

1

Introdução

Uma das maiores motivações, se não a maior, para o uso de ferramentas estatísticas, é a possibilidade de prever resultados futuros com base em informações atuais. Uma economia repleta de incertezas impulsiona ainda mais o desenvolvimento destas ferramentas. Entretanto, para muitas aplicações, além de obter valores preditos, é fundamental a explicação do fenômeno estudado. Para tal, modelos estatísticos devem ser avaliados não somente pela sua capacidade preditiva, mas também pela simplicidade e interpretabilidade.

Modelos probabilísticos simples e facilmente interpretáveis são obtidos a partir da clássica teoria de modelos lineares de regressão e séries temporais. Nestes casos, a metodologia está bastante difundida e os resultados teóricos são bem sedimentados. Para ilustrar esta situação, basta verificar o extenso material bibliográfico e implementações em pacotes computacionais associados aos modelos de regressão linear e aos tradicionais modelos ARIMA em [18] para a análise de séries temporais.

Embora, ainda seja predominante a aplicação destes modelos, há fenômenos na natureza cuja dinâmica é não-linear e, portanto, modelos que levem em conta estes aspectos serão mais apropriados. Para citar alguns exemplos: o caos, a volatilidade em séries financeiras, salto de ressonância, amplitude dependente da frequência, e ciclos limite.

Nos últimos anos, muitos modelos não-lineares para a análise de dados de painel e séries temporais tem sido propostos, tanto na econometria clássica, onde os modelos de múltiplos regimes lineares tais como o STAR [41] tem recebido muita atenção, como na Teoria de Aprendizado, por Máquina cujas redes neurais artificiais (RNA), modelos de particionamento recursivo, especialmente as árvores de classificação e regressão, e métodos não paramétricos de regressão, ocupam espaço.

Modelos de (Auto)Regressão com Transição Suave (STAR - Smooth Transition AutoRegression) são uma alternativa para a modelagem de relações não-lineares entre variáveis, originada dos desenvolvimentos apresentados em Tong [88] e posteriormente em Granger & Teräsvirta [41]. Os últimos autores enfo-

caram relações não-lineares entre variáveis econômicas, especialmente séries temporais. Entretanto, mesmo estando as principais aplicações ainda voltadas para a economia, o aparecimento de desenvolvimentos em outras áreas tem trazido grandes contribuições para esta metodologia. A idéia chave destes modelos está na busca de regiões no espaço das variáveis que caracterizem diferentes relações (regimes) entre a variável dependente e o conjunto de regressores.

Os métodos de partição recursiva têm como principal representante os modelos estruturados por árvores. O trabalho seminal de Breiman et al. [23], intitulado CART (Classification and Regression Trees), representa uma unificação dos métodos de classificação e predição estruturados por árvores que foram desenvolvidos desde uma primeira abordagem apresentada em [73]. Este trabalho transformou as árvores de regressão em uma importante alternativa, não-paramétrica, aos modelos clássicos de regressão. Entre suas vantagens estão: a habilidade para lidar com grande bases de dados contendo variáveis de diversas naturezas (nominal, ordinal e contínua), a exibição gráfica do modelo através da estrutura de uma árvore de decisão binária e a possibilidade de formular regras de associação que são facilmente compreendidas pelo usuário. Apesar destes atrativos, há sérias deficiências intrínsecas ao método tais como: instabilidade na especificação do modelo, ou seja, pequenas mudanças nos dados podem causar grandes mudanças, necessidade de grandes amostras e dificuldade no uso de procedimentos inferenciais que garantam a significância estatística dos parâmetros.

Esta tese apresenta como principal resultado a proposta de construção de um modelo de regressão não-linear, aqui denominado STR-Tree (Smooth Transition Regression Tree-structured), que combina aspectos de duas metodologias: CART e STR, buscando aproveitar suas potencialidades. O modelo herda da metodologia CART a simplicidade e interpretabilidade dos modelos estruturados por árvores, enquanto dos modelos de Regressão com Transição Suave são retiradas ferramentas para realizar inferência estatística.

A extensão da metodologia CART tradicional foi feita através da troca de limiares rígidos por limiares suaves quando executada a divisão de um nó da árvore. Os critérios para dividir um nó estão completamente baseados em testes estatísticos de hipóteses e esta característica diminui a importância de técnicas de poda a posteriori, utilizadas para reduzir a complexidade do modelo final.

