}_t$ , onde  $s_j \in \mathbb{S} = \{1, 2, \dots, q\}$ . Além disso, sejam  $\mathbb{J}$  e  $\mathbb{T}$  os conjuntos de índices dos

nós divisores e terminais, respectivamente; a arquitetura da árvore pode ser determinada integralmente por  $\mathbb{J}$  e  $\mathbb{T}$ .

**Exemplo** Considere um modelo de volatilidade com transições de regimes que permite múltiplos estados associados com efeitos de alavancagem, onde o efeito de um retorno negativo sobre a volatilidade do dia seguinte depende do comportamento dos retornos na semana anterior. Suponha que a série segue uma processo localmente constante, onde a média condicional depende do sinal do retorno do retorno no dia passado, onde esse efeito em si é mais fraco em semanas “boas” (valorização nos últimos cinco dias úteis) do que em semanas “ruins” (retorno negativo no mesmo período):  $v_t = \omega_1 + \varepsilon_t$  se  $r_{t-1} \geq 0$ ,  $v_t = \omega_2 + \varepsilon_t$  se  $r_{t-1} < 0$  e  $ret\ 5\ dias_{t-1} \geq 0$ ,  $v_t = \omega_2 + \varepsilon_t$  se  $r_{t-1} < 0$  e  $ret\ 5\ dias_{t-1} < 0$ , onde  $v_t$  é a volatilidade,  $r_{t-1}$  é o retorno defasado,  $\varepsilon_t$  é um ruído *i.i.d* e  $\omega_3 > \omega_2 > \omega_1$  são constantes. Esse modelo pode ser descrito na notação em árvore de regressão por dois nós divisores nas posições 0 e 2 ( $N = 2$ ,  $\mathbb{J} = \{0, 2\}$ ) associados respectivamente com com retornos diários e semanais defasados (determinando sequencialmente a presença e a intensidade do efeito de alavancagem) e três folhas ou nós terminais nas posições 1, 5 e 6 ( $K = 3$ ,  $\mathbb{T} = \{1, 5, 6\}$ ) representando modelos locais.



## 2.2

### Regressão Estruturada em Árvore com Transições Suaves

O modelo de regressão estruturado em árvore com transições suaves (*Smooth Transition Tree*, abreviado STR-Tree), introduzido por da Rosa, Veiga e Medeiros (2003), tem como refinamento a substituição das divisões discretas na árvore de regressão por transições suaves, prescritas por uma função logística definida como

$$G(x; \gamma, c) = \frac{1}{1 + e^{-\gamma(x-c)}}. \tag{2-2}$$

O parâmetro  $\gamma$ , chamado de *parâmetro de inclinação*, controla a suavidade da função logística, enquanto o parâmetro  $c$ , denominado *parâmetro de locação*, determina a partição do espaço  $\mathbb{X}$ . O modelo de regressão em árvore da seção anterior está contido na especificação com transições suaves como um caso especial obtido quando o parâmetro de inclinação tende ao infinito.

Neste trabalho, a seguinte especificação aumentada do modelo em árvore de da Rosa, Veiga e Medeiros (2003) é considerada:

**Definição 2.1** *Seja  $\mathbf{z}_t \in \mathbb{R}^p$  e  $\mathbf{x}_t \in \mathbb{R}^q$  conjuntos de variáveis explicativas. Defina  $\tilde{\mathbf{z}}_t = [1, \mathbf{z}_t']'$ . Um modelo paramétrico  $\mathcal{M}$  definido pela função  $H_{\mathbb{T}}(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\psi}) : \mathbb{R}^{\max(p,q)} \rightarrow \mathbb{R}$ , indexado pelo vetor de parâmetros  $\boldsymbol{\psi} \in \Psi$ , um conjunto compacto do espaço Euclidiano, é chamado de modelo de regressão em árvore com transições suaves se*

$$H_{\mathbb{T}}(\mathbf{x}_t; \boldsymbol{\psi}) = \boldsymbol{\lambda}' \tilde{\mathbf{z}}_t + \sum_{i \in \mathbb{T}} \beta_i B_{\mathbb{J}_i}(\mathbf{x}_t; \boldsymbol{\theta}_i) \quad (2-3)$$

onde

$$B_{\mathbb{J}_i}(\mathbf{x}_t; \boldsymbol{\theta}_i) = \prod_{j \in \mathbb{J}} G(x_{s_j, t}; \gamma_j, c_j)^{\frac{n_{i,j}(1+n_{i,j})}{2}} [1 - G(x_{s_j, t}; \gamma_j, c_j)]^{(1-n_{i,j})(1+n_{i,j})} \quad (2-4)$$

e

$$n_{i,j} = \begin{cases} -1 & \text{se o caminho até a folha } i \text{ não inclui o nó divisor } j; \\ 0 & \text{se o caminho até a folha } i \text{ inclui o nó direito} \\ & \text{da divisão em } j; \\ 1 & \text{se o caminho até a folha } i \text{ inclui o nó esquerdo} \\ & \text{da divisão em } j. \end{cases} \quad (2-5)$$

