



TCC/UNICAMP
M187p
IE

taxas de juros - maciel

Leandro Dos Santos Maciel

**PRECIFICAÇÃO DAS OPÇÕES DE TAXA DE JUROS NO BRASIL:
UMA AVALIAÇÃO DOS MODELOS DE REDES NEURAIS PROGRESSIVAS E
RECORRENTES**

CEDOC - IE - UNICAMP

Monografia apresentada ao Instituto de Economia
como parte dos requisitos para obtenção do título de
Bacharel em Ciências Econômicas.

Orientadora: Profa. Dra. Rosangela Ballini

Campinas, SP

2010

Leandro Dos Santos Maciel

**PRECIFICAÇÃO DAS OPÇÕES DE TAXA DE JUROS NO BRASIL:
UMA AVALIAÇÃO DOS MODELOS DE REDES NEURAIS PROGRESSIVAS E
RECORRENTES**

Monografia apresentada ao Instituto de Economia como parte
dos requisitos para obtenção do título de Bacharel em Ciências
Econômicas.

Aprovação em 12/07/2010

Banca Examinadora:

Profa. Dra. Rosangela Ballini - IE-UNICAMP

Profa. Dra. Ivette Luna Huamaní - IE-UNICAMP

Campinas, SP

2010

Resumo

Nas últimas décadas, a precificação de derivativos de taxa de juros tem chamado grande atenção dos agentes de mercado em geral, assim como de pesquisadores, devido à sua importância econômica. Gestores de portfólios, *traders*, analistas de risco, *policymakers*, podem extrair informações relevantes de contratos futuros, *swaps* e opções, uma vez que elas desempenham um papel importante em suas estratégias e em seus processos de decisão. Um dos contratos de derivativos de taxa de juros de destaque no mercado brasileiro são as opções sobre o Índice de Depósitos Interfinanceiros de Um Dia (IDI), porém, a liquidez desses papéis é muito baixa, se comparada com o volume de negócios do DI Futuro. Uma das possíveis explicações para essa baixa liquidez seja a falta de um modelo de precificação amplamente aceito pelo mercado e sem deficiências teóricas. Na Bolsa, utiliza-se o modelo de Black (1976) para precificar tais opções. Teoricamente, esse modelo é aplicado para opções sobre futuros e não leva em consideração a estrutura a termo da taxa de juros (ETTJ). Dessa forma, os preços teóricos, obtidos pelo modelo, diferem significativamente dos preços observados no mercado. Vieira e Pereira (1999) e Barbachan e Ornelas (2003) propuseram fórmulas analíticas fechadas para a precificação das opções de IDI, com base, respectivamente, nos modelos de Vasicek (1977) e CIR (1985), para descrever o processo estocástico seguido pela taxa de juros de curto prazo. Entretanto, tais modelos, apesar de apresentarem maior adequação teórica às opções de IDI, em comparação ao modelo de Black (1976), ainda resultam em prêmios que não são explicados pelos observados no mercado. Para superar tais limitações, este trabalho propõe uma avaliação dos modelos de redes neurais progressivas e recorrentes para a valoração das opções de compra de IDI, no período de janeiro de 2003 a julho de 2007. Modelos baseados em redes neurais artificiais (RNAs) têm mostrado grandes vantagens em relação aos modelos tradicionais, com destaque nas aplicações em economia e finanças, sobretudo na precificação de opções. Dessa forma, obtivemos o prêmio das opções de compra de IDI de acordo com os modelos de Black, utilizados pela BM&FBOVESPA, assim como por meio dos modelos de Vasicek e CIR, propostos por Vieira e Pereira (1999) e Barbachan e Ornelas (2003), respectivamente. Esses modelos compuseram nosso *benchmark*, ao qual foi contraposto aos modelos sugeridos neste trabalho: um modelo de rede neural *feedforward* e dois modelos de redes recorrentes, sendo as redes de tipo Elman e Jordan. Os modelos foram avaliados de acordo com as métricas usuais de erro, isto é, MEP, MPE, RMSE, MAE, MAPE e TIC. Além disso, aplicou-se os testes paramétricos AGS e MGN, como também o teste não-paramétrico SIGN que avaliam estatisticamente modelos de predição competitivos. Nossos resultados mostraram que as redes neurais recorrentes geraram os valores dos prêmios para as opções de compra de IDI mais próximos dos observados no mercado. Esses resultados foram indicados tanto pelas métricas de erro quanto pelos testes estatísticos. Além disso, os modelos de redes neurais apresentaram melhores resultados nas opções *out-of-the-money*.

