

Leandro dos Santos Maciel

ESTIMAÇÃO E PREVISÃO DA ESTRUTURA A TERMO DAS TAXAS DE JUROS USANDO TÉCNICAS DE INTELIGÊNCIA COMPUTACIONAL

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica.
Área de concentração: Automação.

Orientador: Prof. Dr. Fernando Antonio Campos Gomide
Coorientadora: Profa. Dra. Rosangela Ballini



Campinas
2012

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

M187e	<p>Maciel, Leandro dos Santos</p> <p>Estimação e previsão da estrutura a termo das taxas de juros usando técnicas de inteligência computacional / Leandro dos Santos Maciel. --Campinas, SP: [s.n.], 2012.</p> <p>Orientador: Fernando Antonio Campos Gomide Coorientador: Rosangela Ballini.</p> <p>Dissertação de Mestrado - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.</p> <p>1. Taxas de juros. 2. Computação evolutiva. 3. Sistemas nebulosos. 4. Sistemas de computação adaptativos. 5. Análise de séries temporais. I. Gomide, Fernando Antonio Campos. II. Ballini, Rosangela. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.</p>
-------	--

Título em Inglês: Term structure of interest rate modeling and forecasting using computational intelligence techniques

Palavras-chave em Inglês: Interest rates , Evolutionary computation, Fuzzy systems, Adaptive computing systems, Time-series analysis

Área de concentração: Automação

Titulação: Mestre em Engenharia Elétrica

Banca examinadora: Romis Ribeiro Faissol Attux, Rodrigo Lanna Franco da Silveira

Data da defesa: 25-06-2012

Programa de Pós Graduação: Engenharia Elétrica

Leandro dos Santos Maciel
Bacharel em Ciências Econômicas – IE/UNICAMP

ESTIMAÇÃO E PREVISÃO DA ESTRUTURA A TERMO DAS
TAXAS DE JUROS USANDO TÉCNICAS DE INTELIGÊNCIA
COMPUTACIONAL

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica.
Área de concentração: Automação.
Aprovada pela banca examinadora no dia 25 de Junho de 2012.

Banca Examinadora:

Prof. Dr. Fernando Antonio Campos Gomide
FEEC/UNICAMP

Prof. Dr. Romis Ribeiro Faissol Attux
FEEC/UNICAMP

Prof. Dr. Rodrigo Lanna Franco da Silveira
IE/UNICAMP

Campinas
2012

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Leandro dos Santos Maciel

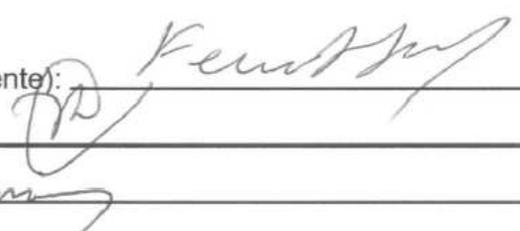
Data da Defesa: 25 de junho de 2012

Título da Tese: "Estimação e Previsão da Estrutura a Termo das Taxas de Juros Usando Técnicas de Inteligência Computacional."

Prof. Dr. Fernando Antônio Campos Gomide (Presidente):

Prof. Dr. Rodrigo Lanna Franco da Silveira:

Prof. Dr. Romis Ribeiro de Faissol Attux:



À MINHA MÃE, IVETE.

Agradecimentos

Agradeço,

aos meus pais, Ivete e Dirceu, o carinho e fundamental apoio em minha jornada. Em especial à minha mãe, imagem pela qual sou encorajado a evoluir constantemente.

ao meu irmão, Junior, minha cunhada, Juliana, e minhas queridas sobrinhas, Giovanna e Sarah, a concessão de um ambiente familiar e de carinho insubstituíveis.

aos Professores Fernando Gomide e Rosangela Ballini a orientação e excepcional prestatividade.

aos Professores da FEEC: Fernando Von Zuben, Romis Attux, Paulo Valente, Akebo Yamakami, os ótimos cursos oferecidos.

aos Professores Romis Attux e Rodrigo Lanna as críticas, sugestões e contribuições, que ajudaram a melhorar a qualidade e a redação final do manuscrito.

aos meus grandes e queridos amigos: Alexandre Martins, Lívia Rodrigues, Aline Galvão, Aquiles Tescari Neto, Adriana Carnizelli, Felipe Germano, Marilene da Silva e Fernanda Araújo, o carinho, ajuda e momentos de descontração.

à Juliane Rosa (*in memoriam*), que também deve estar feliz com mais essa conquista, os sorrisos quando recorro de nossa cumplicidade, infindável.

à CAPES o apoio financeiro concedido durante todo o período do mestrado.

ao DCA a estrutura para o desenvolvimento de pesquisas científicas.

aos amigos de trabalho, Ivette Luna, Daniel Leite e André Paim Lemos, o compartilhamento de seus conhecimentos.

aos Professores e Funcionários da FEEC/UNICAMP a ótima estrutura que oferece aos estudantes e pesquisadores, profissionais que tornam a Faculdade excelência na qualidade de ensino e pesquisa em engenharia elétrica.

a todos que de alguma forma contribuíram para este trabalho.

As complexity rises, precise statements lose meaning
and meaningful statements lose precision.

Lotfi A. Zadeh

The difficulty lies not so much in developing new ideas
as in escaping from old ones.

John Maynard Keynes

Resumo

Este trabalho propõe a utilização de técnicas de inteligência computacional para a estimação e previsão da estrutura a termo das taxas de juros, com base em dados dos mercados de renda fixa dos Estados Unidos e Brasil. Para o problema de estimação da curva de juros, as técnicas de computação evolucionária, Algoritmos Genéticos, Evolução Diferencial e Estratégias Evolutivas, foram comparadas com abordagens tradicionais da literatura, como mínimos quadrados não-lineares e programação quadrática sequencial. A motivação da aplicação de técnicas de computação evolucionária no problema de estimação da estrutura a termo busca superar limitações como não-convergência e elevada instabilidade dos parâmetros à inicialização. Além disso, recentemente, a literatura tem apontado o elevado desempenho dos algoritmos genéticos em problemas de modelagem da curva de rendimentos. Outra contribuição deste trabalho consiste no desenvolvimento de um modelo nebuloso evolutivo de aprendizado participativo estendido, denominado ePL+, que inclui em sua versão original, ePL, mecanismos para aumentar sua autonomia e adaptabilidade na modelagem de sistemas complexos. Dessa forma, o modelo ePL+ e outros modelos nebulosos funcionais evolutivos foram avaliados na questão da previsão das taxas futuras de juros, em contraposição com modelos econométricos baseados em processos autoregressivos e modelos de redes neurais artificiais multi-camadas, uma vez que a evolução das taxas de juros apresenta uma dinâmica altamente não-linear e variante no tempo, justificando a ideia de modelagem adaptativa. O desempenho dos métodos considerados foi avaliado em termos de métricas de erro, complexidade computacional e por meio de testes estatísticos paramétricos e não-paramétricos, MGN e SIGN, respectivamente. Os resultados evidenciaram o elevado potencial dos modelos de inteligência computacional na estimação e previsão da estrutura a termo em ambas economias consideradas, constatado pelo melhor desempenho, em termos de ajuste e significância estatística, em relação às técnicas de otimização de parâmetros e econométricas mais utilizadas na literatura.

Palavras-chave: Taxas de Juros. Computação Evolutiva. Sistemas Nebulosos. Sistemas de Computação Adaptativos. Análise de Séries Temporais.

Abstract

This work proposes the term structure of interest rates modeling and forecasting using computational intelligence techniques, based on data from the US and Brazilian fixed income markets. The yield curve modeling includes the use of some evolutionary computation methods like Genetic Algorithms, Differential Evolution and Evolution Strategies in comparison with traditional optimization techniques such as nonlinear least squares and sequential quadratic programming. The motivation behind the use of evolutionary computation to yield curve estimation aims to overcome limitations like non-convergence and high parameters instability to initialization. Moreover, recently, the literature has been shown the higher performance of genetic algorithms in yield curve modeling problems. This work also contributes by developing an extended participatory learning fuzzy model, called ePL+, which includes on its original version, ePL, mechanisms to improve its autonomy and adaptability in complex systems modeling. Therefore, the ePL+ model and some evolving functional fuzzy approaches were evaluated in the future interest rates forecasting, as opposed to econometric models based on autoregressive processes and multilayer artificial neural networks methodologies, since interest rates evolution shows a high non-linear dynamics and also time-varying, justifying the idea of adaptive modeling. Models' performance were compared in terms of error measures, computational complexity and by parametric and non-parametric statistical tests, MGN and SIGN, respectively. The results reveal the high potential of computational intelligence methods to deal with the term structure modeling and forecasting for both economies considered, as pointed out by their adjustment and statistical superior performance then traditional optimization and econometrics techniques reported in the literature.

Key-words: Interest Rates. Evolutionary Computation. Fuzzy Systems. Adaptive Computing Systems. Time-Series Analysis.

Lista de Figuras

2.1	Persistência dos fatores de NS como função da maturidade, para $\lambda = 0.0609$	13
3.1	Fluxograma do algoritmo genético para estimação da ETTJ.	24
3.2	Fluxograma do algoritmo de evolução diferencial para estimação da ETTJ.	29
5.1	Representação gráfica do modelo ePL+.	50
5.2	Atualização da base de regras no AP. Fonte: (Lima 2008).	52
5.3	Representação esquemática dos estágios para identificação dos modelos ePL+.	57
6.1	Estrutura a Termo das Taxas de Juros do Mercado dos Estados Unidos.	64
6.2	Estrutura a Termo das Taxas de Juros do Mercado do Brasil.	65
6.3	Evolução do fator de nível $\hat{\beta}_1$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.	69
6.4	Evolução do fator de inclinação $\hat{\beta}_2$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.	70
6.5	Evolução do fator de curvatura $\hat{\beta}_3$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.	71
6.6	Evolução do fator de decaimento $\hat{\lambda}$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.	71
6.7	Evolução do fator de nível $\hat{\beta}_1$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.	72
6.8	Evolução do fator de inclinação $\hat{\beta}_2$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.	73
6.9	Evolução do fator de curvatura $\hat{\beta}_3$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.	73
6.10	Evolução do fator de decaimento $\hat{\lambda}$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.	74
6.11	Resíduos da estimação da ETTJ no mercado dos EUA, utilizando o algoritmo de Evolução Diferencial.	76
6.12	Resíduos da estimação da ETTJ no mercado do Brasil, utilizando o algoritmo de Estratégias Evolutivas CMA-ES.	76
6.13	Estimação da ETTJ para o mercado dos EUA em Maio de 1995 de acordo com as técnicas de otimização de MQNL, SQP, ED e EE.	77

6.14	Estimação da ETTJ para o mercado do Brasil em 19 de Setembro de 2007 de acordo com as técnicas de otimização de MQNL, SQP, ED e EE.	78
6.15	Representação esquemática dos modelos nebulosos evolutivos para predição dos fatores latentes da curva de juros.	81
6.16	Evolução do número de regras nos modelos ePL e ePL+ para a previsão da ETTJ com dados do mercado dos Estados Unidos.	90
6.17	Evolução do número de regras nos modelos ePL e ePL+ para a previsão da ETTJ com dados do mercado do Brasil.	91
A.1	Taxas de juros livres de risco e ETTJ estimada.	111
A.2	ETTJ “normal”.	112
A.3	ETTJ “invertida”.	112
A.4	ETTJ com curvatura convexa.	112
A.5	ETTJ com curvatura côncava.	112

Lista de Tabelas

2.1	Modelos de estimação da ETTJ para os bancos centrais que reportam suas estimativas ao BIS.	9
6.1	Estatísticas descritivas para os fatores latentes de NS estimados de acordo com as metodologias de MQNL, SQP, AG, ED e EE, considerando os mercados de renda fixa dos Estados Unidos e do Brasil.	68
6.2	Erro quadrático médio ponderado (WMSE) para estimação da ETTJ nos mercados dos EUA e Brasil.	75
6.3	Complexidade computacional e número de parâmetros dos algoritmos utilizados para estimar a ETTJ nos mercados de renda fixa dos EUA e Brasil.	77
6.4	REQM na precificação de títulos públicos zero-cupom nos mercados dos EUA e Brasil.	79
6.5	Períodos das amostras de treinamento e validação para os dados dos mercados de juros dos Estados Unidos e Brasil.	80
6.6	Parâmetros de controle dos modelos nebulosos evolutivos para a predição da curva de juros nos mercados dos Estados Unidos e Brasil.	82
6.7	REQM para previsão dos fatores latentes da curva de juros na amostra de treinamento nos dados do mercado dos Estados Unidos.	83
6.8	REQM para previsão dos fatores latentes da curva de juros na amostra de treinamento nos dados do mercado do Brasil.	83
6.9	EQM das taxas estimadas nas amostras de treinamento para os dados dos Estados Unidos.	83
6.10	EQM das taxas estimadas nas amostras de treinamento para os dados do Brasil.	84
6.11	EQM das predições da curva de juros no mercado dos Estados Unidos.	85
6.12	EQM das predições da curva de juros no mercado do Brasil.	86
6.13	Testes <i>MGN</i> e <i>SIGN</i> para a previsão da ETTJ no mercado dos Estados Unidos.	88
6.14	Testes <i>MGN</i> e <i>SIGN</i> para a previsão da ETTJ no mercado do Brasil.	89
6.15	Complexidade computacional e número de parâmetros dos modelos utilizados para predição da ETTJ nos mercados de renda fixa dos EUA e Brasil.	89

Lista de Acrônimos e Notação

ETTJ	Estrutura a Termo das Taxas de Juros
ANBIMA	Associação Brasileira das Entidades dos Mercados Financeiros e de Capitais
MQNL	Mínimos Quadrados Não-Lineares
SQP	<i>Sequential Quadratic Programming</i> (Programação Quadrática Sequencial)
AG	Algoritmos Genéticos
ED	Evolução Diferencial
EE	Estratégias Evolutivas

y	taxa de juros nominal
β_1	fator de nível da curva de juros
β_2	fator de inclinação da curva de juros
β_3	fator de curvatura da curva de juros
λ	fator de decaimento da curva de juros / grau de ativação nebulosa normalizado
x	variáveis de entrada
N	número de dados em uma amostra / número de indivíduos em uma população
\mathcal{F}	função objetivo em problemas de otimização
\mathfrak{R}^n	conjunto dos números reais n -dimensional
k	instante, estágio ou etapa em processos iterativos/construtivos
μ	multiplicador da função Lagrangeana / grau de pertinência nebulosa
σ	parâmetro da diversidade da população / raio de dispersão nebulosa
F	fator de escalamento do algoritmo de evolução diferencial
Cr	taxa de crossover do algoritmo de evolução diferencial
C	matriz de covariância / dispersão
Γ	função de pertinência nebulosa
γ	parâmetros dos consequentes das regras nebulosas
R	número de regras nebulosas ou agrupamentos
v	centro dos grupos nebulosos
τ	maturidade / grau de ativação de regras nebulosas
Υ	função potencial do modelo eTS
a	índice de alerta
ρ	índice de compatibilidade

Sumário

1	Introdução Geral	1
1.1	Motivação	1
1.2	Objetivos	4
1.3	Organização do Trabalho	5
2	Estrutura a Termo das Taxas de Juros	7
2.1	Modelos de Estimação da Curva de Juros	7
2.2	Modelo de Fatores Latentes	9
2.2.1	Conceitos Básicos	10
2.2.2	Curva de Juros de Nelson e Siegel (1987)	11
2.3	Problema de Otimização de Parâmetros	13
2.4	Mínimos Quadrados Não-Lineares	16
2.5	Programação Quadrática Sequencial	18
2.6	Resumo	20
3	Computação Evolucionária	21
3.1	Algoritmos Genéticos	21
3.1.1	Inicialização	23
3.1.2	Seleção	25
3.1.3	Crossover	26
3.1.4	Mutação	26
3.2	Evolução Diferencial	27
3.2.1	Inicialização	30
3.2.2	Mutação	30
3.2.3	Crossover	31
3.2.4	Seleção	31
3.3	Estratégias Evolutivas	32
3.4	Resumo	35
4	Previsão da Curva de Juros	37
4.1	Revisão Bibliográfica	37

4.2	Técnica de previsão de Dois Estágios	40
4.3	Modelos Nebulosos Evolutivos	41
4.4	Modelos Nebulosos Funcionais Evolutivos	44
4.4.1	Modelo Takagi-Sugeno Evolutivo	45
4.5	Resumo	48
5	Modelo Nebuloso Evolutivo Participativo Estendido	49
5.1	Modelos Takagi-Sugeno com Múltiplas Saídas	49
5.2	Algoritmo de Aprendizado Participativo	50
5.2.1	Adaptação da Zona de Influência das Regras	55
5.2.2	Avaliação da Qualidade da Base de Regras	55
5.3	Atualização dos Consequentes das Regras	56
5.4	Algoritmo ePL+	60
5.5	Resumo	62
6	Resultados Empíricos	63
6.1	Base de Dados	63
6.2	Modelagem da Curva de Juros	65
6.2.1	Parâmetros de Controle	65
6.2.2	Resultados	67
6.3	Previsão das Taxas de Juros	79
6.3.1	Metodologia	79
6.3.2	Resultados	82
6.3.3	Resumo	91
7	Conclusões e Trabalhos Futuros	93
	Bibliografia	97
A	Conceitos da Estrutura das Taxas de Juros	109
A.1	Juros, Taxa de Juros e Produtos do Mercado Financeiro	109
A.2	Estrutura a Termo das Taxas de Juros	110
A.3	Formas da ETTJ	111
A.4	Formação da ETTJ	112
A.5	As Metas Inflacionárias no Brasil	113
A.6	LTN, <i>Duration</i> e Convexidade	113

Introdução Geral

1.1 Motivação

A estrutura a termo das taxas de juros (ETTTJ) é definida como a relação entre as taxas nominais de juros de títulos públicos que não pagam cupons (zero-cupom), ou taxas consideradas livres de riscos, e diferentes maturidades, i.e., descreve a evolução dinâmica das taxas de juros para distintos prazos em determinado dia. As taxas de títulos zero-cupom que compõem a estrutura a termo são denominadas taxas livres de riscos, uma vez que, partindo da hipótese de insolvência financeira do emissor (governo), representam o rendimento de uma aplicação, por um determinado período, livre de riscos. Diariamente, os Bancos Centrais emitem títulos públicos para um número limitado de vencimentos, o que implica na necessidade da utilização de alguma técnica de interpolação que permita associar uma taxa zero-cupom para todas as maturidades possíveis.

É de extrema importância, principalmente nos mercados financeiros, a modelagem da estrutura a termo das taxas de juros. Instituições financeiras, por exemplo, necessitam controlar a sua exposição ao risco de oscilação de taxas de juros por meio do balanceamento de suas posições em títulos de dívida pública ou em contratos futuros que possuem seu valor atrelado às taxas de juros. A ETTTJ ainda é essencial na precificação de ativos financeiros como títulos públicos, contratos futuros, *swaps*, opções, assim como é uma variável chave na administração de riscos, em decisões corporativas de financiamento, e na composição de portfólios. A estrutura a termo também permite o monitoramento de variáveis econômicas como prêmio pelo risco, inflação e taxa real de atividade econômica.

A modelagem da curva de juros também ocupa papel central na predição das taxas de juros futuras¹. Estratégias de imunização de portfólio necessitam de medidas preditivas dos movimentos da curva de rendimentos, assim como tesoureiros dos governos gerenciam a emissão e manutenção do nível da dívida pública com base na determinação das taxas de juros corrente e futura. Investidores podem avaliar seus portfólios ao longo do tempo levando em consideração o custo de oportunidade ao se aplicar em ativos de baixo-risco. E, numa perspectiva macroeconômica, Bancos Centrais podem reagir aos níveis esperados de inflação e atividade econômica por

¹Os termos estrutura a termo da taxa de juros, curva de juros e curva de rendimentos serão tratados como sinônimos neste trabalho.

meio do ajuste da taxa de curto prazo, o que, por consequência, afeta todo o comportamento da curva de juros. Por fim, como qualquer taxa de crédito é formada pela adição da taxa livre de riscos a um *spread* que remunera o empréstador pelo risco de *default*, a ETTJ impacta a demanda agregada de uma economia.

Na última década, a literatura avançou significativamente na construção de modelos teóricos para descrever a estrutura a termo, assim como na modelagem econométrica desses modelos. Uma literatura importante está relacionada ao uso de modelos estatísticos, isentos de uma interpretação estrutural, mas que se caracterizam por sintetizar o comportamento e permitir a estimação e previsão da curva de juros, sem necessariamente representarem o ajuste de modelos teóricos sob condições econômicas de equilíbrio ou de arbitragem. Nessa classe, destacam-se os modelos paramétricos, também conhecidos como parcimoniosos, baseados na especificação de uma função com poucos parâmetros, de fácil implementação e interpretabilidade.

Nos modelos parcimoniosos, a questão da modelagem da ETTJ se resume a um problema de otimização não-linear de parâmetros que considera um critério de aderência da curva estimada ao número limitado das taxas implícitas observadas nos preços dos títulos zero-cupom. Como a evolução da taxa de juros apresenta um comportamento altamente não-linear, as especificações paramétricas da curva de rendimentos são não-lineares nos parâmetros, com destaque para os modelos de expansões polinomiais, uma vez que suas características se adequam aos fatos estilizados do comportamento das taxas de juros.

Embora a literatura, em geral, produza maiores avanços na obtenção de uma especificação funcional que melhor represente o comportamento da ETTJ, os métodos de otimização envolvidos na estimação dos parâmetros da curva têm sido negligenciados. Na modelagem diária da curva de juros com modelos parcimoniosos, existe a possibilidade de que os parâmetros estimados possam corresponder a uma solução de ótimo local e não global, i.e., verifica-se o risco de não-convergência. Esse problema emerge da alta não-linearidade da função a ser otimizada, que em prática pode resultar em uma aderência inferior a níveis convencionais. A não-convergência pode também ser explicada pela elevada sensibilidade das estimativas de acordo com a condição inicial associada aos algoritmos tradicionais de otimização não-linear (máxima verossimilhança e mínimos quadrados não-lineares, por exemplo). Como, em geral, nos modelos parcimoniosos, seus parâmetros possuem uma interpretação econômica associada às expectativas dos agentes quanto ao comportamento futuro das taxas de juros, é esperada uma baixa variabilidade dos parâmetros, principalmente em condições de estabilidade nos mercados financeiros.

Para solucionar tais problemas, (Gimeno & Nave 2009) propuseram a utilização de Algoritmos Genéticos (AG) para a modelagem da estrutura a termo das taxas de juros, considerando dados do mercado de renda fixa da Espanha. Os autores mostraram que a técnica sugerida superou o método de mínimos quadrados não-lineares em termos de acurácia e, ao mesmo tempo, gerou estimativas mais estáveis. De acordo com a aplicação da ETTJ, pequenas diferenças na aderência da curva aos dados reais podem implicar em elevadas diferenças na precificação de ativos, gerenciamento de riscos, previsão das taxas futuras de juros. Portanto, a aderência dos resultados é fundamental, uma vez que a estrutura a termo reflete o *status* do mercado de juros. Esses resultados evidenciam um elevado potencial dos métodos de computação evolucionária para o problema de modelagem da curva de juros.

Seguindo o trabalho de (Gimeno & Nave 2009), nesta dissertação, busca-se avaliar distintos métodos de computação evolucionária para o problema de estimação da ETTJ. A proposta de aplicação dessas técnicas é justificada pelo fato de elas apresentarem desempenho superior se comparadas com algoritmos tradicionais de otimização em inúmeras aplicações, e também por evitarem o problema da não-convergência e elevada sensibilidade à inicialização, conforme reportado na literatura (Das & Suganthan 2011). Além disso, pequenas diferenças na modelagem da ETTJ podem refletir em significativas alterações quando esta é avaliada em aplicações práticas: por isso, métodos de otimização que resultem em estimativas mais acuradas são potenciais candidatos para o problema de obtenção da curva de juros, incluindo meta-heurísticas, como os algoritmos baseados em computação natural. Adicionalmente, diferente da técnica de algoritmos genéticos, os métodos de Evolução Diferencial e Estratégias Evolutivas, em inúmeros problemas de otimização e considerados neste trabalho para o problema de estimação da curva de juros, possuem um número menor de parâmetros de controle, são algoritmos de simples implementação e, em geral, requerem um custo computacional mais baixo.

Além da estimação da curva de juros, neste trabalho, também é estudado o problema de previsão da estrutura a termo das taxas de juros. A previsão das taxas de juros futuras é de fundamental importância para instituições financeiras, investidores e *policy makers*. Sua aplicação inclui o gerenciamento de riscos, composição de carteiras, precificação de títulos públicos e derivativos, e o auxílio a condução de políticas econômicas (monetária e fiscal). A análise de previsão da curva de juros foi abordada pela literatura mais intensivamente nos anos 2000, sendo o trabalho de (Diebold & Li 2006) uma das principais referências nessa linha de pesquisa. Desde então, inúmeras propostas foram avaliadas para previsão da ETTJ, conforme será descrito posteriormente. É verificado na literatura que as abordagens empregadas, em geral, apresentam resultados satisfatórios. Contudo, frequentemente, suas previsões são melhores no curto prazo ou no longo prazo, não sendo possível realizar boas previsões da curva de juros para prazos curtos e longos simultaneamente. Dessa forma, na tentativa de superar tal limitação, um dos objetivos deste trabalho é avaliar o poder preditivo de sistemas nebulosos (*fuzzy*) evolutivos para a estrutura a termo das taxas de juros, considerando distintos horizontes de previsão.

Modelos nebulosos evolutivos vêm sendo aplicados nos últimos anos na solução de diversos problemas, inclusive previsão de séries temporais, revelando bons resultados e definindo uma promissora abordagem para a identificação de sistemas não-lineares adaptativos. Como a evolução da estrutura a termo das taxas de juros apresenta um comportamento não-linear e dinâmico, em decorrência da condição verificada nos mercados de taxas de juros, essa abordagem apresenta um paradigma relevante para esse tipo de fenômeno. Nesta dissertação, a questão da previsão da curva de juros é avaliada por meio de modelos nebulosos evolutivos, uma vez que esses são adaptativos, permitindo que a dinâmica do mercado de juros seja capturada e refletida em sua estrutura de base de regras. A literatura não apresenta a aplicação desses modelos ao problema considerado, evidenciando mais uma contribuição deste trabalho, que inclui, ainda, a avaliação empírica com dados dos mercados de renda fixa de uma economia desenvolvida (Estados Unidos) e emergente (Brasil).

Adicionalmente, outra contribuição deste trabalho consiste no desenvolvimento de um modelo nebuloso evolutivo de aprendizado participativo estendido, denominado ePL+. Assim como em sua versão inicial, ePL, o modelo ePL+ utiliza o conceito de aprendizado participativo para a construção de bases de regras em sistemas nebulosos. Ainda, o modelo ePL+ é composto por mecanismos que aumentam sua autonomia, por meio da atualização recursiva da zona de influência das regras nebulosas e monitoramento da qualidade da base de regras com base em uma medida de utilidade. Diferentemente do ePL, o modelo ePL+ ainda foi estendido para o caso de sistemas com múltiplas saídas. Dessa forma, o problema empírico de previsão da estrutura a termo das taxas de juros é abordado considerando modelos nebulosos evolutivos relatados na literatura, redes neurais multi-camadas e o modelo ePL+.

1.2 Objetivos

Este trabalho tem como objetivo geral a aplicação de técnicas de inteligência computacional na estimação e previsão da estrutura a termo das taxas de juros considerando dados dos mercados de renda fixa dos Estados Unidos e Brasil.

Na questão de estimação da ETTJ, os objetivos se resumem na aplicação de diferentes algoritmos evolucionários, adequados ao problema, e sua comparação com técnicas de otimização não-linear mais utilizadas, i.e., mínimos quadrados não-lineares e programação quadrática sequencial. Portanto, na modelagem da curva de juros, o objetivo principal é avaliar o desempenho de técnicas de computação evolucionária para a estimação dos parâmetros de modelos parciais que descrevem a estrutura a termo das taxas de juros. Os algoritmos considerados, Algoritmos Genéticos, Evolução Diferencial e Estratégias Evolutivas serão avaliados em termos de acurácia, sensibilidade à inicialização, determinação dos parâmetros de controle e complexidade computacional. Além disso, como aplicação prática, também será realizada uma avaliação dos resultados na precificação de títulos públicos zero-cupom.

Para o problema de previsão da curva de juros, a ideia central consiste na aplicação de modelos nebulosos funcionais evolutivos baseados em regras. O principal objetivo é a aplicação das técnicas nebulosas eTS, xTS, eTS+ e ePL, consolidadas na literatura, para previsão da curva de juros, tomando como *benchmark* o modelo proposto por (Diebold & Li 2006), e ainda, considerando modelos baseados em redes neurais artificiais de múltiplas camadas (MLP), de forma a avaliar os resultados obtidos em termos de acurácia de predição, e por meio de testes estatísticos para modelos de previsão competitivos.

Outro objetivo subjacente a esta dissertação é estender o modelo ePL de forma a incluir mecanismos que permitam uma maior flexibilidade à técnica, e evitar a necessidade de conhecimento prévio do problema a ser tratado pelo usuário, o que permite o aprendizado autônomo por meio da base de dados que representa o comportamento do sistema avaliado. O modelo estendido, denominado ePL+, inclui essencialmente os seguintes procedimentos: adaptação recursiva da zona de influência das regras nebulosas, extensão para a modelagem de sistemas com múltiplas saídas, e inclusão de uma medida de utilidade para avaliar a qualidade da estrutura da base de regras. Além disso, o modelo ePL+ proposto será também avaliado para o problema de predição da estrutura a termo das taxas de juros.

1.3 Organização do Trabalho

Este capítulo apresentou o problema tratado nesta dissertação, incluindo a relevância da pesquisa na área de modelagem e previsão da estrutura a termo das taxas de juros para investidores, gerentes de riscos, administradores de carteiras, *policy makers* e demais participantes do mercado em geral. O trabalho está dividido essencialmente em duas partes: estimação e previsão. Os problemas são apresentados distintamente, seguidos da descrição das abordagens tradicionalmente avaliadas na literatura, e das técnicas alternativas de inteligência computacional sugeridas para o tratamento de cada aplicação. Por fim, os resultados gerais são descritos.

No Capítulo 2, são reportadas as principais abordagens de modelagem da curva de juros, incluindo uma revisão bibliográfica sobre o tema. Em particular, o modelo de fatores latentes, proposto por (Nelson & Siegel 1987), é apresentado como a metodologia considerada nesta dissertação para descrever a estrutura a termo das taxas de juros, transformando a questão de obtenção da curva de rendimentos em um problema de otimização não-linear de parâmetros. Além disso, as técnicas de mínimos quadrados não-lineares e de programação quadrática sequencial são descritas, uma vez que constituem o *benchmark* para a questão de modelagem da curva zero-cupom.

Em seguida, o Capítulo 3 traz a apresentação das técnicas de computação evolucionária aplicadas à modelagem da curva de juros: Algoritmos Genéticos, Evolução Diferencial e Estratégias Evolutivas. Cada metodologia é descrita levando em consideração o problema proposto, de forma a apresentar os algoritmos implementados, destacando suas peculiaridades. Além disso, o capítulo inclui uma revisão da literatura dessas técnicas evolucionárias na consideração de problemas em economia e finanças.

O Capítulo 4, por sua vez, introduz a idéia de previsão da estrutura a termo das taxas de juros. Inicialmente, é reportada a revisão bibliográfica do problema de predição da curva de rendimentos, seguida da apresentação da técnica de (Diebold & Li 2006), metodologia amplamente utilizada. Em seguida, são apresentadas as idéias dos modelos nebulosos funcionais evolutivos, abordagem proposta nesta dissertação para a previsão das taxas de juros, incluindo suas principais características e destacando suas vantagens na área de séries temporais.

No Capítulo 5, é introduzida a nova versão estendida do modelo ePL (Lima 2008), denominada ePL+, que também será avaliada no problema de predição da ETTJ. O modelo ePL+ inclui as seguintes modificações em relação à sua versão original: adaptação recursiva da zona de influência das regras nebulosas, extensão para identificação de sistemas com múltiplas saídas, introdução de uma medida de utilidade para avaliar a qualidade da estrutura da base de regras, e utilização do algoritmo *subtractive clustering* para inicialização dos antecedentes das regras. Essas modificações permitem um maior grau de adaptação do modelo a partir da base de dados, eximindo a necessidade de conhecimento prévio do problema para uma aplicação adequada do modelo.

Os resultados das aplicações das metodologias de inteligência computacional sugeridas para a modelagem e predição da curva de juros são apresentados no Capítulo 6. Para o problema de predição, comparações incluem outros modelos nebulosos evolutivos presentes na literatura, como o eTS, xTS, eTS+, ePL, o modelo proposto nesta dissertação, ePL+, e uma rede neural do tipo MLP. Os modelos são avaliados em termos de acurácia e a partir de testes estatísticos

paramétricos e não-paramétricos, MGN e SIGN, respectivamente.

Por fim, o Capítulo 7 conclui o trabalho resumando suas principais contribuições, sugere tópicos a serem desenvolvidos futuramente, e apresenta os artigos produzidos, diretamente relacionados com esta dissertação.

Estrutura a Termo das Taxas de Juros

Neste capítulo serão apresentados os principais conceitos e métodos de estimação da estrutura a termo das taxas de juros, em conjunto com uma breve revisão bibliográfica sobre o tema. Os modelos parcimoniosos, considerados nesta dissertação, serão descritos em seguida. Por fim, formula-se o problema de otimização associado e suas respectivas propostas de solução verificadas na literatura.

2.1 Modelos de Estimação da Curva de Juros

A estrutura a termo das taxas de juros (ETTJ) é definida como uma relação entre taxas nominais de juros de títulos públicos zero-cupom livres de risco e datas de maturidade¹. A curva de juros reflete as associações entre as taxas nominais de curto, médio e longo prazos dos títulos considerados em um determinado momento no tempo. Embora o conceito de ETTJ esteja associado à taxas implícitas de títulos públicos, em alguns países, como o Brasil, que apresentam uma liquidez relativamente baixa nesse mercado, a estrutura a termo pode ser construída por outras taxas, implícitas em contratos de juros transacionados no mercado de renda fixa².

A ausência de títulos públicos zero-cupom para todas as possíveis maturidades implica que, na prática, a estrutura a termo real não é observada. A curva de juros deve então ser estimada utilizando um número limitado de taxas à vista, obtidas diretamente dos preços dos títulos públicos disponíveis no mercado, e por meio dos preços dos cupons de títulos que pagam tais fluxos periódicos, exigindo assim um método de interpolação. Portanto, a questão central na estimação da estrutura a termo é gerar uma curva que associe uma taxa de juros nominal zero-cupom à todas possíveis maturidades.

A hipótese de não-arbitragem no mercado de renda fixa implica que a curva de rendimentos deve apresentar essencialmente duas características: continuidade e suavidade (*smoothness*). A suavidade garante a continuidade da estrutura a termo implícita das taxas futuras de juros, elemento fundamental em um contexto mais geral de não-arbitragem financeira. (Guthmann

¹Títulos zero-cupom ou de descontos puros são contratos de renda fixa que fazem um único pagamento, i.e., o valor de face do título, na maturidade, sem pagamento de cupons intermediários.

²No Brasil, as Letras do Tesouro Nacional (LTNs) são os títulos públicos zero-cupom transacionados nos mercados, entretanto, apresentam uma baixa liquidez, o que resulta na necessidade da utilização de taxas alternativas como *proxy* para construção da ETTJ.

1929) iniciou as pesquisas na tentativa de formular um modelo capaz de descrever a ETTJ. Posteriormente, empregou-se desde métodos gráficos e subjetivos (Anderson, Breedon, Deacon, Derry & Murphy 1996), até sofisticados modelos não-paramétricos baseados na estimação de funções polinomiais gradualmente suavizadas (Anderson & Sleath 2001). Contudo, modelos mais compactos, descritos com base em poucos parâmetros, vêm sendo priorizados pela literatura e agentes de mercado (Friedman 1977).

Os trabalhos relacionados à estimação da estrutura a termo das taxas de juros mostram que a curva de juros é essencialmente monotônica e com presença de curvatura(s) (*humped*), ou, ocasionalmente, em formato de “S” (Dobson 1978, Chambers, Carleton & Waldman 1984, Shea 1985). Os métodos de interpolação da estrutura a termo podem ser divididos em três principais classes: modelos de equilíbrio, de não-arbitragem e estatísticos ou paramétricos.

Na classe dos modelos de equilíbrio, as taxas de curto prazo são modeladas na forma de um processo de difusão e de uma função de desconto, para então derivar a evolução da estrutura a termo na forma de uma relação entre as taxas das maturidades mais longas em função da taxa de curto prazo. Os principais representantes dessa classe são os modelos de (Vasicek 1977), (Brennan & Schwartz 1979) e (Cox, Ingersoll & Ross 1985). Nesses modelos, a estimação econométrica se dá com a obtenção dos parâmetros do processo de difusão da taxa de juros de curto prazo (Chan, Karolyi, Longstaff & Sanders 1992).

Na segunda classe de modelos da ETTJ, o representante mais importante é o modelo de (Heath, Jarrow & Morton 1992). O ajuste da curva de juros nos modelos de não-arbitragem é realizado de forma a não existirem condições de arbitragem entre as taxas, assim como não envolve diretamente a estimação de parâmetros subjacentes ao processo gerador das taxas de juros. Utilizando árvores binomiais ou trinomiais, em geral, a curva observada é ajustada perfeitamente em cada dia, sem a presença de uma estrutura dinâmica nas taxas de curto prazo. O objetivo desses modelos não está relacionado à previsão da curva de juros, mas à precificação de instrumentos financeiros por meio do ajuste da curva observada no mercado.

Na classe dos modelos paramétricos ou estatísticos, não há uma interpretação estrutural, pois não representam o ajuste de modelos teóricos sob condições de equilíbrio e não-arbitragem. Os principais representantes dessa classe incluem a metodologia de componentes principais (Litterman & Scheinkman 1991), modelos de interpolação como os splines (Shea 1984), *smoothing* splines (Fisher, Nychka & Zervos 1995), regressão por kernel (Linton, Mammen, Nielson & Tanggaard 2001), e modelos parcimoniosos ou de fatores latentes como os de (Nelson & Siegel 1987) e (Svensson 1994).

Os modelos de (Nelson & Siegel 1987), NS daqui em diante, e (Svensson 1994) são os mais utilizados na literatura e na prática para a estimação da curva de rendimentos. Os modelos estruturais de equilíbrio têm um poder preditivo baixo, enquanto modelos de calibração baseados em condições de não-arbitragem não permitem diretamente uma previsão para a curva de juros (Laurini & Hotta 2010). Dessa forma, os modelos parcimoniosos são normalmente aplicados ao ajuste e previsão da ETTJ por serem mais simples de implementação e apresentarem desempenho superior aos demais modelos econométricos. Entretanto, nos modelos parcimoniosos não são impostas restrições teóricas de não-arbitragem financeira, desejável em modelos de taxa de juros (Diebold, Piazzesi & Rudebusch 2005).

As três classes de estimação da estrutura a termo apresentam suas vantagens e desvantagens, sem evidências empíricas da superioridade de uma metodologia em relação às demais (Manousopoulos & Michalopoulos 2009). Entretanto, segundo o relatório do Banco de Compensações Internacionais (BIS - *Bank of International Settlements*), nove dentre treze importantes bancos centrais que reportam suas estimativas da estrutura a termo ao BIS utilizam modelos parcimoniosos, sobretudo os modelos de NS e (Svensson 1994) (BIS 2005), conforme mostrado na Tabela 2.1³.

Tabela 2.1: Modelos de estimação da ETTJ para os bancos centrais que reportam suas estimativas ao BIS.

Banco Central	Modelo de ETTJ
Alemanha	Svensson
Bélgica	Svensson/Nelson-Siegel
Canadá	Splines Exponenciais
Espanha	Svensson
Estados Unidos	Splines Suavizados
Finlândia	Nelson-Siegel
França	Svensson/Nelson-Siegel
Inglaterra	VRP
Itália	Nelson-Siegel
Japão	Splines Suavizados
Noruega	Svensson
Suécia	Splines Suavizados/Svensson
Suiça	Svensson

Dessa forma, nesta dissertação, o modelo de fatores latentes de (Nelson & Siegel 1987) foi considerado para descrever a estrutura a termo das taxas de juros, uma vez que se trata de uma abordagem robusta, no sentido de que é capaz de descrever diversos espectros de estrutura a termo com um número relativamente baixo de parâmetros, e por se tratar de uma abordagem muito utilizada na literatura e na prática (Gürkaynak, Sack & Wright 2006, Nyholm & Rebonato 2008), inclusive no mercado Brasileiro. Além disso, o conselho de governadores do sistema do *Federal Reserve*, Banco Central dos Estados Unidos, publica diariamente as curvas de juros diárias, considerando o modelo de NS, em tempo real⁴, assim como o Banco Central Europeu⁵.

2.2 Modelo de Fatores Latentes

Para apresentação do modelo de (Nelson & Siegel 1987), é necessário descrever alguns conceitos subjacentes à modelagem da estrutura a termo, incluindo as relações entre as principais construções teóricas associadas: curva de desconto, curva de juros futura, e curva de rendimentos.

³O modelo de (Svensson 1994) é caracterizado pela adição de um termo extra na função de (Nelson & Siegel 1987) para permitir a presença de uma segunda curvatura na estrutura a termo das taxas de juros.

⁴Ver: <http://www.federalreserve.gov/econresdata/researchdata.htm>.

⁵Ver: <http://www.ecb.europa.eu/stats/money/yc/html/index.en.html>.

2.2.1 Conceitos Básicos

Os principais ativos de renda fixa nos mercados financeiros utilizados para construção da estrutura a termo das taxas de juros são os títulos públicos⁶. O preço de um título público é definido como o valor presente de seus fluxos de caixa:

$$\mathcal{P} = \mathcal{C} \cdot \sum_{i=1}^m \delta(\tau_i) + \mathcal{A} \cdot \delta(\tau_m) \quad (2.1)$$

onde \mathcal{P} é o preço do título, \mathcal{C} corresponde ao valor do cupom pago pelo título nos períodos τ_i ($i = 1, \dots, m$), \mathcal{A} à amortização (valor de face) do título na maturidade τ_m , e $\delta(\cdot)$ à função de desconto associada.

Na precificação de qualquer ativo de renda fixa, a função de desconto é o ponto de partida. Define-se a função de desconto, no caso nominal, como o valor atual (hoje) de um título que pagará uma unidade monetária \$ 1 na maturidade τ . Em termos discretos, podemos escrever a função de desconto como:

$$\delta(\tau) = \frac{1}{(1 + y(\tau))^\tau} \quad (2.2)$$

em que $y(\tau)$ é a taxa de juros nominal zero-cupom de maturidade τ .

A curva de rendimentos de um título pode então ser descrita por meio de uma única taxa de juros que desconta seus fluxos de caixa futuros ao seus valores presentes:

$$\mathcal{P} = \sum_{i=1}^m \frac{\mathcal{C}}{(1 + y(\tau))^{\tau_i}} + \frac{\mathcal{A}}{(1 + y(\tau))^{\tau_m}} \quad (2.3)$$

No caso dos títulos zero-cupom, i.e., títulos que não pagam cupons, seu rendimento até a maturidade é dado pela taxa à vista (*spot*) de mesma maturidade, denominada rendimento zero-cupom.

Conhecidas as taxas à vista $y(\tau_i)$ e suas funções de desconto correspondentes $\delta(\tau_i)$, podemos calcular as taxas futuras implícitas, para um determinado período, como:

$$(y_{fut})_i = \frac{\delta(\tau_{i-1}) - \delta(\tau_i)}{\delta(\tau_i)} \quad (2.4)$$

Em finanças, em geral, é conveniente considerar as taxas de juros capitalizadas continuamente, ou seja, assumir que os descontos são realizados em períodos infinitesimais. Pode-se então reescrever a função de desconto em (2.2) da seguinte forma:

$$\delta(\tau) = e^{-\tau y(\tau)} \quad (2.5)$$

A representação da curva da taxa instantânea futura em tempo contínuo é obtida pela extensão da expressão (2.4):

$$y_{fut}(\tau) = \frac{-\delta'(\tau)}{\delta(\tau)} \quad (2.6)$$

⁶No Apêndice A são apresentados os principais conceitos relacionados a construção da estrutura a termo das taxas de juros e suas especificidades para o caso do Brasil.

