

5

DESCRIÇÃO DO SISTEMA

5.1

Objetivo

O sistema desenvolvido nesta dissertação tem como objetivo agregar um conjunto de funcionalidades que permita aplicar o modelo STAR em séries temporais, gerar séries artificiais a partir desses modelos, e avaliar a qualidade de suas estimações. Para isso, foram construídos diversos módulos que serão detalhados nas seções a seguir.

Algumas preocupações foram tomadas no desenvolvimento da ferramenta. Devido ao caráter muitas vezes exploratório da análise, foram oferecidas ferramentas de avaliação, parametrização e visualização que visavam facilitar o trabalho do usuário neste sentido.

Os recursos de visualização permitem que o usuário avalie a qualidade do resultado obtido, através de gráficos, ou seqüências de gráficos e tabelas. Foram criados recursos visuais que permitissem mais facilmente a visualização de modelos estruturados em árvore. Também foram incorporados recursos de gravação e carga de avaliações já executadas, para que pudessem ser reutilizadas futuramente.

O sistema também é bastante flexível, pois permite um alto grau de parametrização dos modelos e análises. É possível escolher entre diversos métodos de estimação de parâmetros e de crescimento de árvore, e reapplicá-los facilmente a diferentes configurações do modelo a ser estimado.

5.2

Módulo de estimação de séries temporais

Neste módulo, a estimação é vista em mais detalhes. Ao estimar uma série, seja artificial ou não, o usuário é levado a escolher uma série de parâmetros.

5.2.1

Parâmetros de estimação

Primeiramente, deve escolher o método através do qual os parâmetros não lineares serão estimados. São oferecidos quatro métodos: Gradiente, Newton-Rhaphson, BFGS, e algoritmos genéticos. Em todos os casos, é necessário algum nível de parametrização dos métodos. Os algoritmos genéticos são os que mais oferecem parametrização, e talvez por isso, sejam os mais difíceis de operar.

A imagem mostra uma interface de usuário para a escolha de métodos de otimização. À esquerda, há um grupo de radio buttons sob o título 'Método de Otimização'. As opções são: 'Newton', 'Algoritmo Genético', 'Gradiente' e 'BFGS'. O botão 'BFGS' está selecionado e destacado com um retângulo tracejado. À direita, há quatro controles deslizantes para ajustar parâmetros: 'Tolerância' (0.0001), 'Máx de iterações (Externo)' (20), 'Máx de iterações (Interno)' (300) e 'Passo inicial' (1). Cada controle deslizante possui setas de seta para cima e para baixo.

Figura 5.1: Escolha do método de estimação dos parâmetros não lineares

Nos demais métodos de otimização, só é necessário estabelecer critérios de parada e de busca. Neste tipo de parametrização, é importante encontrar um balanço entre melhorar os resultados e reduzir o tempo total de estimação. Muitas vezes, uma redução da tolerância do critério de parada de um método de otimização pode não trazer muitos resultados, mas implica em um aumento brutal do tempo de estimação.



The image shows a configuration window for genetic algorithms with the following parameters and values:

Parameter	Value
Tolerância	[Empty field]
Máx de iterações (Externo)	20
Gerações	300
Tamanho da População	50
Prob. de Crossover inicial	0,25
Prob. de Crossover final	0,15
Prob. de Mutação inicial	0,02
Prob. de Mutação final	0,02
Steady State inicial	0,50
Steady State final	0,40

Figura 5.2: Exemplo de parametrização para algoritmos genéticos

Uma vez escolhido algum dos métodos mencionados acima, é necessário que se escolha o método de crescimento da árvore STAR. O sistema oferece duas possibilidades: CART e teste de hipótese por multiplicadores de Lagrange. No primeiro caso, é preciso definir um nível de tolerância, que será utilizada no critério de parada do método. No caso do teste de hipótese, é preciso definir um nível α de significância para o teste. A escolha de um α muito tolerante fará com que o sistema sobreparametrize o modelo estimado.