Palavras-chave: Precificação de Opções, Redes Neurais, Opções de Taxa de Juros, Redes Neurais Recorrentes, Opções de IDI.

Agradecimentos

Aos meus pais pelo carinho e apoio, essenciais para realização dessa conquista. Em especial, à minha mãe, que durante todos esses anos de graduação tem sido minha eterna ajudadora e sem a qual nada teria alcançado. Ao meu irmão pela consideração e motivação.

À Professora Rosangela Ballini pela orientação e suporte nas inúmeras reuniões ao longo de três anos de pesquisa.

Ao Professor Rodrigo Lanna F. Da Silveira pelas críticas e sugestões.

À minha eterna amiga, Juliane Rosa (*in memoriam*) que, onde quer que esteja, está feliz com mais essa realização, ao qual tanto apoio me concedeu. Agradeço também aos incontáveis momentos felizes que vivemos juntos, inesquecíveis.

Aos meus grandes e queridos amigos Hildo Sena e Adriana Carnizelli, que permitiram felizes episódios de descontração. À minha nova amiga, Marilene da Silva, que proporcionou discussões esclarecedoras em torno de diversos tópicos matemáticos, assim como ajudou a tornar nossa casa um ambiente mais agradável possível e se mostrou uma companheira de estudo e ouvinte nos momentos de desabafo.

À minha querida amiga, irmã, paixão, Livia Rodrigues, futura grande médica, que mesmo apesar dos poucos encontros pessoais, sempre tem me mostrado o exemplo de determinação, garra e valorização dos princípios que evitam a profunda degradação da sociedade atual. À minha companheira dos momentos de cursinho, Aline Galvão, pessoa admirável, que me conquistou e a quem sempre terei como grande amiga.

Aos professores e funcionários do Instituto de Economia da Unicamp, excelentess profissionais que tornam o Instituto excelência na qualidade de ensino e pesquisa em ciências econômicas.

À minha mãe, Ivete, imagem encorajadora

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	4
1.3 Estrutura do Trabalho	5
2 Revisão Bibliográfica	7
2.1 Opções de Taxa de Juros	7
2.2 Modelos de Estrutura a Termo da Taxa de Juros (ETTJ)	8
2.3 Modelos para Opções de Taxa de Juros no Brasil	9
2.4 Redes Neurais na Precificação de Opções	12
3 Modelos de Taxa de Juros	15
3.1 Precificação Livre de Arbitragem	15
3.2 Modelos Indiretos	16
3.2.1 Modelos Indiretos Endógenos	16
3.2.2 Modelos Indiretos Exógenos	19
3.3 Precificação das Opções de IDI no Brasil	22
3.3.1 Modelo de Black para as Opções de IDI	23
3.3.2 Modelo de Vasicek para as Opções de IDI	24
3.3.3 Modelo CIR para as Opções de IDI	25

4	Redes Neurais Artificiais	27
4.1	Introdução	27
4.2	Rede Perceptron Multi-Camadas	29
4.3	Redes Neurais <i>Feedforward</i>	30
4.4	Redes Neurais Recorrentes	32
4.5	Algoritmo de Retro-Propagação do Erro	32
4.5.1	Taxa de Aprendizagem	36
4.5.2	Modos de Treinamento	36
4.5.3	Critério de Parada e Validação	37
5	Metodologia	39
5.1	Amostra	39
5.2	Estimação dos Modelos	41
5.2.1	Modelo de Black	41
5.2.2	Modelo de Vasicek	41
5.2.3	Modelo CIR	43
5.2.4	Modelos de Redes Neurais	44
5.3	Métricas de Erro	46
5.4	Testes Estatísticos	48
6	Resultados e Discussão	51
6.1	Topologias das Redes Neurais	51
6.2	Comparação das Métricas de Erro	57
6.3	Avaliação dos Testes Estatísticos	62
7	Conclusões	67
7.1	Trabalhos Futuros	68
	Referências bibliográficas	70