Consequentemente, a taxa de juros futura média para uma determinada maturidade no intervalo $[\tau_1, \tau_2]$ é obtida por meio da média de $r(\tau)$:

$$\bar{y}_{fut}(\tau_1, \tau_2) = \frac{1}{\tau_2 - \tau_1} \int_{\tau_1}^{\tau_2} y_{fut}(t) dt \quad (2.7)$$

A taxa de juros à vista com maturidade τ é então a média da taxa futura no intervalo $[0, \tau]$:

$$y(\tau) = \frac{1}{\tau} \int_0^{\tau} y_{fut}(t) dt = \frac{1}{\tau} \int_0^{\tau} \frac{-\delta'(\tau)}{\delta(\tau)} dt = \frac{-\ln(\delta(\tau))}{\tau} \quad (2.8)$$

sendo $\delta(\tau) = 1$, uma vez que o valor presente de um dado montante hoje ($\tau = 0$) é igual ao seu próprio valor.

Dessa forma, pode-se computar o preço de um título que paga cupons da seguinte forma:

$$\mathcal{P} = \sum_{i=1}^m \mathcal{C} \cdot e^{-\tau_i y} + \mathcal{A} \cdot e^{-\tau y} \quad (2.9)$$

O termo *curva de juros* é muito geral e associa-se à curva que descreve a estrutura a termo das taxas de juros. Pode ainda corresponder à curva instantânea de juros futura, função de desconto, curva de amortização de títulos públicos, e curva da taxa de juros à vista, ou curva de juros zero-cupom, termo mais utilizado, e que será considerado nesta dissertação.

2.2.2 Curva de Juros de Nelson e Siegel (1987)

O modelo paramétrico proposto por (Nelson & Siegel 1987) é amplamente utilizado por bancos centrais, investidores e pesquisadores para descrever a estrutura a termo das taxas de juros⁷. O modelo é baseado em uma aproximação exponencial de três componentes, de forma parcimoniosa e conveniente para descrever o comportamento da curva de rendimentos. A curva da taxa futura de juros é descrita pela seguinte função exponencial expandida:

$$(y_{fut})_t(\tau) = \beta_{1t} + \beta_{2t} e^{-\lambda_t \tau} + \beta_{3t} \lambda_t e^{-\lambda_t \tau} \quad (2.10)$$

em que β_{1t} , β_{2t} , β_{3t} e λ_t são os parâmetros do modelo relativos ao dia t .

A curva de NS para a taxa futura pode ser vista como uma constante somada a uma função de Laguerre, conhecida função matemática de aproximação (Courant & Hilbert 1953). Os limites da curva da taxa futura são tais que:

$$\lim_{\tau \rightarrow \infty} (y_{fut})_t(\tau) = \beta_{1t} \quad (2.11)$$

$$\lim_{\tau \rightarrow 0} (y_{fut})_t(\tau) = \beta_{1t} + \beta_{2t} \quad (2.12)$$

que correspondem, respectivamente, às interpretações de taxas instantâneas futuras de longo e curto prazos.

⁷(McCulloch 1971), (Chambers et al. 1984), (Diament 1993) e (Mansi & Phillips 2001) são exemplos de outras propostas funcionais para descrever a curva de juros, contudo, são pouco utilizadas na prática.

Os parâmetros β_3 e λ controlam a forma em que ocorre a transição entre as taxas de curto e longo prazos. A curva de juros pode assumir a forma de “U”, “U” invertido, ou formato senoidal ao invés de apenas um ponto de ótimo local se o parâmetro β_3 é negativo, positivo ou próximo de zero, respectivamente (Gimeno & Nave 2009). O parâmetro λ corresponde à taxa de decaimento exponencial da curva; em que baixos (elevados) valores de λ resultam em lento (rápido) decaimento da curva zero-cupom, e fornecem melhor aderência da curva em maturidades mais longas (curtas). Além disso, λ determina o ponto em que o parâmetro β_3 atinge o valor máximo.

Ao integrar a função da taxa de juros futura em (2.10), a curva de juros zero-cupom, denotada por $y_t(\tau)$, é obtida:

$$y_t(\tau) = \frac{1}{\tau} \int_0^\tau y_{fut}(t) dt = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \quad (2.13)$$

A literatura interpreta os parâmetros β_1 , β_2 e β_3 da função de NS, respectivamente, como os três fatores latentes dinâmicos de nível, inclinação e curvatura (Diebold & Li 2006). A persistência do fator de nível, β_1 , na função é igual a 1, confirmando a idéia de fator de “longo prazo”, uma vez que não decai a zero considerando seu limite. O fator β_2 , de inclinação, tem persistência $\left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} \right)$, função que inicia em 1 e decai monotonicamente a zero, relacionada ao fator de “curto prazo”. Finalmente, a função $\left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} - e^{-\lambda \tau} \right)$, que inicia em zero, aumenta, e então decai novamente a zero, está associada à persistência do fator de “médio prazo”, β_3 , ou fator de curvatura. O comportamento da persistência dos fatores da curva de NS pode ser verificado na Figura 2.1.

Por definição, as taxas de juros nominais à vista são estritamente positivas. Dessa forma, para que o modelo possua sentido econômico, as restrições aos parâmetros são tais que:

$$\begin{aligned} \beta_1 &> 0, \\ \beta_1 + \beta_2 &> 0, \\ \lambda &> 0 \end{aligned} \quad (2.14)$$

As duas primeiras inequações garantem que os extremos da curva de juros sejam positivos, enquanto a terceira inequação permite a existência de um limite finito para a taxa de longo prazo.

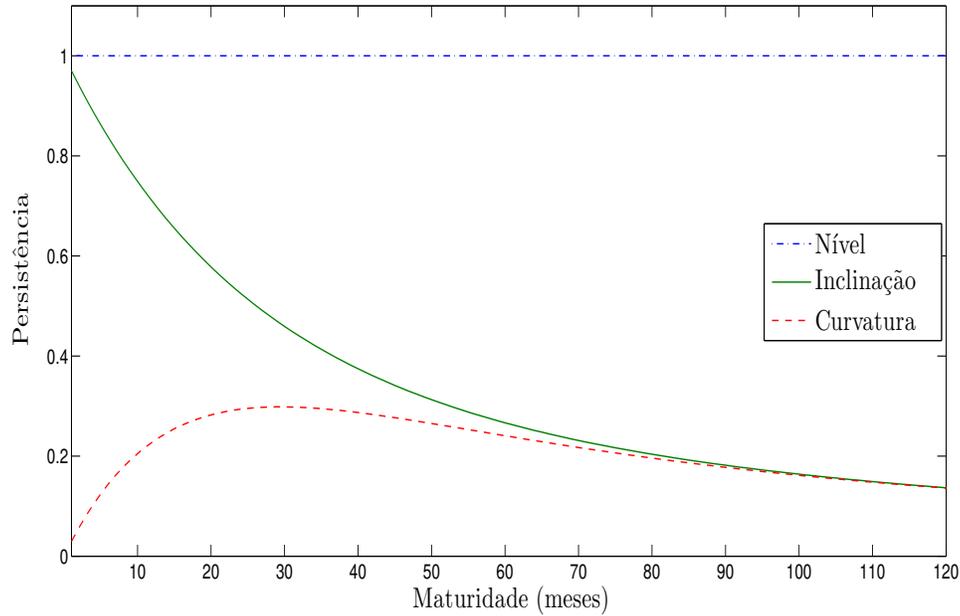


Figura 2.1: Persistência dos fatores de NS como função da maturidade, para $\lambda = 0.0609$.

2.3 Problema de Otimização de Parâmetros

O número de títulos públicos zero-cupom transacionados nos mercados não cobre todas as possíveis maturidades, implicando que a curva de juros deve ser estimada. Portanto, o problema associado consiste na estimação do conjunto de parâmetros $\mathbf{x} = \{\beta_1, \beta_2, \beta_3, \lambda\}$ que minimiza uma medida de erro entre os valores reais (observados nos mercados) e estimados (obtidos da curva de juros parametrizada) em relação aos preços ou taxas dos títulos zero-cupom disponíveis, ou ainda às taxas de outros instrumentos financeiros que são consideradas como *proxy* das taxas livres de risco.

A escolha entre preços ou taxas na função objetivo do processo de otimização não é definitiva. (Svensson 1994), por exemplo, sugere a minimização do erro obtido por meio das taxas, uma vez que produz melhor aderência da curva em maturidades mais curtas. Contudo, segundo (Manousopoulos & Michalopoulos 2009), para escolher o método mais apropriado de definição da função custo, é necessário levar em consideração o objetivo da estimação da curva de rendimentos. Se a curva de juros for aplicada como um indicador de política monetária, ou seja, se o foco são as taxas de juros, então a minimização em relação às taxas deve ser utilizada. Por outro lado, caso as curvas estimadas sejam utilizadas na precificação de títulos, é então preferível a minimização em relação aos preços.

Assim como na maioria dos trabalhos de estimação da estrutura a termo das taxas de juros, considerou-se nesta dissertação a minimização em relação às taxas livres de riscos na construção da curva de juros zero-cupom. Dessa forma, o problema de otimização de parâmetros da função de NS pode ser formulado da seguinte forma:

$$\min_{\mathbf{x}} \mathcal{F}(\mathbf{x}) = \min_{\beta_i, \lambda} \sum_{j=1}^N (y_j - \hat{y}_j(\mathbf{x}))^2 \cdot \frac{1}{\sqrt{D_j}} \quad (2.15)$$

$$\begin{aligned} \text{sujeito a} \quad & \beta_{1t} > 0 \\ & \beta_{1t} + \beta_{2t} > 0 \\ & \lambda_{1t} > 0 \end{aligned} \quad (2.16)$$

em que D é a *duration* dos títulos considerados, y a taxa à vista zero-cupom real observada, $\hat{y}(\mathbf{x})$ é a taxa implícita do título calculado por meio da estrutura a termo construída pelo conjunto de parâmetros estimados \mathbf{x} , $\mathcal{F}(\mathbf{x})$ denota a função objetivo ou também chamada de função de *fitness* e N o número de observações.

A não-linearidade do modelo é introduzida pelo parâmetro de decaimento λ . Vários trabalhos na literatura, tais como (Nelson & Siegel 1987) e (Diebold & Li 2006), costumam fixar o valor do parâmetro λ arbitrariamente para tornar a função que descreve a curva de juros linear em relação aos parâmetros β 's. Todavia, esse método está associado a limitações, uma vez que erros de especificação do modelo podem resultar em soluções não-ótimas para os parâmetros da curva de juros, pois não há uma técnica apropriada para a escolha de um valor adequado para o parâmetro de decaimento. Assim, neste trabalho, todos os parâmetros da função de NS foram considerados variantes no tempo.

Os métodos tradicionais de otimização, i.e., máxima verossimilhança e mínimos quadrados não-lineares, necessitam da determinação dos valores iniciais dos parâmetros β 's e λ para inicializar o processo iterativo em busca da solução ótima. Dessa forma, como a solução de um problema de otimização de parâmetros, considerando técnicas usuais, é muito sensível à determinação dos pontos iniciais, essa escolha é crucial para a acurácia da curva resultante. Neste trabalho, para evitar problemas de convergência, segue-se a proposta de (Manousopoulos & Michalopoulos 2009) para obter uma inicialização adequada:

$$\beta_1 = \frac{1}{l} \sum_{j=1}^l y_j, \quad \beta_2 = y_s - \beta_1, \quad \beta_3 = 0, \quad \lambda = 1 \quad (2.17)$$

sendo que os índices l e s correspondem às taxas com maturidades mais longas e à taxa de maturidade mais curta, respectivamente, presentes na amostra considerada para estimação.

Conforme apontado por (Gimeno & Nave 2009), métodos tradicionais de otimização não-linear possuem duas limitações importantes que podem interferir significativamente nos resultados. Em primeiro lugar, existe a possibilidade de os parâmetros estimados corresponderem a um ponto de mínimo local e não de mínimo global, ou seja, o risco de não-convergência. Esse risco emerge da elevada não-linearidade da função objetivo do problema de programação, e essa questão torna-se visível quando verificada uma baixa aderência dos resultados, considerando níveis de acurácia tradicionais. Além disso, pode-se observar uma elevada sensibilidade das estimativas às condições iniciais impostas na implementação das técnicas usuais de otimização. As consequências empíricas dessa limitação são resultados com alta variabilidade de acordo com o processo adotado para inicialização do algoritmo.

Apesar desses problemas, a literatura, em geral, busca desenvolver formas funcionais que melhor descrevam a estrutura a termo e satisfaçam teorias econômicas e financeiras intrínsecas, negligenciando o processo de otimização associado aos modelos parcimoniosos. (Csa,jbok 1998) sugeriu a utilização do algoritmo de otimização não-linear Gauss-Newton para estimar diversos modelos de curva de juros. A mesma análise foi realizada por (Ioannides 2003), mas considerando o algoritmo Broyden-Fletcher-Goldfarb-Shanno (BFGS)⁸. Por outro lado, (Bolder & Strélisk 1999) avaliaram as estimativas da função de NS em termos do processo de otimização envolvido, considerando algoritmos de busca como Simplex e de Programação Quadrática Sequencial. Análise similar também foi realizada por (Manousopoulos & Michalopoulos 2009), que utilizaram uma variedade de algoritmos de programação não-linear das classes de busca direta, assim como os baseados em gradientes e de otimização global para o problema de modelagem da ETTJ. Os autores mostraram que a escolha do algoritmo para estimar os parâmetros da função de (Nelson & Siegel 1987) não é simples, e que os resultados variam de acordo com a natureza do problema, sendo os algoritmos baseados em gradientes os que apresentam pior desempenho, devido à elevada não-linearidade da função objetivo do problema.

Para superar as limitações das tradicionais técnicas de otimização não-lineares, recentemente (Gimeno & Nave 2009) propuseram a aplicação de algoritmos genéticos para estimar os parâmetros das funções de (Nelson & Siegel 1987) e suas extensões. Os resultados foram promissores, no sentido de que o algoritmo genético proposto pelos autores foi capaz de gerar curvas de juros mais acuradas e, ao mesmo tempo, parâmetros mais estáveis, característica desejável com base na hipótese das expectativas dos agentes nos mercados monetários e de taxa de juros. Desde então, o mercado de renda fixa do Brasil, por meio da ANBIMA (Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais), passou a utilizar tal metodologia para geração das curvas de juros diárias⁹.

A modelagem da curva de juros utilizando algoritmos genéticos também foi considerada por (Fernández-Rodríguez 2006), porém, a curva foi gerada pela técnica de splines cúbicos. Os resultados mostraram uma elevada acurácia no ajuste da estrutura a termo em relação aos dados de mercado, superando técnicas tradicionais de otimização não-linear de parâmetros. Além disso, (Gilli & Winker 2008) avaliaram as diferenças entre taxas reais e estimadas como uma primeira abordagem do método de evolução diferencial para a ETTJ. O algoritmo de computação evolucionária apresentou maior estabilidade na geração das estimativas dos fatores da curva de NS em comparação à técnica de programação quadrática sequencial.

Esses resultados sugerem que a utilização de algoritmos evolucionários pode apresentar melhor desempenho no problema da modelagem da ETTJ em relação à técnicas tradicionais de programação. Dessa forma, conforme mencionado, um dos objetivos dessa dissertação é avaliar as técnicas de algoritmos genéticos, evolução diferencial e de estratégias evolutivas na modelagem dos fatores latentes da função de (Nelson & Siegel 1987). Para avaliação de desempenho, tomou-se como *benchmarks* as técnicas de otimização de mínimos quadrados não-lineares e de programação quadrática sequencial. As idéias básicas dessas técnicas serão descritas a seguir.

⁸O algoritmo BFGS trata-se de um método não-linear de programação da classe dos algoritmos Quasi-Newton (Nocedal & Wright 2006).

⁹Para detalhes da metodologia da ANBIMA para modelagem da ETTJ, ver: http://www.andima.com.br/est_termo/arqs/est-termo_metodologia.pdf.

2.4 Mínimos Quadrados Não-Lineares

O método dos mínimos quadrados não-lineares (MQNL) busca um vetor \mathbf{x}^* que minimize a função objetivo:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \mathcal{F}(\mathbf{x}) = \sum_{i=1}^m r_i^2 = \frac{1}{2} R(\mathbf{x})^T R(\mathbf{x}) \quad (2.18)$$

com $m > n$, em que $R : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ representa uma função residual de forma que:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \rightarrow R(x) = \begin{bmatrix} r_1(\mathbf{x}) \\ \vdots \\ r_m(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}(\mathbf{x}) \\ \vdots \\ y_m - \hat{y}(\mathbf{x}) \end{bmatrix} \quad (2.19)$$

Esses problemas aparecem no contexto de ajuste de dados, como é o caso da modelagem da estrutura a termo das taxas de juros, em que $\mathbf{x} = [\beta_1 \ \beta_2 \ \beta_3 \ \lambda]^T$, sendo $n = 4$.

O processo de obtenção da solução ótima \mathbf{x}^* é iterativo, uma vez que, a partir de um vetor inicial \mathbf{x}_0 , o método produz uma série de vetores $\mathbf{x}_1, \mathbf{x}_2, \dots$ que deverá convergir para \mathbf{x}^* de acordo com um determinado número k de iterações. A condição de convergência é tal que:

$$\mathcal{F}(\mathbf{x}_{k+1}) < \mathcal{F}(\mathbf{x}_k) \quad (2.20)$$

O algoritmo gera a cada iteração k um ajuste Δ ao vetor de parâmetros \mathbf{x} , ou seja:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta_k \quad (2.21)$$

No caso da função objetivo possuir mais de um mínimo local, a solução ótima \mathbf{x}^* obtida irá depender da inicialização em \mathbf{x}_0 (Nocedal & Wright 2006).

O gradiente da função objetivo \mathcal{F} é descrito como:

$$\nabla \mathcal{F}(\mathbf{x}) = \sum_{i=1}^m (r_i(\mathbf{x}) \nabla r_i(\mathbf{x})) = J(\mathbf{x})^T R(\mathbf{x}) \quad (2.22)$$

em que $J(\mathbf{x}) \in \mathfrak{R}^{m \times n}$ é a matriz Jacobiana, em \mathbf{x} , da função vetorial R .

A matriz Hessiana $\nabla^2 \mathcal{F}$ é derivada da seguinte forma:

$$\nabla^2 \mathcal{F}(\mathbf{x}) = \sum_{i=1}^m (\nabla r_i(\mathbf{x}) \nabla r_i(\mathbf{x})^T + \nabla r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x})) = J(\mathbf{x})^T J(\mathbf{x}) + S(\mathbf{x}) \quad (2.23)$$

na qual $S(\mathbf{x}) = \sum_{i=1}^m r_i(\mathbf{x}) \nabla^2 r_i(\mathbf{x})$.

Note ainda que:

$$R(\mathbf{x}^*) = 0 \rightarrow S(\mathbf{x}^*) = 0 \quad e \quad \nabla^2 \mathcal{F}(\mathbf{x}^*) = J(\mathbf{x}^*)^T J(\mathbf{x}^*) \quad (2.24)$$

Existem diversos métodos para solução iterativa do problema de mínimos quadrados não-lineares, dentre eles, o algoritmo de Gauss-Newton, que será considerado nesta dissertação, de Regiões de Confiança e o de Levenberg-Marquardt. O método de Gauss-Newton consiste na aplicação do método de Newton à resolução do sistema de equações não-lineares $\nabla \mathcal{F}(\mathbf{x}) = 0$,

ignorando a contribuição do termo $S(\mathbf{x})$ da matriz Hessiana $\nabla^2 \mathcal{F}(\mathbf{x})$. O algoritmo gera uma sucessão de vetores $\{\mathbf{x}_k\}$ de acordo com a seguinte atualização:

$$J(\mathbf{x})^T J(\mathbf{x}) \Delta = J(\mathbf{x})^T R(\mathbf{x}) \quad (2.25)$$

O método de Gauss-Newton gera iterações de \mathbf{x}_k para os quais $J(\mathbf{x}_k)^T J(\mathbf{x}_k)$ não seja singular. Uma propriedade interessante desse método é a geração de passos que são direções de descida para a função $\mathcal{F}(\mathbf{x})$.

O algoritmo de Gauss-Newton pode ser sumarizado de acordo com os passos descritos abaixo¹⁰.

Método de Gauss-Newton para Mínimos Quadrados Não-Lineares

Inicializar $\mathbf{x}_0 \in \mathfrak{R}^n$

Escolher $\epsilon \sim 0$, positivo

Para $k = 1, 2, \dots$

Resolver o sistema de equações lineares $J(\mathbf{x}_k)^T J(\mathbf{x}_k) \Delta = J(\mathbf{x}_k)^T R(\mathbf{x}_k)$

Atualizar $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta_k$

Calcular $\|\nabla \mathcal{F}(\mathbf{x}_k)\|$

Se $\|\nabla \mathcal{F}(\mathbf{x}_k)\| > \epsilon$

Fim Para

O método de Gauss-Newton é utilizado para resolver problemas irrestritos de otimização de parâmetros. Contudo, a modelagem da estrutura a termo das taxas de juros, com base no modelo de fatores latentes de NS em (2.15), constitui um problema restrito, uma vez que os parâmetros têm interpretação financeira de taxas de juros e por isso devem ser positivos. Para aproximar problemas restritos por problemas irrestritos e então utilizar a técnica de Gauss-Newton, implementou-se um método de penalidade, caracterizado por adicionar um termo de alto custo no caso de violação das restrições da função objetivo. Nesse caso, o problema de otimização em (2.15) pode ser reescrito da seguinte forma:

$$\min_{\mathbf{x} \in \mathfrak{R}^n} \mathcal{F}(\mathbf{x}) + c \cdot \mathcal{B}(\mathbf{x}) \quad (2.26)$$

sendo $c > 0$ e \mathcal{B} uma função contínua tal que $\mathcal{B}(\mathbf{x}) \geq 0$ para todo $\mathbf{x} \in \mathfrak{R}^n$, e $\mathcal{B}(\mathbf{x}) = 0$ se $\mathbf{x} \in \mathcal{D}$, em que $\mathcal{D} = \{\mathbf{x} \in \mathfrak{R}^n : \beta_1 > 0, \beta_1 + \beta_2 > 0, \lambda > 0\}$ representa o domínio da função.

O parâmetro c penaliza o processo de otimização quando as restrições do problema são violadas. Idealmente, quando $c \rightarrow \infty$, as soluções do problema restrito e irrestrito se aproximam ($\mathcal{B}(\mathbf{x}) \rightarrow 0$).

O espaço paramétrico do problema de construção da curva de juros envolve apenas restrições de desigualdade do tipo $g(\mathbf{x}) \leq \mathbf{0}$. Dessa forma, a função de penalização é:

$$\mathcal{B}(\mathbf{x}) = \sum_{i=1}^{m_D} (\max\{0, g_i(\mathbf{x})\})^2 \quad (2.27)$$

¹⁰Para maiores detalhes do algoritmo ver (Nocedal & Wright 2006).

em que m_D representa o número de restrições de desigualdade do problema (no caso da função de NS, $m_D = 3$).

O algoritmo a seguir apresenta o algoritmo de Gauss-Newton associado ao método de penalidade para solução do problema de otimização de parâmetros da curva de juros. Quando o parâmetro c é grande, o mínimo do problema irrestrito está situado numa região onde o termo $c \cdot \mathcal{B}(\mathbf{x})$ é pequeno. Dessa forma, quando $c \rightarrow \infty$, a solução do problema irrestrito tende para a solução do problema restrito. Portanto, o procedimento adequado é resolver uma sequência de problemas irrestritos com $c \rightarrow \infty$, necessitando atualizar o parâmetro c por outro parâmetro κ , conforme mostrado abaixo.

Método de Gauss-Newton com Penalidade

Inicializar $\mathbf{x}_0 \in \mathfrak{R}^n$

Escolher $c_0 > 0$, $\kappa > 1$ e $\epsilon \sim 0$, positivo

Para $k = 1, 2, \dots$

Enquanto $c_k \mathcal{B}(x_k) \geq \epsilon$ **Faça**

Determinar \mathbf{x}_{k+1} resolvendo (2.26)

Atualizar $c_{k+1} = \kappa c_k$

Fim Enquanto

Fim Para

2.5 Programação Quadrática Sequencial

A ideia principal do algoritmo de programação quadrática sequencial (*Sequential Quadratic Programming* - SQP) é obter uma direção de busca em problemas de otimização de parâmetros com restrições, resolvendo um subproblema quadrático, com uma aproximação quadrática para a função objetivo e restrições lineares. Dessa forma, o problema de otimização é transformado em um subproblema mais simples que pode ser resolvido por meio de um processo iterativo e sem restrição. Quando há convergência, o SQP é considerado um dos melhores métodos de otimização em programação não-linear (Rao 1996). Em cada iteração do algoritmo SQP, uma aproximação é feita para a inversa da matriz Hessiana da função do lagrangeano, utilizando uma atualização por meio do método Quasi-Newton, para então gerar um subproblema quadrático no qual se obtém uma nova direção de busca.

Considerando novamente a formulação para o problema de otimização, e incluindo restrições de igualdade e desigualdade, como caso geral, obtemos:

$$\begin{array}{ll}
 \min_{\mathbf{x} \in \mathfrak{R}^n} & \mathcal{F}(\mathbf{x}) \\
 \text{sujeito a} & g_j(\mathbf{x}) \leq 0 \quad j = 1, 2, \dots, m_D \\
 & h_i(\mathbf{x}) = 0 \quad i = 1, 2, \dots, m_I
 \end{array} \tag{2.28}$$

$g_D(\mathbf{x})$ e $h_I(\mathbf{x})$ representam as restrições de desigualdade e igualdade do problema, respectivamente.

O algoritmo SQP é caracterizado essencialmente por três estágios principais: i) atualização da matriz Hessiana; ii) solução do subproblema quadrático e da nova direção de busca, iii) avaliação de uma função de mérito para garantir convergência global.

Para o problema em (2.28), a função lagrangeana \mathcal{L} é representada por:

$$\mathcal{L}(\mathbf{x}, \lambda, \mu) = \mathcal{F}(\mathbf{x}) + \sum_{i=1}^{m_I} \mu_i^I h_i(\mathbf{x}) + \sum_{j=1}^{m_D} \mu_j^D g_j(\mathbf{x}) \quad (2.29)$$

sendo μ^I e μ^D correspondem aos multiplicadores de Lagrange associados às condições de igualdade e desigualdade, respectivamente.

Em cada iteração do algoritmo SQP, um subproblema quadrático é resolvido por meio de uma aproximação quadrática do lagrangiano da função objetivo e linearização das restrições, da seguinte forma:

$$\begin{aligned} \text{minimizar} \quad & \frac{1}{2} d^T \mathcal{L}(\mathbf{x}) d + \nabla \mathcal{L}(\mathbf{x})^T d + \mathcal{L}(\mathbf{x}) \\ \text{sujeito a} \quad & h(\mathbf{x}) + \nabla h(\mathbf{x})^T d = \mathbf{0} \\ & g(\mathbf{x}) + \nabla g(\mathbf{x})^T d \leq \mathbf{0} \end{aligned} \quad (2.30)$$

em que $d = \mathbf{x}_{k+1} - \mathbf{x}_k$.

A principal idéia do algoritmo SPQ é, a cada iteração, gerar uma matriz definida positiva como aproximação \hat{H}_k da inversa da Hessiana da função do lagrangeano, que pode ser gerada, por exemplo, pelo método BFGS:

$$\hat{H}_{k+1} = \hat{H}_k + \frac{q_k q_k^T}{q_k^T d_k} - \frac{\hat{H}_k d_k d_k^T \hat{H}_k}{d_k^T \hat{H}_k d_k} \quad (2.31)$$

sendo

$$q_k = \nabla \mathcal{L}(\mathbf{x}_{k+1}) - \nabla \mathcal{L}(\mathbf{x}_k) \quad (2.32)$$

Dessa forma, a regra de atualização do algoritmo é tal que:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta_k d_k, \quad \Delta_k \in [0, 1] \quad (2.33)$$

O tamanho do passo Δ_k da direção de descida é obtido da seguinte forma:

$$\Delta_k = \min_{\Delta \geq 0} \psi(\mathbf{x}_k + \Delta d_k) \quad (2.34)$$

onde

$$\psi(\mathbf{x}) = \mathcal{F}(\mathbf{x}) + c \left[\sum_{i=1}^{m_I} |h_i(\mathbf{x})| + \sum_{j=1}^{m_D} \max\{0, g_j(\mathbf{x})\} \right], \quad c > 0 \quad (2.35)$$

a função mérito, que permite a obtenção do tamanho do passo na direção de descida, e confere uma condição necessária para convergência global do algoritmo.

Apropriadamente, o algoritmo SQP converge para uma solução ótima satisfazendo as condições de Kuhn-Tucker do problema original. Para detalhes do algoritmo SQP ver (Boggs & Tolle 1995). O algoritmo de programação quadrática sequencial é apresentado a seguir.

Método de Programação Quadrática Sequencial

Inicializar $\mathbf{x}_0 \in \mathfrak{R}^n$

Escolher $c > 0$, $\hat{H}_0 > 0$ e $\epsilon \sim 0$, positivo

Resolver (2.30) com $\mathcal{L}(\mathbf{x}_0) = \hat{H}_0$

Computar d_0 , μ_1^I e $\mu_1^D \geq 0$

Para $k = 1, 2, \dots$

Enquanto $\|d_k\| < \epsilon$ **Faça**

$$\begin{aligned} \text{Determinar } (d_k, \mu_{k+1}^I, \mu_{k+1}^D) = \min & \quad \frac{1}{2} d^T \hat{H}_k(\mathbf{x}_k) d + \nabla \mathcal{F}(\mathbf{x}_k)^T d + \mathcal{F}(\mathbf{x}_k) \\ \text{s.a. } & \quad h(\mathbf{x}) + \nabla h(\mathbf{x})^T d = \mathbf{0} \\ \text{s.a. } & \quad g(\mathbf{x}) + \nabla g(\mathbf{x})^T d \leq \mathbf{0} \end{aligned}$$

Resolver $\Delta_k = \min_{\Delta \geq 0} \psi(\mathbf{x}_k + \Delta d_k)$

Atualizar $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta_k d_k$

Computar $d_k = \mathbf{x}_{k+1} - \mathbf{x}_k$

Obter $q_k = \nabla \mathcal{L}(\mathbf{x}_{k+1}) - \nabla \mathcal{L}(\mathbf{x}_k)$

Atualizar $\hat{H}_{k+1} = \hat{H}_k + \frac{q_k q_k^T}{q_k^T d_k} - \frac{\hat{H}_k d_k d_k^T \hat{H}_k}{d_k^T \hat{H}_k d_k}$

Fim Enquanto

Fim Para

2.6 Resumo

Esse capítulo introduziu o conceito de modelagem da estrutura a termo das taxas de juros por meio do modelo parcimonioso de fatores latentes proposto por (Nelson & Siegel 1987). Além disso, o problema de otimização de parâmetros associado à construção da curva de juros foi formulado, de forma que as abordagens tradicionais utilizadas na literatura e que também serão consideradas nessa dissertação, algoritmos de mínimos quadrados não-lineares e de programação quadrática sequencial, foram brevemente descritos.

O próximo capítulo apresenta as abordagens de computação evolucionária aplicadas ao problema de estimação da estrutura a termo, sendo elas: algoritmos genéticos, de evolução diferencial e de estratégias evolutivas. Esses algoritmos serão formulados para o problema de modelagem da ETTJ, para então avaliar essas metodologias em relação aos *benchmarks* considerados.

Computação Evolucionária

Este capítulo descreve os algoritmos evolucionários implementados na estimação da estrutura a termo das taxas de juros, i.e., Algoritmos Genéticos, de Evolução Diferencial e de Estratégias Evolutivas. Uma breve revisão bibliográfica da aplicação desses métodos em problemas de economia e finanças é reportada nas seções referentes à cada metodologia, seguida das respectivas descrições dos algoritmos, considerando o problema de modelagem da curva de juros.

3.1 Algoritmos Genéticos

As técnicas de computação evolucionária são um dos importantes paradigmas da computação inteligente. Tratam-se de métodos estocásticos de busca e otimização de sistemas complexos que associam as ideias de otimização e conceitos da biologia evolutiva. A computação evolucionária utiliza progressos iterativos em uma população selecionada por meio de uma busca randômica guiada via processamento paralelo para obter um resultado desejável (Das & Suganthan 2011). Os Algoritmos Genéticos (AGs), introduzidos por (Holland 1975), partem de uma população inicial de indivíduos, também chamados de *genes* ou *cromossomos*, sendo cada indivíduo um conjunto de parâmetros que representa um possível candidato à solução do problema de otimização em questão.

A evolução em tais algoritmos ocorre por meio de gerações, i.e., a cada geração alguns indivíduos da população são selecionados e a nova geração é criada por meio da reprodução e mutação desses indivíduos. A aptidão de cada indivíduo é computada para determinar a seleção e composição da população da nova geração. Essa aptidão está associada ao problema de otimização considerado, ou seja, é proporcional ao valor da função objetivo do problema avaliado no conjunto de parâmetros definido por cada indivíduo. Portanto, os cromossomos que representam melhores soluções para o problema irão sobreviver e reproduzir para formar novos indivíduos por meio de recombinação e mutação, e assim sucessivamente (Mitchell 1998). Ao longo das gerações a população convergirá para a solução “ótima” do problema.

Técnicas baseadas em AGs vêm sendo aplicadas em diversos problemas de otimização em economia e finanças, com destaque para a questão de otimização de portfólios (Lin, Wang & Yan 2001, Oh, Kin, Min & Lee 2006, Lin & Liu 2008). (Chang, Yang & Chang 2009), por exemplo, propuseram a aplicação de algoritmos genéticos na otimização de carteiras de acordo

com distintas medidas de riscos, em especial considerando volatilidade estocástica. Composição de carteiras considerando AGs ainda foi avaliada por (Soleimani, Golmakani & Salimi 2009), mas os autores adicionaram restrições associadas a custos mínimos de transação, limites de cardinalidade e capitalização setorial. As restrições associadas tornam o problema de otimização de carteiras mais complexo, de forma a permitir que os algoritmos evolucionários apresentem como mais apropriados, pois não estão associados às condições de regularidade do espaço paramétrico como as técnicas tradicionais de programação não-linear. Nessa mesma linha de pesquisa, (Golmohammadi & Pajoutan 2011) desenvolveram um modelo que inclui custos de transação à seleção de portfólios, tornando a técnica mais realista. Os resultados revelaram que o método de AG proposto apresentou elevado desempenho, produzindo carteiras de maior rentabilidade associadas a um nível mínimo de risco.

Otimização de portfólios com inúmeros ativos, considerados de grande porte, foi abordada por (Chen, Mabu & Hirasawa 2011). Alternativamente, os autores propuseram um modelo de algoritmo genético guiado pelo operador de mutação para coordenar a evolução dos indivíduos da população. Nesse caso, a evolução estava associada à capacidade de gerar populações diversificadas ao longo das gerações. Os resultados dos autores foram promissores, no sentido de acelerar a evolução, quando comparados a tradicionais algoritmos genéticos comumente caracterizados por utilizar uma função de *fitness* para guiar o processo iterativo. Por outro lado, (Anagnostopoulos & Mamanis 2011) compararam o efeito de cinco algoritmos evolutivos multi-objetivo considerados estado-da-arte para otimização de portfólios de acordo com modelos de média-variância. Os autores destacaram os bons resultados obtidos pelos algoritmos genéticos em termos de geração de uma fronteira de Pareto adequada aos dados dos ativos avaliados¹.

Algoritmos genéticos também vêm sendo com sucesso aplicados na previsão de carteiras baseadas em medidas de variabilidade como o Valor-em-Risco (Lin & Ko 2009), desempenho de estratégias em mercados de ações (Ehrentreich 2006), avaliação de mercados de ações artificiais de múltiplos ativos considerando relações entre a dinâmica da riqueza e aversão ao risco dos agentes (Chen & Huang 2007), predição de falência bancária (Shin & Lee 2002), gerenciamento de carteiras financeiras com base em indicadores de análise técnica (Gorgulho, Neves & Horta 2011), treinamento de redes neurais artificiais para tratamento de dados financeiros (Kim 2006), e estimação da estrutura a termo das taxas de juros (Gimeno & Nave 2009).

Os algoritmos genéticos são caracterizados pelos seus operadores de seleção, cruzamento (*crossover*) e mutação (Mitchell 1998). Esses operadores podem ser adaptados de acordo com o problema analisado, definindo assim a estratégia empregada pela técnica evolucionária.

Nesta dissertação, será implementado o algoritmo genético proposto por (Gimeno & Nave 2009) para estimar os parâmetros da estrutura a termo da taxa de juros de acordo com o modelo parcimonioso de (Nelson & Siegel 1987), considerando como *benchmark* dos algoritmos evolucionários para comparações. Dessa forma, os mesmos operadores sugeridos pelos autores serão considerados. Contudo, o processo de inicialização do algoritmo será diferente, de forma a adequar o método à periodicidade dos dados utilizados. O algoritmo genético conforme definido por (Gimeno & Nave 2009) e implementado neste trabalho é apresentado a seguir. Além disso, a

¹(Tollo & Rolli 2008) resumem as principais aplicações de meta-heurísticas para o problema de seleção de portfólios.

Figura 3.1 apresenta o fluxograma que descreve o algoritmo. As subseções seguintes descrevem os operadores genéticos envolvidos no processo evolutivo do algoritmo genético em questão, reportando os passos apresentados abaixo.

Algoritmo Genético para Estimação da ETTJ (Gimeno & Nave 2009)

Determinar β_i^* e λ^*
 Inicializar a população de N indivíduos em torno de β_i^* e λ^*
 Iniciar contador de gerações $G = 0$
Enquanto $G < G_{max}$ **Faça**
 Selecionar soluções viáveis da população inicial
 Computar o *fitness* da população
 Ordenar a população em relação ao *fitness*
 $\mathbf{x}_0^* = \{\hat{\beta}_i^*, \hat{\lambda}^*\}$ é a solução mais apta (maior *fitness*)
 Para $t = 1$ **até** N **Faça**
 Selecionar pais: $r, s = N \cdot \xi$ $\xi \sim B(1, \alpha)$
 Gerar descendentes via crossover $\mathbf{x}_k = \Xi \cdot \mathbf{x}_r + (1 - \Xi) \cdot \mathbf{x}_s$
 Fim Para
 Substituir os piores indivíduos da população pelos descendentes
 Mutar a população (exceto \mathbf{x}_0)
 Se $\mathcal{F}(\mathbf{x}_0^*) < \mathcal{F}(\mathbf{x}_G^*)$ **Então**
 $G = G + 1$
 Caso Contrário
 $G = G$
 Fim Se
Fim Enquanto

3.1.1 Inicialização

Em geral, a população inicial nos algoritmos genéticos é gerada aleatoriamente, sendo composta por soluções candidatas para a estimação da curva de juros com base no modelo de (Nelson & Siegel 1987). No contexto de otimização de parâmetros, as soluções candidatas ao modelo podem ser vistas como diferentes indivíduos ou cromossomos (\mathbf{x}), ou seja, cada cromossomo é um conjunto solução de valores para os parâmetros da curva de NS que geram um melhor ajuste considerando os dados reais². Portanto, cada cromossomo (\mathbf{x}_i) é um candidato à solução do problema de otimização descrito em (2.15) no Capítulo 2, de forma que, para a função de fatores latentes, tem-se:

$$\mathbf{x}_i = (\beta_{1i}, \beta_{2i}, \beta_{3i}, \lambda_i), \quad i = 1, 2, \dots, N \quad (3.1)$$

em que N representa o número de indivíduos que compõem a população.

Como discutido no Capítulo 2, os parâmetros da função de (Nelson & Siegel 1987) possuem uma interpretação financeira específica, o que permite uma inicialização mais apropriada. Em algoritmos evolucionários, é garantido que nenhum ponto no espaço de busca possua probabilidade zero de ser examinado. Dessa forma, ao se escolherem valores iniciais mais “próximos” aos

²Assim como (Gimeno & Nave 2009), os algoritmos foram implementados com codificação em ponto flutuante.

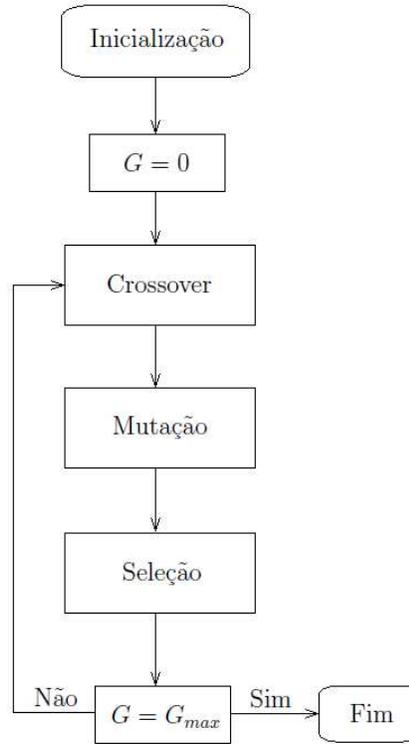


Figura 3.1: Fluxograma do algoritmo genético para estimação da ETTJ.

valores ótimos, o número de gerações (iterações) necessárias para o algoritmo convergir é reduzido drasticamente (Gimeno & Nave 2009). Nas técnicas tradicionais de otimização, o mesmo não se verifica, uma vez que a inicialização determina a qualidade da solução obtida (ótimo local ou ótimo global). Para reduzir a complexidade computacional, em termos de tempo de processamento, neste trabalho, a população inicial, como em (Gimeno & Nave 2009), foi gerada como:

$$\beta_{i,j} = \beta_j^* + \varepsilon_i, \quad i = 1, 2, \dots, N, \quad j = 1, 2, 3 \quad (3.2)$$

$$\lambda_i = \lambda^* + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (3.3)$$

em que os valores “sementes” de β_i^* e λ^* são gerados ao redor de $\varepsilon_i \sim N(0, \sigma)$, cujo desvio padrão, σ , é o parâmetro de controle da diversidade da população inicial na primeira geração.

(Gimeno & Nave 2009) determinaram os valores sementes dos parâmetros com base em conceitos financeiros relacionados às taxas de curto, médio e longo prazos dos títulos utilizados na amostra de dados considerada.

Em (Manousopoulos & Michalopoulos 2009) e em (Gimeno & Nave 2009), depois de realizado o processo de otimização para o primeiro dia da amostra e de terem sido obtidos os valores dos parâmetros ótimos, para o dia seguinte da curva de juros, metade da população inicial foi gerada com base nos valores sementes, seguindo seus métodos específicos de determinação desses valores, e a outra metade da população inicial foi gerada aleatoriamente considerando como valores sementes as estimativas obtidas do primeiro dia. Os autores assumiram que são

verificadas pequenas variações na curva de juros entre dias consecutivos e que os valores ótimos dos parâmetros do dia anterior são pontos iniciais aceitáveis para os algoritmos de otimização.

Nesta dissertação, serão utilizadas duas bases de dados, uma com dados mensais dos Estados Unidos, e dados diários do Brasil. O mesmo método dos autores citados, de geração da população inicial a partir de valores ótimos estimados no dia anterior, será aplicado apenas para os dados diários. Quando utilizada a base de dados mensais, os valores β_j^* e λ^* serão determinados de acordo com a regra proposta por (Manousopoulos & Michalopoulos 2009), conforme descrita no capítulo anterior, metodologia que independe da base de dados considerada e garante o sentido econômico-financeiro do modelo.

3.1.2 Seleção

Determinada a população inicial de cromossomos, o processo iterativo do algoritmo inicia. Como os parâmetros das funções que descrevem a curva de juros estão sujeitos a restrições de positividade, segundo o sentido econômico de taxa de juros, inicialmente, foram selecionados apenas os indivíduos viáveis, i.e., aqueles que satisfazem as restrições definidas no capítulo anterior.