O procedimento para o crescimento da árvore é utilizado aqui como uma ferramenta para a especificação de um modelo paramétrico que pode ser analisado posteriormente como um regressão com transição suave com múltiplos regimes ou como uma regressão nebulosa. No primeiro caso, podem ser obtidos intervalos de confiança para os parâmetros e predições, enquanto o segundo possibilita o uso da inferência nebulosa.

Para o modelo obtido a partir do uso do algoritmo, demonstra-se a consistência dos estimadores de mínimos quadrados e avalia-se, através de um experimento de Monte Carlo, suas propriedades em pequenas e médias amostras. Também é avaliada a capacidade do algoritmo identificar corretamente a arquitetura de árvores com transições suaves entre os modelos ajustados nas folhas. A proposta é aplicada a conjuntos de dados *benchmark* e comparada com os resultados obtidos pelo algoritmo CART

A metodologia desenvolvida é estendida, através de uma aplicação, para a análise de séries temporais. Nesta situação o modelo é denominado STAR-Tree (Smooth Transition AutoRegression Tree-Structured); baseando-se no mesmo algoritmo de especificação do modelo STR-Tree, embora ajustando modelos autoregressivos de primeira ordem para caracterizar os diferentes regimes. Foi utilizada a série de retornos da taxa de câmbio Euro/Dólar (EUR/USD) para avaliar a capacidade preditiva e financeira do modelo.

Em paralelo aos desenvolvimentos citados acima, foi feito um experimento de Monte Carlo para avaliar o vício e o erro quadrático médio nas estimativas dos parâmetros de diferentes processos LSTAR(1) em amostras de tamanhos pequenos e médios. Tal experimento traz informações que podem auxiliar a tomar decisões durante o processo de construção da árvore, pois indica situações em que não é possível estimar corretamente os parâmetros do modelo.

No Capítulo 2, é apresentado um pequeno levantamento dos métodos existentes, na Econometria e Teoria de Aprendizado por Máquina, para a construção de modelos não lineares para dados de painel e análise de séries temporais. Procura-se enfatizar neste capítulo que os modelos lineares são constantemente utilizados como fonte de inspiração. O Capítulo 3 apresenta as principais idéias relacionadas com o método de Árvores de Classificação e Regressão, dando maior atenção ao problema de regressão. Procura-se, neste capítulo, colocar estes modelos dentro de uma mesma formulação matemática para compará-los com os desenvolvimentos feitos ao longo da tese. Aspectos relacionados com a estimação do modelo STAR são abordados no Capítulo 4. Nele, estuda-se, através de simulações de Monte Carlo, o vício e a variabilidade do erro na estimação dos parâmetros do modelo LSTAR(1), um caso particular dos modelos STAR. Os resultados também servem de base para especificação dos modelos desenvolvidos nos capítulos seguintes. O Capítulo 5 constitui o núcleo do trabalho pois introduz de forma detalhada a proposta de um modelo de regressão com transição suave entre múltiplos regimes. Utiliza-se a denominação STR-Tree para este modelo, sendo uma alternativa direta e extensão dos modelos obtidos com o uso do algoritmo CART, utilizados com frequência em problemas de regressão. O modelo proposto traz como inovação a possibilidade de construir

uma árvore de regressão inteiramente baseada em procedimentos da estatística clássica como teste de hipóteses e construção de intervalos de confiança. Discute-se as propriedades dos estimadores de Mínimos Quadrados Não-Lineares, provando algebricamente sua consistência e utilizando um experimento de Monte Carlo para avaliar as propriedades destes estimadores em pequenas amostras. Conjuntos de dados reais são utilizados para comparar a metodologia proposta com o algoritmo CART. O Capítulo 6 propõe a extensão do modelo proposto para a análise de séries temporais, chamando-o de STAR-Tree. Para compará-lo com alternativas do tipo *benchmark* é feita uma aplicação à série temporal de retornos da taxa de câmbio euro/dólar (EUR/USD) e avaliado o desempenho estatístico e financeiro deste modelo após a simulação de negociações que baseiam-se em previsões 1-passo-à-frente. Nos apêndices constam tabelas produzidas nos estudos de Monte Carlo, assim como a prova dos teoremas enunciados ao longo do texto.