Lista de Figuras

4.1	Modelo não-linear de um neurônio artificial	28
4.2	Topologia multi-layer perceptrons com duas camadas ocultas	29
4.3	Exemplos de Funções de Ativação	31
4.4	Rede Neural Recorrente de Elman	33
4.5	Rede Neural Recorrente de Jordan	33
5.1	Evolução das Taxas CDI e Selic	40
6.1	Topologia da rede neural progressiva para o prêmio das opções de compra de IDI	53
6.2	Topologia da rede neural recorrente de Elman para o prêmio das opções de compra de IDI	55
6.3	Topologia da rede neural recorrente de Jordan para o prêmio das opções de compra de IDI	57
6.4	Evolução dos prêmios de mercado e do modelo de Black	61
6.5	Evolução dos prêmios de mercado e do modelo de Vasicek	62
6.6	Evolução dos prêmios de mercado e do modelo de CIR	62
6.7	Evolução dos prêmios de mercado e da RNF	63
6.8	Evolução dos prêmios de mercado e da RNRE	63
6.9	Evolução dos prêmios de mercado e da RNRJ	64

Lista de Tabelas

6.1	Critérios de informação para determinação da estrutura da rede neural <i>feedforward</i> para o apreçamento das opções de compra de IDI	52
6.2	<i>EQM</i> para as diferentes taxas de aprendizagem com base na amostra teste para a rede neural <i>feedforward</i>	53
6.3	Critérios de informação para determinação da estrutura da rede neural recorrente de Elman para o apreçamento das opções de compra de IDI	54
6.4	<i>EQM</i> para as diferentes taxas de aprendizagem com base na amostra teste para a rede neural recorrente de Elman	55
6.5	Critérios de informação para determinação da estrutura da rede neural recorrente de Jordan para o apreçamento das opções de compra de IDI	56
6.6	<i>EQM</i> para as diferentes taxas de aprendizagem com base na amostra teste para a rede neural recorrente de Jordan	57
6.7	Aderência dos preços calculados pelos modelos tradicionais e de Redes Neurais aos preços de fechamento das <i>calls</i> de IDI européias negociadas na BM&FBOVESPA entre 11/12/2006 e 31/07/2007. Amostra com 1.270 observações	58
6.8	Aderência dos preços calculados pelos modelos tradicionais e de Redes Neurais aos preços de fechamento das <i>calls</i> de IDI européias negociadas na BM&FBOVESPA entre 02/01/2003 e 08/12/2006. Amostra com 11.531 observações	60
6.9	Aderência dos preços calculados pelos modelos tradicionais e de Redes Neurais aos preços de fechamento das <i>calls</i> de IDI européias negociadas na BM&FBOVESPA entre 11/12/2006 e 31/07/2007 de acordo com o grau de <i>moneyness</i> . Amostra com 1.270 observações	60

6.10 Testes AGS, MGN e SIGN para avaliação dos modelos de precificação para as opções de compra de IDI européias negociadas na BM&FBOVESPA entre 11/12/2006 e 31/07/2007, amostra com 1.270 observações.	65
--	----

Capítulo 1

Introdução

1.1 Motivação

Durante cerca de vinte anos, a partir de 1974, a economia brasileira conviveu com altas taxas de inflação, da ordem de 10 a 80% ao mês. Com taxas de tal magnitude, além de extremamente voláteis, não se desenvolvia no país um mercado de títulos pré-fixados, uma vez que o risco associado à contratação de tais papéis era muito elevado¹. Com o advento do Plano Real, em meados de 1994, a introdução da nova moeda, que dá nome ao plano, e o controle do processo inflacionário permitiram uma relativa estabilização e fortalecimento do mercado brasileiro de juros. Dessa forma, esse novo cenário permitiu a ampliação dos prazos das operações de títulos pré-fixados nos mercados primário e secundário. Consequentemente, as instituições financeiras abriram espaço para a comercialização de ativos antes considerados inoperáveis para o mercado nacional, dentre eles, os derivativos de taxa de juro.