Nesse caso, a população é composta apenas por indivíduos que satisfazem as restrições, uma vez que a utilização de cromossomos não viáveis apenas prolonga o tempo de processamento do algoritmo. Computa-se o *fitness* dos indivíduos da população, que então é ordenada em ordem crescente de seu valor correspondente. O *fitness* de cada indivíduo corresponde ao valor da função objetivo definida em (2.15) no Capítulo 2, ou seja, avalia-se o conjunto de parâmetros obtidos que definem o quanto a curva de juros gerada adere aos dados reais observados nos mercados. Como se trata de uma medida de minimização de uma diferença quadrática de desvios, quanto menor o valor da função custo, melhor será o valor resultante em relação ao problema. O *fitness* corresponde à aderência da curva gerada pelo conjunto de parâmetros que formam um cromossomo aos dados de mercado correspondentes. Deve-se perceber, ainda, que essa é uma medida de adaptabilidade ou capacidade de sobrevivência dos indivíduos da população.

Conforme (Gimeno & Nave 2009), após a ordenação da população em relação ao *fitness*, a melhor solução da geração atual, ou seja, que fornece o menor valor para a função objetivo, é armazenada. Em seguida, é definida a proporção $\phi \in [0, 1]$ dos indivíduos que irão sobreviver para a próxima geração. De acordo com os autores, se o número de sobreviventes é muito baixo, o processo de convergência do algoritmo pode ser muito lento. Por outro lado, quando essa proporção de sobreviventes é elevada, o algoritmo genético pode resultar em um ponto de mínimo local, uma vez que os indivíduos de baixa capacidade de sobrevivência são mantidos na população, impedindo que novos indivíduos mais bem adaptados componham as gerações subsequentes. (Gimeno & Nave 2009), em particular, afirmam que um valor adequado dessa proporção para os dados avaliados foi $\phi = 0.4$, i.e., apenas 40% dos indivíduos sobreviverão em cada geração.

3.1.3 Crossover

Como medida de adaptabilidade, o *fitness* dos indivíduos da população é utilizado diretamente na evolução, no sentido de selecionar os indivíduos mais aptos para reproduzi-los por meio do operador genético de crossover. Em cada geração, os indivíduos com pior desempenho serão substituídos por cada descendente (\mathbf{x}_k) gerado por meio da reprodução de dois indivíduos (\mathbf{x}_r e \mathbf{x}_s) da geração anterior. O operador de crossover utilizado será o crossover aritmético, i.e., os descendentes resultam da combinação linear convexa de dois progenitores (Goldberg 1991):

$$\mathbf{x}_k = \Xi \cdot \mathbf{x}_r + (1 - \Xi) \cdot \mathbf{x}_s \quad (3.4)$$

na qual Ξ é um vetor coluna composto por variáveis aleatórias uniformes no intervalo $[0, 1]$, com componentes específicos para cada par (r, s) .

A seleção dos progenitores da nova geração levará em consideração a aptidão dos indivíduos, ou seja, de acordo com o *fitness* de cada conjunto ou vetor de parâmetros. Os indivíduos com maior probabilidade de reprodução serão aqueles com maior aptidão (menor valor de *fitness*). Portanto, r e s serão selecionados da seguinte forma:

$$r, s = N \cdot \xi, \quad \xi \sim B(1, \alpha) \quad (3.5)$$

em que ξ é uma variável aleatória com distribuição Beta com parâmetros 1 e α . A distribuição Beta vai de 0 a 1 como a distribuição uniforme, mas sua parametrização permite diversos formatos de distribuição. Se α é maior que 1, os valores mais baixos da distribuição estarão associados a maiores probabilidades. Portanto, quanto maior o valor desse parâmetro, maior a probabilidade de valores baixos, de forma que, no problema da curva de juros, indivíduos com menor *fitness* terão maior probabilidade de serem selecionados como progenitores.

3.1.4 Mutação

Para evitar que a convergência do algoritmo resulte em um mínimo local, os indivíduos foram submetidos ao operador de mutação. Dessa forma, os parâmetros de cada indivíduo (com exceção do melhor indivíduo de cada geração) podem sofrer mutação com uma dada probabilidade π .

Quando a mutação ocorre, é adicionada uma variável aleatória $\varepsilon \sim N(0, \sigma)$ ao parâmetro. Esse processo de mutação é independente para cada parâmetro de cada indivíduo.

A magnitude da mutação é dada pelo parâmetro σ^* , que a cada geração será acrescido. Na primeira geração σ^* será igual ao desvio padrão que controla a diversidade na criação da população. Nas gerações subsequentes o valor será acrescido em 2% (Gimeno & Nave 2009). Esse procedimento é adotado para evitar o risco de não-convergência, devido à elevada não-linearidade da função objetivo (Yang, Tian & Yuan 2007).

Por fim, são diversos os critérios de parada que podem ser adotados em um algoritmo evolutivo, tais como, número máximo de gerações, variação no valor da função objetivo, ou mesmo quando a função objetivo atinge um determinado limiar. Como proposto por (Gimeno & Nave 2009), completada uma geração é determinado um novo melhor indivíduo. Se esse indivíduo for igual a melhor solução da geração anterior, o contador de gerações é acrescido,

caso contrário mantém-se o mesmo. O processo continua até que em G_{max} gerações não ocorra mudança ao indivíduo mais apto. Portanto, o critério de parada escolhido é tal que a função objetivo não deve sofrer alteração após G_{max} iterações.

No problema desta dissertação, a ETTJ será estimada diariamente para os dados Brasileiros e mensalmente para os dados Americanos, de forma que para cada processo de otimização é necessário um número distinto de iterações para atingir a solução ótima. Para evitar alguma restrição ao impor um número máximo de gerações para todos os dias ou meses, o critério de não variação da melhor solução durante um número pré-especificado de gerações é mais adequado. Assim, os parâmetros de controle do algoritmo genético implementado são: N , σ , G_{max} , ϕ , α e π .

3.2 Evolução Diferencial

Metaheurística inspirada também na teoria da evolução biológica, o algoritmo de Evolução Diferencial (ED) foi proposto por (Storn & Price 1997) e é caracterizado por se tratar de um método de otimização eficiente para problemas complexos, que utiliza procedimentos de seleção baseados na aptidão dos indivíduos e operadores de cruzamento e mutação. O método consiste essencialmente na geração aleatória de uma população de indivíduos, em que cada indivíduo representa um ponto de busca no espaço de soluções potenciais do problema em questão. A inicialização da população é definida de acordo com o conhecimento *a priori* do projetista sobre o problema ou, caso contrário, aleatoriamente.

O algoritmo realiza sucessivas modificações na população inicial para dar lugar a uma nova população de mesmo tamanho, constituindo assim uma nova geração até que o critério de parada adotado seja satisfeito. As modificações se verificam com a geração de novos indivíduos, denominados vetores modificados ou doadores, obtidos por meio da adição da diferença ponderada entre dois indivíduos aleatórios da população a um terceiro indivíduo também escolhido aleatoriamente, constituindo assim o operador de mutação do algoritmo de evolução diferencial.

Em seguida, um vetor alvo, também escolhido aleatoriamente, cujas componentes são intercambiadas com as componentes do vetor doador, gera um vetor experimental, caracterizando o operador de cruzamento entre os indivíduos. A composição da nova geração é então descrita pela seleção dos vetores que melhoram o valor da função objetivo ao se comparar o vetor alvo e o vetor experimental obtidos no processo evolutivo.

As principais características que tornam os algoritmos de evolução diferencial uma ferramenta de otimização promissora são: simplicidade de implementação³; elevada acurácia e robustez no tratamento de problemas com funções objetivo separáveis, não-separáveis, modais e multimodais (Zhang & Sanderson 2003); baixo número de parâmetros de controle (no algoritmo de ED clássico são apenas três)⁴; e, finalmente, para problemas de otimização de larga escala e

³Métodos baseados em enxames de partículas (*particle swarm optimization*) também são de implementação simples. Contudo, (Das, Abraham, Chakraborty & Konar 2009) e (Rahnamayan, Tizhoosh & Salama 2008) mostraram que o desempenho dos algoritmos de ED é superior se comparada com a técnica de enxame de partículas em variados problemas de otimização.

⁴Trabalhos como o de (Qin, Huang & Suganthan 2009) mostram que regras adaptativas para alguns parâmetros de controle beneficiam a performance dos algoritmos de ED sem afetar sua complexidade computacional.

computacionalmente custosos, o espaço de complexidade da técnica de ED é muito menor que a maioria dos demais algoritmos evolucionários tradicionais (Das & Suganthan 2011).

Nos últimos anos, inúmeros estudos sobre e/ou utilizando algoritmos de evolução diferencial vêm sendo desenvolvidos em diversos campos de pesquisa como sistemas elétricos de potência, robótica, reconhecimento de padrões, bioinformática, processamento de imagens, treinamento de redes neurais artificiais, assim como em modelagem financeira e econométrica⁵.

Aplicações na área de economia e finanças incluem (Chauhan, Ravi & Chandra 2009), que treinaram uma rede neural *wavelet* com base em um algoritmo de evolução diferencial para previsão de falências bancárias. Os resultados superaram os obtidos por tradicionais algoritmos de aprendizagem supervisionada em termos de acurácia e sensibilidade dos parâmetros aos critérios iniciais configurados. Para solucionar problemas de despacho econômicos⁶, (Lu, Zhou, Qin, Li & Zhang 2010) sugeriram um algoritmo híbrido de evolução diferencial adaptativo que, segundo os autores, solucionou os problemas considerados no trabalho com menor tempo de processamento e ao mesmo tempo de forma mais efetiva e com maior robustez em comparação com outros algoritmos avaliados⁷.

Recentemente, (Bhattacharya & Chattopadhyay 2011) propuseram também um algoritmo híbrido de evolução diferencial para problemas de despacho econômico de cargas. Os resultados indicaram elevada capacidade do algoritmo em fornecer soluções ótimas e que mais se adequam às restrições associadas. Na previsão de preços em mercados de energia elétrica, (Chakravarty & Dash 2011) aplicaram um modelo de rede neural artificial baseado em filtragem dinâmica treinado por um algoritmo de evolução diferencial. Além desses trabalhos, algoritmos de evolução diferencial também vêm sendo aplicados na estimação de modelos econométricos (Gilli & Winker 2008), em seleção de modelos econômicos (Maringer & Meyer 2008), assim como na otimização de portfólios com base em conjuntos não-convexos (Krink & Paterlini 2011).

Os operadores genéticos que constituem a técnica de evolução diferencial são: mutação, cruzamento e seleção. Entretanto, diferentemente dos algoritmos genéticos, a mutação nos algoritmos de evolução diferencial precede o operador de *crossover*, como pode ser notado nos procedimentos do algoritmo implementado nesta dissertação e apresentado abaixo, seguido da Figura 3.2 que apresenta o fluxograma com os principais passos do algoritmo de evolução diferencial. Nas subseções seguintes são descritos os componentes do algoritmo de evolução diferencial clássico utilizado neste trabalho para estimação dos parâmetros que descrevem as curvas de juros de acordo com o modelo parcimonioso de (Nelson & Siegel 1987).

⁵(Das & Suganthan 2011) apresentam uma coletânea dos algoritmos de ED considerados estado-da-arte e suas aplicações em diversos problemas de busca e otimização.

⁶Problemas de despacho econômico envolvem a minimização de custos de produção satisfazendo condições de operação de determinado sistema, como por exemplo, sistemas elétricos de distribuição.

⁷(Yuan, Wang & Huan 2009) e (Yuan et al. 2009) também consideraram a utilização de algoritmos de ED em problemas dinâmicos de despacho econômicos.

Algoritmo de Evolução Diferencial para Estimação da ETTJ

Determinar β_j^* e λ^*

Inicializar a população de N indivíduos em torno de β_j^* e λ^*

Iniciar contador de gerações $G = 0$

Enquanto $G < G_{max}$ **Faça**

Para $i = 1$ **até** N **Faça** para cada indivíduo sequencialmente

$\mathbf{x}_0^* = \{\hat{\beta}_j^*, \hat{\lambda}^*\}$ é a solução mais apta (maior *fitness*)

Gerar um vetor doador $\mathbf{v}_{i,G} = \mathbf{x}_{r_1,G} + F \cdot (\mathbf{x}_{r_2,G} - \mathbf{x}_{r_3,G})$

Gerar um vetor experimental $u_{j,i,G} = \begin{cases} v_{j,i,G} & \text{se } (\varepsilon_{i,j} \leq Cr \text{ or } j = j_{rand}) \\ x_{j,i,G} & \text{caso contrário} \end{cases}$

Se $\mathcal{F}(u_{i,G}) \leq \mathcal{F}(x_{i,G})$ **Então**

$x_{j,i,G+1} = u_{j,i,G}$

Caso Contrário

$x_{j,i,G+1} = x_{j,i,G}$

Fim Se

Fim Para

Se $\mathcal{F}(\mathbf{x}_0^*) < \mathcal{F}(\mathbf{x}_G^*)$ **Então**

$G = G + 1$

Caso Contrário

$G = G$

Fim Se

Fim Enquanto

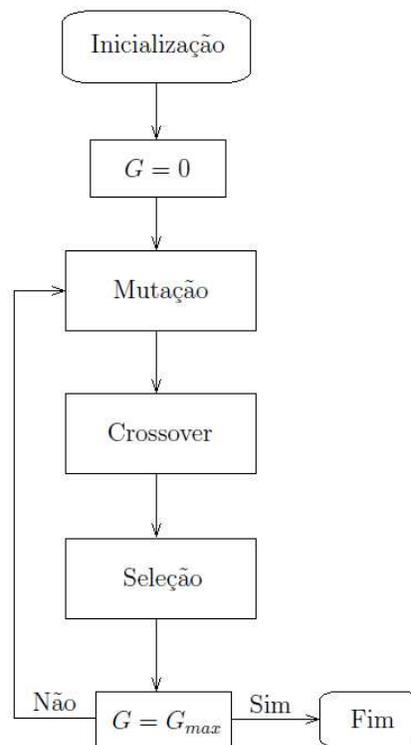


Figura 3.2: Fluxograma do algoritmo de evolução diferencial para estimação da ETTJ.

3.2.1 Inicialização

Considere um problema de otimização de dimensão D em um espaço de parâmetros reais \mathbb{R}^D . A população P é composta por vetores de parâmetros de dimensão $N \times D$, descrita por:

$$P_{\mathbf{x},G} = [\mathbf{x}_{1,G}, \mathbf{x}_{2,G}, \dots, \mathbf{x}_{NP,G}]^T, \quad G = 0, 1, \dots, G_{max} \quad (3.6)$$

$$\mathbf{x}_{i,G} = [x_{1,i,G}, x_{2,i,G}, \dots, x_{D,i,G}], \quad i = 1, 2, \dots, N \quad (3.7)$$

sendo que $G = 0, 1, \dots, G_{max}$ denota o contador de gerações e N o número de vetores (indivíduos) na população (tamanho da população).

Cada vetor da população ou cromossomo constitui uma solução candidata ao problema de otimização. Considerando a questão da estimação dos parâmetros da curva de juros, cada vetor é composto pelos parâmetros da função exponencial de (Nelson & Siegel 1987), conforme descrito na seção anterior no caso do modelo do algoritmo genético implementado.

Na proposta de evolução diferencial clássica de (Storn & Price 1997), a população inicial do algoritmo é gerada aleatoriamente de acordo com:

$$x_{j,i,1} = x_{j,min} + \varepsilon_{i,j} \cdot (x_{j,max} - x_{j,min}), \quad j = 1, 2, \dots, D \quad (3.8)$$

na qual $\varepsilon_{i,j}$ é um número aleatório uniformemente distribuído no intervalo entre 0 e 1, gerada de forma independente para cada i -ésimo componente vetor, enquanto $x_{j,min}$ e $x_{j,max}$ indicam os limites inferior e superior do vetor de parâmetros $\mathbf{x}_{i,j}$, respectivamente.

Assim como no algoritmo genético implementado, a população inicial será gerada aleatoriamente a partir de valores sementes, computados conforme sugerido por (Manousopoulos & Michalopoulos 2009). Esse ainda é um procedimento opcional, uma vez que os resultados do algoritmo de evolução diferencial não são interferidos pela escolha dos valores iniciais. Este passo do algoritmo destina-se apenas a reduzir o tempo de processamento para obter o valor ótimo do problema.

3.2.2 Mutação

A mutação é caracterizada por ser uma perturbação em um gene característico de um cromossomo com base em um elemento aleatório. No algoritmo de evolução diferencial, um vetor progenitor da população é chamado de vetor alvo (*target vector*), um vetor mutado obtido por meio do operador de mutação diferencial é conhecido como vetor doador (*donor vector*), e o descendente formado pela combinação do vetor doador com o vetor alvo é chamado de vetor experimental (*trial vector*) (Das & Suganthan 2011).

No operador de mutação do algoritmo de ED clássico, três vetores de parâmetros distintos $\mathbf{x}_{r_1^i}$, $\mathbf{x}_{r_2^i}$ e $\mathbf{x}_{r_3^i}$ são escolhidos aleatoriamente da atual população, para criar o vetor doador para cada i -ésimo vetor alvo. Os índices r_1^i , r_2^i e r_3^i , diferentes do índice do vetor base i , são inteiros mutuamente exclusivos escolhidos no intervalo $[1, N]$. O operador de mutação pode então ser descrito como:

$$\mathbf{v}_{i,G} = \mathbf{x}_{r_1^i,G} + F \cdot (\mathbf{x}_{r_2^i,G} - \mathbf{x}_{r_3^i,G}) \quad (3.9)$$

em que $\mathbf{v}_{i,G}$ é o vetor doador e F é o fator de mutação, o qual é um número escalar, usualmente definido no intervalo $[0.4, 1]$.

Neste caso, a mutação utiliza um vetor randômico $\mathbf{x}_{r_1^i}$ e apenas um vetor de diferenças ponderado para perturbá-lo: $F \cdot (\mathbf{x}_{r_2^i} - \mathbf{x}_{r_3^i})$ ⁸. As estratégias em algoritmos de evolução diferencial, em geral, estão associadas com o operador de mutação utilizado. Nesta dissertação será utilizado o operador clássico, conforme descrito em (3.9)⁹.

3.2.3 Crossover

Após ter sido gerado um vetor doador por meio do operador de mutação, o próximo passo do algoritmo de evolução diferencial é a aplicação do operador de cruzamento. Nesse estágio, é intensificado o potencial de aumento da diversidade na população por meio da troca entre os parâmetros do vetor que sofreu mutação $\mathbf{v}_{i,G}$ e o vetor alvo $\mathbf{x}_{i,G}$, para então gerar o chamado vetor experimental $\mathbf{u}_{i,G} = [u_{1,i,G}, u_{2,i,G}, \dots, u_{D,i,G}]$. Nesta dissertação, o operador de crossover utilizado é o crossover uniforme ou binomial¹⁰, o qual é um dos mais tradicionalmente aplicados, definido como:

$$u_{j,i,G} = \begin{cases} v_{j,i,G} & , \text{ se } (\varepsilon_{i,j} \leq Cr \text{ or } j = j_{rand}) \\ x_{j,i,G} & , \text{ caso contrário} \end{cases} \quad (3.10)$$

sendo que $\varepsilon_{i,j}$ é um número aleatório com distribuição uniforme com valores entre 0 e 1, considerado para cada j -ésimo componente do i -ésimo vetor de parâmetros; $j_{rand} \in [1, 2, \dots, D]$ é um índice escolhido aleatoriamente, garantindo que $\mathbf{u}_{i,G}$ assumo pelo menos um componente de $\mathbf{v}_{i,G}$; e, finalmente, Cr é denominado de *taxa de crossover*, parâmetro de controle no operador de cruzamento em algoritmos de evolução diferencial, contido no intervalo $[0, 1]$.

O parâmetro de controle do crossover Cr , ou taxa de crossover, gerencia o número esperado de parâmetros que devem ser alterados em um determinado membro da população. Na literatura, (Ronkkonen et al. 2005) argumentam que Cr deve ser selecionado de forma apropriada no intervalo $[0.3, 0.9]$, enquanto afirma que esse parâmetro de controle deve estar contido no intervalo $[0, 0.2]$ para funções separáveis, e no intervalo $[0.9, 1]$ quando os parâmetros das funções são dependentes. Entretanto, nesta dissertação, a taxa de crossover foi determinada de acordo com simulações.

3.2.4 Seleção

Finalmente, a *seleção* determina se o vetor alvo ou experimental irá sobreviver e compor a nova geração $G = G + 1$, de acordo com o seguinte critério guloso:

$$\mathbf{x}_{i,G+1} = \begin{cases} \mathbf{u}_{i,G} & , \text{ se } \mathcal{F}(\mathbf{u}_{i,G}) \leq \mathcal{F}(\mathbf{x}_{i,G}) \\ \mathbf{x}_{i,G} & , \text{ se } \mathcal{F}(\mathbf{u}_{i,G}) > \mathcal{F}(\mathbf{x}_{i,G}) \end{cases} \quad (3.11)$$

⁸(Ronkkonen, Kukkonen & Price 2005) sugerem que F deve ser tal que $0.4 < F < 0.9$ com $F = 0.9$ como uma primeira escolha.

⁹Outros operadores de mutação em evolução diferencial e suas aplicações em distintos problemas de otimização podem ser vistos em (Chakraborty 2008).

¹⁰Crossover exponencial (ou módulo de dois pontos) é outro tipo de operador de crossover relatado na literatura (Price, Storn & Lampinen 2005).

onde $\mathcal{F}(\mathbf{x})$ é a função custo a ser minimizada¹¹.

O número de indivíduos na população é mantido constante ao longo das gerações¹². Dessa forma, o novo vetor experimental substitui seu correspondente vetor alvo na próxima geração apenas quando esse processo mantém no mesmo nível ou melhora o valor da função custo. Caso contrário, o vetor alvo é mantido na população. Deve-se perceber que este critério de seleção nunca deteriora o *status* da atual geração da função objetivo, uma vez que a população é evoluída quando o valor do *fitness* mantém-se no mesmo nível ou melhora. Esse critério permite que o vetores do algoritmo de evolução diferencial se movam ao longo da superfície da função objetivo em questão em busca da solução desejável (Das & Suganthan 2011).

O critério de parada adotado para o algoritmo de evolução diferencial foi o mesmo do algoritmo genético. Como o número necessário de iterações necessário para estimação da curva de juros varia de acordo com a periodicidade dos dados, o algoritmo não é mais iterado quando a melhor solução se mantém constante ao longo de um número fixo de gerações G_{max} . Portanto, os parâmetros de controle são: N , σ , F , Cr e G_{max} .

3.3 Estratégias Evolutivas

As estratégias evolutivas (*Evolution Strategies* - ES) correspondem a uma classe de algoritmos evolucionários para solução de problemas, geralmente, em domínios contínuos. Em contraposição aos algoritmos genéticos, as ES utilizam condificação real, um mecanismo de auto-adaptação, um operador de seleção não-elitista, crossover multi-indivíduo e operador de mutação gaussiano (Schwefel 1981). Um indivíduo k em uma estratégia evolutiva é composto pela tripla $i_k = (\mathbf{x}_k, \mathbf{s}_k, \mathcal{F}(\mathbf{x}_k))$, sendo que \mathbf{x}_k é o vetor de parâmetros a serem otimizados, \mathbf{s}_k são os parâmetros de auto-adaptação e $\mathcal{F}(\mathbf{x}_k)$ é o valor de *fitness* do indivíduo (Beyer & Schwefel 2002).

Inicialmente, uma ES tem uma população ω de “pais” que gerarão ϖ “filhos” via operadores de mutação e/ou crossover¹³. Em geral, faz-se $\omega < \varpi$ para assegurar uma melhor exploração do espaço de busca. Os descendentes são gerados da seguinte forma:

$$\mathbf{x}_k^{G+1} = \mathbf{x}_k^G + N(0, \mathbf{s}_k) \quad (3.12)$$

em que \mathbf{x}_k^{G+1} corresponde ao descendente k na geração $G + 1$, \mathbf{x}_k^G é o pai da geração atual G , e $N(0, \mathbf{s}_k)$ é uma função de densidade gaussiana com média zero e \mathbf{s}_k é o vetor de desvios padrões.

Cada uma das componentes do vetor de parâmetros \mathbf{x}_k recebe um valor gaussiano, de forma que $\mathbf{s}_k = [\sigma_{1k} \ \sigma_{2k} \ \dots \ \sigma_{Dk}]$, em que D corresponde a dimensão do problema de otimização considerado. Esse mecanismo garante que a função de densidade gaussiana distribua melhor ao longo dos eixos correspondentes aos valores que estão sendo otimizados¹⁴. Uma das principais

¹¹No caso da estimação da curva de juros, considerando a função de (Nelson & Siegel 1987), a função de *fitness* é a mesma conforme definida pela expressão (2.15), descrita no Capítulo 2.

¹²Ao longo das gerações, a população foi composta sempre por indivíduos viáveis, assim como no algoritmo genético descrito.

¹³Em ES mais simples, é considerado apenas o operador de mutação.

¹⁴Esse processo de mutação também pode consistir em atribuir o mesmo valor de desvio padrão para todas as componentes do vetor de parâmetros, mas não distribui muito bem a função de densidade gaussiana ao longo

características dos algoritmos de estratégias evolutivas é a auto-adaptação dos parâmetros de dispersão das variáveis do problema.

Embora exista uma vasta literatura que aborda aplicações de algoritmos evolucionários em economia e finanças (Korczak & Lipinski 2004, Korczak & Roger 2001), são poucos os trabalhos que consideram métodos de estratégias evolutivas. (Lipinski, Winczura & Wojcik 2007), por exemplo, empregaram um algoritmo de ES para otimização de portfólios considerando custos de transação, preferências e restrição orçamentária do investidor, de acordo com distintas medidas de riscos. Em comparação com a técnica *buy-é-hold*, a técnica de estratégias evolutivas proposta apresentou maior eficiência, em termos de risco/retorno do investimento. Análises similares também foram realizadas por (Korczak, Lipinski & Roger 2002) utilizando operadores de mercados artificiais. Além disso, a literatura apresenta aplicações de técnicas de estratégias evolutivas na evolução de conhecimento de mercado em empresas (Kuo, Hong & Chen 2008), previsão de evasão de divisas em moedas estrangeiras (Binner, Kendall & Gazely 2004) e otimização multi-objetivo de portfólios (Schlottmann & Seese 2004)¹⁵.

Nesta dissertação, o algoritmo CMA-ES (Covariance Matrix Adaptation) será considerado. O CMA-ES é um método que atualiza não somente os parâmetros de uma estratégia evolutiva, como também uma matriz de covariância gerada por uma distribuição gaussiana multidimensional para a mutação dos indivíduos pais (Hansen & Ostermeier 2001). Esse mecanismo permite calcular as interdependências entre as variáveis do problema de acordo com o cálculo da matriz de covariâncias.

Novos candidatos à solução do problema de otimização são gerados por meio de uma distribuição de probabilidade multidimensional, correspondendo ao operador de mutação, e cada elemento da matriz de covariância descreve a relação entre duas variáveis do problema. O algoritmo CMA-ES, por meio da matriz de covariância, consegue espelhar a interdependência das variáveis do problema, o que permite uma maior flexibilidade a esses tipos de relações como no caso da modelagem da curva de juros, uma vez que os fatores latentes são interdependentes.

O CMA-ES possui duas características básicas. A primeira delas consiste na questão de que se deve aumentar a probabilidade de uma mutação bem sucedida a cada iteração do algoritmo, i.e., adaptar sempre os parâmetros (inclusive a matriz de covariância). A segunda ideia insere um armazenamento do caminho da evolução (*evolution path*) no tempo da estratégia. Esse caminho armazena informações de correlação entre passos contíguos do algoritmo, utilizado para o processamento de atualização da matriz de covariância e para conduzir um controle do tamanho do passo adicional do algoritmo, evitando convergência prematura.

Os algoritmos evolucionários empregados neste trabalho são distintos em suas formulações. O AG é caracterizado pela geração de descendentes pelo operador de crossover, essencialmente pela combinação linear de indivíduos distintos. Por outro lado, a técnica de evolução diferencial é caracterizada pelo operador de mutação, de forma que os descendentes são gerados por meio dos próprios indivíduos da população. Por fim, o algoritmo de estratégias evolutivas evolui tanto os indivíduos da população quanto os parâmetros da evolução, evidenciando uma abordagem mais

dos eixos do espaço de busca (Beyer & Schwefel 2002).

¹⁵(Chen & Kuo 2002) e (Castillo Tapia & Coello 2007), respectivamente, apresentam uma revisão bibliográfica dos trabalhos com aplicações em economia e finanças de algoritmos evolucionários e em problemas de otimização multi-objetivos.

autônoma.

No algoritmo CMA-ES, denotado por $(\omega/\omega, \varpi)$ -CMA-ES, a geração dos descendentes é realizada da seguinte forma:

$$\mathbf{x}_k^{G+1} = \mathbf{z}_k, \quad z \sim N(\langle \mathbf{x} \rangle_\mu^G, (\sigma^2)^G \mathbf{C}^G), \quad k = 1, 2, \dots, \lambda \quad (3.13)$$

em que $\langle \mathbf{x} \rangle_\omega^G = \sum_{i=1}^{\omega} w_i x_i^G$ é o indivíduo médio gerado pelo operador de recombinação intermediária ($w_1 = w_2 = \dots = w_\omega = \frac{1}{\omega}$), e \mathbf{C}^G é a matriz de covariância/dispersão na geração G .

A matriz de covariância é adaptada pelo caminho da evolução p_c^{G+1} , levando em consideração o efeito das mudanças anteriores do vetor de atributos na direção do movimento atual no espaço de busca. Portanto, o caminho da evolução e a matriz de covariância são atualizados da seguinte forma:

$$p_c^{G+1} = (1 - c_c) p_c^G + \sqrt{c_c(2 - c_c)} \frac{\sqrt{\omega}}{\sigma^G} (\langle \mathbf{x} \rangle_\omega^{G+1} - \langle \mathbf{x} \rangle_\omega^G) \quad (3.14)$$

$$\mathbf{C}^{G+1} = (1 - c_{cov}) \mathbf{C}^G + c_{cov} \left[\frac{1}{\omega} p_c^{G+1} (p_c^{G+1})^T + \left(\frac{1}{\omega} + \frac{1}{\omega^2} \right) \sum_{i=1}^{\omega} \frac{1}{(\sigma^2)^G} d_f (d_f)^T \right] \quad (3.15)$$

em que $d_f = (\mathbf{x}_i^{G+1} - \langle \mathbf{x} \rangle_i^G)$, c_c é um parâmetro da estratégia e $c_{cov} \in [0, 1[$ determina a taxa de adaptação da matriz de covariância \mathbf{C} .

A adaptação do tamanho do σ^G está baseada no caminho da evolução conjugado p_σ , atualizado da seguinte forma:

$$p_\sigma^{G+1} = (1 - c_\sigma) p_\sigma^G + \sqrt{c_\sigma(2 - c_\sigma)} \mathbf{P}^G (\mathbf{D}^G)^{-1} (\mathbf{P}^G)^{-1} \frac{\sqrt{\omega}}{\sigma^G} (\langle \mathbf{x} \rangle_\omega^{G+1} - \langle \mathbf{x} \rangle_\omega^G) \quad (3.16)$$

em que as matrizes \mathbf{P} e \mathbf{D} são obtidas de acordo com a decomposição de \mathbf{C} :

$$\mathbf{C}^G = \mathbf{P}^G (\mathbf{D}^G)^2 (\mathbf{P}^G)^T \quad (3.17)$$

em que as colunas de \mathbf{P}^G são os autovetores de \mathbf{C}^G e \mathbf{D}^G é uma matriz diagonal das raízes quadradas dos autovalores de \mathbf{C}^G .

O tamanho do passo global σ^G é atualizado como:

$$\sigma^{G+1} = \sigma^G e^{\frac{c_\sigma}{t} \left(\frac{\|\mathbf{p}_\sigma^{G+1}\|}{\mathbb{E}[\|\mathbf{N}(0, I)\|]} \right)} \quad (3.18)$$

em que (Hansen, Müller & Koumoutsakos 2003):

$$c_c = \frac{4}{D + 4}, \quad c_\sigma = \frac{10}{D + 20}, \quad t = \max \left\{ 1, \frac{3\omega}{D + 10} \right\} + c_\sigma \quad (3.19)$$

$$c_{cov} = \frac{1}{\omega} \frac{2}{(D + \sqrt{2})^2} + \left(1 - \frac{1}{\omega} \cdot \min \left\{ 1, \frac{2\omega - 1}{(D + 2)^2 + \omega} \right\} \right) \quad (3.20)$$

$$E[||N(0, I)||] \cong \sqrt{D} \left(1 - \frac{1}{4D} + \frac{1}{21D^2} \right) \quad (3.21)$$

$$\lambda = 4 + 3\ln(D), \quad \omega = \frac{\lambda}{2} \quad (3.22)$$

A inicialização do algoritmo se dá com $p_c^0 = p_\sigma^0 = 0$ e, para a matriz de covariância \mathbf{C}^0 , considera-se a matriz identidade. Já os valores de \mathbf{x}^0 e σ^0 variam de acordo com o problema em questão. Portanto, um indivíduo no algoritmo CMA-ES é caracterizado por uma tripla $\mathfrak{N} = (\mathbf{x}, \sigma, \mathbf{C})$, sendo que $\mathbf{x} \in \mathfrak{R}^D$ representa um vetor solução para o problema, σ é o desvio padrão e $\mathbf{C} \in \mathfrak{R}^{D \times D}$ é a matriz de covariância, tal que D é a dimensão do problema ou o número de variáveis da função objetivo. Como critério de parada, pode-se utilizar ou um número máximo de gerações G_{max} e/ou assumir um limiar mínimo ϵ para a função objetivo $\mathcal{F}(\mathbf{x})$.

Abaixo é apresentado o algoritmo CMA-ES.

Algoritmo de Estratégias Evolutivas CMA-ES para Estimação da ETTJ

Inicializar \mathbf{x}^0 , $\mathbf{C}^0 = I$, ϖ , ω , $E[||N(0, I)||]$, c_{cov} , c_σ , t e c_c .

Iniciar contador de gerações $G = 0$

Enquanto $G < G_{max}$ **Ou** $\mathcal{F}(\mathbf{x}) < \epsilon$ **Faça**

Para $i = 1$ **até** ϖ **Faça**

Gerar os descendentes:

$$\mathbf{x}_k^{G+1} = \mathbf{z}_k, \quad \mathbf{z} \sim N(\langle \mathbf{x} \rangle_\omega^G, (\sigma^2)^G \mathbf{C}^G)$$

Computar o caminho da evolução:

$$p_c^{G+1} = (1 - c_c)p_c^G + \sqrt{c_c(2 - c_c)} \frac{\sqrt{\omega}}{\sigma^G} (\langle \mathbf{x} \rangle_\omega^{G+1} - \langle \mathbf{x} \rangle_\omega^G)$$

Atualizar a matriz de covariâncias:

$$\mathbf{C}^{G+1} = (1 - c_{cov})\mathbf{C}^G + c_{cov} \left[\frac{1}{\omega} p_c^{G+1} (p_c^{G+1})^T + \left(\frac{1}{\omega} + \frac{1}{\omega^2} \right) \sum_{i=1}^{\omega} \frac{1}{(\sigma^2)^G} d_f (d_f)^T \right]$$

Decompor a matriz de covariâncias:

$$\mathbf{C}^G = \mathbf{P}^G (\mathbf{C}^G)^2 (\mathbf{P}^G)^T$$

Atualizar o caminho da evolução conjugado:

$$p_\sigma^{G+1} = (1 - c_\sigma)p_\sigma^G + \sqrt{c_\sigma(2 - c_\sigma)} \mathbf{P}^G (\mathbf{D}^G)^{-1} (\mathbf{P}^G)^{-1} \frac{\sqrt{\omega}}{\sigma^G} (\langle \mathbf{x} \rangle_\omega^{G+1} - \langle \mathbf{x} \rangle_\omega^G)$$

Adaptar o tamanho do passo global:

$$\sigma^{G+1} = \sigma^G e^{\frac{c_\sigma}{t} \left(\frac{||p_\sigma^{G+1}||}{E[||N(0, I)||]} \right)}$$

Fim Para

Fim Enquanto

3.4 Resumo

Esse capítulo apresentou as abordagens evolucionárias aplicadas na estimação da estrutura a termo das taxas de juros: algoritmos genéticos, evolução diferencial e estratégias evolutivas. Cada algoritmo foi estruturado de acordo com o problema da curva de juros, de forma a obter os melhores resultados para todo o período de análise em termos de ajuste com as taxas observadas no mercado.

Definidas as técnicas para modelagem da curva de juros, o próximo capítulo introduz a questão de predição dos fatores latentes de NS, para obtenção da estrutura a termo prevista. O método de dois estágios, proposto por (Diebold & Li 2006), será descrito, assim como os principais conceitos dos modelos nebulosos funcionais evolutivos.

Previsão da Curva de Juros

Este capítulo insere a questão da previsão da estrutura a termo das taxas de juros e a abordagem proposta de utilização de modelos funcionais nebulosos evolutivos. Uma revisão da literatura relacionada às diversas metodologias de previsão da curva de juros é reportada inicialmente. Em seguida, é apresentada a abordagem considerada como *benchmark* para as análises, i.e., a técnica de dois estágios, conforme proposta por (Diebold & Li 2006). Por fim, as ideias centrais dos modelos nebulosos evolutivos do tipo Takagi-Sugeno são apresentadas.

4.1 Revisão Bibliográfica

A previsão da estrutura a termo das taxas de juros, em uma perspectiva prática, é um dos principais objetivos na modelagem da curva de juros. A previsão da estrutura a termo é essencial para a administração de riscos e portfólios, instituições financeiras, *policy makers*, e para os agentes de mercado. Tesoureiros de bancos centrais gerenciam a emissão e manutenção dos níveis de dívida pública via avaliação contínua das taxas de juros corrente e futuras. Investidores necessitam acompanhar o desempenho de seus portfólios de ativos em relação aos custos de investimentos em títulos públicos de baixo risco, que intrinsecamente estão associados aos níveis futuros de taxas de juros livres de riscos. Por outro lado, bancos centrais podem reagir à inflação e atividade econômica esperadas pelo ajuste da taxa de curto prazo, o que afeta toda a curva de juros, e necessita de previsões quanto ao futuro das taxas correntes para ir de encontro à essas expectativas (Matsumura, Moreira & Vicente 2010). Além disso, prospecções da estrutura a termo das taxas de juros têm-se mostrado ainda como um importante indicador do comportamento futuro da inflação e atividade econômica (Duarte, Venetis & Paya 2005, Venetis, Paya & Peel 2003, Ang, Piazzesi & Wei 2006).

A literatura de estrutura a termo das taxas de juros devotou seus estudos às questões de modelagem, i.e., desenvolver modelos capazes de descrever toda a curva de juros de acordo com as taxas disponíveis no mercado (Vasicek & Fong 1982, Heath et al. 1992, Nelson & Siegel 1987), de forma que a previsão fora da amostra recebeu pouca atenção (Matsumura et al. 2010). (Fama & Bliss 1987) investigaram as relações entre maturidade e taxas de juros futuras como o primeiro trabalho a abordar questões de previsão da estrutura a termo. Os autores sugeriram uma tendência de reversão à média das taxas de juros de 1 ano dos títulos públicos americanos,

capaz de gerar previsões das taxas futuras. (Greer 2003) mostrou que as previsões das taxas de juros apenas são capazes de fornecer as direções de mudanças nas taxas de maturidades mais longas. Por outro lado, (Bidarkota 1998) indica que modelos multivariados de previsão, em especial modelos bivariados, são capazes de realizar boas previsões para taxas de juros no curto prazo para o mercado dos Estados Unidos.

Recentemente, (Duffee 2002) mostrou que modelos afim apresentam baixo desempenho para a previsão da curva de juros considerando o mercado de títulos públicos dos Estados Unidos, assim como (Dai & Singleton 2002) que encontraram os mesmos resultados. Considerando ainda modelos afim, (Ang & Piazzesi 2003) apontam que a inclusão de variáveis macroeconômicas em tais modelos melhora as previsões da curva de juros fora da amostra, superando modelos de vetores autoregressivos. Posteriormente, (Hördahl, Tristani & Vestin 2006) confirmaram a obtenção de melhores previsões da estrutura a termo quando fatores macroeconômicos são inseridos na modelagem da curva de juros.

Baseados no modelo de fatores latentes de (Nelson & Siegel 1987) para a estrutura a termo das taxas de juros, (Diebold & Li 2006), DL daqui em diante, propuseram um modelo de dois estágios para a previsão da curva de juros dos Estados Unidos. Os autores mostraram que o modelo proposto apresentou acuradas previsões, superando metodologias tradicionais como vetores autoregressivos e modelos de passeio aleatório (*random walk*). Entretanto, a técnica de dois estágios apresentou melhor desempenho para previsões de maturidades mais longas, de forma que no curto prazo, modelos de passeio aleatório proporcionam resultados equivalentes e até melhores (Diebold & Li 2006). Contudo, (Almeida & Vicente 2008) indicaram que a inclusão de condições de não-arbitragem no modelo de fatores latentes melhora as previsões fora da amostra, mas em contrapartida implica em maior complexidade computacional. Nessa mesma linha de pesquisa, (Moench 2008) aponta que a inclusão dessas condições em modelos de vetores autoregressivos melhora o desempenho de previsão da estrutura a termo, essencialmente quando horizontes de curto prazo são levados em consideração.

No mercado de renda fixa do Brasil, (Lima, Ludovice & Tabak 2006) avaliaram modelos de vetores autoregressivos na previsão da estrutura a termo das taxas de juros. Os autores mostraram que o modelo de passeio aleatório ainda apresenta melhores previsões para a curva de juros. Além disso, (Vicente & Tabak 2007) mostraram que o modelo de DL apresentou previsões mais acuradas para dados de juros do Brasil quando comparado a modelos afim, principalmente para horizontes de longo prazo.

Análises similares foram observadas por (Matsumura et al. 2010), que testaram modelos lineares para realizar previsões da estrutura a termo no Brasil. Os autores mostraram que os modelos lineares mais simples, baseados em poucos parâmetros, são capazes de realizar melhores previsões, mas ainda assim são superados por modelos de passeio aleatório. Ainda, foi observado que a inclusão de uma variável financeira ao modelo, índice do mercado de ações, aumentou a acurácia dos modelos lineares considerados. Por fim, (Caldeira, Moura & Portugal 2010) avaliaram o modelo de DL estimado por filtro de Kalman no mercado de títulos públicos brasileiro, evidenciando elevada adequação do modelo ao gerar previsões acuradas, principalmente considerando horizontes de um mês a três meses.

Uma recente e distinta abordagem para a questão de previsão da estrutura a termo das taxas de juros é a utilização de modelos de equilíbrio de correção de erro. (Bowsher & Meeks 2008) e (Koopman, Mallee & Wel 2010) avaliaram tal abordagem para o mercado de juros dos EUA e mostraram resultados superiores se comparados aos obtidos pela metodologia de DL e de passeio aleatório, porém, apenas para horizontes de previsão de curto prazo. (McMillan 2009), por exemplo, sugere que modelos de correção de erro são capazes de capturar a dinâmica das taxas de curto prazo, superando técnicas econométricas lineares e não-lineares de previsão. Entretanto, o mesmo não se verifica quando maturidades mais longas são consideradas. No mercado de renda fixa do Brasil, (Cajueiro, Divino & Tabak 2009) realizaram as mesmas análises, e mostraram que a adequação dos modelos de previsão está diretamente correlacionada com o horizonte de previsão, confirmando os resultados da literatura.

Métodos Bayesianos também foram aplicados na análise de previsão da estrutura a termo das taxas de juros. (Laurini & Hotta 2010) e (Hautsch & Yang 2010) são exemplos de aplicações da metodologia Bayesiana de Mínimos Quadrados de Monte Carlo para estimação do modelo de dois estágios de (Diebold & Li 2006). Embora tais metodologias apresentem elevada acurácia na previsão da curva de juros no curto prazo, superando a abordagem tradicional de estimação de DL, são de elevada complexidade computacional, resultando em aumento do tempo de processamento ou mesmo em não-convergência (Hautsch & Yang 2010). Resultados semelhantes também foram obtidos por (Diebold, Rudebusch & Arouba 2006) e (Yu & Zivot 2011), com a modelagem da ETTJ utilizando modelos econométricos de espaço-estado.