Tolerância	<input type="text"/>
Máx de iterações (Externo)	20
Gerações	300
Tamanho da População	50
Prob. de Crossover inicial	0,25
Prob. de Crossover final	0,15
Prob. de Mutação inicial	0,02
Prob. de Mutação final	0,02
Steady State inicial	0,50
Steady State final	0,40

Figura 5.3: Escolha do método de crescimento da árvore

5.2.2

Gerador de séries artificiais

Esta funcionalidade permite que se crie um modelo gerador de séries artificiais da classe STAR de qualquer ordem p . O usuário cria, através de um recurso visual, a árvore do nível desejado e parametriza cada um de seus nós. No caso de nós de decisão, escolhe-se o parâmetro de escala e localização (gama e c), além da variável de transição autoregressiva.



Figura 5.4: Escolha dos parâmetros de um nó de decisão

Em nós terminais, definem-se os parâmetros do modelo AR local.

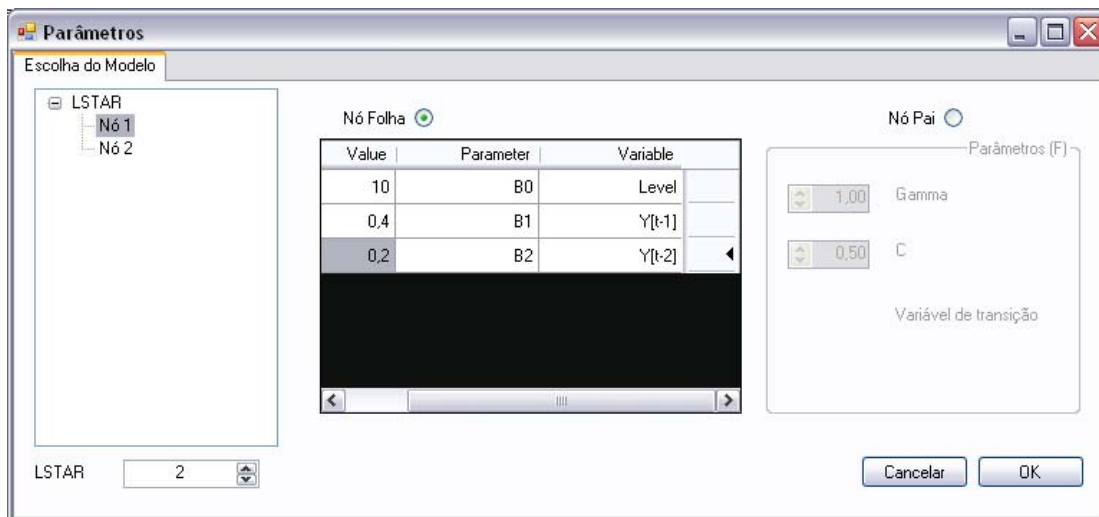


Figura 5.5: Escolha dos parâmetros de um nó de terminal

Uma vez gerado, o modelo pode ser reaproveitado para geração de novas séries, pois ele é armazenado em memória e empilhado com os outros modelos previamente gerados. É possível salvar e carregar os modelos em disco.