Seja  $\mathbb{J}_i$  o subconjunto de  $\mathbb{J}$  contendo os índices dos nós divisores que formam o caminho até a folha  $i$ . Então  $\boldsymbol{\theta}_i$  é o vetor contendo todos os parâmetros  $(\gamma_k, c_k)$  tais que  $k \in \mathbb{J}_i$ ,  $i \in \mathbb{T}$ .  $B_{\mathbb{J}_i}$ ,  $0 < B_{\mathbb{J}_i} < 1$  Note que  $\sum_{j \in \mathbb{J}} B_{\mathbb{J}_j}(\mathbf{x}_t; \boldsymbol{\theta}_j) = 1$ ,  $\forall \mathbf{x}_t \in \mathbb{R}^{q+1}$ .

## 2.3

### Expandindo a Árvore

Esta seção apresenta a teoria de inferência que irá nortear os passos de construção dos modelos de volatilidade descritos no próximo capítulo. A escolha das variáveis relevantes, a seleção dos nós a serem divididos (se for o caso) e a definição das variáveis de transição são concretizadas através de uma seqüência de testes de multiplicadores de Lagrange (LM) seguindo as idéias apresentadas originalmente em Luukkonen, Saikkonen e Terasvirta (1988), extensamente adotadas na literatura.

Considere um modelo em árvore com  $K$  folhas. Queremos testar se o nó terminal  $i^* \in \mathbb{T}$  deve ser segmentado ou não. Escreva o modelo como

$$RV_t = \lambda \tilde{\mathbf{z}}_t + \sum_{i \in \mathbb{T} - \{i^*\}} \beta_i B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i) + \beta_{2i^*+1} B_{\mathbb{J}2i^*+1}(\mathbf{x}_t; \boldsymbol{\theta}_{2i^*+1}) + \beta_{2i^*+2} B_{\mathbb{J}2i^*+2}(\mathbf{x}_t; \boldsymbol{\theta}_{2i^*+2}) + \varepsilon_t, \quad (2-6)$$

onde

$$B_{\mathbb{J}2i^*+1}(\mathbf{x}_t; \boldsymbol{\theta}_{2i^*+1}) = B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) G(x_{i^*t}; \gamma_{i^*}, c_{i^*})$$

$$B_{\mathbb{J}2i^*+2}(\mathbf{x}_t; \boldsymbol{\theta}_{2i^*+2}) = B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) [1 - G(x_{i^*t}; \gamma_{i^*}, c_{i^*})].$$

A equação (2-6) pode ser escrita mais compactamente como

$$RV_t = \sum_{i \in \mathbb{T} - \{i^*\}} \beta_i B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i) + \phi_1 B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) + \phi_2 B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) G(x_{i^*t}; \gamma_{i^*}, c_{i^*}) + \varepsilon_t, \quad (2-7)$$

onde  $\phi_1 = \beta_{2i^*+2}$  e  $\phi_2 = \beta_{2i^*+1} - \beta_{2i^*+2}$ .

Para testar a significância estatística da divisão, a hipótese nula conveniente é  $H_0 : \gamma_{i^*} = 0$ , contra a alternativa  $H_a : \gamma_{i^*} > 0$ . Uma outra hipótese nula possível é dada por  $H'_0 : \phi_2 = 0$ . Contudo, fica claro pela equação (2-7) que sob  $H_0$ , os parâmetros  $\phi_2$  e  $c_{i^*}$  podem assumir diferentes valores sem qualquer impacto na função de quase-máxima verossimilhança, o que interpõe um problema de identificação (Davies, 1977, 1987).

Uma solução para essa dificuldade, proposta em Luukkonen, Saikkonen e Terasvirta (1988), está em aproximar a função logística por uma expansão de Taylor de terceira ordem em volta de  $\gamma_{i^*} = 0$ . Com um pouco de álgebra,

$$RV_t = \sum_{i \in \mathbb{T} - \{i^*\}} \beta_i B_{\mathbb{J}i}(\mathbf{x}_t; \boldsymbol{\theta}_i) + \alpha_0 B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) + \alpha_1 B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) x_{i^*t} + \alpha_2 B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) x_{i^*t}^2 + \alpha_3 B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) x_{i^*t}^3 + e_t, \quad (2-8)$$

onde  $e_t = \varepsilon_t + \phi_2 B_{\mathbb{J}i^*}(\mathbf{x}_t; \boldsymbol{\theta}_{i^*}) R(x_{i^*t}; \gamma_{i^*}, c_{i^*})$  e  $R(x_{i^*t}; \gamma_{i^*}, c_{i^*})$  é o resto. Os parâmetros  $\alpha_k$ ,  $k = 0, \dots, 3$  são funções dos parâmetros originais do modelo.