Apesar do desenvolvimento de diversas técnicas para previsão da curva de juros, uma limitação de tais técnicas é apresentar apenas bons resultados em horizontes de previsão específicos (curto ou longo prazos). Além disso, modelos estimados por meio de mínimos quadrados de Monte Carlo, apesar de apresentarem elevada acurácia, são computacionalmente ineficientes, o que prejudica sua implementação na prática, principalmente em ambientes que requerem uma capacidade de processamento de dados mais rápida, ou mesmo em tempo real.

Os modelos nebulosos evolutivos mostram-se como uma solução para o problema da previsão da curva de juros, pois não necessitam de uma amostra de treinamento, podendo fornecer previsões em tempo real e ao mesmo tempo são de baixa complexidade computacional. Para aplicação de modelos nebulosos evolutivos na previsão da ETTJ será considerada a mesma proposta da metodologia de (Diebold & Li 2006), i.e., método de dois estágios. Entretanto, os fatores latentes da curva de juros serão modelados por sistemas nebulosos, e não por meio de um processo autoregressivo. Dessa forma, as comparações tomarão como *benchmark* a técnica de DL, descrita a seguir.

4.2 Técnica de previsão de Dois Estágios

(Diebold & Li 2006) propuseram uma metodologia em dois estágios para previsão da estrutura a termo das taxas de juros de acordo com o modelo de fatores latentes da curva de (Nelson & Siegel 1987) conforme apresentado na Seção 2.2.2 no Capítulo 2:

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right)$$

O primeiro estágio da técnica de previsão de DL compreende essencialmente a estimação do conjunto de fatores latentes $\mathbf{x} = \{\beta_1, \beta_2, \beta_3, \lambda\}$ para o período correspondente à amostra de dados disponível. Ou seja, por meio de uma técnica de otimização de parâmetros são obtidas as estimativas dos fatores da função de NS para cada amostra de dados t , seja ela de periodicidade diária ou mensal, i.e., são obtidas as séries temporais dos fatores latentes da curva de juros.

(Diebold & Li 2006) fixaram o parâmetro de decaimento λ em um valor estabelecido, e obtiveram os demais parâmetros da função de NS utilizando o método de mínimos quadrados ordinários (OLS - *Ordinary Least Squares*). Recordando o fato de que o fator λ determina a maturidade em que o fator de curvatura atinge seu máximo, (Diebold & Li 2006) sugeriram escolher a média do número de meses para maturidades de um e três anos, ou seja 30 meses e, então, determinaram o valor de λ que maximiza o carregamento do fator de médio prazo (β_3) com maturidade de 30 meses. Entretanto, conforme apontado pela literatura (Laurini & Hotta 2010), esse procedimento é inadequado, pois, apesar da técnica ser intuitiva, é uma limitação manter tal parâmetro constante ao longo do período em questão, uma vez que o decaimento apresenta considerável variabilidade, conforme evidenciado em inúmeras análises empíricas da estrutura a termo (Rezende 2011, Yu & Zivot 2011, Caldeira et al. 2010). Nesta dissertação, optou-se por manter o parâmetro de decaimento variante no tempo, de forma a evitar limitações e permitir maior realismo ao modelo. Dessa forma, aplicando uma técnica de otimização não-linear de parâmetros para os dados das taxas de juros para cada período t , obtem-se as séries temporais das estimativas de $\hat{\mathbf{x}} = \{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\lambda}\}$.

O segundo estágio da técnica de previsão de (Diebold & Li 2006) é modelar os fatores por meio de um processo autoregressivo de ordem unitária, i.e., $AR(1)$. Portanto, a curva de juros prevista para h passos à frente é descrita da seguinte forma:

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t} \left(\frac{1 - e^{-\hat{\lambda}_{t+h/t} \tau}}{\hat{\lambda}_{t+h/t} \tau} \right) + \hat{\beta}_{3,t+h/t} \left(\frac{1 - e^{-\hat{\lambda}_{t+h/t} \tau}}{\hat{\lambda}_{t+h/t} \tau} - e^{-\hat{\lambda}_{t+h/t} \tau} \right) \quad (4.1)$$

em que:

$$\hat{\beta}_{i,t+h/t} = \hat{a}_i + \hat{b}_i \hat{\beta}_{i,t}, \quad i = 1, 2, 3 \quad (4.2)$$

$$\hat{\lambda}_{t+h/t} = \hat{a}_4 + \hat{b}_4 \hat{\lambda}_t \quad (4.3)$$

sendo que a_i e b_i , com $i = 1, 2, 3, 4$, são os parâmetros dos modelos autoregressivos associados aos fatores latentes da curva de juros.

A abordagem de (Diebold & Li 2006) chama a atenção pela sua simplicidade, capaz de gerar previsões razoáveis, principalmente quando o horizonte de previsão é de longo prazo. Nesta dissertação, a metodologia proposta por DL será considerada como *benchmark* na questão de previsão da estrutura a termo. Alternativamente, propomos uma abordagem nebulosa evolutiva seguindo a técnica de DL, de forma que a diferença principal concerne o segundo estágio, no qual, ao invés de modelar os fatores latentes de NS via um processo autoregressivo, eles serão previstos em um modelo de múltiplas-entradas-múltiplas-saídas conjuntamente. Dessa forma, não é necessário estimar um modelo autoregressivo para cada fator latente, diminuindo o tempo de processamento e evidenciando a interação temporal entre os parâmetros da curva de juros.

4.3 Modelos Nebulosos Evolutivos

Nesta dissertação serão considerados modelos nebulosos ou fuzzy funcionais evolutivos para a previsão da estrutura a termo das taxas de juros. Os modelos nebulosos podem ser caracterizados como modelos funcionais ou modelos linguísticos. Modelos nebulosos linguísticos são modelos baseados em regras nebulosas onde tanto o antecedente como o consequente são definidos por termos linguísticos. Nos modelos nebulosos funcionais, o antecedente é definido por termos linguísticos, mas o consequente é uma função das variáveis dos antecedentes (Pedrycz & Gomide 2007).

Os modelos nebulosos funcionais evolutivos são aqueles que utilizam um mecanismo de adaptação da base de regras de acordo com o estado atual do sistema. Dessa forma, a base de regras pode ser reduzida ou expandida, alterando sua estrutura. O modelo se adapta à cada novo dado de entrada que lhe é apresentado, de forma que o próprio sistema indica a necessidade de criação de uma nova regra, atualização de regras existentes e/ou exclusão de regras redundantes ou que reduzem a qualidade da base de regras atual, sendo não mais necessárias. Deve-se notar que o termo evolutivo utilizado neste contexto difere do termo *evolucionário*, comumente usado em algoritmos evolucionários, uma vez que o evolutivo refere ao processo contínuo de auto-desenvolvimento de uma entidade individual, enquanto o termo evolucionário está associado à geração de populações de indivíduos por reprodução, mutação e seleção natural (Baruah & Angelov 2011).

Os modelos nebulosos considerados nesta dissertação serão aqueles caracterizados por possuírem consequentes que são funções da variável de entrada, e a saída da base de regras é obtida a partir dessas funções e dos respectivos graus de ativação, conhecidos como modelos do tipo Takagi-Sugeno (Takagi & Sugeno 1985). Os antecedentes, por outro lado, são constituídos por regras nebulosas, obtidas por meio de um algoritmo de agrupamento.

A emergência dos modelos nebulosos evolutivos decorreu da elevada necessidade que surgiu na última década de processamento eficiente de informações, além da extração de conhecimento, elemento chave na modelagem de sistemas complexos. Em sistemas complexos de grande porte, os dados são obtidos em quantidades elevadas, dinamicamente, na forma de fluxo de dados (*streams of data*) (Angelov, Filev & Kasabov 2010). Dessa forma, a análise de tais sistemas era construída de acordo com o conhecimento do especialista (*expert knowledge*) (Mamdani & Assilian 1975), e, a partir dos anos de 1990, passaram a ser desenvolvidas metodologias capazes

de extrair conhecimento apenas de dados sobre seu funcionamento em tempo real, conhecidos como *online data-driven rule/knowledge systems* (Kasabov 1996).

Os modelos nebulosos evolutivos se inserem na categoria de técnicas adaptativas que são auto-desenvolvidas por meio de extração de conhecimento em fluxos de dados. De acordo com os dados que são apresentados ao modelo, caracterizando o estado do sistema atual, as técnicas nebulosas evolutivas permitem o aprendizado simultâneo de sua estrutura e parametrização. Esses modelos podem ser vistos, genericamente, como uma mistura nebulosa adaptativa de Gaussianas para aproximação da densidade da distribuição dos dados reais (Angelov & Zhou 2008). Esse procedimento, por sua vez, pode ser realizado em tempo real (*online*), sem a necessidade de um período de treinamento (Angelov et al. 2010).

Os principais modelos nebulosos evolutivos reportados na literatura consideram modelos funcionais baseados em regras nebulosas, de forma que sua estrutura (número de regras e parâmetros) evolua continuamente por meio de algoritmos recursivos de agrupamento (Lemos, Caminhas & Gomide 2011). O modelo Takagi-Sugeno evolutivo (eTS), proposto por (Angelov & Filev 2004), mostrou-se como a primeira abordagem para a identificação em tempo real de modelos de inferência nebulosa baseados em regras. A base de regras e parâmetros no modelo eTS continuamente evoluem pela adição de novas regras com maior poder de sumarização, e pela atualização de regras já existentes e parâmetros. A base de regras é construída por meio de uma versão estendida do algoritmo *subtractive clustering* (Chiu 1994). A avaliação recursiva do potencial de informação contido em cada novo dado que é apresentado ao sistema permite a criação de novas regras, ou ainda revisar as regras existentes. Os parâmetros do consequente são atualizados por meio do algoritmo de mínimos quadrados recursivos (Ljung 1999).

Para reduzir a complexidade computacional no cálculo do potencial dos dados no eTS, o modelo Simpl_eTS foi desenvolvido posteriormente em (Angelov & Filev 2005). Simpl_eTS corresponde à uma versão simplificada do modelo eTS, que substitui a noção de potencial de informação pelo conceito de dispersão para gerar um algoritmo similar, porém, computacionalmente mais eficaz. Posteriormente, uma versão estendida do modelo eTS (xTS) foi desenvolvida por (Angelov & Zhou 2008), introduzindo a idEia de adaptação do raio dos clusters (zona de influência) nos modelos eTS. O parâmetro que define a dispersão dos clusters é muito importante, pois é parte das funções de pertinência dos conjuntos nebulosos e, conseqüentemente, implica no nível de ativação das regras nebulosas. Recentemente, (Angelov et al. 2010) propuseram o modelo eTS+, como uma extensão aos modelos eTS e xTS, caracterizado pela atualização da base de regras considerando critérios como idade, utilidade, densidade local e zona de influência.

Uma abordagem distinta, mas computacionalmente similar, para modelagem de sistemas Takagi-Sugeno é o modelo neuro-nebuloso dinâmico evolutivo, proposto por (Kasabov & Song 2002). A abordagem utiliza um método recursivo de clusterização baseado no conceito de distância para adaptar a estrutura da base de regras, enquanto que os parâmetros dos consequentes são atualizados por meio de um algoritmo de mínimos quadrados ponderados com fator de esquecimento. Um algoritmo recursivo de agrupamento derivado de uma modificação de técnicas de quantização vetorial, conhecido como quantização vetorial evolutiva, é uma outra metodologia para construção do modelo flexível de inferência nebulosa (FLEXFIS - *flexible fuzzy inference system*) (Lughofer 2008).

Outro exemplo de sistemas nebulosos evolutivos, sugerido por (Rubio 2009), inserido na classe dos modelos neuro-nebulosos tipo-2, é a modelagem nebulosa auto-organizada e com método de mínimos quadrados modificado (SOFMLS - *self-organizing fuzzy modified least squares*). O modelo SOFMLS inclui um algoritmo de agrupamento baseado no conceito de proximidade de vizinhança. Por outro lado, o modelo de inferência nebulosa sequencial adaptativo (SAFIS - *sequential adaptive fuzzy inference system*), proposto por (Rong, Sundarajan, Huang & Saratchandran 2006), utiliza um critério de distância em conjunção com uma medida de influência dos novos clusters criados¹.

Conforme apontado por (Lemos et al. 2011), uma limitação dos algoritmos recursivos de agrupamento adotados pela maioria dos modelos nebulosos evolutivos é a falta de robustez na presença de ruídos e *outliers*. Nessas situações, tais algoritmos podem criar um novo cluster/agrupamento quando ruídos ou *outliers* são apresentados, ao invés de identificá-los. Para melhorar a robustez da modelagem nebulosa evolutiva, um novo método foi sugerido por (Lima 2008), chamado modelagem nebulosa evolutiva com aprendizado participativo (ePL). Essa abordagem une os conceitos de aprendizado participativo (Yager 1990) com a ideia de sistemas nebulosos evolutivos (Angelov & Filev 2004, Angelov & Zhou 2008). Em sistemas nebulosos evolutivos, o conceito de aprendizado participativo é implementado como um algoritmo não-supervisionado de agrupamento (Silva, Gomide & Yager 2005), e apresenta-se como um candidato natural para identificar estruturas em bases de regras considerando sistemas dinâmicos.

Posteriormente, com base no modelo ePL, (Lemos et al. 2011) desenvolveram o modelo nebuloso evolutivo Gaussiano multivariado (eMG). O eMG utiliza um algoritmo de agrupamento Gaussiano baseado no conceito de aprendizado participativo, em que cada cluster é representado por uma função de pertinência Gaussiana multivariada. A estrutura dos clusters (centro e dispersão) e o número de clusters são recursivamente atualizados a cada iteração do algoritmo, e os limiares dos parâmetros são determinados automaticamente. Como nos modelos eTS e suas variações e o modelo ePL, o consequente é parametrizado de acordo com o algoritmo de mínimos quadrados recursivos.

A literatura de economia e finanças relacionada apresenta poucos e recentes trabalhos envolvendo modelos nebulosos evolutivos baseados em regras. (Ballini, Mendonça & Gomide 2009a) utilizaram diferentes estruturas nebulosas evolutivas para estimar a medida de risco de mercado Valor-em-Risco (VaR) e compararam os resultados em termos de acurácia com modelos GARCH utilizando dados do índice da bolsa de valores de São Paulo. Os resultados mostraram que as abordagens de computação inteligente superaram tradicionais *benchmarks* para estimação do VaR, considerando o número de falhas de cada modelo. Resultados similares foram obtidos na modelagem de títulos soberanos (Ballini, Mendonça & Gomide 2009b).

A previsão de séries temporais financeiras envolvendo modelos nebulosos evolutivos foi considerada por (McDonald & Angelov 2010), utilizando um modelo eTS com memória para a modelagem e previsão da taxa de câmbio GBP/EUR e dos valores do Produto Interno Bruto da economia dos Estados Unidos. (McDonald & Angelov 2010) indicaram que o poder preditivo do

¹Além disso, a rede neural nebulosa auto-organizada (Leng, McGinnity & Prasad 2005), como um sistema evolutivo alternativo, adota um critério de erro como medida para avaliar a capacidade de generalização da rede.

modelo eTS com memória é maior, e que seus benefícios podem ser apropriadamente explorados para um entendimento linguístico do sistema. (Maciel, Gomide & Ballini 2011) sugeriram uma abordagem nebulosa evolutiva na precificação de opções de renda fixa. Tendo por base medidas tradicionais de erro e testes estatísticos, os autores mostraram a elevada capacidade dos modelos nebulosos evolutivos na precificação de tais papéis, superando métodos de fórmulas fechadas de avaliação, como a de Black-Scholes, e redes neurais artificiais.

Os modelos eTS, suas extensões (xTS e eTS+), e ePL serão aplicados para a previsão da estrutura a termo das taxas juros. Além disso, nesta dissertação, será proposta uma versão estendida para o modelo ePL, denominado ePL+, de forma a inserir no modelo um mecanismo para adaptação da zona de influência dos clusters e incluir uma medida de utilidade para avaliar a estrutura da base de regras com o objetivo de manter no modelo as regras com maior capacidade de sumarização. Tal modelo será apresentado no Capítulo 5. O modelo ePL+ proposto é uma extensão do modelo ePL para sistemas de múltiplas-entradas-múltiplas-saídas. Na próxima seção serão apresentados os principais conceitos dos modelos nebulosos evolutivos utilizados nesta dissertação.

4.4 Modelos Nebulosos Funcionais Evolutivos

O modelo eTS, suas extensões, e o modelo ePL assumem a classe dos modelos Takagi-Sugeno (TS), sendo os antecedentes das regras nebulosas termos linguísticos, enquanto que os consequentes são funções lineares das variáveis de entrada. Nesses modelos as regras nebulosas são descritas da seguinte forma:

$$\mathcal{R}^i : \mathbf{SE} \ x_1 \text{ is } \Gamma_1^i \ \mathbf{E} \ x_2 \text{ is } \Gamma_2^i \ \mathbf{E} \ \dots \ \mathbf{E} \ x_n \text{ is } \Gamma_n^i \ \mathbf{ENTÃO} \ y^i = \gamma_0^i + \sum_{j=1}^n \gamma_j^i x_j \quad (4.4)$$

sendo que \mathcal{R}^i corresponde a i -ésima regra nebulosa ($i = 1, 2, \dots, R$), $x_j \in \mathfrak{R}$ é o dado de entrada ($j = 1, 2, \dots, n$), Γ_j^i é o conjunto nebuloso associado à j -ésima variável de entrada da i -ésima regra nebulosa, $y^i \in \mathfrak{R}$ é a saída linear da i -ésima regra, e γ_j^i são os i -ésimos parâmetros dos consequentes da regra i .

As regras descrevem comportamentos locais do sistema de dados, de forma que o consequente de cada regra nebulosa é definido por um modelo linear em cada um dos R grupos (regras). Dessa forma, em um modelo TS, cada região nebulosa está associada a um sub-sistema linear. O sistema não-linear é formado pela união dos múltiplos modelos lineares (Angelov & Filev 2004). A contribuição de cada modelo linear local na saída total é proporcional ao grau de ativação de cada regra.

Para definir os antecedentes das regras nebulosas, os modelos estudados nessa dissertação consideram funções de pertinência Gaussianas, que garantem uma elevada capacidade de generalização ao modelo:

$$\mu^i(x_j) = \exp\left(-\frac{(x_j - v^j)^2}{2(\sigma_j^i)^2}\right) \quad (4.5)$$

na qual $\mu^i(x_j)$ é o grau de pertinência da entrada atual x_j em Γ_j^i , v_j^i é o respectivo centro do cluster associado ou ponto focal, e σ_j^i define o raio ou dispersão do conjunto nebuloso associado, ou a zona de influência do conjunto nebuloso i .

O grau de ativação de cada regra nebulosa i pode ser definido de acordo com o operador algébrico de produto como t -norma:

$$\tau^i(x_k) = \prod_{j=1}^n \mu_j^i(x_k) = \mu_1^i(x_k) \cdot \mu_2^i(x_k) \cdot \dots \cdot \mu_n^i(x_k) \quad (4.6)$$

A saída k do modelo TS é obtida por meio da média ponderada da contribuição de cada regra:

$$y = \sum_{i=1}^R \lambda_i y_i, \quad \lambda_i = \frac{\tau_i}{\sum_{l=1}^R \tau_l} \quad (4.7)$$

em que λ_i é o grau de ativação normalizado da regra i , e R é o número de regras nebulosas.

O problema de identificação de modelos nebulosos TS inclui, essencialmente, dois procedimentos (Angelov & Filev 2004): (i) aprendizado dos antecedentes do modelo por meio de um algoritmo de agrupamento nebuloso, (ii) aprendizado dos parâmetros dos consequentes lineares. A estrutura descrita é a mesma para modelos nebulosos evolutivos funcionais. Os modelos eTS e ePL diferem na escolha do algoritmo de clusterização considerado, enquanto os parâmetros do consequente são obtidos por meio da técnica de mínimos quadrados não-lineares recursivos. A seguir, o aprendizado dos antecedentes em modelos do tipo eTS será descrito, enquanto que o modelo ePL, em sua versão estendida, ePL+, será detalhado no próximo capítulo.

4.4.1 Modelo Takagi-Sugeno Evolutivo

O modelo Takagi-Sugeno evolutivo (eTS) (Angelov & Filev 2004) é caracterizado por atualizar o sistema de forma recursiva. Os antecedentes são determinados por meio de um algoritmo de agrupamento não-supervisionado, adicionando-se regras ou modificando as regras existentes. De acordo com o dado mais recente que é apresentado ao modelo, a base de regras e os parâmetros do consequente são atualizados. O eTS baseia-se em uma noção onde a representatividade de um dado é medida a partir de uma função que define o potencial relativo à cada centro de grupo (Yager & Filev 1994).

Essencialmente, caso o potencial de um novo dado de entrada seja maior que o potencial dos centros dos grupos já existentes, esse dado substituirá algum centro existente ou será o centro de um novo cluster e, assim, uma nova regra será criada. O processo de clusterização é realizado em cada instante, de acordo com cada novo dado de entrada que é apresentado ao sistema, mantendo os grupos existentes, atualizando algum centro ou mesmo criando um novo grupo (Angelov 2002). Os parâmetros das funções lineares dos consequentes são calculados utilizando o algoritmo de mínimos quadrados recursivos (Young 1984).

O agrupamento pode ser *online*, de forma a processar os dados em tempo real, não requerendo um conjunto fixo de dados de treinamento. A clusterização se dá de forma recursiva sob a assunção de gradual mudança na base de regras nebulosas. O primeiro dado de entrada x_1 ,

$k = 1$, é definido como o ponto focal da primeira regra, $i = 1$, e o potencial dessa regra é assumido igual a 1 (Angelov 2002). Portanto, a inicialização do modelo é tal que:

$$x^{i*} = x_k, \Upsilon_1(z^{i*}) = 1, R = 1 \quad (4.8)$$

em que x^{i*} corresponde ao ponto focal da primeira regra nebulosa, $z = ([x^T, y^T]^T)$ é o vetor composto pelas entradas e saídas (y), e $\Upsilon(\cdot)$ indica a função potencial.

O próximo dado ($k = k + 1$) é então apresentado ao modelo. O potencial dos novos dados de entrada é calculado recursivamente por meio de uma função potencial obtida a partir da função de Cauchy dada por (Angelov 2002, Angelov & Filev 2004):

$$\Upsilon_k(z_k) = \frac{k - 1}{(k - 1) \cdot (\psi_k + 1) + \iota_k - 2\phi_k} \quad (4.9)$$

na qual $\psi_k = \sum_{j=1}^{n+m} (z_k^j)^2$; $\iota_k = \sum_{l=1}^{k-1} \sum_{j=1}^{n+m} (z_l^j)^2$; $\phi_k = \sum_{j=1}^{n+m} z_k^j \beta_k^j$; $\beta_k^j = \sum_{l=1}^{k-1} z_l^j$.

Na expressão (4.9), os parâmetros ψ_k e ϕ_k são calculados a partir do dado de entrada atual z_k , enquanto β_k^j e ι_k são atualizados recursivamente:

$$\iota_k = \iota_{k-1} + \sum_{j=1}^{n+m} (z_{k-1}^j)^2 \quad \text{e} \quad \beta_k^j = \beta_{k-1}^j + z_{k-1}^j \quad (4.10)$$

com $j = 1, 2, \dots, n$ sendo n a dimensão dos dados de entrada.

No procedimento recursivo, o novo dado disponível é apresentado ao sistema e afeta o potencial dos pontos focais dos centros existentes (z^{l*} , $l = 1, 2, \dots, R$). A medida de potencial depende da distância dos pontos focais em relação à todos os demais pontos, incluindo os novos que são apresentados. Assim, por definição, os potenciais dos agrupamentos já existentes são atualizados recursivamente de acordo com (Angelov & Filev 2004):

$$\Upsilon_k(z^{l*}) = \frac{(k - 1) \cdot \Upsilon_{k-1}(z^{l*})}{k - 2 + \Upsilon_{k-1}(z^{l*}) \cdot \left[1 + \sum_{j=1}^{n+m} (d_{lk}^j)^2\right]} \quad (4.11)$$

sendo que $\Upsilon_k(z^{l*})$ é o potencial do i -ésimo cluster z^{l*} , que representa o protótipo da i -ésima regra no período k , e $d_{lk}^j = z_l^j - z_k^j$ denota a distância entre o agrupamento existente (z_l^j) e o novo dado apresentado (z_k^j).

A evolução na base de regras é derivada pela comparação do potencial do novo dado apresentado ao sistema com o potencial atualizado dos centros/clusters existentes. Em situações onde os dados são obtidos continuamente, o algoritmo de agrupamento deverá determinar se um dado irá modificar um determinado grupo ou se um novo centro de grupo será definido. (Angelov & Filev 2004) definiram os seguintes princípios para a atualização da base de regras em modelos nebulosos TS evolutivos:

(i) Substituição: Se o potencial do novo dado (z_k) é maior que o potencial atualizado dos centros das regras já existentes:

$$\Upsilon_k(z_k) > \max_{i=1}^R \Upsilon_k(z^{i*}) \quad (4.12)$$

e z_k está próximo, em termos de distância Euclidiana, ao centro de algum grupo existente:

$$\frac{\Upsilon_k(z_k)}{\max_{i=1}^R \Upsilon_k(z^{l*})} - \frac{d_{min}}{\|\sigma\|} > 1 \quad (4.13)$$

em que $\delta_{min} = \min_{i=1}^R \|z_k - z^{l*}\|$ é a medida de distância do novo dado ao centro mais próximo das regras existentes, e $\|\sigma\| = \sqrt{\frac{1}{n} \sum_{i=1}^n \sigma_i^2}$ é a norma da dispersão do antecedente.

Então o novo dado (z_k) substitui o centro mais próximo. Seja j o índice associando ao centro mais próximo:

$$x^{j*} = x_k; \quad \Upsilon_k(z^{j*}) = \Upsilon_k(z_k) \quad (4.14)$$

Esse princípio permite que a estrutura da base de regras seja mantida constante, em termo de número de regras, mas que o centro mais próximo ao novo dado seja substituído caso seu potencial seja o maior dentre todos dos agrupamentos existentes.

(ii) Adição: Se apenas (4.12) é satisfeito, mas não (4.13), então o novo dado (z_k) é adicionado à base de regras como um novo centro e assim uma nova regra ou agrupamento é formado:

$$R = R + 1; \quad x^{R*} = x_k; \quad \Upsilon_k(z^{R*}) = \Upsilon_k(z_k) \quad (4.15)$$

Caso contrário, se nenhum dos princípios é verificado com o novo dado que é apresentado ao sistema, a base de regras não sofre alteração. A adaptação recursiva da base de regras garante que o modelo resultante será formado apenas por uma base de regras compacta e informativa, uma vez que as informações espaciais entre os dados não são esquecidas, mas fazem parte do processo de decisão de como atualizar a estrutura da base de regras atual (Angelov 2002).

Abaixo é descrito o algoritmo de agrupamento para o modelo eTS.

Algoritmo de Agrupamento do eTS
Inicializar $x^{i*} = x_k, \Upsilon_1(z^{i*}) = 1, R = 1$
Para $k = 1, 2, \dots$ Faça
Calcular o potencial $\Upsilon_k(z_k) = \frac{k-1}{(k-1) \cdot (\psi_k+1) + \iota_k - 2\phi_k}$
Para $i = 1$ até R Faça :
Atualizar $\Upsilon_k(z^{l*}) = \frac{(k-1) \cdot \Upsilon_{k-1}(z^{l*})}{k-2 + \Upsilon_{k-1}(z^{l*}) \cdot [1 + \sum_{j=1}^{n+m} (d_{lk}^j)^2]}$
Fim Para
Se $\Upsilon_k(z_k) > \max_{i=1}^R \Upsilon_k(z^{l*})$ Então
Se $\frac{\Upsilon_k(z_k)}{\max_{i=1}^R \Upsilon_k(z^{l*})} - \frac{d_{min}}{\ \sigma\ } > 1$ Então
x_k substitui o centro do grupo i : $x^{j*} = x_k; \quad \Upsilon_k(z^{j*}) = \Upsilon_k(z_k)$
Caso Contrário
$R = R + 1; \quad x^{R*} = x_k; \quad \Upsilon_k(z^{R*}) = \Upsilon_k(z_k)$
Fim Se
Fim Se
Fim Para

Nos consequentes, os parâmetros são calculados de acordo com a alteração feita nos centros de grupo. Caso um novo grupo seja adicionado, os parâmetros da função linear do respectivo consequente são obtidos a partir dos parâmetros das regras não modificadas. Por outro lado,

caso algum grupo seja alterado, os parâmetros são atualizados utilizando mínimos quadrados recursivos (Young 1984).

O modelo xTS apresenta as mesmas características que o eTS, exceto por uma fórmula recursiva para o cálculo da dispersão dos agrupamentos. Esse processo de atualização da zona de influência das regras nebulosas será discutido no desenvolvimento do modelo ePL+ no Capítulo 5, que adota o mesmo princípio que o modelo xTS. Por outro lado, o modelo eTS+ (Angelov et al. 2010), que também será avaliado nesta dissertação para a questão de previsão da estrutura a termo das taxas de juros, considera medidas de utilidade, idade, zona de influência e densidade local para a avaliação da qualidade da base de regras.

4.5 Resumo

Esse capítulo introduziu a questão de previsão da estrutura a termo das taxas juros, incluindo uma revisão bibliográfica sobre o tema, a abordagem considerada como *benchmark* para comparações e a ideia da modelagem nebulosa evolutiva. O modelo nebuloso Takagi-Sugeno evolutivo foi descrito. Esse modelo será a base de comparação para a metodologia proposta neste trabalho.

O capítulo seguinte descreverá o algoritmo nebuloso de aprendizado participativo evolutivo estendido, ePL+, que utiliza o agrupamento participativo na determinação de um modelo nebuloso funcional evolutivo do tipo Takagi-Sugeno.

Modelo Nebuloso Evolutivo Participativo Estendido

Este capítulo descreve o modelo nebuloso evolutivo participativo estendido, ePL+. Essa abordagem combina os conceitos de aprendizado participativo e a ideia da modelagem nebulosa evolutiva. As principais diferenças em relação à versão original do modelo ePL são: atualização recursiva do raio de dispersão dos conjuntos nebulosos, utilização de medidas de avaliação de qualidade da estrutura da base de regras e extensão para a modelagem de sistemas com múltiplas saídas. A estrutura do modelo ePL+ é descrita inicialmente, seguida da apresentação do algoritmo de agrupamento, e da estimação dos parâmetros dos consequentes por meio da técnica de mínimos quadrados recursivos.

5.1 Modelos Takagi-Sugeno com Múltiplas Saídas

A versão inicial do algoritmo ePL (Lima 2008) possui uma estrutura baseada em regras nebulosas funcionais do tipo Takagi-Sugeno, conforme apresentado no Capítulo 4. O modelo considera apenas a modelagem de sistemas de múltiplas-entradas-única-saída (MISO - *multi-input-single-output*). Em inúmeras aplicações empíricas, a interação dos fenômenos modelados em sistemas complexos implica a análise de múltiplas decisões (Laurí, Salcedo, García-Nieto & Martínez 2010). Dessa forma, nesta dissertação estendemos o modelo ePL para o caso geral de sistemas de múltiplas-entradas-múltiplas-saídas (MIMO - *multi-input-multi-output*). O modelo ePL+, portanto, assume a seguinte estrutura Takagi-Sugeno para descrever as regras nebulosas:

$$\begin{aligned}
 R^i & : \quad \mathbf{SE} \ x_1 \text{ é } \Gamma_1^i \ \mathbf{E} \ x_2 \text{ é } \Gamma_2^i \ \mathbf{E} \ \dots \ \mathbf{E} \ x_n \text{ é } \Gamma_n^i \ \mathbf{ENTÃO} \\
 & \quad y_1^i = \gamma_{01}^i + \sum_{j=1}^n \gamma_{j1}^i x_j \ \mathbf{E} \ y_2^i = \gamma_{02}^i + \sum_{j=1}^n \gamma_{j2}^i x_j \ \mathbf{E} \ \dots \ \mathbf{E} \ y_m^i = \gamma_{0m}^i + \sum_{j=1}^n \gamma_{jm}^i x_j
 \end{aligned}
 \tag{5.1}$$

Os consequentes assumem forma vetorial $y^i = [y_1^i, y_2^i, \dots, y_m^i]^T$, em que m é a dimensão da saída de cada sub-sistema linear associado a i -ésima regra nebulosa. Os parâmetros do

consequente podem ser escritos na forma matricial:

$$\gamma^i = \begin{bmatrix} \gamma_{01}^i & \gamma_{02}^i & \cdots & \gamma_{0m}^i \\ \gamma_{11}^i & \gamma_{12}^i & \cdots & \gamma_{1m}^i \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1}^i & \gamma_{n2}^i & \cdots & \gamma_{nm}^i \end{bmatrix} \quad (5.2)$$

Nesse caso, o sistema nebuloso evolutivo ePL+ apresenta saída multidimensional (Figura 5.1). Essa extensão permite a adequação do modelo evolutivo participativo para sistemas de múltiplas saídas, incluindo a aplicação de predição da estrutura a termo das taxas de juros, tratada neste trabalho.

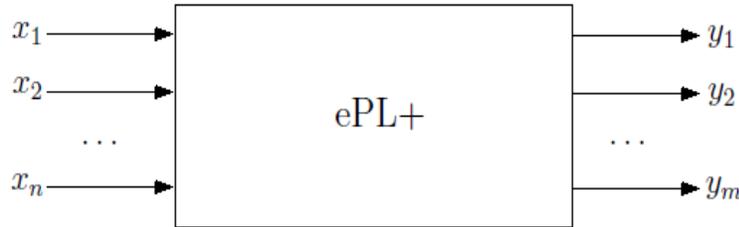


Figura 5.1: Representação gráfica do modelo ePL+.

Assim como no modelo eTS, as funções de pertinência dos antecedentes são Gaussianas, e o grau de ativação de cada regra é calculado por meio da t-norma produto. Os consequentes da base de regras são construídos criando um modelo linear para cada agrupamento nebuloso, e os parâmetros são estimados pelo método de mínimos quadrados recurvisos. A seguir, será detalhado a aprendizagem dos antecedentes do modelo ePL+, incluindo suas extensões em relação ao sistema ePL.

5.2 Algoritmo de Aprendizado Participativo

O Aprendizado Participativo (AP), proposto por (Yager 1990), considera uma estrutura atual como forma de representar o conhecimento de um sistema a ser modelado, de forma que cada nova informação (entrada) é utilizada para atualizar esse conhecimento. O AP representa um importante paradigma na construção de algoritmos de agrupamento (Silva 2003, Silva et al. 2005). A percepção do sistema depende de quanto o mecanismo de aprendizagem participativa é capaz de representar o conhecimento já estabelecido. O conhecimento atual faz parte do processo de aprendizagem, que depende da coerência entre o estado atual do sistema e a informação mais recente (Lima 2008). O aprendizado é realizado por meio da interpretação da nova informação no contexto atual. Com a nova informação, o algoritmo é capaz de avaliar recursivamente a necessidade de atualização do modelo atual.

Assim como em (Lima 2008), o conceito de AP é utilizado neste trabalho no contexto de algoritmos de agrupamento, onde o objetivo é dividir um conjunto de dados em subconjuntos, e associar cada subconjunto à um ponto representativo chamado centro de grupo, ponto focal ou protótipo do grupo. O algoritmo de agrupamento participativo se assemelha aos tradicionais

algoritmos de partição como Fuzzy C-*means* (Bezdek 1981) e suas extensões (Bezdek, Tsao & Pal 1992, Kaymak & Setnes 2000), em que o agrupamento é formado a partir de um processo de otimização cujas soluções são os centros dos grupos e os graus de pertinência de cada dado em relação aos subconjuntos determinados.

O algoritmo de agrupamento no modelo ePL+, assim como na versão inicial do modelo ePL, utiliza a noção de aprendizado participativo para a interpretação dos dados e construção dos antecedentes das regras nebulosas. A base de regras é atualizada recursivamente a cada novo dado apresentado ao sistema, na forma de fluxo de dados, modificando o centro que seja mais compatível com esse novo dado, ou mesmo criando uma nova regra caso seja verificada a necessidade de uma revisão na estrutura da base de regras.

A principal diferença entre os modelos ePL+ e eTS está relacionada ao procedimento de construção e atualização da base de regras. Diferente do eTS, o modelo ePL+ utiliza uma medida nebulosa de similaridade para determinar a proximidade do novo dado apresentado ao modelo com os centros das regras já existentes. O algoritmo de aprendizado participativo assume que o aprendizado do modelo depende do conhecimento atual sobre o sistema considerado. Dessa forma, a estrutura do modelo atual faz parte e contribui para o processo evolutivo, e define como as novas informações do sistema serão utilizadas em sua auto-organização. Uma característica essencial do aprendizado participativo é que o novo dado/informação fornecido ao sistema impacta na atualização da estrutura da base de regras de acordo com uma medida de compatibilidade com a base construída.

Dessa forma, seja $v_k^i \in [0, 1]^n$ a variável que associa o centro ao i -ésimo agrupamento ($i = 1, \dots, R$) na iteração k . O objetivo do processo de aprendizado participativo é aprender o valor de v_k^i , utilizando o fluxo de dados de entrada $x_k \in [0, 1]^n$ apresentado ao sistema. Em outras palavras, cada dado x_k ($k = 1, 2, \dots$) é utilizado como um meio informativo para aprender sobre v_k^i . O processo de aprendizado é participativo se a contribuição de cada novo dado do sistema x_k depende de sua aceitação da estimativa atual de v_k^i como válida, caso isso não ocorra, a estrutura da base de regras deverá ser revisitada. Implicitamente, a ideia central é que o novo dado deve ser útil e contribuir no aprendizado de v_k^i , de forma que os dados observados x_k devem ser compatíveis com as estimativas atuais de v_k^i .

No modelo ePL+, o objetivo do aprendizado é a estrutura dos agrupamentos ou *clusters*, definidas essencialmente por seus pontos focais. Formalmente, dada uma estrutura inicial de regras nebulosas, o conjunto de vetores $v_k^i \in [0, 1]^n$ é atualizado de acordo com uma medida de compatibilidade, $\rho_k^i \in [0, 1]$, e um índice de alerta, $a_k^i \in [0, 1]$. Enquanto ρ_k^i mede o quanto um dado é compatível com a estrutura atual da base de regras, o índice de alerta a_k^i representa uma medida crítica capaz de detectar o grau com que novas informações são coerentes com o modelo atual. O alerta, portanto, é capaz de medir a necessidade de revisão do modelo e adaptação da base de regras. Além disso, o aprendizado participativo evolutivo possui um mecanismo que indica a presença de regras redundantes. Nesse caso, centros muito similares detectados no agrupamento participativo são redefinidos para representar uma única regra com o objetivo de aumentar a eficiência do algoritmo e descrever o sistema com um número reduzido de regras. A Figura 5.2 sumariza os principais constituintes e funcionamento do aprendizado participativo na construção de base de regras nebulosas.

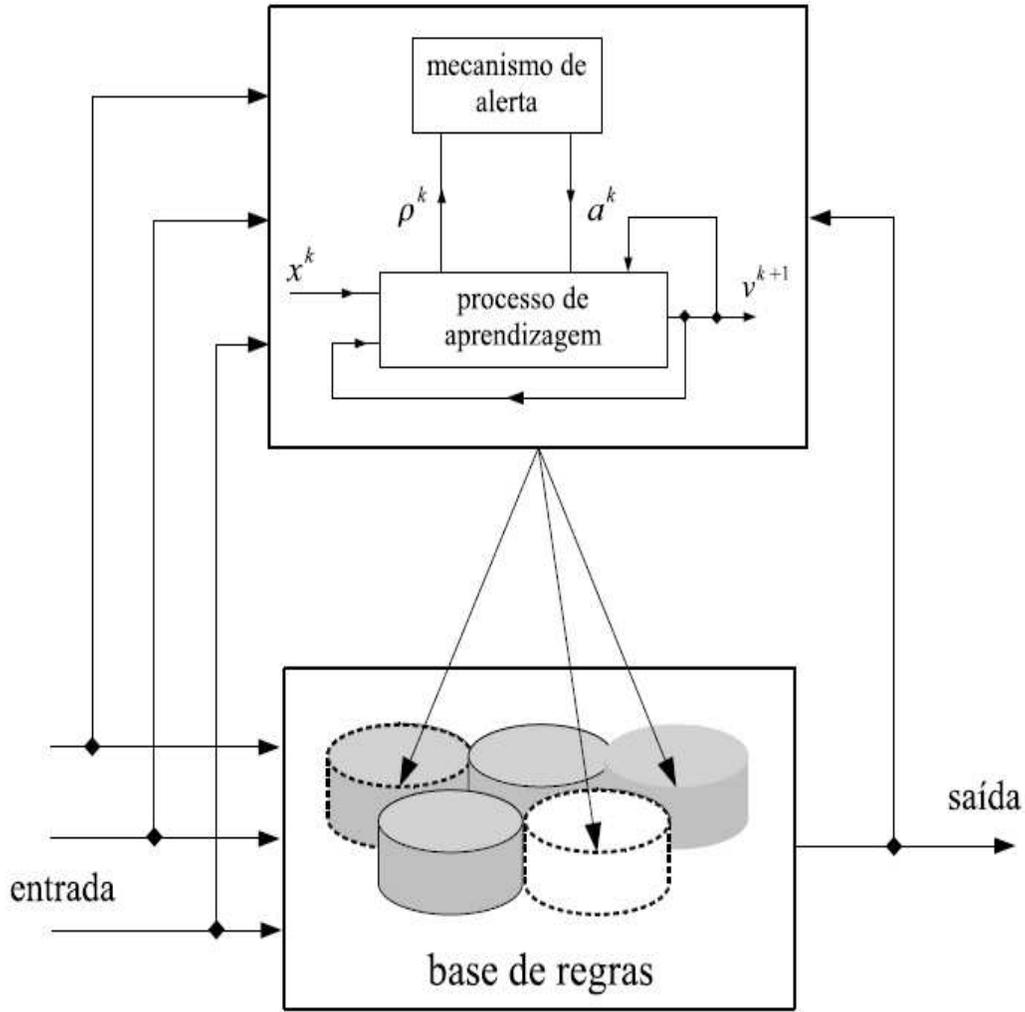


Figura 5.2: Atualização da base de regras no AP. Fonte: (Lima 2008).

O agrupamento participativo possui a capacidade de auto-organização por se tratar de uma técnica não-supervisionada, de forma a criar novas regras ou modificar as regras existentes a cada instante, de acordo com a disponibilidade de novas informações a respeito do sistema. Se o índice de alerta é maior que um determinado limiar $\eta \in [0, 1]$, então uma nova regra ou agrupamento é criado:

$$v_k^{R+1} = x_k \quad (5.3)$$

Por outro lado, caso o índice de alerta não seja ativado, o i -ésimo cluster, mais compatível com o novo dado x_k , é ajustado da seguinte forma:

$$v_{k+1}^i = v_k^i + G_k^i(x_k - v_k^i) \quad (5.4)$$

sendo que

$$G_k^i = \alpha \rho_k^i \quad (5.5)$$

representa uma função que depende da compatibilidade entre os vetores x_k e v_k , tal que $\alpha \in [0, 1]$ é a taxa de aprendizagem do sistema, e

$$\rho_k^i = 1 - \frac{\|x_k - v_k^i\|}{n} \quad (5.6)$$

em que $\|\cdot\|$ é o operador de norma, n a dimensão do espaço de entrada, e

$$i = \arg \max_j \{\rho_k^j\} \quad (5.7)$$

Nesse caso, quando um novo dado x_k é apresentado ao modelo, calcula-se sua compatibilidade em relação aos centros v_k^i ($i = 1, 2, \dots, R$) dos grupos existentes por meio da expressão (5.6). Selecionado o grupo de maior compatibilidade com o novo dado x_k (5.7), esse centro é atualizado por meio de uma combinação convexa do novo dado apresentado ao modelo x_k e seu ponto focal, de acordo com (5.4).

