

**Joel Maurício Corrêa da
Rosa**

**Modelos de Regressão com
Transição Suave
Estruturados por Árvores**

TESE DE DOUTORADO

**DEPARTAMENTO DE ENGENHARIA ELÉTRICA
Programa de Pós-graduação em
Engenharia Elétrica**

Rio de Janeiro
Fevereiro de 2005

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



Joel Maurício Corrêa da Rosa

**Modelos de Regressão com Transição
Suave Estruturados por Árvores**

Tese de Doutorado

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Engenharia Elétrica

Orientador: Prof. Dr. Álvaro Veiga Lima Filho
Co-Orientador: Prof. Dr. Marcelo Cunha Medeiros

Rio de Janeiro
Fevereiro de 2005



Joel Maurício Corrêa da Rosa

**Modelos de Regressão com Transição
Suave Estruturados por Árvores**

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Engenharia Elétrica Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Dr. Álvaro Veiga Lima Filho

Orientador

Departamento de Engenharia Elétrica — PUC-Rio

Prof. Dr. Marcelo Cunha Medeiros

Co-Orientador

Departamento de Economia — PUC-Rio

Prof. Dr. Alexandre Xavier Ywata de Carvalho

IPEA - Instituto de Pesquisa Econômica e Aplicada

Prof. Dr. Alexandre Pinto Alves da Silva

COPPE/UFRJ

Prof. Dr. Caio Ibsen Rodrigues de Almeida

IBMEC

Prof. José Eugênio Leal

Coordenador Setorial do Centro Técnico Científico —

PUC-Rio

Rio de Janeiro, 2 de Fevereiro de 2005

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Joel Maurício Corrêa da Rosa

Graduou-se pela Escola Nacional de Ciências Estatísticas (ENCE) no Rio de Janeiro. cursou o Mestrado na Universidade de Campinas (UNICAMP) onde obteve o título de Mestre em Estatística. Trabalhou durante o mestrado com a análise de séries temporais para dados de contagem. Foi admitido em 1994 como Professor Auxiliar do Departamento de Estatística da Universidade Federal do Paraná onde trabalha até os dias atuais. Durante o doutorado passou o período de outubro de 2001 a abril de 2002 na Escola de Economia de Estocolmo onde fez doutorado sanduiche sob a supervisão do Professor Timo Terasvirta.

Ficha Catalográfica

Corrêa da Rosa, J. M.

Modelos de Regressão com Transição Suave Estruturados por Árvores/ Joel Maurício Corrêa da Rosa; orientador: Dr. Álvaro Veiga Lima Filho; co-orientador: Dr. Marcelo Cunha Medeiros. — Rio de Janeiro : PUC–Rio, Departamento de Engenharia Elétrica, 2005.

v., 157 f: il. ; 29,7 cm

1. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui referências bibliográficas.

Árvores de Regressão, Modelos Não-Lineares, Regressão com Transição Suave, Análise de Séries Temporais, Particionamento Recursivo. I. Veiga, A.. II. Medeiros, M.C.. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Agradecimentos

A Deus pelo seu amor e fidelidade durante toda esta caminhada.

Ao Prof. Dr. Álvaro Lima Veiga Filho por toda a compreensão e incentivo e valiosa orientação

Ao Prof. Dr. Marcelo Cunha Medeiros pelo constante apoio e incentivo que foram fundamentais para a elaboração deste trabalho.

Ao Prof. Dr. Timo Teräsvirta que tornou possível um período de valioso aprendizado na Escola de Economia de Estocolmo na Suécia.

À meus pais, irmãos e tios pelo apoio incondicional.

A todo corpo docente do Departamento de Estatística da Universidade Federal do Paraná que assumiu todos os encargos didáticos neste período de afastamento e prestou admirável incentivo ao longo de todo percurso.

A CAPES pelo auxílio financeiro concedido através do extinto programa PICDT.