Nas décadas recentes, a precificação de derivativos de taxa de juros tem chamado grande atenção dos agentes de mercado em geral, assim como de pesquisadores, devido à sua importância econômica. Gestores de portfólios, *traders*, analistas de risco, *policymakers*, podem extrair informações relevantes de contratos futuros, *swaps* e opções, uma vez que elas desempenham um papel importante em suas estratégias e processos de decisão. Além disso, a curva de rendimentos, ou *yield curve*, é, sem dúvida, uma importante variável econômica, podendo ser modelada por meio das taxas implícitas desses contratos.

¹Garcia e Barcinski (1998) afirma que, durante esse período, as aplicações pré-fixadas no Brasil dificilmente ultrapassavam um mês.

De acordo com o Banco de Compensações Internacionais (BIS - *Bank for International Settlements*), no mercado de derivativos mundial, os contratos relacionados à níveis da taxa de juros são os que apresentaram maior volume financeiro. Em 2009, no mercado secundário, estima-se que o valor nacional para contratos de taxa de juros atingiu US\$ 437 trilhões, além de US\$ 342 trilhões em *swaps* de taxa de juros. Ainda, segundo a Associação Internacional de *Swaps* e Derivativos (ISDA - *International Swaps and Derivatives Association*), 80% das 500 maiores companhias do mundo utilizaram derivativos de taxa de juros para controle de seus fluxos de caixa. Dessas mesmas companhias, 75% transacionaram opções de taxas cambiais, 25% opções sobre commodities e 10% aplicaram seus recursos em opções sobre índices de ações.

Os principais contratos de derivativos relacionados a taxas de juros transacionados na Bolsa de Valores, Mercadorias e Futuros de São Paulo (BM&FBOVESPA) são: Contrato Futuro de Taxa Média de Depósitos Interfinanceiros de Um Dia (DI Futuro), Contrato de Opção de Compra/Venda sobre Índice de Taxa Média de Depósitos Interfinanceiros de Um Dia (Opções de IDI) e o Contrato Futuro de Cupom Cambial². Em termos de volume financeiro, o DI Futuro é o derivativo de taxa de juros mais líquido da Bolsa, possuindo contratos vencendo a cada mês. Segundo a BM&FBOVESPA, na década de 2000, o número médio de contratos negociados de DI Futuro foi em torno de 13 milhões, enquanto que as opções de IDI apresentaram o valor médio de 700 mil contratos negociados. Uma das possíveis explicações para a baixa liquidez dos contratos de opções de IDI seja a falta de um modelo de precificação amplamente aceito pelo mercado e sem deficiências teóricas.

Neste trabalho, realizaremos um estudo relacionado à valoração das opções de IDI. Para tanto, iremos, inicialmente, justificar a relevância econômica desses contratos, por meio de uma comparação com os *swaps* de taxa de juros, para então descrever os objetivos de pesquisa³.

Contratos de *swaps* podem permitir que empresas se protejam contra oscilações inesperadas na taxa de juros, pois, por exemplo, permite que se transforme uma dívida contraída a taxa pós-fixada numa pré-fixada, eliminando a incerteza com relação ao valor futuro do passivo e permitindo um melhor planejamento estratégico. As opções de IDI permitem uma alternativa aos *swaps*. Neto e Pereira (1999) afirmam que uma empresa com passivo pós-fixado consegue fixar um teto para a taxa de juros com a compra de uma *call* de IDI. Assim, com a alta da taxa de juros, o aumento de seu passivo será compensado pelo excesso do IDI sobre seu preço de exercício e, em caso de uma queda na taxa, se

²Podemos citar ainda o Contrato de *Swap* Cambial com Ajuste Periódico, o Contrato Futuro de Cupom de IGP-M e o Contrato Futuro de Cupom IPCA.

³A comparação com o contrato de Futuro DI não se mostra necessária, uma vez que os *swaps* se assemelham aos futuros, porém, são mais simples devido à ausência de ajustes diários.

beneficiará com a redução do valor de seu passivo, enquanto que a opção deixará de existir⁴.

Avaliando estrategicamente os demais fatores que afetam suas atividades, as empresas podem escolher entre o padrão simétrico do *swap* (com custo inicial nulo) ou o padrão assimétrico da opção (com custo inicial positivo).