Similar a expressão (5.4), o índice de alerta a_k^i é atualizado da seguinte forma:

$$a_{k+1}^i = a_k^i + \beta(1 - \rho_{k+1}^i - a_k^i) \quad (5.8)$$

O parâmetro $\beta \in [0, 1]$ controla a taxa de mudança do nível de alerta: quanto mais próximo o parâmetro β é da unidade, mais rapidamente o sistema estará sensível à variações do índice de compatibilidade.

O mecanismo de alerta é considerado no aprendizado participativo com a incorporação do índice de alerta (5.8) em (5.5), i.e., assume-se que:

$$G_k^i = \alpha(\rho_k^i)^{1-a_k^i} \quad (5.9)$$

Quando $a_k^i = 0$, tem-se $G_k^i = \alpha\rho_k^i$, formando o mecanismo de aprendizado participativo sem índice de alerta. Deve-se perceber que, quando o nível de alerta aumenta, a medida de similaridade tem seu efeito reduzido. O índice de alerta pode ser interpretado como um complemento na confiança na atual estrutura da base de regras. Dessa forma, o mecanismo de alerta monitora o desempenho do sistema pela observação da compatibilidade do modelo atual com as novas informações que são adicionadas. O aprendizado é dinâmico, no sentido de que (5.4) pode ser visto como uma estratégia de revisão, afetada pela taxa de aprendizagem (5.9) que depende da compatibilidade entre a nova informação ou dado, a atual estrutura da base de regras, e da própria capacidade de sumarização do modelo.

Deve-se notar ainda que a taxa de aprendizagem é modulada pelo índice de compatibilidade. Em modelos de aprendizado convencionais, não há a questão participativa e a taxa de aprendizagem é usualmente baixa para evitar oscilações indesejáveis geradas por valores de dados que situam longe dos pontos focais dos clusters existentes. Valores baixos associados à taxa de aprendizagem, apesar de resultar em lento aprendizado, permite que o modelo não sofra influência de dados ruidosos ou *outliers*. O aprendizado participativo permite a utilização de elevados valores para a taxa de aprendizagem, uma vez que o índice de compatibilidade atua de forma a reduzir a taxa de aprendizagem quando desvios são verificados. Por outro lado, quando o índice de compatibilidade é elevado, o tempo de aprendizado é acelerado.

Além da adaptação da base de regras por meio da atualização dos centros já existentes e da criação de novas regras, o algoritmo nebuloso de aprendizado participativo também verifica se regras redundantes são criadas. Ao se atualizar a base de regras utilizando (5.4) ou inserindo uma nova regra ao modelo (5.3), pode haver a ocorrência da criação de um cluster redundante, caracterizado por estar próximo de outro agrupamento já existente. Um mecanismo simples de verificação desse problema é avaliar se saídas similares são formadas quando obtidas por diferentes regras. No agrupamento participativo, um centro de uma regra é declarado redundante se possuir similaridade superior ou igual a um determinado limiar ϑ , em relação a um outro centro. Neste caso, pode-se manter o centro original ou substituí-lo pela média entre o novo centro e o cluster atual. Assim como em (5.6), o índice de compatibilidade entre centros de agrupamentos é obtido da seguinte forma:

$$\rho_k^{v^i} = 1 - \sum_{l=1}^n |v_k^{i,l} - v_k^{j,l}| \quad (5.10)$$

Portanto, se

$$\rho_k^{v^i} \geq \vartheta \quad (5.11)$$

então os agrupamentos i e j são declarados redundantes.

Os parâmetros de controle do algoritmo de aprendizado participativo são α , β , ϑ e η . A escolha desses parâmetros deve ser de forma apropriada com o mecanismo de inclusão e atualização dos centros dos agrupamentos. Suponha que o centro de índice s , de maior compatibilidade com o dado de entrada x_k , é atualizado. Neste caso, $a_{k+1}^s < \eta$, uma vez que o índice de alerta não ultrapassou o limiar η para gerar sua ativação. Com base na expressão (5.8), tem-se:

$$a_k^s + \beta(1 - \rho_{k+1}^s - a_k^s) = (1 - \beta)a_k^s + \beta d_k^s < \eta \rightarrow d_k^s < \frac{\eta - (1 - \beta)a_k^s}{\beta} < \frac{\eta}{\beta} \quad (5.12)$$

em que $d_k^s = d(v_k^s, x_k)$ é a distância entre o centro do cluster s e o novo dado x_k na iteração k .

Por outro lado, ao se considerar a medida de compatibilidade $\rho_k^{i,j}$, para quaisquer centros i, j distintos:

$$\rho_k^{i,j} = \rho_k^{i,j}(v_k^i, v_k^j) = 1 - d_k^{i,j} \vartheta \rightarrow d_k^{i,j} > 1 - \vartheta \quad (5.13)$$

Se $d_k^{i,j} < 1 - \vartheta$, então os agrupamentos i e j são declarados redundantes e redefinidos como um único agrupamento. Neste caso, $d_k^{i,j}$ é a distância entre os centros v_k^i e v_k^j .

A partir das expressões (5.12) e (5.13), para garantir que um novo centro, não redundante, seja adicionado ao modelo, deve-se escolher os valores de β , ϑ e η de forma que:

$$0 < \frac{\eta}{\beta} \leq 1 - \vartheta \leq 1$$

sendo

$$\eta \leq \beta \quad \text{e} \quad \vartheta \leq 1 - \eta$$

Por outro lado, a inicialização da taxa de aprendizagem α deve ser um valor baixo, tipicamente $\alpha \in [10^{-1}, 10^{-5}]$. Para maiores detalhes da escolha dos parâmetros de controle do

algoritmo de agrupamento participativo e suas questões de convergência ver (Lima 2008). A seguir são apresentadas as extensões que caracterizam o modelo ePL+.

5.2.1 Adaptação da Zona de Influência das Regras

Um importante parâmetro associado às funções de pertinência Gaussianas dos antecedentes das regras nebulosas é a dispersão ou raio dos clusters σ , que define a zona de influência das respectivas regras. Este parâmetro define a dispersão das funções de pertinência e os respectivos termos linguísticos associados na projeção dos agrupamentos em relação aos eixos correspondentes às variáveis de entrada/saída. Por constituir parte das funções de pertinência, o raio dos clusters influencia na determinação do grau de ativação das regras nebulosas (Angelov et al. 2010), revelando sua importância na determinação da saída do modelo. Como em (Angelov & Zhou 2006), o modelo ePL+ considera a zona de influência das regras nebulosas variável, atualizada recursivamente da seguinte forma:

$$\sigma_{j,k}^i = \varsigma \sigma_{j,k}^i + (1 - \varsigma) \varrho_{j,k}^i \quad (5.14)$$

em que ς é a taxa de aprendizagem (valores sugeridos estão no intervalo [0.3,0.5]) (Angelov et al. 2010), e $\varrho_{j,k}^i$, $j = 1, 2, \dots, n + m$, denota a densidade local, definida como:

$$(\varrho_{j,k}^i)^2 = \frac{1}{S_k^i - 1} \sum_{l=1}^{S_k^i - 1} (z_j^i - z_{j,l}^i)^2 \quad (5.15)$$

no qual $z = [x^T, y^T]^T$ é o vetor entrada/saída e $S_k^i = S_{k-1}^i + 1$, para $i = \arg \max_{i=1}^R \lambda^i(x_k)$, é o suporte do cluster i , que associa o número de dados da amostra ao cluster i , e indica a capacidade de generalização que a regra nebulosa em específico proporciona.

Esse mecanismo de adaptação da zona de influência das regras nebulosas permite uma avaliação recursiva de tal parâmetro, evitando a necessidade de determinação pelo usuário e ao mesmo tempo supera a limitação de considerar esse parâmetro fixo ao longo das iterações e igual para todas as variáveis de entrada/saída.

5.2.2 Avaliação da Qualidade da Base de Regras

No agrupamento participativo, os clusters ou grupos são gerados ao associar o dado da amostra atual como ponto focal do novo agrupamento. O índice de compatibilidade permite que os centros sejam ajustados de acordo com as informações que são apresentadas ao modelo, assim como o mecanismo de eliminação de agrupamentos redundantes atualiza a base de regras de forma a evitar a geração de agrupamentos que representem um mesmo conjunto de dados. Entretanto, a versão inicial do modelo ePL (Lima 2008) não possui um mecanismo que monitore, de forma recursiva, a qualidade dos antecedentes das regras nebulosas. No modelo eTS+, (Angelov et al. 2010) propuseram a utilização de distintas medidas para monitoramento dos agrupamentos que constituem a base de regras em modelos nebulosos evolutivos, sendo elas: suporte, utilidade, idade, zona de influência e densidade local.

O modelo ePL+, proposto nesta dissertação, adotará a medida de utilidade como mecanismo para avaliação recursiva da qualidade da estrutura da base de regras. Utiliza-se apenas essa medida para não comprometer a complexidade computacional do modelo e, além disso, conforme análises empíricas (Angelov et al. 2010), a medida de utilidade é capaz de manter uma base de regras com elevada capacidade de sumarização em sistemas complexos.

A utilidade de uma regra nebulosa i corresponde ao seu grau de ativação relativo acumulado (Angelov & Zhou 2006):

$$U_k^i = \frac{\sum_{l=1}^k \lambda^l}{k - I^{i*}}, \quad i = 1, \dots, R, \quad k = 1, 2, \dots \quad (5.16)$$

em que I^{i*} denota o índice temporal (*time tag*) do dado da amostra que foi associado como ponto focal da regra nebulosa i , i.e., indica quando a regra associada ao índice i foi gerada.

Desde que uma determinada regra nebulosa foi gerada, a medida de utilidade indica o quanto essa regra foi utilizada. Essa medida evita a manutenção na estrutura da base de regras de agrupamentos não mais utilizados, i.e., aqueles que possuem um grau de ativação muito baixo. Dessa forma, mantém-se uma base de regras mais compacta, apenas considerando as regras necessárias para descrever o sistema em seu estado atual (Angelov et al. 2010).

A estrutura da base de regras no modelo ePL+ pode variar de acordo com a geração de novas regras, mas pode, ao mesmo tempo, ter a estrutura reduzida ao remover regras redundantes e com baixa utilidade (quando o padrão de comportamento dos dados se distancia do ponto focal de determinadas regras), desconsiderando regras marginais. O mecanismo de verificação da qualidade da base de regras se dá de acordo com o seguinte princípio (Angelov et al. 2010):

$$\mathbf{SE} \ U_k^i < \epsilon \quad \mathbf{ENTÃO} \ R \leftarrow R - 1 \quad (5.17)$$

sendo que $\epsilon \in [0.03; 0.1]$ representa o limiar que controla a utilidade de cada regra nebulosa.

Ao associar essa medida de qualidade, avaliada recursivamente, à estrutura da base de regras, é garantido que a cada novo dado, a qualidade da base de regras seja monitorada, e aquelas com utilidade menor que o limiar ϵ são declaradas desnecessárias para sumarizar o estado atual do sistema. Definidos os mecanismos para composição dos antecedentes das regras no modelo ePL+, a seguir é descrito a estimação dos parâmetros dos consequentes.

5.3 Atualização dos Consequentes das Regras

Uma vez que a estrutura do modelo nebuloso evolutivo ePL+ está definida, por meio da constituição de sua base de regras, o próximo passo é a identificação recursiva dos parâmetros dos consequentes. Escrevendo a saída do modelo ePL+ na forma vetorial:

$$y = \Psi^T \theta \quad (5.18)$$

em que $y = [y_1, y_2, \dots, y_m]^T$ é a saída do modelo MIMO ePL+, $\Psi = [\lambda^1 x_e^T, \lambda^2 x_e^T, \dots, \lambda^R x_e^T]^T$ é o vetor das entradas ponderadas pelo grau de ativação normalizado das regras nebulosas, $x_e = [1, x_1, x_2, \dots, x_n]^T$ é o vetor expandido das entradas, e $\theta = [(\gamma^1)^T, (\gamma^2)^T, \dots, (\gamma^R)^T]^T$ é o vetor de parâmetros da base de regras (associados ao i -ésimo sub-sistema linear local).

A identificação do modelo ePL+ inclui agrupamento não-supervisionado de dados e aprendizado supervisionado dos parâmetros dos consequentes (Figura 5.3).

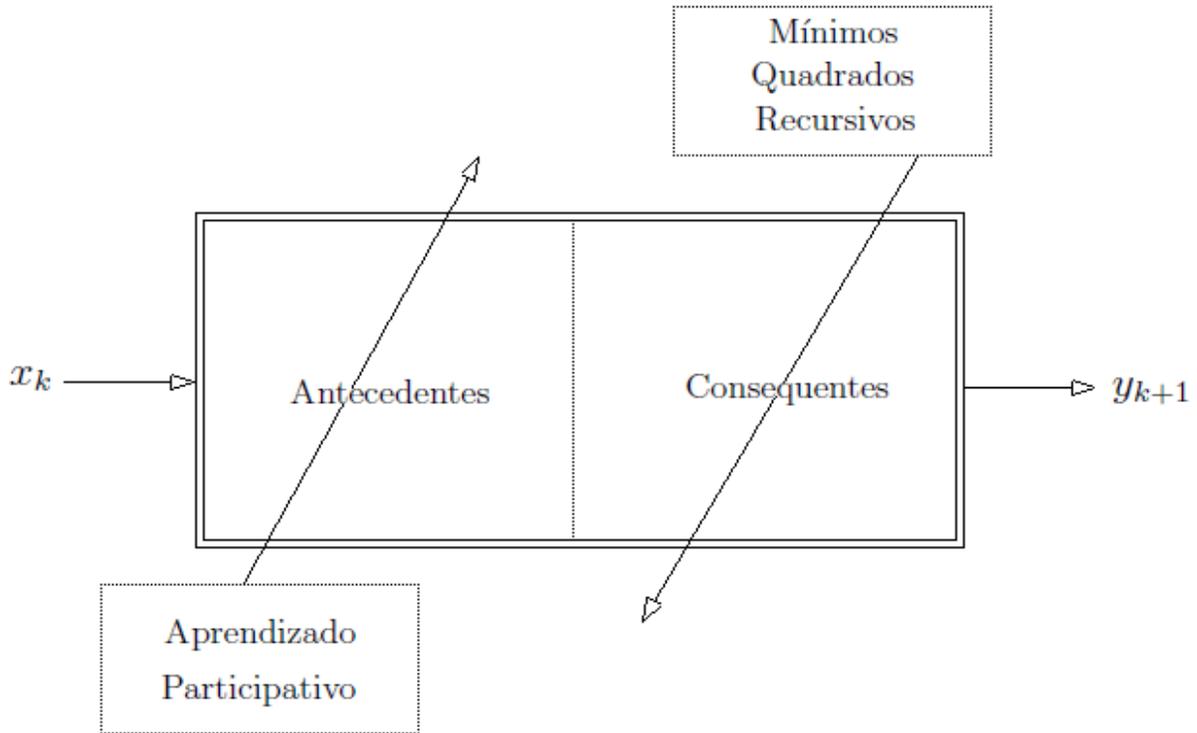


Figura 5.3: Representação esquemática dos estágios para identificação dos modelos ePL+.

A cada instante de tempo k , ou de acordo com o fluxo de dados fornecido ao sistema, as saídas \hat{y}_t são calculadas, de forma que os parâmetros dos consequentes são otimizados recursivamente por meio do algoritmo de mínimos quadrados recursivos. A atualização dos parâmetros pode ser realizada de forma global ou local, exigindo distintos recursos computacionais. A função de erro em termos globais pode ser escrita da seguinte forma (Angelov & Filev 2004):

$$\min E_G = \min \sum_{k=1}^t (\hat{y}_k - \Psi_k^T \theta_k)^2 \quad (5.19)$$

Dessa forma, no caso global, as soluções para os parâmetros são obtidas por meio do algoritmo de mínimos quadrados recursivos como (Angelov & Filev 2004):

$$\hat{\theta}_k = \hat{\theta}_{k-1} + C_k \Psi_{k-1} (y_k - \Psi_{k-1}^T \hat{\theta}), \quad k = 2, 3, \dots \quad (5.20)$$

$$C_k = C_{k-1} - \frac{C_{k-1} \Psi_{k-1} \Psi_{k-1}^T C_{k-1}}{1 + \Psi_{k-1}^T C_{k-1} \Psi_{k-1}} \quad (5.21)$$

sendo que C é a matrix de dispersão de dimensão $R(n+1) \times R(n+1)$.

As condições iniciais são tais que $\hat{\theta}_1 = 0$ e $C_1 = \Omega I$, em que Ω corresponde a um número da ordem de 10^3 para inicialização da matriz de dispersão e I a matriz identidade de mesma ordem de C .

Na otimização em termos globais, no caso de adição de uma nova regra, o algoritmo é atualizado em dois passos.

i) Os parâmetros da nova regra são obtidos pela média ponderada dos parâmetros das demais regras:

$$\hat{\theta}_k = [(\hat{\gamma}_{k-1}^1)^T, (\hat{\gamma}_{k-1}^2)^T, \dots, (\hat{\gamma}_{k-1}^R)^T, (\hat{\gamma}_{k-1}^{R+1})^T]^T \quad (5.22)$$

sendo $\hat{\gamma}_k^{R+1} = \sum_{i=1}^R \lambda^i \hat{\gamma}_{k-1}^i$.

ii) A matriz de dispersão é atualizada da seguinte forma:

$$C_k = \begin{bmatrix} \nu \xi_1 & \dots & \nu \xi_{R(n+1)} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \nu \xi_{R(n+1)1} & \dots & \nu \xi_{R(n+1)R(n+1)} & 0 & \dots & 0 \\ 0 & 0 & 0 & \Omega & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & \Omega \end{bmatrix} \quad (5.23)$$

na qual ξ_{ij} , $i = 1, 2, \dots, R(n+1)$; $j = 1, 2, \dots, R(n+1)$ corresponde aos elementos da matriz de dispersão, e $\nu = \frac{R^2+1}{R^2}$ são os coeficientes.

Por outro lado, quando uma regra é substituída por outra, a matriz de dispersão é mantida constante em relação ao período anterior. Por fim, o valor da saída é obtido recursivamente como:

$$\hat{y}_{k+1} = \Psi_k^T \hat{\theta}_k, \quad k = 2, 3, \dots \quad (5.24)$$

Uma maneira alternativa para o aprendizado dos parâmetros dos consequentes em modelos nebulosos evolutivos é a otimização local. Nesse caso, o critério de erro para a otimização é escrito como:

$$\min E_L^i = \min \sum_{k=1}^t \lambda^i (x_k) (\hat{y}_k - (x_e^T)_k \gamma_k^i)^2 \quad (5.25)$$

Deve-se notar que o critério de erro local é dado por regra ($i = 1, 2, \dots, R$). Para se obter o critério de erro total, faz-se $E_L = \sum_{i=1}^R E_L^i$. De acordo com o algoritmo de mínimos quadrados recursivos, a atualização dos parâmetros dos consequentes é dado por cada regra i , correspondente a cada sub-sistema linear (Angelov et al. 2010):

$$\gamma_{k+1}^i = \gamma_k^i + C_k^i(x_e)_k \lambda_k^i (\hat{y}_k - (x_e^T)_k \gamma_k^i), \quad \gamma_1^i = 0 \quad (5.26)$$

$$C_{k+1}^i = C_k^i - \frac{\lambda_k^i C_k^i (x_e)_k (x_e^T)_k C_k^i}{1 + \lambda_k^i (x_e^T)_k C_k^i (x_e)_k}, \quad C_1^i = \Omega I \quad (5.27)$$

em que I denota a matriz identidade de ordem $(n+1) \times (n+1)$.

Quando uma nova regra é inserida no sistema, seus parâmetros correspondentes são inicializados com base nos parâmetros das R regras existentes:

$$\gamma_k^{R+1} = \sum_{i=1}^R \lambda^i \gamma_{k-1}^i \quad (5.28)$$

A matriz de dispersão é inicializada conforme anteriormente $C_k^{R+1} = \Omega I$. No caso de inserção de uma nova regra, os parâmetros das regras existentes são mantidos constantes em relação ao passo anterior:

$$\gamma_k^i = \gamma_{k-1}^i, \quad i = 1, 2, \dots, R \quad (5.29)$$

Quando dois centros, i e j , são redundantes, apenas uma regra é mantida no sistema, de forma que o centro v_k^R do grupo gerado é determinado como o valor médio dos centros dos grupos redundantes:

$$v_k^R = \frac{1}{2}v_k^i + \frac{1}{2}v_k^j \quad (5.30)$$

ou ainda, simplesmente, pode-se manter um dos grupos e excluir o outro redundante.

Os parâmetros dos consequentes e a matriz de dispersão do centro mantido são associados ao centro redundante excluído:

$$\gamma_k^R = \gamma_{k-1}^j \quad (5.31)$$

$$C_k^R = C_{k-1}^j \quad (5.32)$$

Em geral, o conjunto de parâmetros dos consequentes que satisfazem o critério de otimização global são distintos dos parâmetros correspondentes ao critério local (Angelov et al. 2010). Entretanto, o efeito da adição ou remoção de uma regra nebulosa na estrutura do modelo terá efeitos mais significativos no processo de otimização em termos globais, uma vez que esse efeito é refletido apenas nos valores normalizados dos graus de ativação das regras. No caso local, as matrizes de dispersão C^i , $i = 1, 2, \dots, R$, não mudam com a adição de novas regras ao sistema, uma vez que elas não estão associadas a uma única matriz de dispersão como no caso global. As atualizações dos parâmetros dos consequentes no caso global requerem um maior número de cálculos que no caso local, pois a dimensão da matriz de dispersão no caso local é $(n+1) \times (n+1)$, enquanto que na otimização global, a matriz assume uma dimensão maior, $R(n+1) \times R(n+1)$.

(Angelov et al. 2010) avaliaram as diferenças entre a utilização dos critérios de otimização global e local em modelos nebulosos evolutivos para inúmeros *benchmarks* da literatura, e mostraram que ambas as abordagens exibem resultados similares, com destaque para quando a otimização é realizada em termos locais. Dessa forma, nesta dissertação, foi considerado o critério local, uma vez que possui menor complexidade computacional e garante bons resultados de convergência.

Para maiores detalhes da derivação do algoritmo de mínimos quadrados recursivos associado ao problema de identificação de modelos nebulosos Takagi-Sugeno evolutivos, ver (Lima 2008), (Angelov & Filev 2004), (Angelov & Zhou 2006) e (Angelov et al. 2010).

5.4 Algoritmo ePL+

Uma questão importante na implementação do modelo ePL+ consiste na inicialização dos antecedentes da base de regras. Existem duas formas de realizar tal procedimento. Uma delas é quando não existe nenhum conhecimento *a priori* do sistema a ser modelado, i.e., não existe uma amostra para treinamento. Nesse caso, a inicialização dos antecedentes consiste em associar o primeiro dado de entrada x_1 como ponto focal da primeira regra nebulosa:

$$x_1 = v_1^1 \rightarrow i = R = 1 \quad (5.33)$$

A partir de então, a cada novo dado de entrada x_k ($k = 2, 3, \dots$), calculam-se recursivamente os índices de alerta e de compatibilidade para atualização da base de regras com o algoritmo de aprendizado participativo, conforme descrito na seção 5.2.

Por outro lado, quando existe uma amostra de treinamento, uma solução para inicialização da base de regras é a utilização de um algoritmo de agrupamento nebuloso tal como o algoritmo Fuzzy C-Means, conforme sugerido e utilizado na versão inicial do modelo ePL (Lima 2008). Entretanto, nesta dissertação, será utilizado o algoritmo de agrupamento *Subtractive Clustering* (SC) (Chiu 1994) para inicialização da base de regras no modelo ePL+, quando uma amostra de treinamento está disponível. No algoritmo Fuzzy C-Means, o usuário deve fornecer o número de agrupamentos a ser realizado, o que implica em limitações, exigindo conhecimento específico do problema para determinação de tal parâmetro. Distintamente, o algoritmo SC requer apenas a determinação do parâmetro de dispersão dos agrupamentos e, então, o número de clusters é obtido automaticamente.

O algoritmo SC avalia o potencial de cada dado de entrada tornar o ponto focal de um agrupamento como uma medida de proximidade espacial entre um dado x_i em particular, em relação aos demais dados da amostra de treinamento:

$$\Upsilon_t = \frac{1}{TD} \sum_{j=1}^{TD} e^{-\sigma \|x_t - x_j\|^2}, \quad t = 1, 2, \dots, TD \quad (5.34)$$

sendo que Υ_t denota o potencial do t -ésimo dado de entrada e TD é o número de dados utilizados no treinamento.

O mecanismo de agrupamento no algoritmo SC é descrito pelos seguintes passos:

i) Inicialmente, o dado com o maior potencial é o ponto focal do primeiro agrupamento:

$$\Upsilon_1^* = \max_{j=1}^{TD} \Upsilon_j \quad (5.35)$$

em que Υ_1^* corresponde ao potencial do primeiro cluster.

ii) Reduz-se então o potencial de todos os dados restantes proporcionalmente ao potencial do primeiro ponto focal, e inversamente proporcional à distância desse centro:

$$\Upsilon_t = \Upsilon_t - \Upsilon_k^* e^{-\frac{1}{\sigma} \|x_t - x_j\|^2} \quad (5.36)$$

tal que Υ_k^* indica o potencial do k -ésimo ponto focal.

iii) Determina-se o intervalo de decisão $[\underline{\epsilon}\Upsilon_1^*, \bar{\epsilon}\Upsilon_1^*]$. Se o potencial de um dado for maior que o limite superior do intervalo de decisão ($\bar{\epsilon}\Upsilon_1^*$), então esse dado é definido como o ponto focal de uma nova regra ou agrupamento.

iv) Alternativamente, se o potencial de um dado x_t estiver contido no intervalo de decisão e satisfizer a condição:

$$\frac{d_{min}}{\sigma} + \frac{\Upsilon_t}{\Upsilon_1^*} \leq 1 \quad (5.37)$$

em que d_{min} indica a distância entre o dado x_t e o ponto focal mais próximo.

Neste caso, o dado de entrada x_t substitui o centro mais próximo.

Dessa forma, a utilização do algoritmo SC permite uma inicialização mais autônoma, que não requer conhecimento específico sobre o problema em questão (*expert knowledge*). Além disso, trata-se de um algoritmo menos custoso em termos de complexidade computacional se comparado com o algoritmo clássico Fuzzy C-Means.

A seguir é apresentado o algoritmo de agrupamento do ePL+.

Algoritmo de Agrupamento Participativo Evolutivo Estendido ePL+

Inicializar a base de regras usando o algoritmo SC

Escolher os parâmetros α , β , ϑ e η , e fazer $a_1^i = 0$

Para $k = 1, 2, \dots$ **Faça**

Para $i = 1, 2, \dots, R$ **Faça**

Calcular a compatibilidade: $\rho_k^i = 1 - \frac{\|x_k - v_k^i\|}{n}$

Calcular o alerta: $a_{k+1}^i = a_k^i + \beta(1 - \rho_{k+1}^i - a_k^i)$

Fim Para

Se $a_{k+1}^i \geq \eta \forall i = 1, 2, \dots, R$ **Então**

Acrescentar um novo grupo: $R = R + 1$

x_k é o centro do novo grupo: $v_k^{R+1} = x_k$

Caso Contrário

Encontrar o grupo mais compatível: $s = \arg \max_{j=1}^R \rho_k^j$

Atualizar o grupo s : $v_{k+1}^s = v_k^s + \alpha(\rho_k^s)^{1-a_{k+1}^s}(x_k - v_k^s)$

Sim Se

Para $i = 1, 2, \dots, R - 1$ **E** $j = (i + 1), (i + 2), \dots, R$

Calcular a compatibilidade entre os centros v_k^i e v_k^j : $\rho_k^{i,j} = 1 - d(v_k^i, v_k^j)$

Se $\rho_k^{i,j} \geq \vartheta$ para algum j **Então**

Redefinir v_k^i usando os grupos i e j e exclua v_k^j

Um grupo é excluído: $R = R - 1$

Sim Se

Fim Para

Atualize a medida de utilidade: $U_k^i = \frac{\sum_{l=1}^k \lambda^l}{k - I^{i*}}$

Se $U_k^i \leq \epsilon$ para algum i **Então**

O grupo i é excluído: $R = R - 1$

Fim Se

Fim Para

As principais vantagens dos modelos nebulosos evolutivos são: a estrutura do modelo evolui de acordo com as modificações no padrão do sistema considerado; é possível inicializar o processo de evolução com apenas um dado e melhorar o desempenho do modelo com previsões em tempo real; e trata-se de uma técnica não-iterativa e recursiva, o que garante uma maior efetividade em termos computacionais.

5.5 Resumo

Neste capítulo foi apresentado o modelo nebuloso participativo evolutivo estendido, ePL+, proposto nesta dissertação. Na proposta, as principais extensões da versão inicial do modelo ePL incluem o mecanismo de adaptação da zona de dispersão das regras nebulosas, assim como acrescenta a medida de utilidade para o monitoramento da qualidade da estrutura da base de regras. Além disso, a utilização do algoritmo SC na inicialização dos antecedentes das regras quando uma amostra de treinamento está disponível foi considerado.

O próximo capítulo apresenta os resultados empíricos para os problemas de modelagem e previsão da estrutura a termo das taxas de juros nos mercados de renda fixa dos Estados Unidos e do Brasil. A questão de estimação da curva de juros foi tratada com a utilização de algoritmos de computação evolucionária, enquanto a previsão da estrutura a termo inclui a aplicação de modelos nebulosos evolutivos. O desempenho das técnicas de computação inteligente será comparado com o de abordagens econométricas tradicionais verificadas na literatura.

Resultados Empíricos

Esse capítulo reporta os resultados computacionais na modelagem e previsão da estrutura a termo das taxas de juros. Inicialmente, a base de dados, para os mercados dos Estados Unidos e do Brasil, é descrita. Em seguida, as comparações entre as técnicas de computação evolucionária e abordagens tradicionais de otimização não-linear de parâmetros são apresentadas considerando a ideia da modelagem da estrutura a termo de acordo com o modelo de fatores latentes. Por fim, são descritos e avaliados os resultados das técnicas nebulosas evolutivas para a previsão da curva de rendimentos, em comparação com modelos autoregressivos e de redes neurais artificiais.

6.1 Base de Dados

Neste trabalho, os mercados de juros dos Estados Unidos e do Brasil são avaliados. Para o mercado dos EUA, a base de dados corresponde a taxas zero-cupom dos títulos públicos emitidos pelo *Federal Reserve*, banco central norte-americano, para as maturidades de 1, 3, 6, 12, 24, 36, 60, 84 e 120 meses, no período de Janeiro de 1987 a Agosto de 2011¹. A amostra de dados inicia poucos meses antes da sucessão de Paul Vocker por Alan Greenspan como presidente do *Federal Reserve* no verão de 1987. Esse acontecimento é considerado por muitos autores como uma mudança na política monetária dos EUA (Bernanke & Woodford 2006). Por fim, o final da amostra é marcado pela saída do movimento recessivo da atual crise do *Subprime*.

Para o mercado brasileiro foram consideradas as taxas do contrato de *swap* DI contra taxa pré-fixada (*swap* PRÉ×DI), um dos instrumentos mais importantes e líquidos do mercado de renda fixa. A taxa DI é a taxa média das operações interbancárias de empréstimos por um dia, apuradas pela CETIP (Central de Custódia e Liquidação Financeira de Títulos). A taxa DI em um dia t é definida como a média ponderada das taxas às quais foram efetuadas operações interbancárias nesse mercado, sendo então divulgada como uma taxa composta calculada a partir de uma base de 252 dias úteis por ano. O contrato *swap* DI com taxa pré-fixada em T ($T > t$) é um derivativo que no dia corrente t é negociado a preço zero, mas que tem o preço de referência dado pelo valor esperado sob à medida neutra ao risco de 100 mil reais, descontado pela taxa DI de um dia, considerada como taxa livre de risco. Esse tipo de contrato apresenta

¹Os dados foram obtidos no banco de dados do Federal Reserve e estão disponíveis em: <http://www.federalreserve.gov/econresdata/releases/statisticsdata.htm>.

as mesmas características de um título zero-cupom, uma vez que esses contratos são registrados e contam como garantia dadas pela BM&FBOVESPA, podendo assim ser consideradas suas taxas como livres de riscos. Dessa forma, a ETTJ brasileira foi extraída dos contratos de swap PRÉ×DI para as maturidades de 1, 3, 6, 12, 24, 36, 60, 84 e 120 meses, no período de Janeiro de 2004 a Dezembro de 2008². O período em questão foi escolhido de acordo com a disponibilidade dos dados, uma vez que, nos anos 2000, o mercado de renda fixa no Brasil apresentou maior liquidez para taxas com maturidades mais longas e, a partir de 2008, tal mercado sofreu elevadas instabilidades em decorrência da crise do *Subprime*.

As Figuras 6.1 e 6.2 apresentam as estruturas a termo das taxas de juros dos mercados de renda fixa dos EUA e Brasil ao longo dos períodos considerados, respectivamente. Pode-se perceber que em ambas economias as curvas de juros apresentam um comportamento diversificado, com diferentes formatos de nível, inclinação e curvatura. Para o mercado brasileiro, apesar de a amostra cobrir um período, em anos, menor que para os dados dos EUA, a estrutura a termo apresentou dinâmicas variadas. Deve-se notar que os dados norte-americanos são amostrados mensalmente, enquanto, para o caso do Brasil, a periodicidade é diária. Além disso, as taxas no mercado brasileiro são, em geral, mais altas que nos Estados Unidos, uma vez que o custo do capital em economias emergentes é maior, pois o risco associado é também maior. Contudo, pode-se perceber ainda o reflexo da crise do subprime em ambas economias, marcada pela acentuada queda nos níveis de taxas de juros para distintas maturidades.

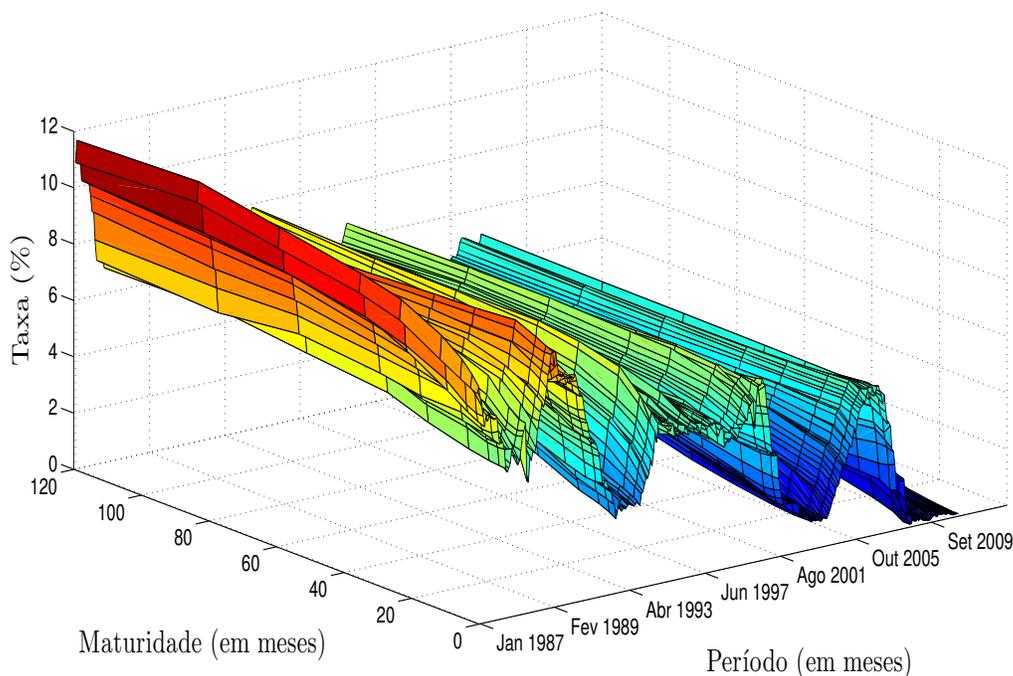


Figura 6.1: Estrutura a Termo das Taxas de Juros do Mercado dos Estados Unidos.

²Os dados foram fornecidos pela BM&FBOVESPA.

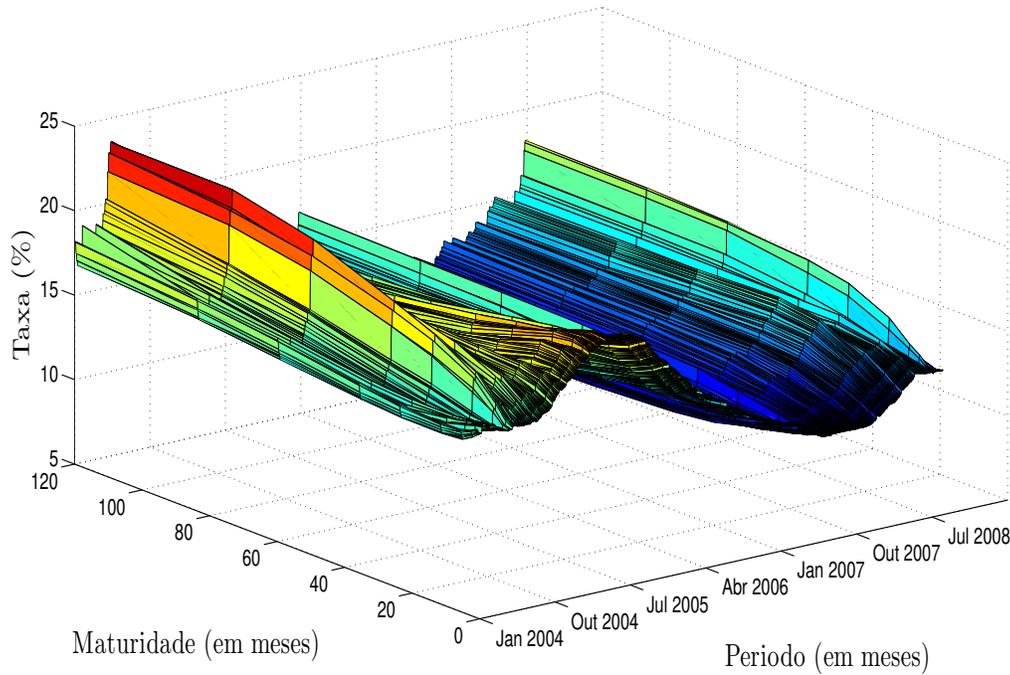


Figura 6.2: Estrutura a Termo das Taxas de Juros do Mercado do Brasil.

6.2 Modelagem da Curva de Juros

Nesta seção, serão apresentados os resultados de estimação da estrutura a termo das taxas de juros de acordo com o modelo parcimonioso de (Nelson & Siegel 1987), considerando dados dos mercados de renda fixa dos Estados Unidos e do Brasil. A seguir, é apresentada a parametrização das metodologias empregadas, seguida das comparações em termos de ajuste.

6.2.1 Parâmetros de Controle

Apesar de cada técnica avaliada possuir seus parâmetros de controle específicos, todas necessitam da determinação de um vetor de parâmetros $\mathbf{x}_0 = \{\beta_{1,0}, \beta_{2,0}, \beta_{3,0}, \lambda_0\}$ para a inicialização do processo iterativo. Para todos os algoritmos foi empregada a mesma inicialização, proposta por (Manousopoulos & Michalopoulos 2009):

$$\beta_{1,0} = \frac{1}{l} \sum_{j=1}^l y_j, \quad \beta_{2,0} = y_s - \beta_{1,0}, \quad \beta_{3,0} = 0 \text{ e } \lambda_0 = 1 \quad (6.1)$$

em que os índices l e s correspondem às taxas com as maturidades mais longas e mais curta das amostras, respectivamente.

Embora os algoritmos evolucionários forneçam boas estimativas para os parâmetros da curva de NS quando inicializados aleatoriamente, um método mais apropriado de inicialização permite elevado ganho em termos de complexidade computacional (tempo de processamento), diferente

de técnicas tradicionais de otimização não-linear, em que esse procedimento tem papel crucial na obtenção de soluções adequadas, diminuindo o risco de não-convergência. Os algoritmos de mínimos quadrados não-lineares (MQNL daqui em diante) e de programação quadrática sequencial (SQP) correspondem às abordagens consideradas como *benchmarks* para objetivos de comparações.

Os parâmetros de controle do algoritmo de MQNL são: limiar associado ao critério de parada ϵ , coeficiente associado à função de penalização c_0 , e parâmetro de atualização da penalização κ . Como o problema de modelagem da curva de NS corresponde a um problema de otimização não-linear de parâmetros restritos, os parâmetros de controle estão associados à função de penalidade inserida ao problema para implementação da técnica de Gauss-Newton no algoritmo de MQNL. Portanto, os parâmetros de controle utilizados foram: $\epsilon = 10^{-4}$, $c_0 = 100$ e $\kappa = 4$. Essa parametrização evitou a parada prematura do algoritmo em pontos de mínimos locais, assim como foi capaz de fornecer estimativas razoáveis, em termos de ajuste da curva aos dados reais, sem implicar em aumento expressivo do tempo de processamento.

Semelhantemente a técnica de MQNL, o algoritmo de programação quadrática sequencial possui os mesmos parâmetros c (só que agora considerado fixo ao longo das iterações), e ϵ , cuja interpretação é idêntica. Contudo, tem-se ainda a inicialização da matriz B_0 , uma matriz definida positiva como aproximação da Hessiana da função do lagrangeano. Tomou-se então: $\epsilon = 10^{-4}$, $c = 100$, e $B_0 = I$, em que I representa a matriz identidade de ordem $D \times D$, em que D representa a dimensão do problema em questão, i.e., $D = 4$. Deve-se notar que, a parametrização dos algoritmos tradicionais de otimização foi a mesma para ambos os mercados, i.e., Estados Unidos e Brasil.

Os algoritmos de computação evolucionária avaliados nessa dissertação foram: Algoritmos Genéticos (AG), de Evolução Diferencial (ED) e de Estratégias Evolutivas (EE), mais precisamente, o algoritmo CMA-ES. A determinação dos parâmetros de controle das abordagens evolucionárias foi feita com base em simulações, como em (Gimeno & Nave 2009). Selecionaram-se distintas amostras em cada base de dados, de forma aleatória, e então computou-se o ajuste, em termos de erro quadrático médio, em função dos parâmetros. Os parâmetros de controle escolhidos foram aqueles que resultaram em melhor ajuste e menor tempo de processamento, correspondendo a um critério de parcimônia, uma vez que, como cada amostra de dados é extensa, seria infactível em termos computacionais determinar os parâmetros “ótimos” para cada dia ou mês da amostra, de acordo com a periodicidade dos dados.

O AG implementado possui seis parâmetros de controle: número de indivíduos da população de cada geração N , variabilidade da população de indivíduos inicial medida pelo desvio padrão σ , probabilidade de mutação π , parâmetro de dispersão da distribuição beta, α , utilizada na seleção dos pais de cada geração, proporção ϕ dos indivíduos sobreviventes de cada geração para a próxima, e número máximo de gerações sem modificação do melhor indivíduo G_{max} . Dessa forma, a parametrização para os dados do mercado de renda fixa dos Estados Unidos foi: $N = 800$, $\sigma = 0.5$, $\pi = 0.65$, $\alpha = 6$, $\phi = 0.4$ e $G_{max} = 200$. Para a base de dados da estrutura a termo do Brasil, tem-se: $N = 700$, $\sigma = 0.5$, $\pi = 0.50$, $\alpha = 6$, $\phi = 0.4$ e $G_{max} = 300$.