Ao CNPQ pelo auxílio concedido durante o período de intercâmbio na Suécia.

A Maria Aparecida Thiengo pela compreensão, apoio, incentivo e ajuda na elaboração deste material.

Aos amigos Annistina e Mika pelo companheirismo nas horas de lazer que foram revitalizantes.

À Pernilla Watson pelo apoio "logístico" que auxiliou a minha estadia em Estocolmo.

À Denize Oliveira e Celso Sá pelos ensinamentos sobre a vida acadêmica.

A Cláudio Sá pelos momentos de apoio e companheirismo.

Aos grandes amigos Jorge e Marcel cuja amizade será eterna.

Aos amigos Hans Malmstaden, Gianluigi Rech, Anna Persson, Markus, Jana, Bruno, Andres Gonzales,

À amiga Malin Hornell pela ajuda no período de adaptação à Suécia.

Aos muitos amigos que fiz em Estocolmo.

Aos vizinhos amigos Cristina, Eduardo e seus filhos Álvaro e Maria Eduarda pela agradável companhia na cidade do Rio de Janeiro.

A Glaucy Ortiz pelos momentos de companheirismo e apoio.

Aos amigos José Vicente, Alexandre Zanini, Mayte, Kátia, Roberto, Eliana, Dênis, Christian, Lúcio, André Cunha, Ana Paula Sobral pelos momentos de descontração e companheirismo.

Lia, Danny e Ana Alice pelo companheirismo nos meses que antecederam o término deste trabalho.

A todos os não citados que contribuíram direta ou indiretamente para a conclusão deste trabalho.

Resumo

Corrêa da Rosa, J. M.; Veiga, A.; Medeiros, M.C.. **Modelos de Regressão com Transição Suave Estruturados por Árvores**. Rio de Janeiro, 2005. 157p. Tese de Doutorado — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo principal desta tese é introduzir um modelo estruturado por árvores que combina aspectos de duas metodologias: CART (Classification and Regression Tree) e STR (Smooth Transition Regression). O modelo é aqui denominado STR-Tree. A idéia é especificar um modelo não-linear paramétrico através da estrutura de uma árvore de decisão binária. O modelo resultante pode ser analisado como uma regressão com transição suave entre múltiplos regimes. As decisões sobre as divisões dos nós são inteiramente baseadas em testes do tipo Multiplicadores de Lagrange. Uma especificação alternativa baseada em validação cruzada também é utilizada. Um experimento de Monte Carlo é utilizado para avaliar o desempenho da metodologia proposta comparando-a com outras técnicas comumente utilizadas. Como resultado verifica-se que o modelo STR-Tree supera o tradicional CART quando seleciona a arquitetura de árvores simuladas. Além do mais, utilizar testes do tipo Multiplicadores de Lagrange gera resultados melhores do que procedimentos de validação cruzada. Quando foram utilizadas bases de dados reais, o modelo STR-Tree demonstrou habilidade preditiva superior ao CART. Através de uma aplicação, estende-se a metodologia para a análise de séries temporais. Neste caso, o modelo é denominado STAR-Tree, sendo obtido através de uma árvore de decisão binária que ajusta modelos autoregressivos de primeira ordem nos regimes. A série de retornos da taxa de câmbio Euro/Dólar foi modelada e a capacidade preditiva e o desempenho financeiro do modelo foi comparado com metodologias padrões como previsões ingênuas e modelos ARMA. Como resultado é obtido um modelo parcimonioso que apresenta desempenho estatístico equivalente às estratégias convencionais, porém obtendo resultados financeiros superiores.

Palavras-chave

Árvores de Regressão, Modelos Não-Lineares, Regressão com Transição Suave, Análise de Séries Temporais, Particionamento Recursivo.