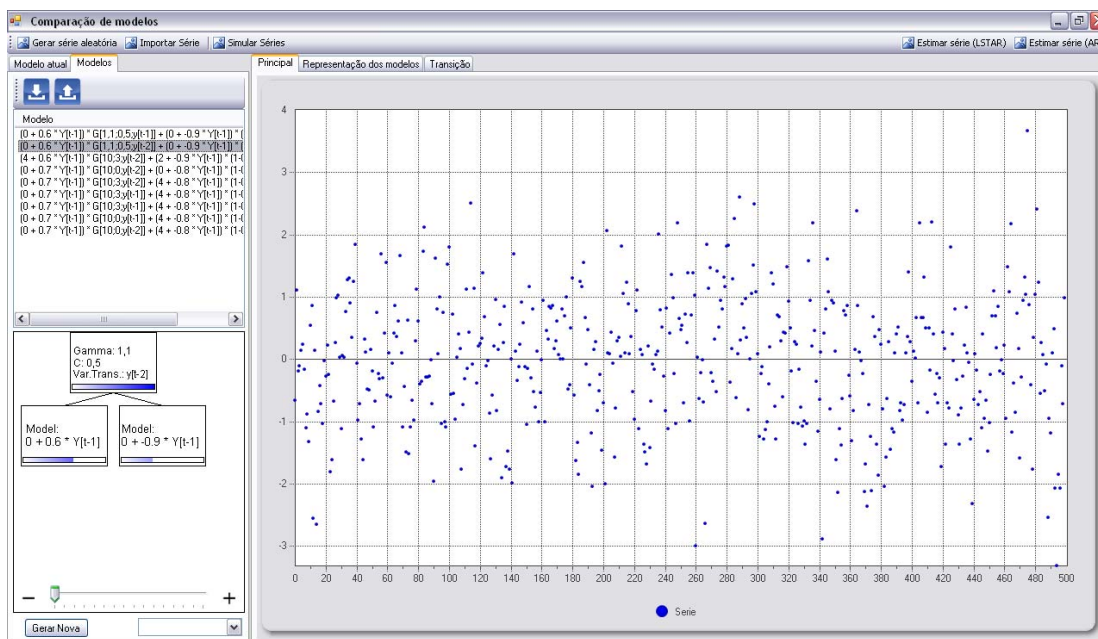


Figura 5.6: Geração de séries a partir de modelos previamente montados

Também é possível importar séries externas e estima-las por modelos da classe STAR. Neste caso, obviamente não é possível comparar os resultados da mesma forma por seu modelo gerador ser desconhecido, mas é possível diagnosticar a eficácia de sua estimação através dos instrumentos que serão descritos a seguir.

5.2.3

Estimação e visualização de resultados

Seja gerando uma série artificial, seja importando uma série, estima-se a série através da parametrização descrita em 5.2.1. É possível acompanhar através da ferramenta todo o processo de estimação. O processo se dará por terminado quando algum dos critérios de parada for atingido (número máximo de iterações ou redução insignificante do erro entre iterações).

Uma vez terminado o processo de crescimento da árvore, uma série de resultados é exibido ao usuário. A estimação final propriamente dita, onde se plota a série estimada sobre a série de dados original. A seqüência de modelos estimados ao longo do processo, e o somatório dos erros quadráticos relacionados; gráfico dos erros individuais do modelo escolhido perfilados em ordem crescente, ilustrando sua distribuição; um gráfico de área que mostra a predominância dos

modelos ao longo do processo de estimação; gráfico que plota no eixo X a variável de transição e no Y a função de pertinência, tanto do modelo estimado quanto do modelo gerador, no caso de ser uma série gerada pelo sistema.