Os parâmetros de controle do algoritmo de evolução diferencial são: N , σ , G_{max} , como no algoritmo genético, mais o parâmetro que controla a mutação F e a taxa de crossover Cr . De acordo com (Zaharie 2002), o parâmetro F possui um valor crítico F_{crit} no qual a variabilidade da população decresce quando $F < F_{crit}$, e cresce quando $F > F_{crit}$. Nas simulações, foram considerados os valores possíveis de F tal que $F > F_{crit}$, para mantermos uma população diversificada ao longo das gerações. Dessa forma, F_{crit} é definido por (Zaharie 2002):

$$F_{crit} = \sqrt{\frac{(1 - Cr)}{2}} \cdot N \quad (6.2)$$

Portanto, para o mercado dos Estados Unidos, os parâmetros de controle do algoritmo de evolução diferencial são: $N = 40$, $\sigma = 0.5$, $Cr = 0.6$, $F = 1$, e $G_{max} = 200$. Já para o mercado de renda fixa do Brasil, temos: $N = 50$, $\sigma = 0.5$, $Cr = 0.5$, $F = 1.2$, e $G_{max} = 200$

Por fim, no algoritmo de estratégias evolutivas CMA-ES, além do vetor inicial de parâmetros, é apenas necessário determinar o parâmetro de dispersão da população de indivíduos ou também denominado como tamanho do passo σ , G_{max} e ϵ , que possuem a mesma interpretação dos demais algoritmos. Para ambos os mercados de renda fixa adotou-se $\sigma = 0.5$, $G_{max} = 500$ e $\epsilon = 10^{-4}$. Determinados os parâmetros de controle dos algoritmos considerados, a subseção seguinte apresenta os resultados computacionais para o ajuste da curva de juros de acordo com o modelo de fatores latentes de (Nelson & Siegel 1987).

6.2.2 Resultados

Conforme discutido nesta dissertação, as estruturas a termo das taxas de juros dos mercados de renda fixa dos Estados Unidos e Brasil foram estimadas por meio do modelo de fatores latentes de NS utilizando técnicas tradicionais de otimização não-linear de parâmetros (MQNL e SQP) e metodologias de computação evolucionária (AG, ED e EE).

A Tabela 6.1 apresenta as estatísticas descritivas dos parâmetros da curva de juros de NS para as diferentes técnicas de otimização consideradas³. Os resultados, para os diferentes algoritmos, foram semelhantes, entretanto, considerando a medida de desvio padrão, os algoritmos evolucionários apresentaram parâmetros menos voláteis, para ambas economias avaliadas, especialmente considerando os algoritmos de evolução diferencial e de estratégias evolutivas. Como os fatores latentes da curva de NS têm interpretação econômica de taxas de juros, em termos de expectativas, é esperada uma menor variabilidade para esses fatores, uma vez que os agentes de mercado não mudam suas expectativas abruptamente no curto prazo, exceto quando instabilidades atingem o mercado de juros e alteram as taxas esperadas. Embora nos mercados avaliados os algoritmos utilizados apresentaram resultados similares, na economia do Brasil isso ficou mais evidente, uma vez que o período amostral é mais curto e com menos instabilidades nas curvas de juros em termos de nível, inclinação e curvatura.

Para o mercado dos Estados Unidos, considerando o coeficiente de autocorrelação dos fatores da curva de juros, a Tabela 6.1 mostra que o fator de nível (β_1) apresenta maior persistência (maior índice de autocorrelação), evidenciado pelo elevado valor do coeficiente de autocorrelação

³As estatísticas descritivas consideradas foram: média, desvio padrão e coeficiente de autocorrelação.

para os lags de 12 e 30. Os fatores de inclinação (β_2) e curvatura (β_3) apresentam uma memória mais recente, em que a persistência decresce com o crescimento do lag considerado. Por fim, o parâmetro de decaimento (λ) da curva de juros é o menos persistente, com um decrescimento considerável de seu coeficiente de correlação (Tabela 6.1). As análises são distintas quando o mercado brasileiro de renda fixa é considerado, uma vez que nele é verificada uma elevada memória para todos os fatores latentes, com exceção do fator de decaimento que, ao lag 30, já possui uma memória mais curta, conforme Tabela 6.1. Essas diferenças são decorrentes da periodicidade dos dados: no Brasil tem-se dados diários, enquanto nos EUA, a periodicidade é mensal e, conseqüentemente, dados diários terão uma memória mais longa, em termos de coeficiente de autocorrelação.

Tabela 6.1: Estatísticas descritivas para os fatores latentes de NS estimados de acordo com as metodologias de MQNL, SQP, AG, ED e EE, considerando os mercados de renda fixa dos Estados Unidos e do Brasil.

Estados Unidos / Brasil						
Fator de Nível (β_1)						
Técnica	Média	Desv. Pad.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
MQNL	6,658/16,163	1,724/2,567	0,959/0,933	0,711/0,837	0,672/0,709	-1,564/-1,066
SQP	6,765/16,284	1,982/2,488	0,977/0,981	0,801/0,906	0,684/0,800	-1,600/-0,837
AG	6,562/16,333	1,804/2,417	0,949/0,985	0,693/0,907	0,654/0,805	-1,611/-0,872
ED	6,631/16,334	1,731/2,361	0,966/0,981	0,711/0,903	0,676/0,798	-1,762/-0,830
EE	6,622/16,332	1,720/2,440	0,975/0,980	0,709/0,900	0,655/0,797	-1,811/-0,851
Fator de Inclinação (β_2)						
Técnica	Média	Desv. Pad.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
MQNL	-2,531/0,826	1,887/3,709	0,914/0,948	0,446/0,895	-0,224/0,801	-1,939/-2,138
SQP	-2,447/0,559	1,833/3,695	0,932/0,989	0,502/0,941	-0,208/0,857	-2,019/-1,908
AG	-2,385/0,619	1,972/3,686	0,843/0,992	0,397/0,944	-0,255/0,861	-3,176/-1,636
ED	-2,544/0,606	1,772/3,690	0,952/0,989	0,478/0,941	-0,257/0,858	-1,324/-1,858
EE	-2,476/0,619	1,788/3,701	0,977/0,990	0,480/0,942	-0,265/0,859	-2,736/-1,796
Fator de Curvatura (β_3)						
Técnica	Média	Desv. Pad.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
MQNL	-1,687/-0,272	2,052/3,908	0,832/0,982	0,311/0,911	-0,296/0,734	-3,731/-2,590
SQP	-1,701/-0,323	2,198/3,921	0,872/0,980	0,327/0,909	-0,266/0,734	-4,826/-2,671
AG	-1,777/-0,372	1,927/3,926	0,657/0,978	0,325/0,907	-0,183/0,732	-5,519/-2,799
ED	-1,753/-0,304	1,974/3,888	0,864/0,965	0,364/0,901	-0,297/0,725	-3,204/-3,476
EE	-1,786/-0,298	1,955/3,926	0,798/0,966	0,338/0,900	-0,199/0,726	-3,872/-3,430
Fator de Decaimento (λ)						
Técnica	Média	Desv. Pad.	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
MQNL	0,234/0,097	0,457/0,063	0,483/0,954	0,012/0,751	-0,101/0,425	-8,928/-1,990
SQP	0,245/0,127	0,501/0,058	0,476/0,944	0,065/0,750	-0,113/0,419	-9,179/-1,815
AG	0,240/0,114	0,416/0,059	0,444/0,924	0,105/0,734	-0,119/0,418	-9,032/-2,271
ED	0,256/0,113	0,503/0,055	0,788/0,926	-0,005/0,734	-0,108/0,411	-5,324/-2,280
EE	0,250/0,116	0,404/0,054	0,465/0,923	-0,094/0,727	-9,106/0,393	-7,625/-2,324

Notas: $\rho(k)$ representa a autocorrelação parcial para o lag k e ADF é a estatística do teste de Dickey-Fuller aumentado para hipótese de raiz unitária (Dickey & Fuller 1979).

Por fim, as estatísticas do teste aumentado de Dickey-Fuller para a hipótese de raiz unitária, tomando o mercado dos EUA, mostrou que o fator de nível da curva de NS (β_1) possui raiz

unitária para todas as técnicas de otimização de parâmetros consideradas (Tabela 6.1)⁴. Além disso, considerando as metodologias de MQNL, SQP, e ED, o parâmetro de inclinação (β_2) também possui raiz unitária. Para o caso do Brasil, para todas as metodologias, apenas o fator de curvatura (β_3) não possui raiz unitária (Tabela 6.1)⁵.

As Figuras 6.3-6.6 apresentam a evolução temporal dos fatores latentes de NS da estrutura a termo das taxas de juros para o mercado dos Estados Unidos, estimados de acordo com as técnicas de otimização de MQNL, SQP, AG, ED e EE.

Avaliando, inicialmente, as estimativas para o mercado dos EUA, o parâmetro de nível da curva de juros (β_1) apresentou uma tendência de queda ao longo do período avaliado, revelando expectativas de tendência de baixa nas taxas de juros de curto prazo (Figura 6.3). Os algoritmos apresentaram resultados muito semelhantes, essencialmente no período pré crise do subprime. Durante a crise, as estimativas são mais voláteis, evidenciando o nervosismo no mercado de renda fixa durante esse momento. Considerando o fator de inclinação (β_2), pode-se perceber uma elevada similaridade entre as estimativas, porém, o algoritmo genético, em determinados momentos, apresentou maior variabilidade (Figura 6.4).

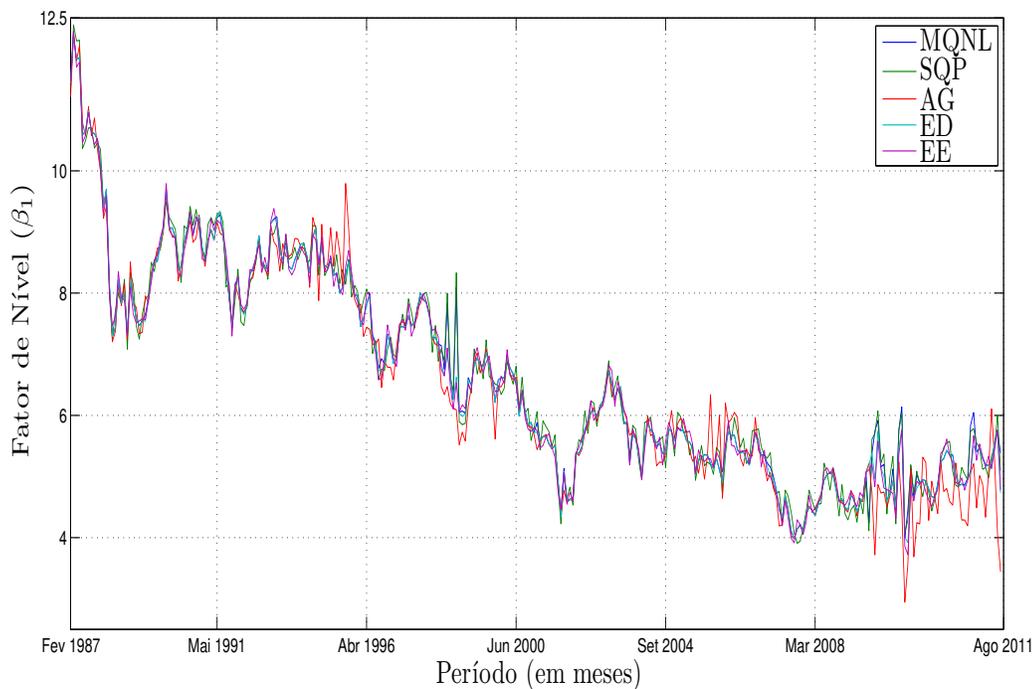


Figura 6.3: Evolução do fator de nível $\hat{\beta}_1$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.

⁴Os valores críticos para rejeição da hipótese de raiz unitária são $-3,4518$ ao nível de 1% , $-2,8704$ e $-2,5714$ aos níveis de 5% e 10% , respectivamente.

⁵A verificação de raiz unitária indica que um processo estocástico é não-estacionário.

Resultados foram semelhantes para as técnicas de MQNL, AG, ED, e EE na estimação do fator de curvatura (β_3), de forma que o algoritmo SQP apresentou resultados com maior variabilidade, conforme verificado na Figura 6.5. Por outro lado, maiores diferenças entre as técnicas de otimização podem ser observadas na estimação do parâmetro de decaimento da curva (λ), mas, ainda, os algoritmos capturaram a mesma tendência da série. Contudo, vale notar a limitação de se fixar um valor determinado *a priori* para o parâmetro de decaimento, uma vez que seu comportamento evidencia que esse fator da curva de juros não pode ser considerado fixo, mas sim variante no tempo.

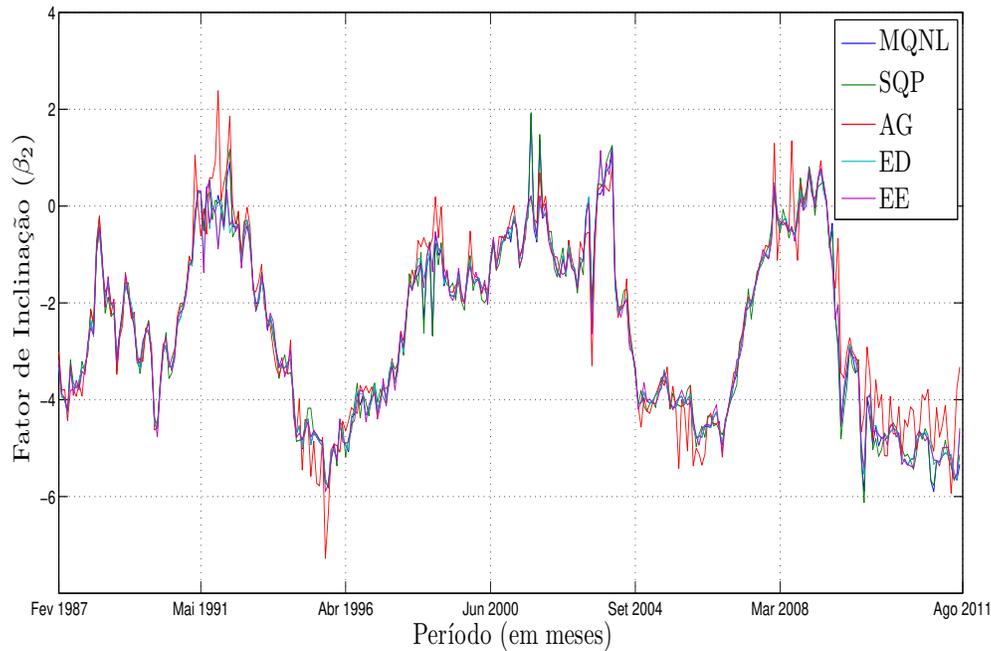


Figura 6.4: Evolução do fator de inclinação $\hat{\beta}_2$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.

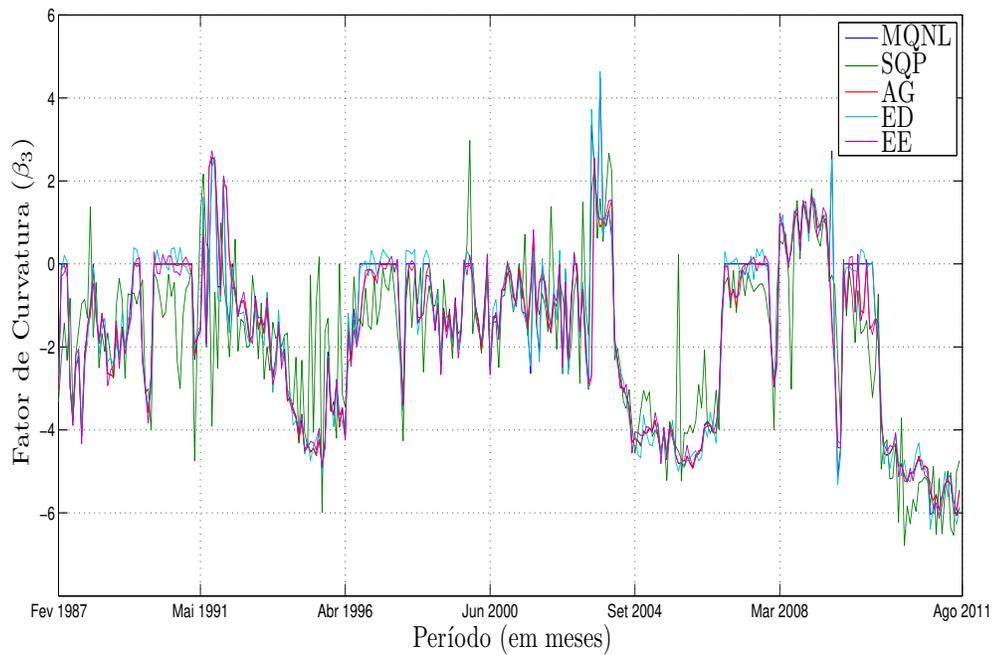


Figura 6.5: Evolução do fator de curvatura $\hat{\beta}_3$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.

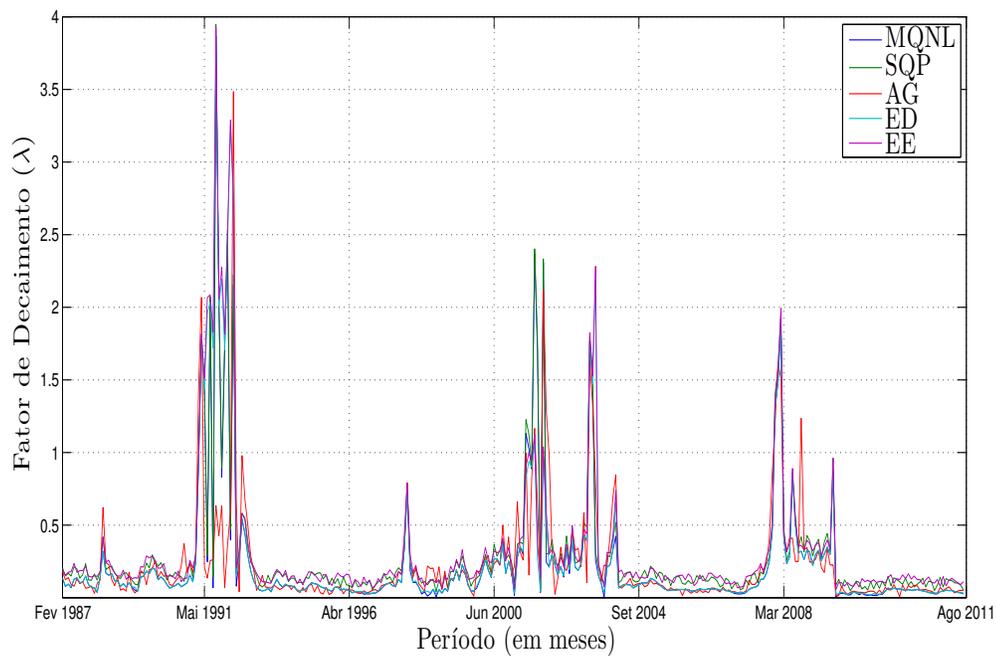


Figura 6.6: Evolução do fator de decaimento $\hat{\lambda}$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa dos EUA.

Para o mercado de renda fixa no Brasil, os resultados foram semelhantes aos dos Estados Unidos. As Figuras 6.7-6.10 apresentam as estimativas para os parâmetros da curva de NS para a estrutura a termo das taxas de juros no Brasil. Considerando o fator de nível (β_1), pode-se perceber um maior nervosismo no mercado de juros, evidenciado pela maior presença de picos e vales abruptos ao longo da evolução do parâmetro, se comparado com os dados norte-americanos (Figura 6.7). No início do período, a técnica de mínimos quadrados não-lineares apresentou resultados distintos as demais metodologias de otimização. Além disso, a partir de setembro de 2007, é verificada uma tendência de queda das expectativas das taxas de juros de curto prazo, decorrente das proporções globais que a crise do subprime tomou no mundo todo, inclusive no Brasil. O mesmo comportamento foi observado na evolução temporal do fator latente de inclinação (β_2), em que os algoritmos apresentaram elevada similaridade nos resultados (Figura 6.8).

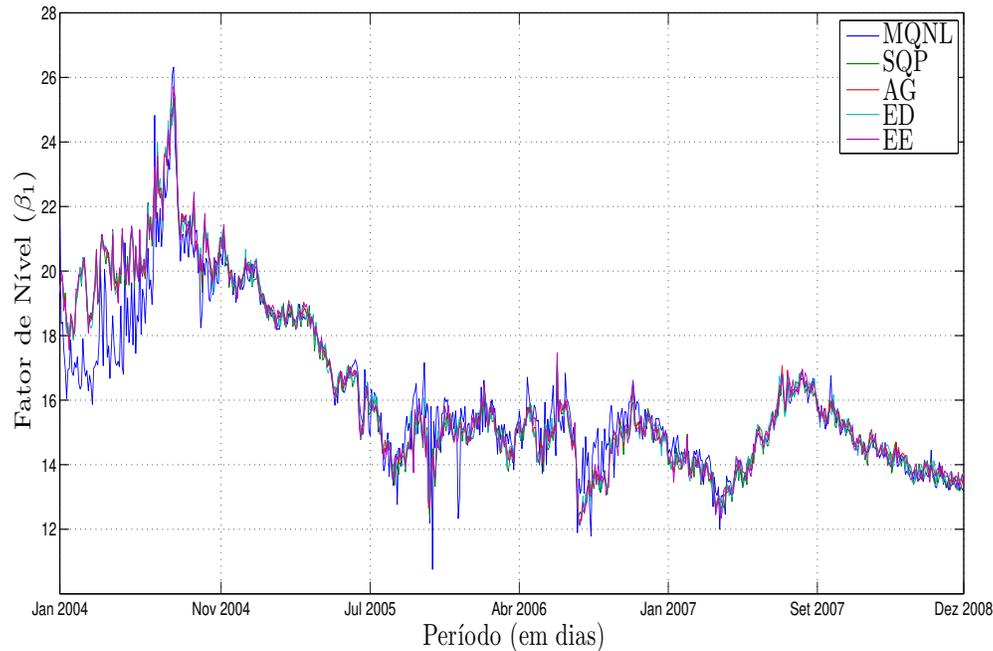


Figura 6.7: Evolução do fator de nível $\hat{\beta}_1$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.

Os algoritmos avaliados apresentaram resultados muito semelhantes para o parâmetro de curvatura (β_3) da estrutura a termo das taxas de juros no mercado brasileiro (Figura 6.9), exceto para o algoritmo de estratégias evolutivas, CMA-ES, que apresentou alguns vales. Nos períodos de setembro de 2004 a maio de 2005 e de abril de 2006 a meados de junho 2007, o parâmetro de curvatura apresentou maior estabilidade, em que, a partir de julho de 2007 a variabilidade de tal fator iniciou um período de elevação, em decorrência da atual crise mundial. Finalmente, como no mercado dos EUA, o fator de decaimento da curva de NS (λ) apresentou um elevado nervosismo, em termos de variabilidade, conforme apresentado na Figura 6.10.

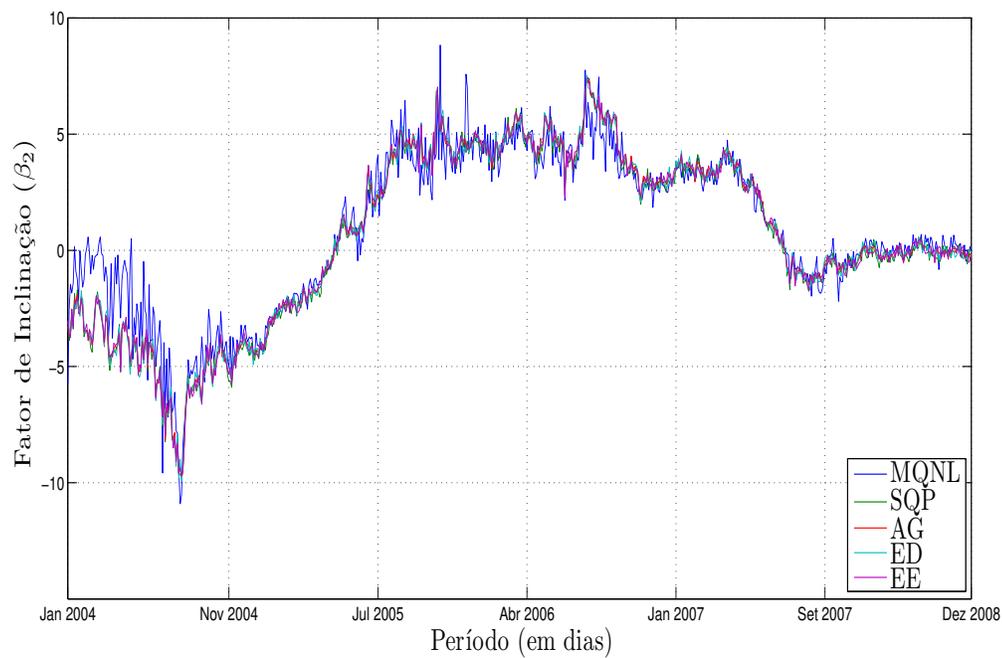


Figura 6.8: Evolução do fator de inclinação $\hat{\beta}_2$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.

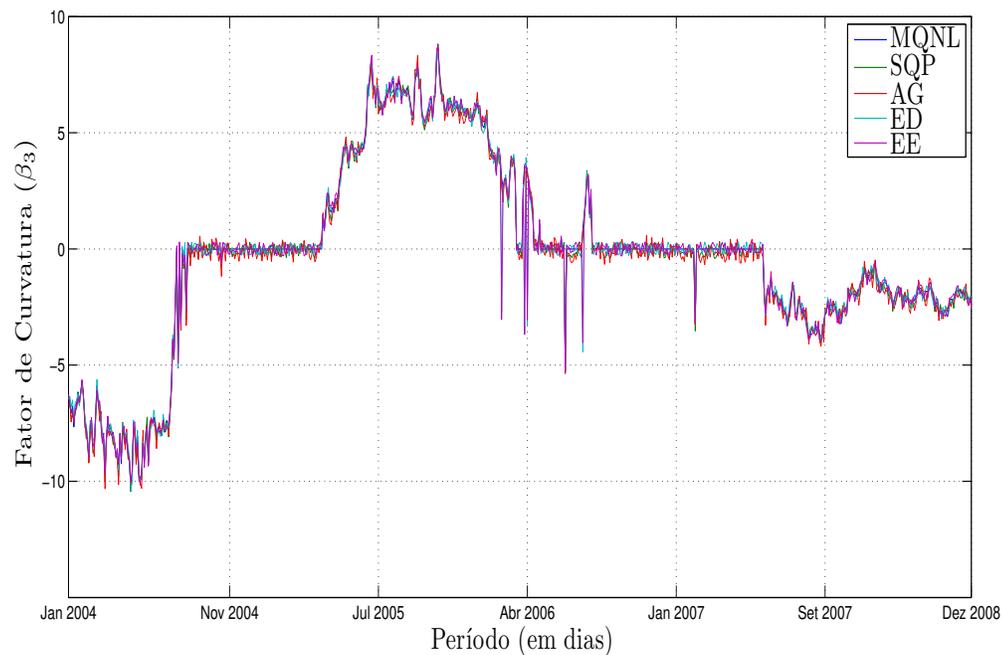


Figura 6.9: Evolução do fator de curvatura $\hat{\beta}_3$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.

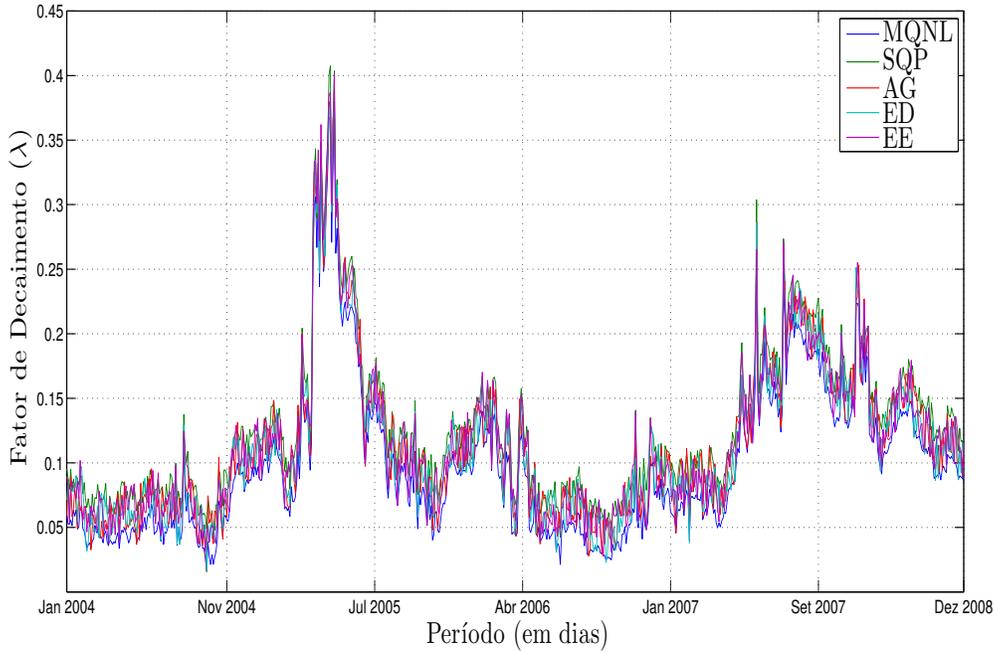


Figura 6.10: Evolução do fator de decaimento $\hat{\lambda}$ estimado de acordo com as técnicas de MQNL, SQP, AG, ED e EE, considerando o mercado de renda fixa do Brasil.

Em termos mais gerais, mesmo considerando amostras de períodos distintos, pode-se perceber que, em média, o mercado de renda fixa no Brasil apresenta níveis mais elevados de taxas de juros que no mercado dos EUA, o que era esperado, ao ser considerada uma economia emergente. Sobretudo, em ambos mercados foi evidente os efeitos da crise do subprime, ao verificar a queda nos fatores de nível (Figuras 6.3 e 6.7) e inclinação (Figuras 6.4 e 6.8) a partir do desencadeamento da crise, inicialmente no mercado dos EUA e, conseqüentemente, nas demais economias do mundo, como o Brasil. Os parâmetros β_1 e β_2 refletem, implícitamente, as expectativas dos agentes do mercado de renda fixa em relação às taxas de curto e longo prazos, respectivamente. Dessa forma, como uma característica da crise foi a queda nos níveis de juros, como tentativa de contenção dos efeitos negativos da crise nas principais economias afetadas, pode-se perceber que os algoritmos foram capazes de capturar esse comportamento das expectativas.

Para avaliar o desempenho dos algoritmos em termos de ajuste da curva parametrizada aos dados de mercado, considerou-se a medida de erro quadrático médio ponderada pelo inverso do quadrado da maturidade (WMSE - *weighted mean squared error*), definida como (Gimeno & Nave 2009):

$$WMSE = \sum_{\tau=1}^{120} \frac{1}{\tau^2} \cdot \left[\frac{1}{N} \sum_{j=1}^N (y_j(\tau) - \hat{y}_j(\tau))^2 \right] \quad (6.3)$$

em que $y(\tau)$ representa as taxas observadas, $\hat{y}(\tau)$ as taxas estimadas de acordo com a curva de NS, e $\tau = 1, 3, 6, 12, 24, 36, 60, 84, 120$ são as maturidades consideradas (em meses).

Na Tabela 6.2 são apresentados os ajustes de cada metodologia de otimização para os mercados do Brasil e dos Estados Unidos. Os resultados são promissores para as técnicas de computação evolucionária, pois apresentaram o melhor ajuste da curva em ambas economias avaliadas. No mercado dos Estados Unidos, o algoritmo de evolução diferencial implementado resultou em um melhor ajuste para a estrutura a termo, enquanto, para os dados do Brasil, os melhores resultados correspondem ao algoritmo de estratégias evolutivas CMA-ES. As técnicas tradicionais de otimização não-linear de parâmetros, MQNL e SQP, apresentaram os piores ajustes. Os níveis de erro no mercado brasileiro foram maiores devido à elevada instabilidade, em termos de intensa variabilidade na série temporal dos fatores latentes⁶. Além disso, as diferenças entre as metodologias tradicionais e evolucionárias é mais evidente no mercado dos EUA (Tabela 6.2).

Tabela 6.2: Erro quadrático médio ponderado (WMSE) para estimação da ETTJ nos mercados dos EUA e Brasil.

Técnica	EUA	Brasil
MQNL	0,0904	1,2076
SQP	0,0769	0,9962
AG	0,0603	0,5390
DE	0,0033	0,3268
EE	0,0061	0,2143

As Figuras 6.11 e 6.12 apresentam os resíduos para os mercados dos EUA e Brasil, obtidos pelas técnicas que proporcionaram o melhor ajuste em termos de WMSE, respectivamente Evolução Diferencial e Estratégias Evolutivas. O comportamento dos resíduos de estimação foi similar em ambos mercados, com erros em torno de zero. Entretanto, no mercado do Brasil os níveis de erros foram, em geral, maiores se comparados com os dos EUA, assim como é observado uma presença maior de erros mais elevados. Entretanto, ainda assim é verificado um ajuste adequado à ETTJ no mercado de renda fixa brasileiro (Figura 6.12).

Além de considerar o ajuste, os algoritmos foram avaliados em termos de complexidade computacional (tempo de processamento) e número de parâmetros de controle. A Tabela 6.3 apresenta o desempenho das metodologias implementadas, de forma que o tempo de processamento foi computado como uma média do tempo total para estimação de todas as curvas em relação ao número de dados em cada amostra⁷. As técnicas tradicionais de otimização (MQNL e SQP), em ambas economias, apresentaram o menor tempo de processamento. Entretanto, apesar de as técnicas evolucionárias levarem um tempo maior, ele ainda não é tão elevado, e não implica em limitações para utilização em ambientes que requerem maior dinamismo. Além disso, dentre os algoritmos evolucionários, destaca-se a abordagem de estratégias evolutivas, pois apresentou rapidez na estimação dos parâmetros da curva de juros, e, ao mesmo tempo, requer

⁶Deve-se notar, adicionalmente, que os resultados refletem a periodicidade das amostras, uma vez que dados diários são mais voláteis que dados mensais.

⁷Os algoritmos foram implementados em MatLab[®] em um laptop equipado com 4GB e processador Intel[®]i3CPU.

um menor número de parâmetros de controle (apenas 3), o que resulta em um modelo mais autônomo, em que a necessidade do conhecimento sobre o problema é reduzida.

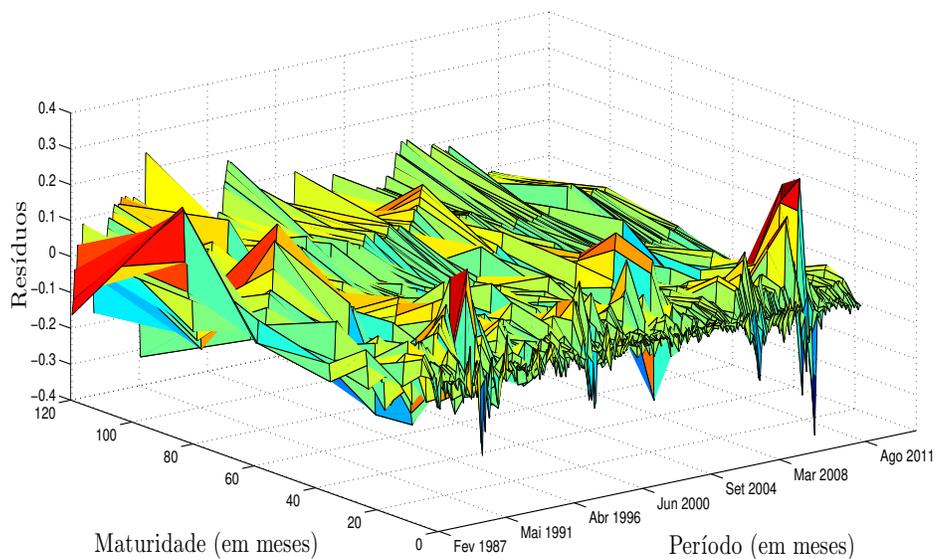


Figura 6.11: Resíduos da estimação da ETTJ no mercado dos EUA, utilizando o algoritmo de Evolução Diferencial.

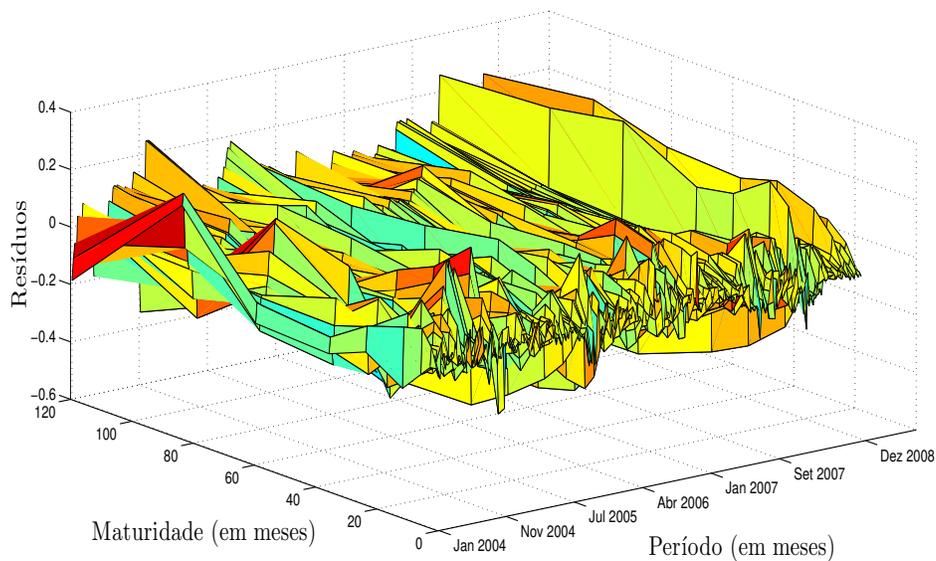


Figura 6.12: Resíduos da estimação da ETTJ no mercado do Brasil, utilizando o algoritmo de Estratégias Evolutivas CMA-ES.

Tabela 6.3: Complexidade computacional e número de parâmetros dos algoritmos utilizados para estimar a ETTJ nos mercados de renda fixa dos EUA e Brasil.

Estados Unidos / Brasil		
Técnica	Tempo de Processamento (segundos)	No. de Parâmetros de Controle
MQNL	45,7236 / 39,2830	3 / 3
SQP	40,0218 / 36,1947	3 / 3
AG	296,8722 / 221,0013	6 / 6
ED	135,9827 / 87,5574	5 / 5
EE	77,1956 / 51,6798	3 / 3

Para elucidar os resultados dos algoritmos, foram selecionadas algumas curvas da estrutura a termo do mercado dos EUA e do Brasil obtidas por cada metodologia implementada, em amostras em que as diferenças entre as estimativas são mais evidentes. A Figura 6.13 apresenta as curvas de juros estimadas pelas técnicas de otimização e de computação evolucionária consideradas para o mercado de renda fixa dos Estados Unidos em Maio de 1995. Pode-se notar que as curvas geradas apresentam comportamento distintos, de forma que essas diferenças podem gerar implicações inadequadas de acordo com a utilização da ETTJ. No Brasil, as curvas de juros estimadas no dia 19 de Setembro de 2007 são apresentadas na Figura 6.14. Conforme já discutido anteriormente, para o mercado brasileiro, as similaridades entre os resultados foram maiores que no mercado norte-americano, entretanto, as diferenças ainda são visíveis. Além disso, a Figura 6.14 já indica expectativas de quedas nas taxas de juros no longo prazo, refletindo os efeitos da crise do *Subprime*.

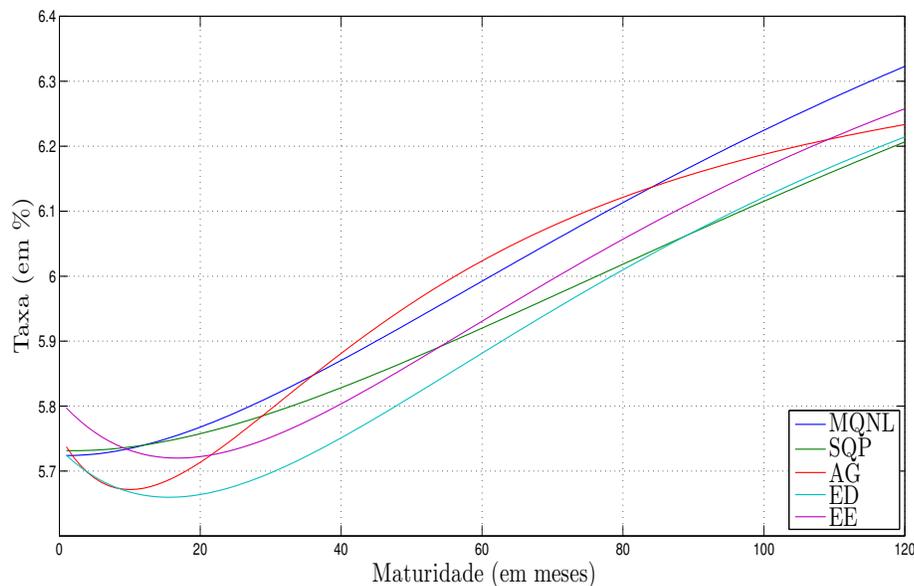


Figura 6.13: Estimação da ETTJ para o mercado dos EUA em Maio de 1995 de acordo com as técnicas de otimização de MQNL, SQP, ED e EE.

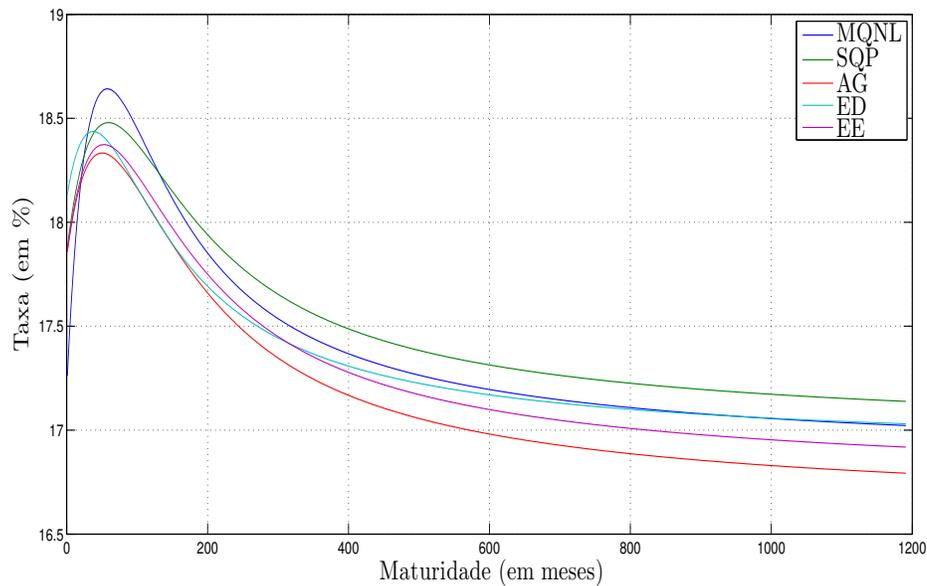


Figura 6.14: Estimação da ETTJ para o mercado do Brasil em 19 de Setembro de 2007 de acordo com as técnicas de otimização de MQNL, SQP, ED e EE.

Em uma avaliação empírica, as estimativas da estrutura a termo das taxas de juros foram aplicadas à precificação de títulos públicos zero-cupom, um dos objetivos principais na obtenção da curva de juros. Os dados do mercado de renda fixa dos Estados Unidos fornecem conjuntamente com as taxas, os preços indicativos dos papéis de mesma maturidade, que servirão como base na comparação dos resultados. Para o caso do Brasil, como obtivemos a estrutura a termo a partir das taxas de swaps PRE×DI, coletaram-se os preços dos títulos das Letras do Tesouro Nacional (LTNs) para o mesmo período considerado, incluindo todas as maturidades disponíveis, para comparar com as estimativas obtidas por meio das curvas de juros modeladas⁸. As LTNs são títulos públicos com rentabilidade definida (taxa pré-fixada) no momento da compra. O pagamento é único e feito na data de vencimento do título ou de seu resgate, sendo um título público zero-cupom. Os preços teóricos (obtidos pelos modelos) foram comparados em relação aos preços dos títulos observados no mercado utilizando a métrica de raiz do erro quadrático médio (REQM):

$$REQM = \left(\frac{1}{N} \sum_{i=1}^N \left(P_i(\tau) - \hat{P}_i(\tau) \right)^2 \right)^{\frac{1}{2}} \quad (6.4)$$

em que $P_i(\tau)$ representa o preço de mercado do título i com maturidade τ , $\hat{P}_i(\tau)$ seu preço teórico correspondente, obtido pela utilização da taxa livre de risco estimada, e N o número de títulos considerados.