Abstract

Corrêa da Rosa, J. M.; Veiga, A.; Medeiros, M.C.. T. Rio de Janeiro, 2005. 157p. PhD. Thesis — Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The main goal of this Thesis is to introduce a tree-structured model that combines aspects from two methodologies: CART (Classification and Regression Trees) and STR (Smooth Transition Regression). The model is called STR-Tree. The idea is to specify a nonlinear parametric model through the structure of a binary decision tree. The resulting model can be analyzed as a smooth transition regression model with multiple regimes. The decisions for splitting the nodes of the tree are entirely based on Lagrange Multipliers tests. An alternative specification that uses cross-validation is also tried. A Monte Carlo Experiment is used to evaluate the performance of the proposed methodology and to compare with other techniques that are commonly used. The results showed that the STR-Tree model outperformed the traditional CART when specifying the architecture of a simulated tree. Moreover, the use of Lagrange Multipliers tests gave better results than a cross-validation procedure. After applying the model to real datasets, it could be seen that STR-Tree showed superior predictive ability when compared to CART. The idea was extended to time series analysis through an application. In this situation, we call the model as STAR-Tree which is obtained through a binary decision tree that fits first-order autoregressive models for different regimes. The model was fitted to the returns of Euro/Dolar exchange rate time series and then evaluated statistically and financially. Comparing with the naive approach and ARMA methodology, the STAR-Tree was parsimonious and presented statistical performance equivalent to others. The financial results were better than the others.

Keywords

Regression Trees, Nonlinear Models, Smooth Transition Regression, Time Series Analysis, Recursive Partitioning.

Conteúdo

1	Introdução	14
2	Modelos Não-Lineares	18
2.1	Introdução - Modelos Lineares como Inspiração	18
2.2	Abordagem Econométrica	25
2.3	Abordagem nas Ciências da Computação	31
2.4	Abordagens Híbridas	34
3	Árvores de Classificação e Regressão	36
3.1	Introdução	36
3.2	Formulação Matemática	39
3.3	O Algoritmo CART	43
3.4	Podagem	46
3.5	Outros Desenvolvimentos	47
4	Um olhar mais Próximo da Estimação do modelo LSTAR(1)	49
4.1	Introdução	49
4.2	Experimento de Monte Carlo	53
4.3	Resultados	58
4.4	Resultados Condicionais ao Teste ML	79
4.5	Conclusões	85
5	O Modelo STR-Tree	86
5.1	Introdução	86
5.2	Construção do Modelo	87
5.3	Estimação de Parâmetros	88
5.4	Divisão dos Nós	93
5.5	Experimento de Monte Carlo	101
5.6	Aplicação à Dados Reais	108
5.7	Conclusões	114
6	O Modelo STAR-Tree	116
6.1	Introdução	116
6.2	Modelos de Múltiplos Regimes para Séries Temporais	117
6.3	O Modelo STAR-Tree	118
6.4	Aplicação à Série de Retornos Euro/Dolar	123
6.5	Conclusões	133
	Referências Bibliográficas	134
A	Vício nos Parâmetros do LSTAR(1)	144
B	Demonstrações dos Teoremas	147
B.1	Demonstração do Teorema 5.1	147

B.2	Demonstração do Teorema 5.2	147
B.3	Prova do Teorema 5.3	148
B.4	Prova do Teorema 5.4	150
C	Especificação de Modelos	151
C.1	Resultados de Simulações	151