As opções de IDI começaram a ser negociadas na BM&FBOVESPA, antiga BM&F, em 1997. O volume de negócios desses contratos foi aumentando aos poucos, porém, devido às instabilidades financeiras do final da década, sofreu fortes reduções, retomando um aumento de liquidez apenas partir dos anos 2000. Em comparação com os demais contratos derivativos de taxa de juros, as opções de IDI apresentam um volume ainda muito baixo de negociações. Conforme mencionado, esse fato pode estar relacionado à utilização, pela Bolsa, do modelo de Black (1976) para a precificação desses instrumentos, uma vez que se trata de um modelo teoricamente inconsistente para esse tipo de opção e, portanto, não é amplamente aceito pelos agentes do mercado Almeida et al. (2003). Teoricamente, esse modelo é aplicado para opções sobre futuros e não leva em consideração a estrutura a termo da taxa de juros. Dessa forma, os preços teóricos, obtidos pelo modelo, diferem significativamente dos preços observados no mercado, o que gera um receio entre os investidores em se tomar posições nesses papéis.

As opções de taxa de juros, transacionadas nos mercados dos EUA e Europeu, em geral, têm como ativo objeto um título de renda fixa com vencimento posterior ao vencimento da opção. Assim, ela reflete a expectativa do comportamento futuro da estrutura a termo da taxa de juros entre o período de vencimento da opção e o vencimento do título. Como as opções de IDI refletem o comportamento da taxa de juros de curto prazo na data de avaliação até o vencimento da opção, torna-se necessário adequar os modelos de precificação para a realidade brasileira.

Neste contexto, trabalhos recentes têm apresentado diversas alternativas para precificar essas opções, com base principalmente nos modelos de Black e Scholes (1973), Vasicek (1977), Cox et al. (1985), Ho e Lee (1986), Hull e White (1990) e Heath et al. (1992). São metodologias baseadas em equações diferenciais estocásticas que permitem a derivação de uma fórmula analítica fechada ou com auxílio de metodologias como árvores binomiais ou trinomiais, para então obter o prêmio ou preço das opções. Entretanto, os resultados mostram que ainda os preços teóricos distanciam-se dos preços observados no mercado, além de apresentarem possibilidades de arbitragem.

Na literatura de opções, a utilização de modelos de Redes Neurais Artificiais (RNAs) tem se

⁴Deve-se lembrar que a posse do contrato de opção exige um custo inicial para a empresa, ao contrário do contrato de *swap*.

apresentado como um método alternativo para a precificação de tais instrumentos⁵. As RNAs se destacam devido sua alta performance ao tratar relações não-lineares de tipo entrada-saída, em especial sua habilidade de aprendizado e capacidade de generalização, associação e busca paralela. Além disso, esses modelos não requerem nenhuma hipótese restritiva quando às entradas, permitindo que a rede seja capaz de aprender o processo gerador da relação entre as variáveis avaliadas⁶.

Embora a superioridade dos modelos de redes neurais, em comparação aos métodos usuais de apreamento de opções, seja quase um consenso entre os pesquisadores, Callen et al. (1996) afirmam que estas técnicas não apresentam, necessariamente, os melhores resultados em relação aos modelos lineares de séries temporais tradicionais. Além disso, as aplicações de RNAs para precificação de opções se concentram nos papéis sobre ações, índices, moedas, futuros e *commodities*, fazendo necessário um estudo para as opções relacionadas à taxa de juros, até o momento escasso para o mercado brasileiro.

1.2 Objetivos

Verificada a importância das opções de IDI no mercado brasileiro e as limitações para obtenção dos prêmios de tais contratos, este trabalho tem o objetivo principal de valorar as opções de compra de IDI, negociadas na BM&FBOVESPA, para o período de janeiro de 2003 a julho de 2008, com base em modelos não-lineares de redes neurais *feedforward* e recorrentes⁷. Como *benchmark*, precificaremos as mesmas opções por meio do modelo de Black (1976), uma vez que é utilizado pela Bolsa para obtenção dos prêmios das opções de IDI, e por meio das equações analíticas derivadas dos trabalhos de Neto e Pereira (1999) e Barbachan e Ornelas (2003), por se tratarem dos modelos analíticos baseados em fórmulas fechadas que mais se adequam à realidade brasileira.