A Tabela 6.4 apresenta os erros REQM para as técnicas de otimização de parâmetros para a precificação de títulos públicos nos mercados dos EUA e Brasil. Em termos da aplicação

⁸Os dados foram obtidos junto a ANBIMA.

da estrutura a termo das taxas de juros na precificação de títulos públicos, os resultados se situam entre os de DE e EE, sendo estes significativamente melhores que as técnicas MQNL e SQP, tanto para os dados dos EUA como Brasil. Por outro lado, no caso dos dados dos Estados Unidos, as diferenças dos erros entre os algoritmos de evolução diferencial e estratégias evolutivas são mais similares, enquanto, no caso do Brasil, o método de estratégias evolutivas demonstrou melhor desempenho.

Tabela 6.4: REQM na precificação de títulos públicos zero-cupom nos mercados dos EUA e Brasil.

Técnica	EUA	Brasil
MQNL	18,9827	27,7162
SQP	16,0229	24,1328
AG	12,1622	19,4430
DE	7,0019	15,3829
EE	7,6211	12,5367

Os resultados apontados na Tabela 6.4 evidenciam a importância da obtenção de técnicas mais acuradas para a geração da curva de juros, uma vez que, de acordo com o objetivo de estimação, pequenas diferenças entre as estimativas dos parâmetros da estrutura a termo podem gerar significativos e distintos resultados na aplicação.

Determinada a questão de modelagem da curva de juros, a seção seguinte reporta os resultados referentes à predição da estrutura a termo considerando a abordagem de sistemas de infêrencia nebulosos evolutivos. Os modelos nebulosos serão avaliados em contraposição com abordagens tradicionais verificadas na literatura, como modelos autoregressivos e de redes neurais artificiais.

6.3 Previsão das Taxas de Juros

Os experimentos computacionais para a previsão da estrutura a termo das taxas de juros nos mercados de renda fixa dos Estados Unidos e do Brasil serão abordados nesta seção. O método de dois estágios de (Diebold & Li 2006) será considerado como *benchmark*, pois é a metodologia amplamente utilizada na literatura. Alternativamente, consideraram-se os métodos nebulosos evolutivos eTS, xTS, eTS+, ePL e ePL+, assim como uma estrutura de rede neural multi-camadas, para previsão dos movimentos futuros das taxas de juros. Os resultados serão avaliados em termos de acurácia e testes estatísticos para modelos competitivos.

6.3.1 Metodologia

Para a previsão da curva de rendimentos, as amostras foram divididas em dois subconjuntos: treinamento e validação. Esse procedimento é necessário para estimar os parâmetros dos processos autoregressivos da metodologia de DL e para o treinamento da rede neural multi-camadas (MLP), implementados neste trabalho. Deve ser destacado que essa divisão em sub-amostras

não é necessária nos modelos nebulosos evolutivos, uma vez que podem iniciar o processo recursivo com apenas um dado e então desenvolver sua estrutura de acordo com a disponibilidade de informação.

Para os dados dos mercados de renda fixa dos Estados Unidos e do Brasil, a amostra de treinamento corresponde a 60% do total dos dados, enquanto os 40% restantes compõem o conjunto de validação. A Tabela 6.5 apresenta os períodos referentes às amostras de treinamento e previsão para cada uma das economias consideradas. Na amostra de validação, para ambas economias, é incluído o período da crise atual do *Subprime*.

Tabela 6.5: Períodos das amostras de treinamento e validação para os dados dos mercados de juros dos Estados Unidos e Brasil.

Mercado	Treinamento	Validação
EUA	Jan/1985 a Jul/2000	Ago/2000 a Ago/2011
Brasil	1/Jan/2004 a 29/Dez/2006	1/Jan/2007 a 31/Dez/2008

Todas os métodos avaliados na previsão da estrutura a termo das taxas de juros foram construídos de acordo com a técnica de dois estágios, sugerida por (Diebold & Li 2006). No primeiro estágio, as séries temporais dos parâmetros serão obtidas de acordo com o algoritmo de mínimos quadrados não-lineares para todo o período.

No modelo de DL, o segundo estágio consiste essencialmente na estimação dos parâmetros de um processo autoregressivo $AR(1)$ para cada fator latente da curva de NS $(\beta_1, \beta_2, \beta_3, \lambda)$. Esse procedimento utiliza a amostra de treinamento, de forma que, com os parâmetros dos processos $AR(1)$ obtidos, as previsões são avaliadas no conjunto de dados de validação. Uma das limitações dessa metodologia é a necessidade de tratar distintamente a modelagem de cada fator latente, sendo que eles, por possuírem uma interpretação econômica que diretamente os relaciona, podiam ser considerados em apenas um sistema de estimação conjunta. Além disso, uma amostra de treinamento representativa do comportamento da curva de juros torna-se fundamental nessa abordagem.

Distintamente, os modelos nebulosos evolutivos implementados são considerados em suas versões de múltiplas saídas (MIMO), conforme é apresentado na Figura 6.15. Como os parâmetros da curva de juros tem interpretação econômica de taxas de juros de curto, médio e longo prazos, essas taxas apresentam uma relação de dependência temporal entre si, uma vez que a formação das taxas de maturidades mais longas são baseadas nas taxas correntes de curto prazo. Dessa forma, essa estrutura em múltiplas saídas permite uma maior interação entre as variáveis modeladas pelo sistema. Portanto, os modelos nebulosos evolutivos realizaram a previsão dos fatores latentes da estrutura a termo conjuntamente, tornando a aplicação mais dinâmica e computacionalmente eficiente.

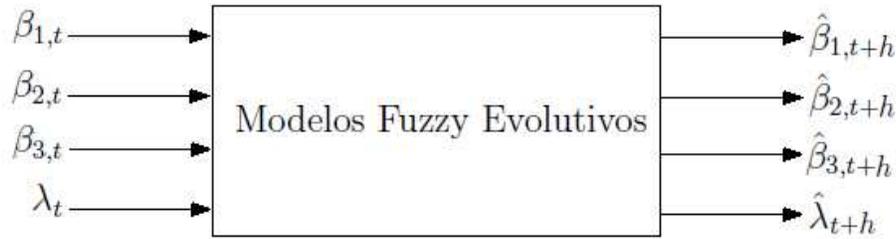


Figura 6.15: Representação esquemática dos modelos nebulosos evolutivos para previsão dos fatores latentes da curva de juros.

Os modelos nebulosos evolutivos consideraram a amostra de treinamento para a inicialização da estrutura de base de regras e parâmetros dos consequentes. As entradas são as primeiras defasagens dos fatores da curva de NS, enquanto as saídas correspondem às previsões h passos à frente dos parâmetros (ver Figura 6.15). Embora a inclusão de mais defasagens aos modelos evolutivos tenha sido considerada, foi verificado que a utilização de apenas uma defasagem é suficiente para gerar resultados satisfatórios.

O modelo de rede neural adotado, uma rede de múltiplas camadas, apresenta as mesmas entradas e saídas dos modelos nebulosos evolutivos, compondo também um sistema de múltiplas saídas. Para o treinamento da rede neural foi usado o método do gradiente, de acordo com a técnica de retropropagação do erro (*backpropagation*). A topologia da MLP foi determinada de forma construtiva, i.e., avaliaram-se os resultados de uma rede com apenas uma camada intermediária com um neurônio oculto, e, posteriormente, novas camadas e neurônios foram inseridos, computando-se o desempenho. Quando a incrementação da complexidade do modelo não mais promoveu ganhos de acurácia, a topologia foi estabelecida. Para os dados de juros dos Estados Unidos, a rede neural implementada foi composta por duas camadas intermediárias com três neurônios em cada uma. Para os dados do mercado de renda fixa do Brasil, a rede neural apresentou uma topologia com apenas uma camada intermediária com cinco neurônios ocultos.

As funções de ativação dos neurônios ocultos foram tangentes hiperbólicas, enquanto a camada de saída foi composta por neurônios lineares. A inicialização dos pesos e bias foi selecionada de forma aleatória com valores em torno de 10^{-2} , os dados de entrada foram normalizados no intervalo $[-1, 1]$ ⁹, e a taxa de aprendizagem foi determinada pelo processo de secção áurea¹⁰. Por fim, estabeleceu-se como critério de parada o processo de validação cruzada, um número máximo de iterações e um valor mínimo para a função de erro. O critério de validação cruzada foi estabelecido da seguinte forma: a cada k iterações durante o treinamento, avaliou-se o erro na amostra de treinamento e validação; quando o erro na amostra de validação for maior que o

⁹Os dados foram normalizados de acordo com a seguinte expressão:

$$\bar{x} = 2 \cdot \left(\frac{x - x_{min}}{x_{max} - x_{min}} \right) - 1$$

em que \bar{x} é o valor normalizado da entrada/saída x , enquanto x_{min} e x_{max} são os valores mínimo e máximo da entrada/saída, respectivamente.

¹⁰Secção áurea é um método de busca unidimensional baseado no conceito de razão áurea. Para maiores detalhes ver (Teixeira, Braga, Sadanha, Takahashi & Medeiros 2007).

erro na mesma amostra em k iterações passadas, o treinamento é interrompido e o conjunto de pesos anteriores obtidos são definidos como o resultado do treinamento. Esse processo evita a ocorrência do problema de *overfitting*, amplamente tratado na literatura de *machine learning* (Haykin 2009).

Descrita a metodologia adotada em cada técnica para a predição da estrutura a termo das taxas de juros, a seguir são apresentadas as comparações entre os resultados obtidos.

6.3.2 Resultados

Os modelos nebulosos evolutivos (eTS, xTS, eTS+, ePL e ePL+) e a rede neural MLP foram comparados com a técnica de (Diebold & Li 2006) para a previsão da estrutura a termo das taxas de juros, com base nos dados dos mercados de renda fixa dos Estados Unidos e Brasil. Para os modelos nebulosos, a Tabela 6.6 apresenta os respectivos parâmetros de controle. A determinação desses parâmetros foi realizada por meio de simulações. Entretanto, foi verificada uma baixa sensibilidade dos resultados em relação à parametrização adotada, mantendo-os iguais para os dados dos Estados Unidos e Brasil.

Tabela 6.6: Parâmetros de controle dos modelos nebulosos evolutivos para a predição da curva de juros nos mercados dos Estados Unidos e Brasil.

Parâmetro	eTS	xTS	eTS+	ePL	ePL+
σ	0,05	-	0,04	0,04	-
Ω	1000	1000	1000	1000	1000
ϵ	-	-	0,1	-	0,1
η	-	-	-	0,16	0,18
ϑ	-	-	-	0,84	0,82
β	-	-	-	0,16	0,18
α	-	-	-	0,01	0,01

A avaliação do desempenho das metodologias adotadas foi considerada tanto na amostra de treinamento quanto na amostra de validação. Inicialmente, avaliou-se o ajuste de cada técnica em modelar e capturar o movimento das curvas de juros, considerando a amostra de treinamento. As Tabelas 6.7 e 6.8 reportam o ajuste dos modelos para previsão dos fatores latentes da curva de (Nelson & Siegel 1987) na amostra de treinamento em termos da métrica de REQM, nas economias dos Estados Unidos e Brasil, respectivamente. Os melhores resultados estão indicados em negrito.

Os resultados de ajuste dos modelos, nas amostras de treinamento, revelam a elevada capacidade dos modelos nebulosos evolutivos em capturar a dinâmica dos fatores latentes da curva de (Nelson & Siegel 1987). As técnicas de inteligência computacional apresentaram os mais baixos níveis de erros considerando todos parâmetros, superando a metodologia de DL em ambas as economias (Tabelas 6.7 e 6.8). Para os fatores de nível (β_1), inclinação (β_2) e curvatura (β_3), os modelos eTS+ e ePL+ diferem significativamente dos resultados obtidos pelo modelo de DL, proporcionando, em média, redução de 100% nos níveis de erros nos mercados dos Estados

Tabela 6.7: REQM para previsão dos fatores latentes da curva de juros na amostra de treinamento nos dados do mercado dos Estados Unidos.

Fatores	DL	MLP	eTS	xTS	eTS+	ePL	ePL+
$\hat{\beta}_1$	0.134	0.028	0.014	0.013	0.011	0.014	0.011
$\hat{\beta}_2$	0.461	0.134	0.127	0.125	0.112	0.124	0.099
$\hat{\beta}_3$	1.244	0.307	0.243	0.297	0.206	0.242	0.212
$\hat{\lambda}$	0.014	0.015	0.019	0.013	0.010	0.017	0.014

Tabela 6.8: REQM para previsão dos fatores latentes da curva de juros na amostra de treinamento nos dados do mercado do Brasil.

Fatores	DL	MLP	eTS	xTS	eTS+	ePL	ePL+
$\hat{\beta}_1$	0.182	0.102	0.029	0.058	0.026	0.044	0.036
$\hat{\beta}_2$	0.191	0.095	0.035	0.065	0.034	0.048	0.029
$\hat{\beta}_3$	0.466	0.287	0.222	0.251	0.217	0.243	0.219
$\hat{\lambda}$	0.011	0.009	0.008	0.005	0.007	0.009	0.006

Unidos e Brasil. Quando o fator de decaimento (λ) é considerado, as diferenças são menores, mas ainda são significativas. O modelo ePL+ evidenciou um melhor ajuste quando comparado com sua versão original, ePL, em especial no mercado de renda fixa do Brasil. Além disso, a rede multi-camadas apresentou resultados similares aos modelos nebulosos.

Avaliou-se também a acurácia dos modelos testados em relação às taxas de juros para distintas maturidades. Dessa forma, com os parâmetros estimados, as curvas de juros zero-cupom associadas foram obtidas. As taxas estimadas foram comparadas com as taxas reais, presentes nas amostras de dados. As Tabelas 6.9 e 6.10 apresentam, respectivamente, os valores do erro quadrático médio nas taxas para a amostra de treinamento com base nos dados das taxas zero-cupom dos mercados de renda fixa dos Estados Unidos e Brasil.

Tabela 6.9: EQM das taxas estimadas nas amostras de treinamento para os dados dos Estados Unidos.

Modelos	Maturidade (em meses)					
	1	3	6	12	60	120
DL	0.202	0.164	0.183	0.275	0.265	0.154
MLP	0.129	0.095	0.106	0.125	0.209	0.121
eTS	0.126	0.081	0.092	0.143	0.211	0.106
xTS	0.112	0.092	0.102	0.153	0.197	0.118
eTS+	0.108	0.079	0.097	0.131	0.188	0.100
ePL	0.121	0.094	0.101	0.129	0.201	0.096
ePL+	0.111	0.076	0.099	0.120	0.190	0.103

Quando os modelos de previsão são avaliados em relação às taxas estimadas, as diferenças entre os métodos de inteligência computacional e a técnica de DL ficam mais evidentes (Tabelas 6.9 e 6.10), sobretudo nas taxas de maturidades de curto prazo. Os modelos nebulosos evolutivos eTS+ e ePL+, em geral, resultaram em níveis mais baixos de erro. Quando maturidades mais longas são consideradas, a similaridade dos resultados das técnicas avaliadas aumenta, o que era esperado, uma vez que a literatura confirma a elevada capacidade do modelo de (Diebold & Li 2006) nas previsões das taxas com prazos mais longos. Entretanto, mesmo nas maturidades de longo prazo, os modelos eTS+ e ePL+ apresentam uma redução de mais de 40% da métrica de EQM, em média, quando os resultados do método de DL são considerados. A rede neural multi-camadas também apresentou resultados superiores ao *benchmark*, mas inferiores em relação aos modelos nebulosos evolutivos.

Tabela 6.10: EQM das taxas estimadas nas amostras de treinamento para os dados do Brasil.

Modelos	Maturidade (em meses)					
	1	3	6	12	60	120
DL	0.106	0.214	0.124	0.095	0.136	0.264
MLP	0.055	0.142	0.060	0.071	0.129	0.215
eTS	0.053	0.113	0.065	0.067	0.111	0.192
xTS	0.047	0.121	0.055	0.059	0.116	0.184
eTS+	0.044	0.123	0.057	0.062	0.101	0.179
ePL	0.050	0.128	0.061	0.069	0.124	0.191
ePL+	0.039	0.119	0.050	0.063	0.117	0.183

Com base nos dados do mercado dos Estados Unidos, a proporção na redução dos níveis de erros obtidos por meio dos modelos de inteligência computacional em relação à técnica de DL foram, em média, maiores quando essa mesma relação é considerada no mercado brasileiro. Os melhores ganhos de desempenho nos dados do mercado de renda fixa dos Estados Unidos estão associados à periodicidade mensal dos dados, pois esta apresenta uma dinâmica menos volátil em comparação com os dados diários do Brasil. Sobretudo, independentemente da base de dados, os modelos evolutivos foram significativamente superiores ao modelo de DL quanto ao ajuste da curva de rendimentos.

Entretanto, a análise fora da amostra assume o papel mais importante na previsão da estrutura a termo das taxas de juros. As Tabelas 6.11 e 6.12, respectivamente, apresentam os resultados de previsão na amostra de validação para horizontes de 1, 3 e 6 meses à frente para os mercado de renda fixa dos Estados Unidos e Brasil.

Considerando o horizonte de previsão de 1 mês (curto prazo), as diferenças entre as abordagens consideradas são mais evidentes, com destaque para os modelos nebulosos evolutivos (os valores em negrito indicam os menores níveis de erro quadrático médio), que apresentaram os menores índices de erro. Os modelos nebulosos e de rede neural apresentaram resultados superiores que o modelo de (Diebold & Li 2006) no curto prazo, com uma significativa redução dos níveis de erro (Tabelas 6.11 e 6.12). Adicionalmente, a rede multi-camadas ainda apresentou resultados, em geral, inferiores aos dos modelos nebulosos, porém, as diferenças entre tais

técnicas é menos acentuada quando a amostra de validação é considerada.

O modelo ePL+, para o horizonte de 1 mês à frente, apresentou resultados superiores à sua versão original, ePL, principalmente nas maturidades mais curtas, em ambas economias. Além disso, nesse mesmo horizonte de previsão, o modelo eTS foi capaz de reduzir os níveis de erro significativamente, quando comparado com o modelo de DL.

Tabela 6.11: EQM das previsões da curva de juros no mercado dos Estados Unidos.

Modelos	Maturidade (em meses)					
	1	3	6	12	60	120
1 mês à frente						
DL	0.189	0.133	0.117	0.123	0.110	0.094
MLP	0.151	0.108	0.111	0.050	0.083	0.035
eTS	0.160	0.126	0.108	0.035	0.059	0.032
xTS	0.142	0.102	0.105	0.044	0.103	0.033
eTS+	0.134	0.096	0.100	0.039	0.066	0.027
ePL	0.148	0.113	0.113	0.041	0.101	0.039
ePL+	0.136	0.089	0.097	0.044	0.074	0.020
3 meses à frente						
DL	0.190	0.135	0.120	0.206	0.120	0.103
MLP	0.177	0.130	0.111	0.190	0.088	0.114
eTS	0.156	0.125	0.103	0.201	0.027	0.094
xTS	0.162	0.122	0.104	0.162	0.078	0.100
eTS+	0.147	0.120	0.093	0.164	0.030	0.091
ePL	0.169	0.129	0.107	0.181	0.064	0.107
ePL+	0.131	0.119	0.090	0.166	0.036	0.096
6 meses à frente						
DL	0.224	0.171	0.182	0.213	0.144	0.136
MLP	0.206	0.183	0.190	0.207	0.169	0.155
eTS	0.187	0.149	0.171	0.209	0.165	0.147
xTS	0.191	0.187	0.183	0.192	0.154	0.145
eTS+	0.183	0.152	0.174	0.184	0.146	0.139
ePL	0.194	0.168	0.186	0.197	0.163	0.151
ePL+	0.188	0.155	0.175	0.188	0.148	0.138

No horizonte de médio prazo (3 meses à frente), para os dados dos Estados Unidos (Tabela 6.11), os resultados foram similares, de tal forma que as abordagens nebulosas apresentaram os menores valores para a métrica de EQM. Nesse caso, o modelo ePL+ obteve, em média, os melhores resultados. Similarmente, no caso dos dados do Brasil (Tabela 6.12), o modelo eTS+ resultou em menores níveis de erro. As diferenças entre as técnicas de DL e de rede neural artificial foram menores para horizontes de médio prazo. Entretanto, o modelo de DL foi superado pelas técnicas de inteligência computacional na previsão da ETTJ para 3 meses à frente. Os ganhos de desempenho do modelo ePL+ em relação ao ePL, para o horizonte de médio prazo, apresentaram-se mais significativos nos dados do mercado de renda fixa dos

Estados Unidos.

Tabela 6.12: EQM das predições da curva de juros no mercado do Brasil.

Modelos	Maturidade (em meses)					
	1	3	6	12	60	120
1 mês à frente						
DL	0.074	0.195	0.111	0.102	0.112	0.285
MLP	0.061	0.166	0.113	0.090	0.107	0.182
eTS	0.066	0.147	0.098	0.067	0.103	0.145
xTS	0.059	0.165	0.104	0.082	0.089	0.166
eTS+	0.034	0.157	0.100	0.079	0.073	0.172
ePL	0.049	0.161	0.109	0.084	0.091	0.181
ePL+	0.042	0.149	0.093	0.075	0.084	0.159
3 meses à frente						
DL	0.096	0.211	0.119	0.112	0.136	0.289
MLP	0.086	0.207	0.098	0.111	0.127	0.202
eTS	0.074	0.195	0.087	0.099	0.122	0.165
xTS	0.080	0.203	0.084	0.106	0.119	0.192
eTS+	0.077	0.183	0.080	0.097	0.114	0.184
ePL	0.081	0.201	0.092	0.103	0.120	0.197
ePL+	0.079	0.189	0.083	0.095	0.123	0.189
6 meses à frente						
DL	0.126	0.234	0.135	0.133	0.154	0.187
MLP	0.113	0.192	0.119	0.124	0.139	0.211
eTS	0.101	0.165	0.100	0.120	0.114	0.193
xTS	0.098	0.187	0.116	0.119	0.136	0.201
eTS+	0.086	0.159	0.107	0.116	0.109	0.190
ePL	0.101	0.177	0.116	0.121	0.134	0.205
ePL+	0.093	0.151	0.104	0.113	0.121	0.199

Conforme o horizonte de predição aumenta, as diferenças entre as metodologias diminuem. Em horizontes de longo prazo, 6 meses à frente, os resultados obtidos com os modelos de DL, eTS+ e ePL+ são muito similares. Esse é o único caso em que os resultados do modelo de DL se aproximam dos obtidos pelas técnicas nebulosas, uma vez que tal metodologia econométrica apresenta elevada acurácia de predição da estrutura a termo em horizontes de longo prazo.

Para o mercado de renda fixa dos Estados Unidos, os modelos nebulosos e de redes neurais superaram a técnica de DL em todas as maturidades nos horizontes de predição de 1 e 3 meses (ver Tabela 6.11). Para previsões de 6 meses à frente, o modelo autoregressivo de (Diebold & Li 2006) e os modelos eTS+ e ePL+ geraram resultados muito similares, em termos de erro quadrático médio, principalmente nas maturidades de 60 e 120 meses. Dentre os modelos nebulosos eTS, xTS, ePL e a rede neural multi-camadas, os resultados no longo prazo foram similares. Além disso, o modelo ePL+ ainda apresentou ganhos significativos de desempenho em relação à sua versão original.

Em horizontes de longo prazo, para o mercado brasileiro, os resultados indicaram a superioridade das técnicas de inteligência computacional, principalmente nas maturidades de 1, 3, 6, 12 e 60 meses (Tabela 6.12). Quando a maturidade de 120 meses é considerada, os modelos de DL, eTS e eTS+ apresentaram resultados muito similares, dificultando a identificação da superioridade de uma metodologia em relação à outra. Dessa forma, esse fato evidencia a necessidade da avaliação mais detalhada da diferença entre os modelos. Para tanto, neste trabalho foi realizada a comparação dos modelos em termos de significância estatística. Foram aplicados dois testes de hipóteses: um paramétrico e outro não paramétrico.

O teste paramétrico, denominado de teste de Morgan-Granger-Newbold (teste *MGN*), baseado inicialmente no trabalho de (Granger & Newbold 1977), é recomendado para avaliar modelos de predição competitivos, sendo amplamente aplicado quando a assunção de não existência de autocorrelação entre os resíduos é relaxada. O valor da estatística *MGN* é:

$$MGN = \frac{\hat{\rho}_{s,d}}{(1 - \hat{\rho}_{s,d}^2)^{\frac{1}{2}}} (N - 1)^{\frac{1}{2}} \quad (6.5)$$

em que $\hat{\rho}_{sd}$ é o coeficiente de correlação estimado entre $s = e_1 + e_2$ e $d = e_1 - e_2$, tal que e_1 e e_2 correspondem aos erros de distintos modelos 1 e 2, respectivamente¹¹, e N representa o número de dados na amostra. A estatística para o teste *MGN* tem distribuição *t-Student* com $(N - 1)$ graus de liberdade. Para este teste, se as previsões são igualmente acuradas, entre os modelos competitivos, então a correlação entre s e d será zero (hipótese nula).

Por fim, o teste não-paramétrico *SIGN* (*Significance Test*) (Lehmann 1988), é uma avaliação não-paramétrica que não requer suposições como normalidade e não-correlação serial entre os erros das técnicas consideradas. A estatística *SIGN* é por:

$$SIGN = \frac{(\psi(N) - \frac{N}{2})}{2(N)^{\frac{1}{2}}} \quad (6.6)$$

em que o termo ψ_t denota o número de vezes em que os erros de um determinado modelo superam os erros derivados de outro modelo competitivo.

A hipótese nula é $H_0 : \psi(N) = (\frac{N}{2})$, contra a hipótese alternativa de que $H_1 : \psi(N) > (\frac{N}{2})$. Assume-se que N é o número de observações previstas por ambos os modelos e, se os erros de previsão de uma dada abordagem são da mesma proporção que os obtidos por outra em exatamente $(\frac{N}{2})$ observações, de acordo com o teste *SIGN*, a acurácia dos modelos pode ser considerada equivalente, sem diferença estatisticamente significativa. A estatística para o teste *SIGN* tem distribuição Gaussiana padrão.

As Tabelas 6.13 e 6.14 reportam os resultados dos testes estatísticos *MGN* e *SIGN* para as comparações entre as técnicas de previsão, avaliadas para a estrutura a termo das taxas de juros nos mercados dos Estados Unidos e Brasil, respectivamente, considerando a amostra de validação.

Os resultados mostram que, em termos estatísticos, as técnicas de inteligência computacional consideradas neste trabalho para a predição da curva de juros nas economias avaliadas são superiores à técnica de DL. Os valores em negrito indicam a significância estatística dos

¹¹Nesse caso, o erro é definido pela diferença entre as taxas observadas nos mercados e as obtidas pelos modelos.

modelos em níveis de 1, 5 e 10%, i.e., aceita-se a hipótese nula de ambos os testes que indica a superioridade dos modelos alternativos (nebulosos evolutivos e MLP) em relação ao *benchmark* (método de DL), em termos de significância estatística. Para o mercado de juros dos Estados Unidos (Tabela 6.13), as estatísticas *MGN* e *SIGN* foram elevadas para as comparações entre as técnicas de inteligência computacional e a econométrica. Entretanto, quando os modelos nebulosos e de rede neural são avaliados entre si, as estatísticas *MGN* e *SIGN* indicam que as abordagens podem ser consideradas competitivas, não apresentando diferenças significativas, em termos estatísticos, na previsão da ETTJ no mercado dos Estados Unidos.

Tabela 6.13: Testes *MGN* e *SIGN* para a previsão da ETTJ no mercado dos Estados Unidos.

Modelos	Teste	eTS	xTS	eTS+	ePL	ePL+	MLP
DL	<i>MGN</i>	8,2716	9,2101	7,3213	8,2209	7,3027	9,1273
	<i>SIGN</i>	-5,1092	-4,9825	-6,1171	7,4462	-8,0021	6,0233
eTS	<i>MGN</i>	-	-0,8790	0,9559	-0,0998	-0,0222	0,9822
	<i>SIGN</i>	-	-1,3726	-1,7861	-0,4791	-1,0267	-1,2861
xTS	<i>MGN</i>	-	-	0,9811	-0,7635	-1,0019	-0,9899
	<i>SIGN</i>	-	-	-0,8725	-1,2928	0,8726	-0,9992
eTS+	<i>MGN</i>	-	-	-	-0,6608	-0,7631	-0,7005
	<i>SIGN</i>	-	-	-	-1,9375	-0,4635	-0,9385
ePL	<i>MGN</i>	-	-	-	-	-0,5005	0,4442
	<i>SIGN</i>	-	-	-	-	1,9826	1,2981
ePL+	<i>MGN</i>	-	-	-	-	-	-0,9811
	<i>SIGN</i>	-	-	-	-	-	0,4873

Os valores em negrito não indicam significância estatística a 1, 5 e 10%, i.e., aceita-se a hipótese nula.

Quando o mercado de renda fixa do Brasil é considerado, o teste estatístico paramétrico, *MGN*, indicou que os modelos nebulosos evolutivos e de rede neural multi-camadas podem ser considerados melhores preditores das taxas de juros futuras que o modelo de DL (Tabela 6.14), evidenciado pelos elevados valores dessas estatísticas. O teste não-paramétrico, *SIGN*, indicou que os modelos xTS, eTS+ e ePL+ são superiores, em termos de acurácia de previsão, que o modelo econométrico de DL. Entretanto, segundo a Tabela 6.14, o teste *SIGN* nos permite inferir que os modelos de DL, eTS, ePL e de redes neurais multi-camadas podem ser considerados modelos competitivos de previsão, uma vez que a estatística obtida nesses casos não permite a rejeição da hipótese nula de competitividade entre as técnicas consideradas. Por outro lado, as estatísticas avaliadas indicam que os modelos nebulosos evolutivos e de rede neural multi-camadas podem ser considerados competitivos, em termos de acurácia.

Tabela 6.14: Testes *MGN* e *SIGN* para a previsão da ETTJ no mercado do Brasil.

Modelos	Teste	eTS	xTS	eTS+	ePL	ePL+	MLP
DL	<i>MGN</i>	7,3837	6,9826	7,0029	7,5244	8,4039	8,0153
	<i>SIGN</i>	-1,0991	-5,3189	-7,5532	0,6541	-7,9934	0,1182
eTS	<i>MGN</i>	-	-0,7141	0,8763	0,1007	-0,1752	0,1148
	<i>SIGN</i>	-	1,1337	-0,6450	0,2243	-1,0185	-0,7762
xTS	<i>MGN</i>	-	-	-0,7654	0,4735	-0,4600	-1,8871
	<i>SIGN</i>	-	-	-0,1281	0,0192	-0,2761	0,2229
eTS+	<i>MGN</i>	-	-	-	1,0766	-0,1118	-0,2019
	<i>SIGN</i>	-	-	-	-1,2615	0,1273	-0,0013
ePL	<i>MGN</i>	-	-	-	-	0,0911	-0,2531
	<i>SIGN</i>	-	-	-	-	0,0251	-0,8735
ePL+	<i>MGN</i>	-	-	-	-	-	1,0092
	<i>SIGN</i>	-	-	-	-	-	-0,7624

Os valores em negrito não indicam significância estatística a 1, 5 e 10%, i.e., aceita-se a hipótese nula.

Alternativamente, os modelos foram avaliados em termos de tempo de processamento (complexidade computacional) e número de regras ou neurônios para a predição da estrutura a termo das taxas de juros, conforme apresentado na Tabela 6.15.

Tabela 6.15: Complexidade computacional e número de parâmetros dos modelos utilizados para predição da ETTJ nos mercados de renda fixa dos EUA e Brasil.

Técnica	Estados Unidos / Brasil	
	Tempo de Processamento (segundos)	No. de Regras ou Neurônios
DL	14,2634 / 26,9386	- / -
eTS	1,7465 / 3,0034	11 / 12
xTS	1,5273 / 2,9981	7 / 9
eTS+	2,0982 / 3,1129	5 / 6
ePL	1,6663 / 2,8195	6 / 7
ePL+	1,9900 / 3,0912	4 / 6
MLP	78,8736 / 164,8733	5 / 6

O tempo de processamento nos modelos nebulosos evolutivos foi muito similar, mas em contrapartida, a técnica de DL e, em especial, a MLP, apresentaram uma maior complexidade computacional. No caso do modelo de DL, esses resultados expressam a necessidade de estimação

individual de cada fator latente da curva de juros. Por outro lado, os resultados da MLP refletiram a limitação do tempo elevado de treinamento presente em modelos de redes neurais.

Em relação ao número de regras, a Tabela 6.15 mostra que as técnicas eTS+ e ePL+ geraram uma estrutura com o menor número de regras nebulosas, em comparação com os demais modelos nebulosos evolutivos, uma vez que tais métodos incluem mecanismos que avaliam a qualidade da estrutura da base de regras, mantendo uma estrutura mais compacta.

Nesta dissertação, o modelo ePL+ foi proposto como uma versão estendida da versão inicial do ePL. Os resultados reportados até então, na questão da previsão da estrutura a termo das taxas de juros, evidenciaram os ganhos de desempenho do modelo ePL+, que forneceu estimativas mais acuradas para os mercados de juros dos Estados Unidos e Brasil. Além disso, os modelos ePL e ePL+ foram comparados em termos da evolução do número de regras ao longo do período avaliado na aplicação proposta. As Figuras 6.16 e 6.17 apresentam, respectivamente, a variação do número de regras dos modelos ePL e ePL+ no problema de predição ETTJ.

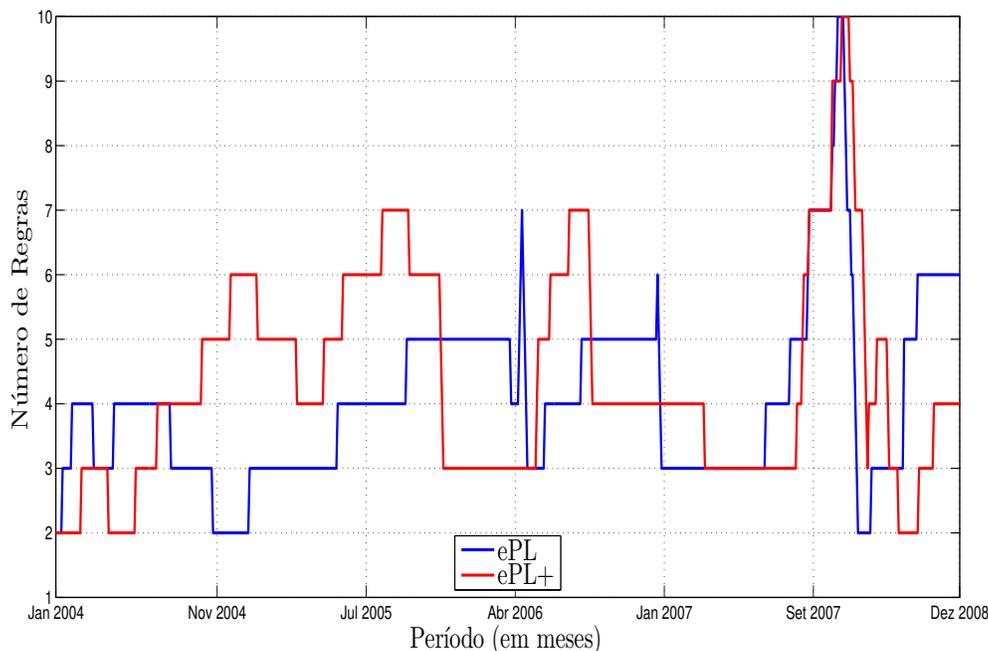


Figura 6.16: Evolução do número de regras nos modelos ePL e ePL+ para a previsão da ETTJ com dados do mercado dos Estados Unidos.

Em ambas economias, o modelo ePL+ gerou um menor número de regras quando comparado com o ePL para a predição da curva de juros, o que garante uma elevada capacidade de constituir uma estrutura de base de regras mais compacta e com maior capacidade de sumarização do sistema. As Figuras 6.16 e 6.17 atentam ao fato de que o modelo ePL+ é caracterizado por um maior dinamismo, de forma que a base de regras varia mais acentuadamente com a disponibilidade de informações na tentativa de sumarizar tal conhecimento com um número baixo de regras, o que é explicado pela utilização da medida de utilidade para avaliar a qualidade da base de regras. Além disso, considerando o período de início e desencadeamento da crise do

mercado. Entre as técnicas de computação evolucionária, os algoritmos de evolução diferencial e estratégias evolutivas (CMA-ES) tiveram melhor desempenho, com erros de previsão menores, baixo custo computacional. Essas duas abordagens, ainda, demandam um menor número de parâmetros que o usuário necessita fornecer.

Por outro lado, nesta seção também foram descritos os resultados das técnicas de inteligência computacional para a questão de previsão das taxas de juros, finalizando este capítulo, constituído pelos resultados gerais da dissertação. Na questão de predição, os resultados elucidaram o potencial dos modelos nebulosos evolutivos para o problema da estrutura a termo das taxas de juros. Além disso, o modelo proposto neste trabalho, ePL+, apresentou resultados superiores à sua versão original, ePL.

O próximo capítulo, por fim, conclui a dissertação resumizando suas principais contribuições, sugerindo perspectivas para futuras investigações e apresentando os artigos produzidos diretamente relacionados com este trabalho.

Conclusões e Trabalhos Futuros

A estimação e previsão da estrutura a termo das taxas de juros assume um papel importante para formuladores de política monetária, analistas, investidores e participantes do mercado financeiro. A curva de juros serve como base para precificação de outros instrumentos de renda fixa, análise e gestão de riscos, avaliação de contratos derivativos, e também reflete movimentos futuros dos níveis de inflação e atividade econômica. Nesta dissertação os problemas de estimação e previsão da curva de juros foram abordados com base em técnicas de inteligência computacional tais como computação evolucionária e modelagem nebulosa evolutiva, respectivamente. Os mercados de renda fixa dos Estados Unidos e do Brasil foram considerados para a análise empírica.

A modelagem da estrutura a termo das taxas de juros em modelos parcimoniosos de fatores latentes se resume basicamente em um problema de otimização não-linear de parâmetros, uma vez que não é verificado nos mercados taxas livres de riscos para todas as possíveis maturidades, o que exige um mecanismo de interpolação. Neste trabalho, o desempenho de algoritmos de computação evolucionária (Algoritmos Genéticos, de Evolução Diferencial e de Estratégias Evolutivas) foi comparado com técnicas tradicionais de programação não-linear como mínimos quadrados não-lineares e programação quadrática sequencial na estimação da curva de juros.

As técnicas de evolução diferencial e estratégias evolutivas superaram, em termos de acurácia, as demais metodologias avaliadas. Alternativamente, quando as estimativas das taxas livres de riscos foram aplicadas na precificação de títulos públicos, o melhor desempenho das técnicas evolucionárias ficou mais evidente. A utilização dos algoritmos baseados em computação natural supera determinadas limitações em abordagens tradicionais de otimização não-linear de parâmetros, uma vez que minimiza o risco de não-convergência e não gera oscilações significativas nas estimativas dos parâmetros em relação à inicialização adotada. Além disso, os resultados foram semelhantes em ambos mercados de renda fixa avaliados, sendo os melhores resultados associados à economia dos Estados Unidos.

Por outro lado, o problema de predição da curva de juros foi abordado distintamente com a utilização de modelos nebulosos evolutivos (eTS, xTS, eTS+, ePL) e uma rede neural do tipo multi-camadas, em comparação com a tradicional modelagem verificada na literatura, i.e., via processos autoregressivos. Além dessa aplicação, esta dissertação sugeriu uma extensão ao modelo ePL, denominada ePL+. O modelo ePL+, assim como o ePL, utiliza o conceito de agru-

pamento participativo em modelos nebulosos evolutivos como um algoritmo não-supervisionado de agrupamento. Em adição ao ePL, a modelagem ePL+ inclui: adaptação recursiva da zona de influência das regras nebulosas, inserção de uma medida de utilidade para monitoramento da qualidade da estrutura da base de regras, extensão para o caso de sistemas com múltiplas saídas, e utilização do algoritmo *subtractive clustering* para inicialização dos antecedentes das regras nebulosas, quando um conjunto de treinamento está disponível. Essas modificações garantem mais autonomia ao modelo, na tentativa de torná-lo totalmente recursivo e livre da exigência de conhecimento prévio do usuário sobre o sistema para uma modelagem adequada.

Na previsão da estrutura a termo das taxas de juros, os modelos nebulosos evolutivos e de rede neural apresentaram um desempenho superior à abordagem autoregressiva, em termos de erro quadrático e testes estatísticos. A rede neural, em termos de significância estatística, apresentou resultados competitivos aos das abordagens evolutivas. Entretanto, métodos baseados em redes neurais são do tipo “caixa-preta” e exigem um processo de treinamento em batelada, o que implica em custos computacionais elevados quando os modelos nebulosos são levados em consideração, uma vez que esses últimos são capazes de operar em tempo real, sendo algoritmos não-iterativos e de treinamento recursivo. A evolução da base de regras nos modelos nebulosos evolutivos evidenciou a capacidade de atualização de suas estruturas de acordo com as mudanças nos padrões dos dados, de forma a capturar os movimentos nos mercados de renda fixa das economias avaliadas. Além disso, tanto no mercado dos Estados Unidos e do Brasil, as técnicas de inteligência computacional tiveram destaque nas previsões de curto prazo. Por fim, na análise de predição da estrutura a termo das taxas de juros, os resultados revelaram uma maior acurácia do modelo ePL+ proposto em relação à sua versão inicial, ePL, e resultados superiores ou similares aos outros modelos como DL, MLP, eTS, xTS e eTS+.

Trabalhos Futuros

Extensões para futuras investigações incluem:

- Comparação entre distintas abordagens parcimoniosas de fatores latentes.
- Estimação da estrutura a termo das taxas de juros por meio de outras meta-heurísticas de otimização como, por exemplo: Enxame de Partículas, Recozimento Simulado, Sistemas Imunológicos Artificiais, Otimização por Colônia de Formigas ou Técnicas Baseadas em Cardumes.
- Modelagem e estimação conjunta da curva de juros com modelos de espaço-estado variantes no tempo.
- Avaliação de modelos nebulosos evolutivos em ambientes dinâmicos de decisões como análise de riscos.
- Aplicação de técnicas de redes neurais com Estados de Eco na previsão da curva de juros, uma vez que tais modelos requerem estágio de treinamento apenas na camada de saída.

- Incorporação de seleção recursiva de variáveis de entrada no modelo ePL+.
- Descrição dos antecedentes das regras no modelo ePL+ com funções de pertinência Gaussianas multivariadas.
- Inclusão de variáveis macroeconômicas para previsão da ETTJ.
- Comparação de modelos de equilíbrio e de não-arbitragem para estimação e previsão da curva de juros.
- Extração e avaliações de relações macroeconômicas em países desenvolvidos e emergentes por meio de sistemas nebulosos evolutivos.
- Construção de funções de reação de bancos centrais considerando relações nebulosas baseadas em estimativas da ETTJ e em indicadores macroeconômicos.

Publicações

Revistas

MACIEL, L. S.; LEMOS, A. P.; GOMIDE, F.; BALLINI, R. Evolving Fuzzy Systems For Pricing Fixed Income Options. *Evolving Systems*, v. 3, p. 5–18, 2012.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Evolving Functional Fuzzy Model for Interest Rate Term Structure Forecasting. *Learning and Nonlinear Models*. Artigo aceito para publicação.

Congressos Internacionais

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Modeling the Term Structure of Government Bond Yields with a Differential Evolution Algorithm. Proceedings of the IEEE Computational Intelligence for Financial Engineering & Economics (CIFEr 2012), New York, NY, v. 1, p. 1-8, 2012. Trabalho recebeu o prêmio de *Honorable Mention* na categoria *Best Paper Award*.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. MIMO Evolving Functional Fuzzy Models for Interest Rate Forecasting. Proceedings of the IEEE Computational Intelligence for Financial Engineering & Economics (CIFEr 2012), New York, USA, v. 1, p. 1–8, 2012.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. An Enhanced Approach for Evolving Participatory Learning Fuzzy Modeling. Proceedings of the 2012 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS 2012), Madrid, Spain, v. 1, p. 1–6, 2012.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Evolving Fuzzy Modeling for Stock Market Forecasting. Proceedings of the 14th International Conference on Information Processing and

Management of Uncertainty in Knowledge-Based Systems (IPMU 2012), Catania, Italy, v.1, p. 1–8, 2012.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. MIMO Evolving Participatory Learning Fuzzy Modeling. Proceedings of the 2012 IEEE World Congress on Computational Intelligence (WCCI 2012), Brisbane, Australia, v. 1, p. 1–8, 2012.