Lista de Figuras

2.1	Sistema de Inferência Nebulosa	33
3.1	Terminologia de um Modelo Estruturado por Árvore	38
3.2	Enumeração dos Nós de uma Árvore	38
3.3	Saída Gráfica de uma Árvore de Regressão.	40
3.4	Arquitetura da Árvore de Regressão Mais Simples .	41
3.5	Árvore de Regressão com 3 Folhas Representando (3-6).	42
3.6	Árvore de Regressão com 4 Folhas Representando(3-7).	42
3.7	Algoritmo de Crescimento da Árvore	45
3.8	Algoritmo de Divisão do Nó	46
4.1	Superfície Suporte do Modelo LSTAR1 para 3000 Valores Simulados para $(\gamma, c, \alpha_0, \alpha_1, \beta_0, \beta_1) = (1, 0, 0, 0, -0.5, 0.5)$	52
4.2	Diferentes Níveis de Suavidade para a Função Logística	53
4.3	Valores de β_0 e β_1 nas Simulações	54
4.4	Simulação de 1000 Valores do processo LSTAR(1) Para os Coeficientes $\beta_0 = -0.5$ e $\beta_1 = 0.5$	55
4.5	Simulação de 1000 Valores de Processos LSTAR(1) com $\beta_0 < 0$ e $\beta_1 > 0$	56
4.6	Função Logística para o Processo LSTAR(1) com $(\gamma, c, \beta_0, \beta_1) = (2.5, 0, -0.5, 0.995)$ T=150	60
4.7	Diagrama de Dispersão entre as Estimativas de Locação e β_0 para Simulações com $(\gamma, c, \beta_0, \beta_1) = (1, 0, 0.9, 0.1)$ and $T = 150$	62
4.8	Vício Mediano na Estimação dos Parâmetros de Locação	66
4.9	Histograma das Estimativas de Locação para $(\gamma, c, \beta_0, \beta_1) = (50, 0, 0.9, 0.1)$ e $(\gamma, c, \beta_0, \beta_1) = (50, 0, 0.1, 0.9)$ com $T = 150$	67
4.10	Desvio Absoluto Mediano na Estimação dos Parâmetros de Locação	69
4.11	Diagrama de Dispersão da Pertinência(1º Regime) vs as Estimativas de Locação em 1000 Processos Simulados com $(\beta_0, \beta_1) = (-0.5, 0.995)$ $T = 150$	70
4.12	Vício Relativo Mediano na Estimação de β_0	73
4.13	Vício Mediano Relativo na Estimação de β_1	74
4.14	Desvio Absoluto Mediano na Estimação de β_0	76
4.15	Desvio Absoluto Mediano na Estimação de β_1	78
4.16	Histograma da Estimativas da locação segundo Resultados do Teste ML para Simulações do Processo $(\gamma, c, \beta_0, \beta_1) = (5, 0, -0.5, 0.995)$ com $T = 150$	81
4.17	Histograma da Estimativas de β_0 segundo Resultados do Teste ML para Simulações do Processo $(\gamma, c, \beta_0, \beta_1) = (5, 0, -0.5, 0.995)$ com $T = 150$	84

4.18	Histograma da Estimativas de β_1 segundo Resultados do Teste ML para Simulações do Processo $(\gamma, c, \beta_0, \beta_1) = (5, 0, -0.5, 0.995)$ com $T = 150$	84
5.1	Ciclo de Modelagem do modelo STR-Tree	89
5.2	Nível de significância durante a seqüência de testes para divisão dos nós - um exemplo hipotético	100
5.3	Arquiteturas Simuladas	102
5.4	Características Geométricas dos Modelos Simulados (Arquitetura I)	104
5.5	Diagrama de dispersão entre as estimativas do parâmetro de locação c_2 e da constante dentro do nó para o modelo 1.1 (T=150)	106
5.6	Modelo STR-Tree para o Conjunto <i>Boston Housing</i>	110
5.7	Regra(nebulosa) de predição associada as observações no interior do nó 10	111
6.1	Parâmetros Do Modelo STAR-Tree para Árvore mais Simples	119
6.2	Possibilidades de arquiteturas na árvore com 3 folhas	122
6.3	Valor do Euro para 1 Dólar Americano- 1994 a 2001	124
6.4	Série de retornos de EUR/USD 1994 a 2001	125
6.5	Fitted STAR-Tree Model to Euro/Dollar Returns	128
6.6	Seqüência de Resíduos	129
6.7	Histograma dos Resíduos	130
6.8	Gráfico de Probabilidades Normais dos Resíduos	130