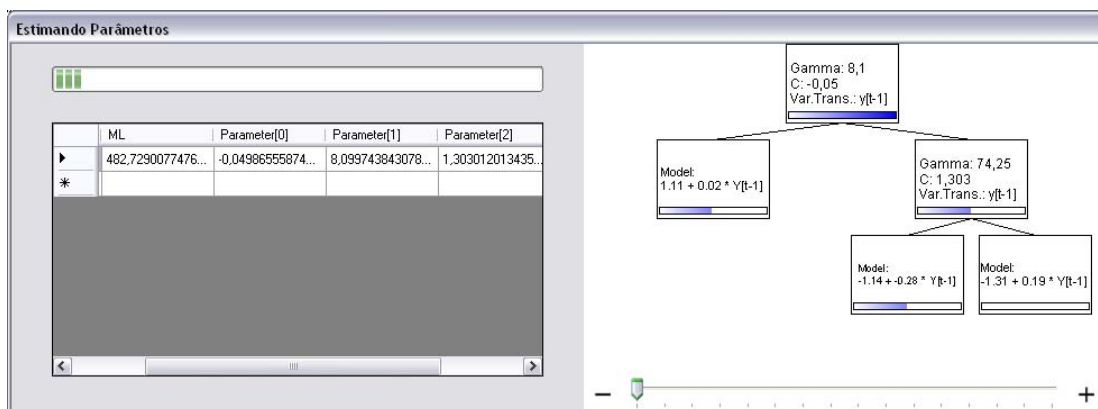


Figura 5.7: Processo de estimação da série

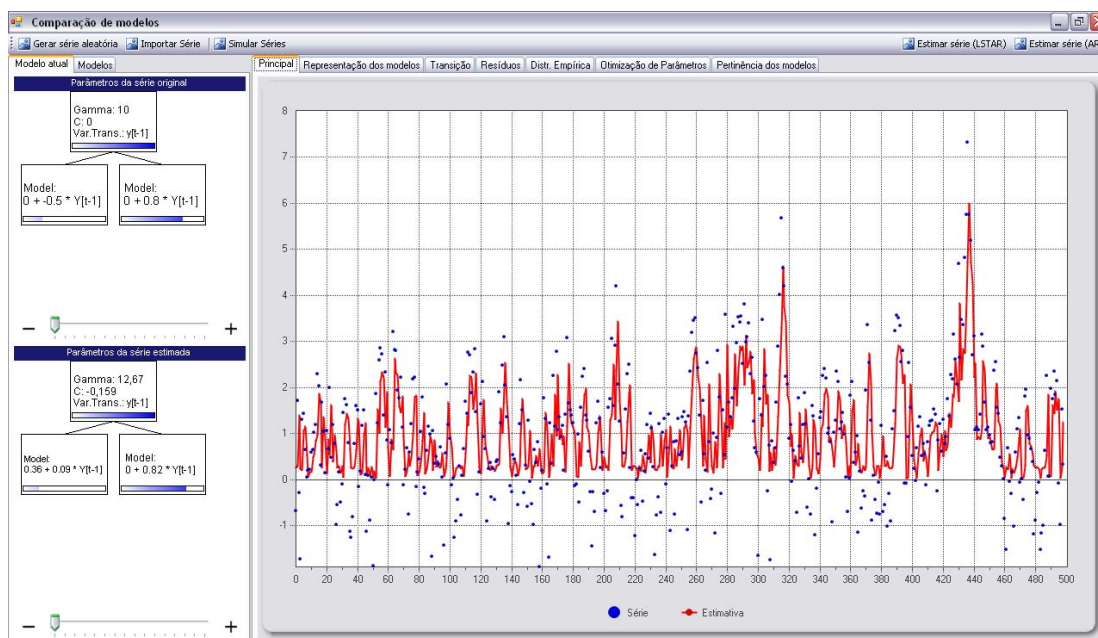


Figura 5.8: Exemplo de estimação de série

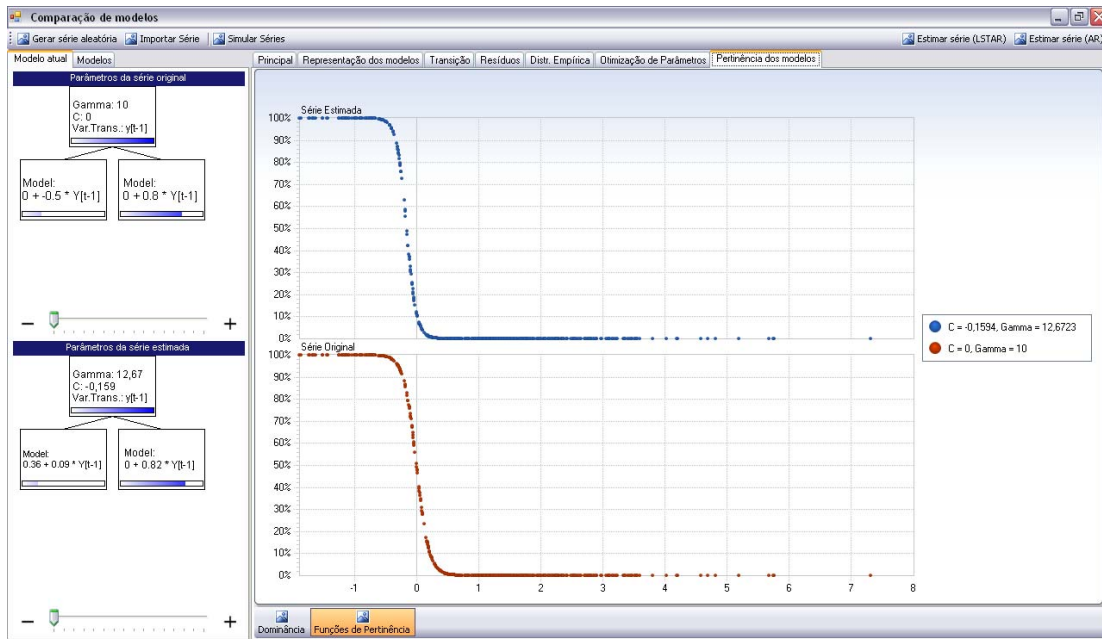


Figura 5.9: Funções de pertinência nos modelos gerador e estimado

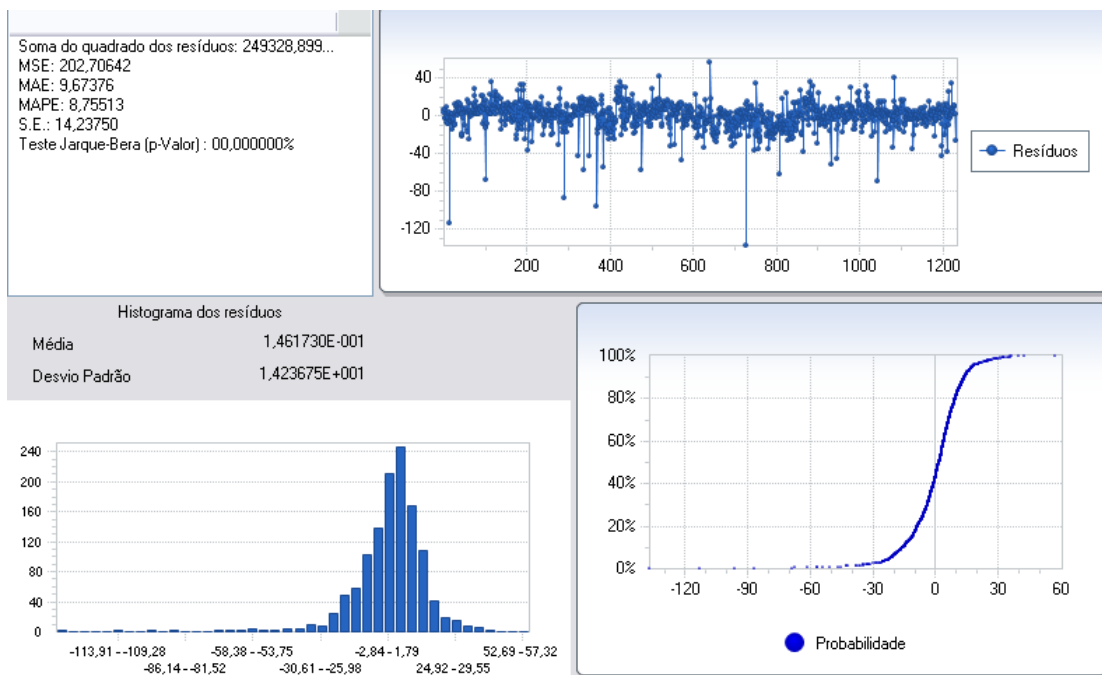


Figura 5.10: Análise dos resíduos decorrentes da estimação

5.2.4

Simulação através de modelos gerados artificialmente

Este módulo permite que se teste a eficácia da estimação sob diversas parametrizações. Para isso, é preciso definir um modelo gerador, com uma estrutura de árvore, parâmetros não lineares e lineares, o método de estimação dos parâmetros não-lineares, o tamanho da série, e a quantidade de rodadas da simulação. Cada rodada irá gerar, a partir do modelo construído, uma nova série de dados, que será estimada pelo método selecionado, utilizando-se da mesma estrutura de geração.

O objetivo deste módulo não é avaliar a detecção de estrutura da árvore, e sim, uma vez conhecida esta estrutura, avaliar a estimação de todos os seus parâmetros. Como pôde ser visto em capítulos anteriores, existem parametrizações mais difíceis de estimar do que outras, assim como séries curtas em comparação com séries longas; alguns métodos convergem mais rápido do que outros, e conforme a complexidade dos modelos aumenta, isto se torna mais crítico. Através de simulações, é possível avaliar a eficácia das estimações e tirar conclusões sobre os modelos. Para cada combinação de estrutura, parâmetros e método utilizados, é gerado um relatório que ilustra o desvio padrão da estimação, a média e a variância dos parâmetros, das pertinências efetivas de cada nó estimado, o tempo médio gasto nas estimações, as relações entre parâmetros, e as relações entre parâmetros e pertinências.



Figura 5.11: Exemplo de relatório gerado pelo módulo de simulações