Os preços teóricos, obtidos pelos modelos, serão então comparados com os prêmios de fechamento das opções verificados no mercado, para elucidar o modelo que mais se aproximou à realidade. Essa avaliação será baseada em métricas tradicionais de erro de previsão, isto é: Erro Percentual Médio (MPE - *Mean Percentage Error*), Erro Percentual Máximo (MPE - *Maximum Percentage Error*), Raiz do Erro Quadrático Médio (RMSE - *Root Mean Squared Error*), Erro Absoluto Médio (MAE -

⁵Shapiro (2003) apresenta uma coletânea dos trabalhos baseados em inteligência artificial aplicados a economia e finanças, sobretudo na precificação de opções.

⁶Bennel e Sutcliffe (2004) e Andreou et al. (2008) afirmam que os modelos de Black e Scholes (1973) e seus derivados, por apresentarem hipóteses restritivas, acabam decorrendo em erros sistemáticos, o que resulta em preços teóricos mais irrealistas, por isso a aplicação de modelos não-lineares alternativos tem se mostrado relevante.

⁷As opções de venda, ou *puts*, não serão consideradas nesta pesquisa devido sua baixa liquidez na Bolsa, uma vez que não permite uma adequada validação dos modelos considerados.

Maximum Absolute Error), Erro Médio Absoluto Percentual (MAPE - *Maximum Absolute Percentage Error*), e o coeficiente de desigualdade de Theil (TIC). Além disso, testes estatísticos paramétricos e não-paramétricos de performance para comparação de modelos de predição competitivos serão aplicados, sendo eles, os testes AGS, MGN e SIGN.

A partir desse objetivo que norteia o presente trabalho, verificaremos também os seguintes pontos:

- O modelo analítico do *benchmark* considerado que resultou em preços que mais se aderiram aos observados no mercado.
- Avaliar o melhor modelo de rede neural, *feedforward* ou recorrente, para precificar as opções de IDI, no sentido de verificar uma dependência temporal na formação dos preços desses papéis.
- Selecionar as variáveis que se mostram relevantes, segundo os métodos de redes neurais, para a obtenção dos prêmios das opções.
- Definir a estrutura de rede, em termos de camadas e neurônios intermediários, que apresenta melhor desempenho de acordo com o objetivo proposto.
- Analisar se a diferença entre os modelos é significativa em termos estatísticos, considerando que os trabalhos na literatura abordam, em sua maioria, apenas as métricas usuais de erros para comparação entre diferentes métodos na precificação de opções.

Este estudo se destaca ao aplicar o método de redes neurais artificiais em um segmento de mercado ainda não avaliado no Brasil, o das opções de taxa de juros. Além disso, utilizamos duas estruturas diferentes de redes neurais recorrentes, uma vez que a literatura, em geral, aplica modelos do tipo progressivas ou *feedforward*, que possuem menos potencial em modelar dados temporais ou espaciais.

1.3 Estrutura do Trabalho

Com os objetivos definidos, este trabalho está organizado como segue. Além desse capítulo de introdução, o capítulo 2 apresenta a revisão da literatura sobre o tema, abordando os trabalhos relacionados aos modelos de taxa de juros, precificação de opções e redes neurais aplicadas à economia e finanças.

O capítulo 3 apresenta os modelos que descrevem o comportamento da estrutura a termo da taxa de juros, base para a construção de metodologias para obtenção de fórmulas analíticas para a precificação

de opções relacionadas à níveis da taxa de juros. Ainda, serão descritas as características das opções sobre IDI, assim como os modelos analíticos de precificação para este ativo, sendo eles: modelo de Black (1976), e os derivados dos trabalhos de Neto e Pereira (1999) e Barbachan e Ornelas (2003), que serão aplicados neste trabalho. Em seguida, no capítulo 4, apresenta-se uma descrição geral dos modelos de redes neurais, com relação a suas estruturas e algoritmo de treinamento.

A metodologia para estimação e validação dos modelos é descrita no capítulo 5. A estimação dos parâmetros das fórmulas fechadas será descrita nesse mesmo capítulo, assim como a metodologia para escolha do melhor modelo de rede para valoração das opções.