LEMOS, A. P.; LEITE, D. F.; MACIEL, L. S.; BALLINI, R.; CAMINHAS, W.; GOMIDE, F. Evolving Fuzzy Linear Regression Tree Approach for Forecasting Sales Volume of Petroleum Products. Proceedings of the 2012 IEEE World Congress on Computational Intelligence (WCCI 2012), Brisbane, Australia, v. 1, p. 1–8, 2012.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Evolving Fuzzy Systems For Pricing Fixed Income Options. Proceedings of the 2011 IEEE Symposium Series on Computational Intelligence (SSCI 2011), Paris, France, v. 1, p. 54–61, 2011.

Congressos Nacionais

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Estimaco dos Parâmetros da Estrutura a Termo das Taxas de Juros Utilizando um Algoritmo de Evoluo Diferencial. Anais do XXXIX Encontro Nacional de Economia da Anpec, Foz do Iguau, Paran, 2011.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Algoritmo Gentico Para Estimar Parâmetros da Estrutura a Termo de Taxa de Juros. Anais do X Simpsio Brasileiro de Automao Inteligente (SBAI), So Joo Del Rei, Minas Gerais, 2011.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Evolving Functional Fuzzy Model for Interest Rate Term Structure Forecasting. Anais do X Congresso Brasileiro de Inteligncia Computacional, Fortaleza, Cear, 2011.

Submisses

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. A Differential Evolution Algorithm for Yield Curve Estimation. *Expert Systems with Applications*.

MACIEL, L. S.; GOMIDE, F.; BALLINI, R. Evolving Fuzzy Modeling for Yield Curve Forecasting. *Algorithmic Finance*.

MACIEL, L. S.; BALLINI, R.; GOMIDE, F. An Evolving Fuzzy-GARCH Approach for Financial Volatility Modeling and Forecasting. *European Journal of Operational Research*.

Bibliografia

- Almeida, C. & Vicente, J. (2008). The role of no-arbitrage on forecasting: Lessons from a parametric term structure model, *Journal of Banking and Finance* **32**(12): 2695–2705.
- Anagnostopoulos, K. P. & Mamanis, G. (2011). The meanvariance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms, *Expert Systems with Applications* **38**(11): 14208–14217.
- Anderson, N., Breedon, F., Deacon, M., Derry, A. & Murphy, G. (1996). *Estimating and interpreting the yield curve*, Wiley, New York.
- Anderson, N. & Sleath, J. (2001). New estimates of the UK real and nominal yield curves, *Technical Report 126*, Bank of England Working Papers, London.
- Ang, A. & Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* **50**(4): 74–787.
- Ang, A., Piazzesi, M. & Wei, M. (2006). What does the yield curve tell us about gdp growth?, *Journal of Econometrics* **131**(1-2): 359–403.
- Angelov, P. (2002). *Evolving Rule-based Models: A Tool for Design of Flexible Adaptive Systems*, New York, NY: Springer-Verlag.
- Angelov, P. & Filev, D. (2004). An approach to on-line identification of evolving Takagi-Sugeno fuzzy models, *IEEE Transactions on Systems, Man and Cybernetics, part B* **34**(1): 484–498.
- Angelov, P. & Filev, D. (2005). SimPLeTS: A simplified method for learning evolving takagi-sugeno fuzzy models, *Proceedings of the 2005 International Conference on Fuzzy Systems, FUZZ-IEEE* **1**: 1068–1073.
- Angelov, P., Filev, D. P. & Kasabov, N. (2010). *Evolving Systems: Methodology and Applications*, Hoboken, NJ: John Wiley & Sons Inc.
- Angelov, P. & Zhou, X. (2006). Evolving fuzzy systems from data streams in real-time, *International Symposium on Evolving Fuzzy Systems*, IEEE Press, Ambleside, Lake District, UK, pp. 29–35.

- Angelov, P. & Zhou, X. (2008). On-line learning fuzzy rule-based system classifiers from data streams, *IEEE Transactions on Fuzzy Systems* **16**(6): 1462–1475.
- Ballini, R., Mendonça, A. R. R. & Gomide, F. (2009a). Evolving fuzzy modeling in risk analysis, *Intelligent Systems in Accounting, Finance and Management* **16**: 71–86.
- Ballini, R., Mendonça, A. R. R. & Gomide, F. (2009b). Evolving fuzzy modeling of sovereign bonds, *Journal of Financial Decision Making* **16**: 3–15.
- Baruah, R. D. & Angelov, P. (2011). Evolving fuzzy systems for data streams: A survey, *Data Mining and Knowledge Discovery* **1**(6): 461–476.
- Bernanke, B. & Woodford, M. (2006). *The inflation-targeting debate*, National Bureau of Economic Research, University of Chicago Press.
- Beyer, H. & Schwefel, H. (2002). Evolution strategies: A comprehensive introduction, *Natural Computing* **1**(1): 3–52.
- Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Norwell, MA: Kluwer Academic Publishers.
- Bezdek, J., Tsao, E. & Pal, N. (1992). Fuzzy Kohonen clustering networks, *IEEE International Conference on Fuzzy Systems*.
- Bhattacharya, A. & Chattopadhyay, P. K. (2011). Hybrid differential evolution with biogeography-based optimization algorithm for solution of economic emission load dispatch problems, *Expert Systems with Applications* **38**(11): 14001–14010.
- Bidarkota, P. V. (1998). The comparative forecast performance of univariate and multivariate models: An application to real interest rate forecasting, *International Journal of Forecasting* **14**(4): 457–468.
- Binner, J. M., Kendall, G. & Gazely, A. (2004). Co-evolving neural networks with evolutionary strategies: A new application to divisia money, *Applications of Artificial Intelligence in Finance and Economics* **19**: 127–143.
- BIS (2005). Zero-coupon yield curves: Technical documentation, *Technical Report 25*, Bank for International Settlements Papers, Basel.
- Boggs, P. T. & Tolle, J. W. (1995). Sequential quadratic programming, *Acta Numerica* **4**: 1–51.
- Bolder, D. J. & Strélsk, D. (1999). Yield curve modelling at the bank of Canada, *Technical Report 84*, Bank of Canada Technical Report.
- Bowsher, C. G. & Meeks, R. (2008). The dynamics of economic functions: Modeling and forecasting the yield curve, *Journal of the American Statistical Association* **103**(484): 1419–1437.

- Brennan, M. J. & Schwartz, E. J. (1979). A continuous time approach to the pricing of bonds, *Journal of Banking and Finance* **3**(2): 133–155.
- Cajueiro, D. O., Divino, J. A. & Tabak, B. M. (2009). Forecasting the yield curve for Brazil, *Technical Report 197*, Central Bank of Brazil Working Paper Series.
- Caldeira, J. F., Moura, G. V. & Portugal, M. S. (2010). Efficient yield curve estimation and forecasting in Brazil, *Economia* **11**(1): 27–51.
- Castillo Tapia, M. & Coello, C. (2007). Applications of multi-objective evolutionary algorithms in economics and finance: A survey, *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, pp. 532–539.
- Chakraborty, U. K. (2008). *Advances in Differential Evolution*, Springer, Heidelberg, New York.
- Chakravarty, S. & Dash, P. K. (2011). Dynamic filter weights neural network model integrated with differential evolution for day-ahead price forecasting in energy market, *Expert Systems with Applications* **38**(9): 10974–10982.
- Chambers, D. R., Carleton, W. T. & Waldman, D. W. (1984). A new approach to the estimation of the term structure of interest rates, *Journal of Financial and Quantitative Analysis* **19**(3): 233–252.
- Chan, K. G., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992). An empirical comparison of alternative models of term structure of interest rates, *Journal of Finance* **47**(3): 1209–1227.
- Chang, T. J., Yang, S. C. & Chang, K. J. (2009). Portfolio optimization problems in different risk measures using genetic algorithm, *Expert Systems with Applications* **36**(7): 10529–10537.
- Chauhan, N., Ravi, V. & Chandra, D. K. (2009). Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks, *Expert Systems with Applications* **36**(4): 7659–7665.
- Chen, S. H. & Huang, Y. C. (2007). Relative risk aversion and wealth dynamics, *Information Sciences* **177**(5): 1222–1229.
- Chen, S. H. & Kuo, T. W. (2002). *Evolutionary Computation in Economics and Finance*, Heidelberg: Physica-Verlag.
- Chen, Y., Mabu, S. & Hirasawa, K. (2011). Genetic relation algorithm with guided mutation for the large-scale portfolio optimization, *Expert Systems with Applications* **38**(4): 3353–3363.
- Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation, *Journal of Intelligent Fuzzy Systems* **2**: 267–278.
- Courant, E. & Hilbert, D. (1953). *Methods of mathematical physics*, John Wiley, New York.

- Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica* **53**(2): 385–408.
- Csajbok, A. (1998). Zero-coupon yield curve estimation from a central bank perspective, *Technical Report 2*, National Bank of Hungary Working Paper.
- Dai, Q. & Singleton, K. (2002). Expectation puzzles, time-varying risk premia and affine models of the term structure, *Journal of Financial Economics* **63**(3): 415–441.
- Das, S., Abraham, A., Chakraborty, U. K. & Konar, A. (2009). Differential evolution using a neighborhood based mutation operator, *IEEE Transactions on Evolutionary Computation* **13**(3): 526–553.
- Das, S. & Suganthan, P. N. (2011). Differential evolution: A survey of the state-of-the-art, *IEEE Transactions on Evolutionary Computation* **15**(1): 4–31.
- Diamant, P. (1993). Semi-empirical smooth fit to the treasury yield curve, *The Journal of Fixed Income* **3**(1): 55–70.
- Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* **74**(266): 427–431.
- Diebold, F. X. & Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**(2): 337–364.
- Diebold, F. X., Piazzesi, M. & Rudebusch, G. D. (2005). Modeling bond yields in finance and macroeconomics, *American Economic Review* **95**(2): 415–420.
- Diebold, F. X., Rudebusch, G. D. & Arouba, B. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach, *Journal of Econometrics* **131**(1-2): 309–338.
- Dobson, S. W. (1978). Estimating term structure equations with individual bond data, *Journal of Finance* **33**(1): 75–92.
- Duarte, A., Venetis, I. A. & Paya, I. (2005). Predicting real growth and the probability of recession in the euro area using the yield spread, *International Journal of Forecasting* **131**(2): 261–277.
- Duffee, G. (2002). Term premia and interest rate forecasts in affine models, *Journal of Finance* **57**(1): 405–443.
- Ehrentreich, N. (2006). Technical trading in the santa fe institute artificial stock market revisited, *Journal of Economic Behavior and Organization* **61**(4): 599–616.
- Fama, E. F. & Bliss, R. R. (1987). The information in long-maturity forward rates, *The American Economic Review* **77**(4): 680–692.
- Fernández-Rodríguez, F. (2006). Interest rate term structure modelling using free knot splines, *The Journal of Business* **79**(6): 3083–3099.

- Fisher, M., Nychka, D. & Zervos, D. (1995). Fitting the term structure of interest rates with smoothing splines, *Technical Report 1995-1*, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System.
- Friedman, M. (1977). Time perspective in demand for money, *The Scandinavian Journal of Economics* **79**(4): 397–416.
- Gilli, M. & Winker, P. (2008). A review of heuristic optimization methods in econometrics, *Swiss Institute Research Paper Series* **1**(8): 8–12.
- Gimeno, R. & Nave, J. M. (2009). A genetic algorithm estimation of the term structure of interest rates, *Computational Statistics and Data Analysis* **53**(6): 2236–2250.
- Goldberg, D. E. (1991). Real-coded genetic algorithms, virtual alphabets, and blocking, *Complex Systems* **5**: 139–167.
- Golmohammadi, A. & Pajoutan, M. (2011). Meta heuristics for dependent portfolio selection problem considering risk, *Expert Systems with Applications* **38**(5): 5642–5649.
- Gorgulho, A., Neves, R. & Horta, N. (2011). Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition, *Expert Systems with Applications* **38**(11): 14072–14085.
- Granger, J. W. C. & Newbold, P. (1977). *Forecasting economic time series*, New York: Academic Press.
- Greer, M. (2003). Directional accuracy tests of long-term interest rate forecasts, *International Journal of Forecasting* **19**(2): 291–298.
- Gürkaynak, R. S., Sack, B. & Wright, J. H. (2006). The U.S. Treasury yield curve: 1961 to the present, *Journal of Monetary Economics* **54**(8): 2291–2304.
- Guthmann, H. G. (1929). *The relation of the maturity factor to yield curve differences among investment securities*, Tese de Doutorado, Universidade de Chicago.
- Hansen, N., Müller, S. D. & Koumoutsakos, P. (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es), *Evolutionary Computation* **11**(1): 1–18.
- Hansen, N. & Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies, *Evolutionary Computation* **9**(2): 159–195.
- Hautsch, N. & Yang, F. (2010). Bayesian inference in a stochastic volatility Nelson-Siegel model, *Computational Statistics & Data Analysis* **In Press, Corrected Proof**: 1–19.
- Haykin, S. (2009). *Neural Networks and Learning Machines*, New Jersey: Pearson Prentice Hall.
- Heath, D., Jarrow, R. & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation, *Econometrica* **60**(1): 77–105.

- Holland, J. H. (1975). *Adaptation in natural and artificial systems*, University of Michigan Press.
- Hördahl, P., Tristani, O. & Vestin, D. (2006). A joint econometric model of macroeconomic and term structure dynamics, *Journal of Econometrics* **131**(1-2): 405–440.
- Ioannides, M. (2003). A comparison of yield curve estimation techniques using UK data, *Journal of Banking and Finance* **27**(1): 1–26.
- Kasabov, N. (1996). *Foundations of Neural Networks, Fuzzy Systems, and Knowledge-Engineering*, Cambridge, MA: MIT Press.
- Kasabov, N. & Song, Q. (2002). DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time series prediction, *IEEE Transactions on Fuzzy Systems* **10**(2): 144–154.
- Kaymak, U. & Setnes, M. (2000). Extended fuzzy clustering algorithms, *Technical report*, Erasmus Research Institute of Management (ERIM).
- Kim, K. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting, *Expert Systems with Applications* **30**(3): 519–526.
- Koopman, S. J., Mallee, M. I. & Wel, M. (2010). Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters, *Journal of Business & Economic Statistics* **28**(3): 329–343.
- Korczak, J. & Lipinski, P. (2004). Evolutionary building of stock trading experts in a real-time system, *Proceedings of the 2004 Congress on Evolutionary Computation (CEC 2004)*, USA, pp. 940–947.
- Korczak, J., Lipinski, P. & Roger, P. (2002). Evolution strategy in portfolio optimization, in P. Collet, C. Fonlupt, J.-K. Hao, E. Lutton & M. Schoenauer (eds), *Artificial Evolution*, Vol. 2310 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 173–193.
- Korczak, J. & Roger, P. (2001). Stock timing using genetic algorithms, *Applied Stochastic Models in Business and Industry* **18**(2): 121–134.
- Krink, T. & Paterlini, S. (2011). Multiobjective optimization using differential evolution for real-world portfolio optimization, *Computational Management Science* **8**(1): 157–179.
- Kuo, C.-S., Hong, T.-P. & Chen, C.-L. (2008). A knowledge-evolution strategy based on genetic programming, *Convergence and Hybrid Information Technology, 2008. ICHIT '08. International Conference on*, pp. 43–48.
- Laurí, D., Salcedo, J. V., García-Nieto, S. & Martínez, M. (2010). Model predictive control relevante identification: Multiple input multiple output against multiple input single output, *IET Control Theory and Applications* **4**(9): 1756–1766.

- Laurini, M. P. & Hotta, L. K. (2010). Bayesian extensions to Diebold-Li term structure model, *International Review of Financial Analysis* **19**(5): 342–350.
- Lehmann, E. L. (1988). *Nonparametric statistical methods based on ranks*, New Jersey: Prentice Hall.
- Lemos, A., Caminhas, W. & Gomide, F. (2011). Multivariable gaussian evolving fuzzy modeling system, *IEEE Transactions on Fuzzy Systems* **19**(1): 91–104.
- Leng, G., McGinnity, T. M. & Prasad, G. (2005). An approach for online extraction of fuzzy rules using a self-organizing fuzzy neural network, *Fuzzy Sets Systems* **150**(2): 211–243.
- Lima, E. (2008). *Modelagem Fuzzy Funcional Evolutiva Participativa*, Dissertação de Mestrado, Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas.
- Lima, E. A., Ludovice, F. & Tabak, B. M. (2006). Forecasting interest rates: An application for Brazi, *Technical Report 120*, Working Paper Series of Banco Central do Brasil.
- Lin, C. C. & Liu, Y. T. (2008). Genetic algorithms for portfolio selection problems with minimum transaction lots, *European Journal of Operational Research* **185**(1): 393–404.
- Lin, D., Wang, S. & Yan, H. (2001). A multiobjective genetic algorithm for portfolio selection problem, *Proceedings of ICOTA*, Hong Kong.
- Lin, P. & Ko, P. (2009). Portfolio value-at-risk forecasting with GA-based extreme value theory, *Expert Systems with Applications* **36**(2): 2503–2512.
- Linton, O., Mammen, E., Nielson, J. & Tanggaard, C. (2001). Estimating yield curves by kernel smoothing methods, *Journal of Econometrics* **105**(1): 185–223.
- Lipinski, P., Winczura, K. & Wojcik, J. (2007). Building risk-optimal portfolio using evolutionary strategies, in M. Giacobini (ed.), *Applications of Evolutionary Computing*, Vol. 4448 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 208–217.
- Litterman, R. & Scheinkman, J. (1991). Common factors affecting bond returns, *Journal of Fixed Income* **1**(1): 54–61.
- Ljung, L. (1999). *System Identification*, Englewood Cliffs, NJ: Prentice-Hall.
- Lu, Y., Zhou, J., Qin, H., Li, Y. & Zhang, Y. (2010). An adaptive hybrid differential evolution algorithm for dynamic economic dispatch with valve-point effects, *Expert Systems with Applications* **37**(7): 4842–4849.
- Lughofer, E. D. (2008). FLEXFIS: A robust incremental learning approach for evolving takagi-sugeno fuzzy models, *IEEE Transactions on Fuzzy Systems* **16**(6): 1393–1410.
- Maciel, L., Gomide, F. & Ballini, R. (2011). Evolving fuzzy systems for pricing fixed income options, *IEEE Workshop on Evolving and Adaptive Intelligent Systems EAIS, Paris, France* **1**: 54–61.

- Mamdani, E. H. & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller, *International Journal of Man-Machine Studies* **7**(1): 1–13.
- Manousopoulos, P. & Michalopoulos, M. (2009). Comparison of non-linear optimization algorithms for yield curve estimation, *European Journal of Operational Research* **192**(2): 594–602.
- Mansi, S. A. & Phillips, J. (2001). Modelling the term structure from the on-the-run treasury yield curve, *Journal of Financial Research* **24**(4): 545–564.
- Maringer, D. G. & Meyer, M. (2008). Smooth transition autoregressive models: New approaches to the model selection problem, *Studies in Nonlinear Dynamics & Econometrics* **12**(1): 1–19.
- Matsumura, M. S., Moreira, A. R. B. & Vicente, J. V. M. (2010). Forecasting the yield curve with linear factor models, *Technical report*, Central Bank of Brazil Working Paper Series, Tech. Rep. 223.
- McCulloch, J. H. (1971). Measuring the term structure of interest rates, *The Journal of Business* **44**(1): 19–31.
- McDonald, S. & Angelov, P. (2010). Evolving Takagi Sugeno modeling with memory for slow process, *International Journal of Knowledge-Based and Intelligence Engineering Systems* **14**(1): 11–19.
- McMillan, D. G. (2009). Non-linear interest rate dynamics and forecasting: evidence for US and Australian interest rates, *International Journal of Finance & Economics* **14**(2): 139–155.
- Mitchell, M. (1998). *An introduction to genetic algorithms*, The MIT Press, Cambridge, MA.
- Moench, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach, *Journal of Econometrics* **146**(1): 26–43.
- Nelson, C. R. & Siegel, A. F. (1987). Parsimonious modeling of yield curves, *Journal of Business* **60**(4): 473–489.
- Nocedal, J. & Wright, S. J. (2006). *Numerical Optimization*, 2nd edn, Berlin, New York: Springer-Verlag.
- Nyholm, K. & Rebonato, R. (2008). Long-horizon yield curve projections: Comparison of semi-parametric and parametric approaches, *Applied Financial Economics* **18**(20): 1579–1611.
- Oh, K. J., Kin, T. Y., Min, S. H. & Lee, H. Y. (2006). Portfolio algorithm based on portfolio beta using genetic algorithm, *Expert Systems with Applications* **30**(3): 527–534.
- Pedrycz, W. & Gomide, F. (2007). *Fuzzy Systems Engineering: Toward Human-Centric Computing*, Wiley - IEEE Press.

- Price, K., Storn, R. & Lampinen, J. (2005). *Differential Evolution - A practical approach to global optimization*, Springer, Berlin, Germany.
- Qin, A. K., Huang, V. L. & Suganthan, P. N. (2009). Differential evolution algorithm with strategy adaptation for global numerical optimization, *IEEE Transactions on Evolutionary Computation* **13**(2): 398–417.
- Rahnamayan, S., Tizhoosh, H. R. & Salama, M. M. A. (2008). Opposition-based differential evolution, *IEEE Transactions on Evolutionary Computation* **12**(1): 64–79.
- Rao, S. S. (1996). *Engineering Optimization – Theory and Practice*, third edn, John Wiley & Sons.
- Rezende, R. B. (2011). Giving flexibility to the Nelson-Siegel class of term structure models, *Revista Brasileira de Finanças* **9**(1): 27–49.
- Rong, H., Sundarajan, N., Huang, G. & Saratchandran, P. (2006). Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction, *Fuzzy Sets Systems* **157**(9): 1260–1275.
- Ronkkonen, J., Kukkonen, S. & Price, K. V. (2005). Real parameter optimization with differential evolution, *Proceedings of IEEE Congress on Evolutionary Computation*, Vol. 1, pp. 506–513.
- Rubio, J. D. J. (2009). SOFMLS: Online self-organizing fuzzy modified least-squares network, *IEEE Transactions on Fuzzy Systems* **17**(6): 1296–1309.
- Schlottmann, F. & Seese, D. (2004). A hybrid heuristic approach to discrete multi-objective optimization of credit portfolios, *Computational Statistics and Data Analysis* **47**(2): 373–379.
- Schwefel, H. P. (1981). *Numerical Optimization of Computer Models*, Wiley & Sons.
- Securato, J. R. (ed.) (2008). *Cálculo Financeiro das Tesourarias: Bancos e Empresas*, Saint Paul.
- Shea, G. (1984). Pitfalls in smoothing interest rate structure data: Equilibrium models and spline approximation, *Journal of Financial and Quantitative Analysis* **19**(3): 253–269.
- Shea, G. S. (1985). Interest rate term structure estimation with exponential splines: A note, *Journal of Finance* **11**(1): 319–325.
- Shin, K. & Lee, Y. (2002). A genetic algorithm application in bankruptcy prediction modeling, *Expert Systems with Applications* **23**(3): 321–328.
- Silva, L. (2003). *Aprendizado Participativo em Agrupamento Nebuloso de Dados*, Dissertação de Mestrado, Faculdade de Engenharia Elétrica e Computação, Universidade Estadual de Campinas.

- Silva, L., Gomide, F. & Yager, R. R. (2005). Participatory learning inf fuzzy clustering, *Proc. of the 2005 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.
- Soleimani, H., Golmakani, H. R. & Salimi, M. H. (2009). Markowitz-based portfolio selection with minimum transaction lots, cardinality constraints and regarding sector capitalization using genetic algorithm, *Expert Systems with Applications* **36**(3): 508–5063.
- Storn, R. & Price, K. (1997). Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization* **11**(4): 341–359.
- Svensson, L. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994, *Technical Report 4871*, NBER Working Paper Series.
- Takagi, T. & Sugeno, M. (1985). Fuzzy identification of systems and its applications in modeling and control, *IEEE Transactions on Systems, Man and Cybernetics* **15**(1): 116–132.
- Teixeira, R. A., Braga, A. P., Sadanha, R. R., Takahashi, R. H. C. & Medeiros, T. H. (2007). The usage of golden section in calculating the efficient solution in artificial neural networks training by multi-objective optimization, *Lectures Notes in Computer Science* **4668**: 289–298.
- Tollo, G. D. & Rolli, A. (2008). Metaheuristics for the portfolio selection problem, *International Journal of Operational Research* **5**(5): 13–35.
- Vasicek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**(2): 177–188.
- Vasicek, O. A. & Fong, H. G. (1982). Term structure modeling using exponential splines, *Journal of Finance* **37**(2): 339–348.
- Venetis, I. A., Paya, I. & Peel, D. A. (2003). Re-examination of the predictability of economic activity using the yield spread: A nonlinear approach, *International Review of Economics & Finance* **12**(2): 187–206.
- Vicente, J. & Tabak, B. M. (2007). Forecasting bond yields in the brazilian fixed income market, *Technical Report 141*, Working Paper Series of the Banco Central do Brasil.
- Yager, R. R. (1990). A model of participatory learning, *IEEE Transactions on Systems, Man and Cybernetics* **20**: 1229–1234.
- Yager, R. R. & Filev, D. P. (1994). Generation of fuzzy rules by mountain clustering, *Journal of Intelligent and Fuzzy Systems* **3**(3): 209–219.
- Yang, Z., Tian, Z. & Yuan, Z. (2007). Gsa-based maximum likelihood estimation for threshold vector error correction model, *Computational Statistics and Data Analysis* **52**(1): 109–120.
- Young, P. C. (1984). *Recursive Estimation and Time-Series Analysis: An Introduction*, New York, NY: Springer-Verlag.

-
- Yu, W. & Zivot, E. (2011). Forecasting the term structures of treasury and corporate yields using dynamic Nelson-Siegel models, *International Journal of Forecasting* **27**(2): 579–591.
- Yuan, X., Wang, L. & Huan, Y. (2009). A hybrid differential evolution method for dynamic economic dispatch with valve-point effects, *Expert Systems with Applications* **36**(2): 4042–4048.
- Zaharie, D. (2002). Critical values for the control parameters of differential evolution algorithms, *Proceedings of MENDEL 2002 - 8th International Conference on Soft Computing*, Brno, Czech Republic, pp. 62–67.
- Zhang, M. & Sanderson, A. C. (2003). JADE: Adaptive differential evolution with optional external archive, *IEEE Transactions on Evolutionary Computation* **13**(5): 945–958.

Conceitos da Estrutura das Taxas de Juros

Neste apêndice são apresentados os conceitos básicos associados a definição e construção da estrutura a termo das taxas de juros.

A.1 Juros, Taxa de Juros e Produtos do Mercado Financeiro

O conceito de juros pode ser definido como sendo a remuneração pelo uso do capital (um dos fatores de produção ou expressão monetária de um bem ou serviço). Em um dado intervalo de tempo, a medida de juros, \mathcal{J} , é obtida pela diferença entre o capital no final do intervalo, \mathcal{F} , e o capital no início do intervalo, \mathcal{P} , tal que:

$$\mathcal{J} = \mathcal{F} - \mathcal{P} \quad (\text{A.1})$$

Os juros são expressos como uma fração do capital inicialmente empregado numa dada unidade de tempo. Portanto, em uma operação financeira, a taxa de juros, y , num dado intervalo de tempo, pode ser entendida como a remuneração da unidade de capital inicial, ou seja:

$$y = \frac{\mathcal{J}}{\mathcal{P}} \quad (\text{A.2})$$

O intervalo de tempo, τ , associado às taxas de juros, indica a *maturidade* de tais taxas, ou seja, o período correspondente ao montante da remuneração do capital.

Nos mercados financeiros, são inúmeros os produtos relacionados às taxas de juros. No Brasil, por exemplo, temos o Certificado de Depósito Interbancário (CDI), títulos de emissão de instituições financeiras que servem de lastro para as operações interbancárias no mercado monetário; Certificado de Depósito Bancário (CDB), títulos emitidos por bancos e utilizados para a captação de recursos entre os investidores, posteriormente, repassados aos clientes nas operações de financiamento tradicionais do mercado de crédito; e Títulos Públicos, lançados nos mercados pela primeira vez por meio de leilões realizados pela Secretaria do Tesouro Nacional, entre as instituições interessadas (conhecido também como *mercado primário*)¹.

¹Os títulos públicos adquiridos no mercado primário não são carregados necessariamente pelas instituições

A.2 Estrutura a Termo das Taxas de Juros

Implicitamente, a taxa de juros mede o *custo do capital*. Como toda mercadoria, inclusive o capital, varia com o tempo de acordo com uma taxa de inflação, dinâmica, temos que na prática, a taxa de juros varia com o tempo. Portanto, a função que expressa a evolução da taxa de juros de acordo com a maturidade é definida como Estrutura a Termo das Taxas de Juros (ETTTJ). A ETTJ associa as taxas de juros y às suas respectivas maturidades τ :

$$y = f(\tau) \tag{A.3}$$

em que $f(\cdot)$ representa a relação função entre a taxa de juros e maturidade².

A ETTJ relaciona as taxas de juros livres de riscos e maturidade. A taxa de juros livre de risco atua como uma referência do custo de oportunidade do capital, caracterizada pelos fatores: taxa conhecida e fixada, maturidade associada e risco zero. Alternativamente, um ativo livre de risco é aquele em que o investidor sabe exatamente o valor que receberá ao final do prazo de investimento. Em geral, as taxas livres de riscos estão associadas aos títulos públicos zero-cupom, ou seja, títulos de rentabilidade prefixada, sem o pagamento de cupons intermediários. Por hipótese de insolvência financeira do emissor (governo) as taxas dos títulos zero-cupom são consideradas livres de riscos nos mercados financeiros.

Mesmo em mercados de renda fixa em que a liquidez de títulos públicos zero-cupom seja elevada, não é verificado um número suficientemente grande de taxas zero-cupom governamentais para todas as possíveis maturidades, o que implica a necessidade de uma técnica de interpolação para a construção da ETTJ. A Figura A.1 apresenta uma ilustração da curva de juros. A maturidade das taxas transacionadas no mercado são conhecidas como *vértices* da estrutura a termo. Portanto, com os vértices disponíveis, a ETTJ é interpolada e extrapolada para maturidades mais longas.

Em economias que apresentam uma baixa liquidez de títulos públicos transacionados no mercado monetário, a estrutura a termo das taxas de juros é construída a partir de taxas consideradas *proxy* para a taxa de juros livre de risco. No Brasil, em geral, tais taxas são as obtidas nos contratos futuros de DI e taxas dos *swaps* DI×PRÉ. Os Depósitos Interfinanceiros (DIs) são títulos de emissão de instituições financeiras que servem de lastro para as operações interbancárias no mercado monetário, de forma que as taxas desses depósitos, ou “taxa do DI”, é uma taxa muito utilizada como *proxy* para a taxa de juros livre de risco no Brasil. Entretanto, como essas taxas são divulgadas diariamente pela Câmara de Custódia e Liquidação (Cetip), na Bolsa de Valores de São Paulo (BM&FBOVESPA) são comercializados contratos futuros da taxa DI para distintas maturidades, de forma que as taxas desses contratos e seus respectivos vértices são utilizados da construção da ETTJ pois possuem uma elevada liquidez. Alternativamente, as taxas de contratos de *swaps* que comercializam taxas prefixadas e DI também constituem uma fonte para a determinação da curva de juros no Brasil.

até o vencimento, podendo ser eventualmente negociados com outras instituições no *open market* ou *mercado secundário*.

²Obviamente não é apenas a maturidade que determina as taxas de juros. Posteriormente, vamos identificar os principais fatores que influenciam a construção da ETTJ.

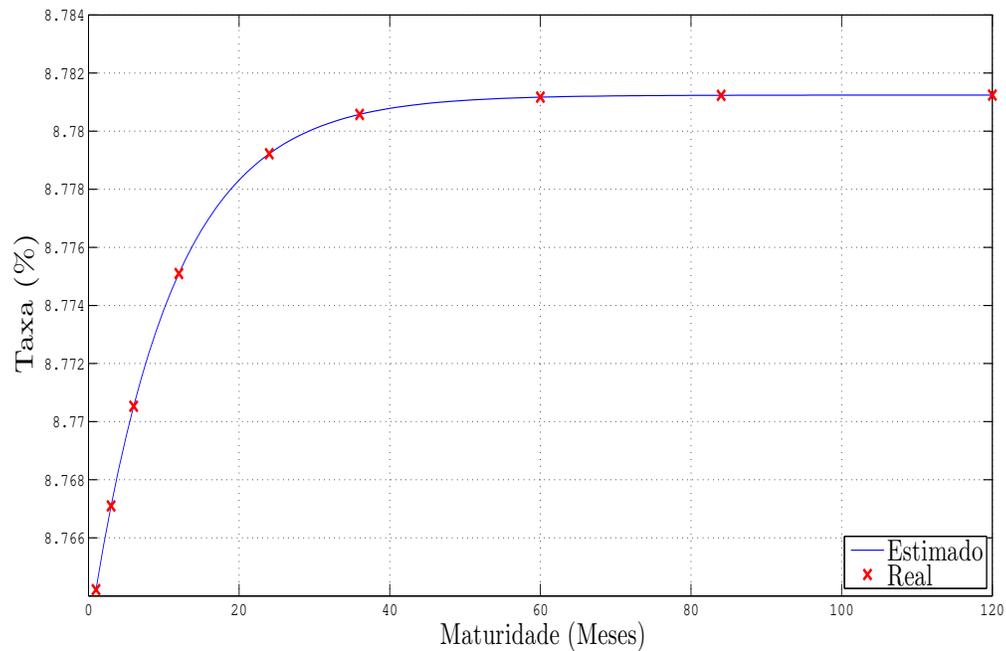


Figura A.1: Taxas de juros livres de risco e ETTJ estimada.

A.3 Formas da ETTJ

A curva que descreve a estrutura a termo das taxas de juros pode assumir diversas formas. A forma “normal” da ETTJ é apresentada na Figura A.2 em que a taxa de juros é uma função crescente da maturidade. Esse formato é dito normal devido à própria definição de juros, uma vez que quanto maior o prazo ou a maturidade, maior o custo do capital. Entretanto, nem sempre o formato da curva é normal. Em alguns momentos, a ETTJ está invertida, ou seja, os juros para os prazos mais longos estão abaixo daqueles para maturidades curtas. A Figura A.3 apresenta uma ETTJ invertida. Além desses dois casos, a curva de juros pode ser caracterizada por uma curvatura côncava ou convexa, conforme mostrado nas Figuras A.4 e A.5, respectivamente. Dessa forma, pode-se perceber que o comportamento da curva de juros pode oscilar ao longo do tempo, de forma que seu formato e oscilação podem ser explicadas por diversos fatores, conforme será descrito a seguir.

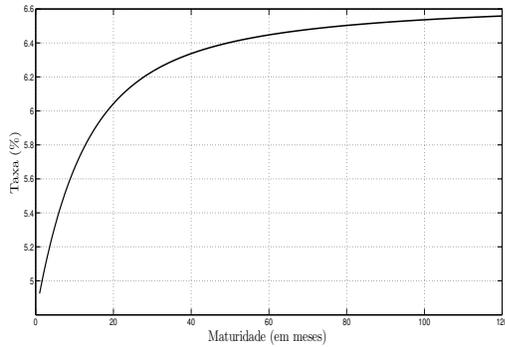


Figura A.2: ETTJ “normal”.

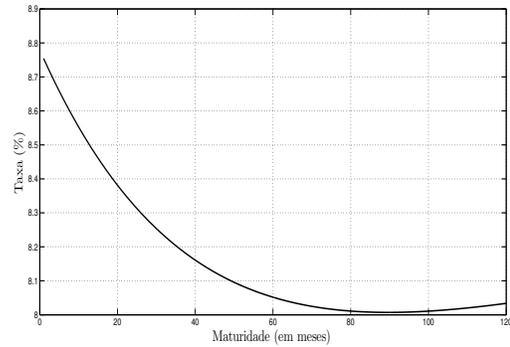


Figura A.3: ETTJ “invertida”.

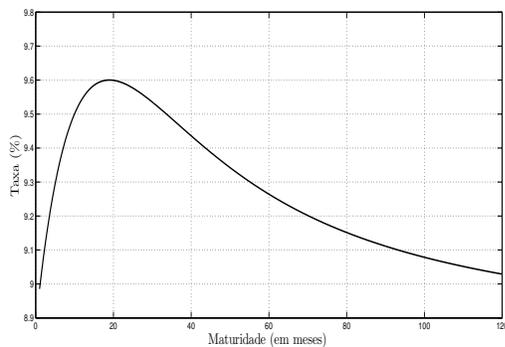


Figura A.4: ETTJ com curvatura convexa.

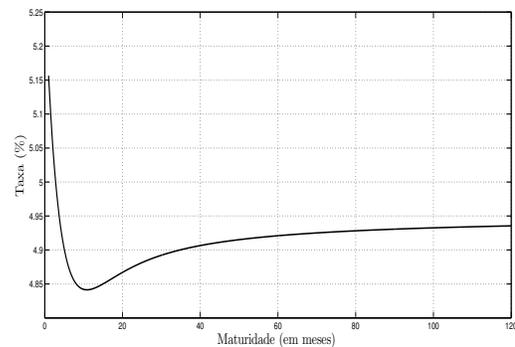


Figura A.5: ETTJ com curvatura côncava.

A.4 Formação da ETTJ

A curva de juros é construída com base nas taxas de juros consideradas livres de risco dos contratos de juros transacionados no mercado de renda fixa, portanto, um dos fatores principais que influenciam a formação da ETTJ são as expectativas do mercado. Essas expectativas refletem a forma da estrutura a termo, podendo ser expectativas de queda ou alta na taxa de juros. Dessa forma, a curva expressa as expectativas do mercado quanto ao futuro das taxas de juros. Entretanto, a fragilidade da hipótese das expectativas na formação da curva de juros reside na assunção implícita de que os agentes econômicos são indiferentes à incerteza, o que na prática não ocorre.

A hipótese de preferência pela liquidez também é muito utilizada para explicar a formação da estrutura a termo, justificando uma estrutura positivamente inclinada, que é o caso mais comum na prática. Essa assertiva implica que existem mais investidores interessados em aplicar dinheiro no curto prazo, de forma a exigir um prêmio maior para investir por prazos mais longos, o chamado *prêmio de liquidez*³. Contudo, essa hipótese se baseia na suposição de que

³O prêmio de liquidez ou prêmio pelo risco está associado ao fato de que o investidor recebe uma determinada taxa de juros por abrir mão da liquidez atual por uma liquidez futura, ou ainda, recebe a remuneração pelo risco associado ao investimento, e como o risco é resultado de incertezas, quanto maior o prazo, maior as incertezas e, por consequência, maior o risco.

os investidores são neutros ao risco.

Em termos menos teóricos, a forma da estrutura a termo das taxas de juros também é influenciada por fatores macroeconômicos como inflação, taxa de câmbio e hiato do produto⁴. Esses indicadores são fundamentais no mercado de renda fixa de uma economia, uma vez que a curva de juros pode ser reflexo dos agentes quanto aos movimentos desses indicadores macroeconômicos, ou mesmo o Governo pode influenciar o mercado de renda fixa com o objetivo de alterar os níveis de inflação e atividade econômica.

A.5 As Metas Inflacionárias no Brasil

No Brasil, o Decreto nº3.088, de 21 de junho de 1999, institui o conceito de meta de inflação (*inflation target*), que é definida pelo Conselho Monetário Nacional. Com isso, a política monetária subordinou-se a essas metas inflacionárias e o Banco Central passou, então, a fixar a chamada meta Selic e seu eventual viés (Securato 2008).

O Sistema especial de liquidação e custódia (Selic) consiste em uma associação entre o Banco Central e a ANBIMA, tendo por finalidade a custódia de títulos públicos e a liquidação financeira da negociação entre compradores e vendedores. Dessa forma, o Comitê de Política Monetária (Copom), em suas reuniões, fixa regularmente a taxa Selic no Banco Central do Brasil, por meio dos principais títulos de renda fixa.

Dessa forma, no Brasil, a estrutura a termo das taxas de juros é formada principalmente com o intuito de atingir as metas de inflação, em que a política monetária é realizada por meio do ajuste das taxas de curto prazo. Assim, o formato da curva de juros no Brasil é atrelada às expectativas quanto ao ajuste das metas de inflação, não sendo construída apenas pela livre interação dos agentes no mercado de renda fixa.

A.6 LTN, *Duration* e Convexidade

No mercado de renda fixa brasileiro, as Letras do Tesouro Nacional (LTNs) são os títulos públicos zero-cupom da economia. A LTN é um título emitido pelo Tesouro Nacional, para a cobertura do déficit orçamentário e para a realização de operações de crédito por antecipação de receitas orçamentárias. É um título prefixado, com prazo mínimo de 28 dias, resgatado no vencimento pelo valor nominal de R\$ 1.000,00. A colocação às instituições interessadas é feita por meio de um desconto em relação ao valor nominal, e o cálculo dos preços de tais papéis são regidos pelo regime de capitalização composta. A disponibilidade de vértices das LTN é muito baixa, mesmo com uma relativa liquidez para os prazos transacionados. Além disso, principalmente, as LTNs comercializadas possuem prazos curtos (em geral, com maturidades de até um ano). Portanto, a construção da ETTJ a partir das taxas das LTNs gera uma curva mais apropriada para análises das taxas de curto prazo, enquanto a qualidade das extrapolações da curva (longo prazo) é deteriorada.

⁴O hiato do produto é definido como a diferença entre o Produto Interno Bruto (PIB) efetivo e o PIB potencial de uma economia.

Quando tratamos da estrutura a termo das taxas de juros e de títulos públicos, é de fundamental importância introduzir os conceitos de *duration* e *convexidade*. Considere um fluxo de caixa F_1, F_2, \dots, F_n correspondente a recursos dos quais conhecemos os prazos respectivos d_1, d_2, \dots, d_n até uma origem de tempos, que em geral é a data hoje do problema, e as taxas de juros y_1, y_2, \dots, y_n de cada fluxo em relação à data de origem de tempos. Podemos então calcular o valor presente do fluxo, PV , da seguinte forma:

$$PV = \frac{F_1}{(1 + y_1)^{d_1}} + \frac{F_2}{(1 + y_2)^{d_2}} + \dots + \frac{F_n}{(1 + y_n)^{d_n}} = \sum_{i=1}^n \frac{F_i}{(1 + y_i)^{d_i}} \quad (\text{A.4})$$

A *duration* procura levar em conta o valor do dinheiro no tempo. Assim, a partir da fórmula do prazo médio ponderado:

$$d = \frac{F_1 d_1 + F_2 d_2 + \dots + F_n d_n}{F_1 + F_2 + \dots + F_n} \quad (\text{A.5})$$

corrigi-se o valor de cada fluxo F_j por uma taxa y_j até uma data fixada. A data escolhida é a data de origem da contagem de tempo, i.e., a data hoje. Portanto, a fórmula da *duration*, D , é dada por:

$$D = \frac{\sum_{i=1}^n \frac{F_i}{(1 + y_i)^{d_i}} \times d_i}{PV} \quad (\text{A.6})$$

Pode-se perceber então que a *duration* é a fórmula do prazo médio levando em conta o valor do dinheiro ao longo do tempo. Nos caso dos títulos zero-cupom, como não existe o pagamento de cupons intermediários, a *duration* é igual a maturidade correspondente do título. Por fim, outro conceito importante no mercado de renda fixa é o de *convexidade*, que mede a variação de uma carteira ou título público em relação à variação das taxas de juros, medida pela derivação da função de valor presente da carteira ou título.

Para maiores detalhes do funcionamento do mercado de renda fixa no Brasil e de seus principais produtos, ver (Securato 2008).