Assim sendo, a hipótese nula é convertida em

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0. \quad (2-9)$$

Note que sob  $H_0$  o resto da expansão de Taylor desaparece e  $e_t = \varepsilon_t$ , de forma que as propriedades do processo de erro permanecem inalteradas sob a hipótese nula – a inferência assintótica pode ser utilizada. A estatística de teste é dada

por: <sup>1</sup>

$$LM = \frac{1}{\widehat{RV}^2} \sum_{t=1}^T \hat{u}_t \hat{\nu}_t' \left\{ \sum_{t=1}^T \hat{\nu}_t \hat{\nu}_t' - \sum_{t=1}^T \hat{\nu}_t \hat{\mathbf{h}}_t' \left( \sum_{t=1}^T \hat{\mathbf{h}}_t \hat{\mathbf{h}}_t' \right)^{-1} \sum_{t=1}^T \hat{\mathbf{h}}_t \hat{\nu}_t' \right\}^{-1} \sum_{t=1}^T \hat{\nu}_t \hat{u}_t \quad (2-10)$$

onde  $\hat{u}_t = y_t - H_{\text{JT}}(\mathbf{z}_t, \mathbf{x}_t; \hat{\boldsymbol{\psi}})$ ,  $\hat{\mathbf{h}}_t = \left. \frac{\partial H_{\text{JT}}(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\psi})'}{\partial \boldsymbol{\psi}} \right|_{H_0}$  e

$$\boldsymbol{\nu}_t = [\tilde{\mathbf{z}}_t B_{\text{JJ}i^*} x_{i^*t}, \tilde{\mathbf{z}}_t B_{\text{JJ}i^*} x_{i^*t}^2, \tilde{\mathbf{z}}_t B_{\text{JJ}i^*} x_{i^*t}^3],$$

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

Como a premissa de normalidade e homocedasticidade dos erros é claramente violada nos dados, uma versão robusta do teste LM baseada nos resultados de Wooldridge (1990) é aplicada. O teste é implementado sequencialmente da seguinte maneira:

1. Estima-se o modelo com  $K$  regimes. Se o tamanho da amostra é reduzido e portanto o modelo é difícil de ser estimado, problemas numéricos na aplicação do algoritmo de máxima verossimilhança podem conduzir a uma solução onde o vetor de resíduos não é exatamente ortogonal à matriz gradiente de  $H_{\text{JT}}(\mathbf{x}_t; \hat{\boldsymbol{\psi}})$ , com efeitos adversos sobre o tamanho empírico do teste. Para contornar esse problema, os resíduos  $\hat{u}_t$  são regredidos em  $\hat{\mathbf{h}}_t$  e a soma dos quadrados dos resíduos  $SSR_0 = \sum_{t=1}^T \tilde{u}_t^2$  é computada. Os novos resíduos  $\tilde{u}_t$  são ortogonais a  $\hat{\mathbf{h}}_t$ .
2.  $\hat{\nu}_t$  é regredido em  $\hat{\mathbf{h}}_t$  e os resíduos  $\mathbf{r}_t$  computados.
3. Um vetor de uns é regredido em  $\tilde{\varepsilon}_t \mathbf{r}_t$  e a soma dos quadrados dos resíduos  $SSR_1$  é calculada.
4. O valor da estatística de teste é dado por

$$LM_{\chi^2}^r = T - SSR_1. \quad (2-11)$$

Sob  $H_0$ ,  $LM_{\chi^2}^{hn}$  tem distribuição assintótica  $\chi^2$  com  $m$  graus de liberdade.

Os parâmetros do modelo são estimados por mínimos quadrados não-lineares (NLS), que nesse caso equivalem à estimação por quase-máxima verossimilhança (QML). O estimador de quase-máxima verossimilhança (QMLE) de  $\boldsymbol{\psi}$ ,  $\hat{\boldsymbol{\psi}}$ , é dado por

<sup>1</sup>Veja Medeiros, da Rosa e Veiga (2005) acerca das condições técnicas para a validade da estatística de teste.

$$\begin{aligned}\hat{\psi} &= \underset{\psi \in \Psi}{\operatorname{argmin}} \mathcal{Q}_T(\psi) = \underset{\psi \in \Psi}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^T q_t(\psi) \\ &= \underset{\psi \in \Psi}{\operatorname{argmin}} \left\{ \frac{1}{T} \sum_{t=1}^T [RV_t - H_{\mathbb{J}\mathbb{T}}(\mathbf{z}_t, \mathbf{x}_t; \psi)]^2 \right\}.\end{aligned}\tag{2-12}$$