Lista de Tabelas

4.1	Desvio Padrão em Processos LSTAR	60
4.2	Pertinência ao Primeiro Regime em Processos LSTAR	61
4.3	Coefficiente de Correlação entre y_t e $F(y_t; \theta)$	61
4.4	Vício Médio nas Estimativas da Locação	63
4.5	Vício Mediano nas Estimativas da Locação	64
4.6	Erro Quadrático Médio nas Estimativas da Locação	68
4.7	Desvio Absoluto Mediano nas Estimativas da Locação	68
4.8	Vício Mediano Relativo nas Estimativas de β_0	72
4.9	Vício Mediano Relativo nas Estimativas de β_1	72
4.10	Desvio Absoluto Mediano nas Estimativas de β_0	77
4.11	Desvio Absoluto Mediano nas Estimativas de β_1	79
4.12	Tamanho Empírico do Teste ML	80
4.13	Estimativas Medianas dos Parâmetros de Acordo com o Resultado do Teste ML Processos em que $\beta_0 > 0, \beta_1 > 0$	82
4.14	Estimativas Medianas dos Parâmetros de Acordo com o Resultado do Teste ML Processos em que $\beta_0 < 0, \beta_1 < 0$	83
5.1	Suavidade das Divisões nas Simulações do STR-Tree	102
5.2	Parâmetros do Modelo STR-Tree Simulado	103
5.3	Estatísticas Descritivas para a Estimação na Arquitetura I	105
5.4	Estatísticas Descritivas para a Estimação na Arquitetura II	106
5.5	Percentual de Especificações Corretas para Árvore Simuladas na Arquitetura I	107
5.6	Percentual de Especificações Corretas para Árvore Simuladas na Arquitetura II	107
5.7	Estimativas dos Parâmetros Não-Lineares do Modelo STR-Tree Ajustado ao Conjunto <i>Boston Housing</i>	111
5.8	Erro Quadrático Médio Fora-da-Amostra dos algoritmos CART e STR-Tree baseados em 100 Observações. observations.	112
5.9	Número de Folhas Especificados pelo CART e STR-Tree baseados em 100 Observações.	113
5.10	Tempo (em segundos) gasto pelos algoritmos CART e STR-Tree com base em 100 observações.	114
6.1	Parâmetros do Modelo MRSTAR	118
6.2	Descrição da Série Temporal EUR/USD	125
6.3	Descrição das Variáveis Exógenas	126
6.4	Medidas Estatísticas para Avaliar o Desempenho dos Modelos	126
6.5	Medidas Financeiras para Avaliação do Modelo	127
6.6	Modelos AutoRegressivos dentre dos Nós Terminais	127
6.7	Erro Padrão e Significância dos Parâmetros Lineares	129
6.8	Desempenho Estatístico (dentro-da-amostra)	131

6.9	Desempenho Estatístico (fora-da-amostra)	132
6.10	Desempenho Financeiro (fora-da-amostra)	133
A.1	Vício Mediano nas Estimativas de α_0	145
A.2	Vício Mediano nas Estimativas de α_1	145
A.3	Desvio Absoluto Mediano nas Estimativas de α_0	146
A.4	Desvio Absoluto Mediano nas Estimativas de α_1	146
C.1	Identificação de Arquiteturas para Árvore Simulada pelo Modelo 1.1	152
C.2	Identificação de Arquiteturas para Árvore Simulada pelo Modelo 1.2	153
C.3	Identificação de Arquiteturas para Árvore Simulada pelo Modelo 1.3	154
C.4	Identificação de Arquiteturas para Árvore Simulada pelo Modelo 1.4	155
C.5	Identificação de Arquiteturas para Árvore Simulada pelo Modelo 2.1	156
C.6	Identificação de Arquiteturas para Árvore Simulada pelo Modelo 2.2	157