Os resultados e as discussões constituem o capítulo 6, que elucida as principais vantagens e limitações dos modelos considerados. Segue-se, por fim, o capítulo 7 com as conclusões, como também sugestões para pesquisas futuras.

Capítulo 2

Revisão Bibliográfica

Este capítulo apresentará a discussão verificada na literatura com relação aos seguintes tópicos: principais métodos para modelagem da estrutura a termo das taxas de juros, modelos para obtenção dos prêmios das opções de taxa de juros no Brasil, e as abordagens nas aplicações das redes neurais artificiais no mercado de opções.

2.1 Opções de Taxa de Juros

Opções de taxa de juro são instrumentos cujos *payoffs* dependem de alguma forma do nível da taxa de juro. Os instrumentos mais transacionados desse tipo de ativo são as opções embutidas em títulos de renda fixa, em que seu preço reflete a expectativa do comportamento futuro de uma parte da estrutura a termo da taxa de juro, mais precisamente, o período entre o vencimento da opção e o vencimento do título. A avaliação de opções de taxa de juro é mais difícil se comparada a de opções sobre ações, pois modelar o comportamento do preço de uma ação se mostra menos complexo que o comportamento da taxa de juro Hull (2003).

De acordo com Neto e Pereira (1999), as características da modelagem de instrumentos de renda fixa os distinguem da modelagem de ações, moeda estrangeira e *commodities*. Elas podem ser resumidas a seguir:

- As taxas de juro apresentam o fenômeno de reversão à média, isto é, verifica-se uma tendência para um estado de equilíbrio. Essa característica se engessa ainda na teoria econômica, pois forças políticas e econômicas pressionam as taxas de juros para baixo em momentos que se encontra em patamares excessivamente altos e, com isso, impactam negativamente o déficit público, o

consumo em nível agregado e o nível de emprego¹.

- Os modelos de taxa de juros englobam a descrição do processo estocástico de toda uma “curva de juros” por motivo da variedade de prazos de vencimentos de títulos existentes no mercado, o que não se verifica no caso de uma ação, taxa de câmbio ou *commodity*².
- As volatilidades nos diversos vértices da estrutura a termo da taxa de juro costumam diferir; além disso, as correlações nos vértices também não são perfeitas, decrescendo na medida em que se aumenta a distância entre eles.
- O preço de um título de renda fixa possui valor fixo e conhecido em sua data de vencimento, diferente do preço de ações, moedas estrangeiras e *commodities*.

Dessa forma, modelar o comportamento do ativo objeto de uma opção de taxa de juro, isto é, o movimento da taxa de juros de curto prazo, requer metodologias que sejam capazes de descrever a estrutura a termo descrita por tal taxa.

2.2 Modelos de Estrutura a Termo da Taxa de Juros (ET TJ)

Segundo Neto e Pereira (1999), são duas as abordagens que descrevem os modelos de estrutura a termo da taxa de juro (ET TJ): direta e indireta. Os modelos de abordagem direta partem da especificação direta do processo estocástico dos preços dos títulos primários, em vez de derivarem este processo de conceitos mais primitivos, como taxa de juro de curto prazo ou a relação de preferência dos agentes econômicos. Já os modelos de abordagem indireta partem da especificação do processo estocástico seguido pela taxa de juro de curto prazo ou pela taxa *forward*.

A abordagem indireta, além de ser aplicada pela maior parte dos modelos de estrutura a termo, também se subdivide em duas classes: modelos indiretos endógenos (ou modelos de equilíbrio) e modelos indiretos exógenos (ou modelos de não-arbitragem). Nos modelos de equilíbrio, a estrutura a termo da data corrente e os preços dos títulos são obtidos endogenamente, ou seja, como função dos parâmetros do processo da taxa de juro de curto prazo e do valor corrente dessa variável. Os modelos mais conhecidos que pertencem a essa classe são as dinâmicas de Rendleman e Bartter (1980), Vasicek (1977) e Cox et al. (1985) ou CIR³. Em contrapartida, os modelos de não-arbitragem não derivam a

¹No sentido inverso, a mesma lógica se verifica, com taxas de juros excessivamente baixas.

²Ainda, o formato dessa curva pode se modificar ao longo do tempo.

³Esses modelos serão descritos no Capítulo 3.

estrutura a termo, mas a tomam como um dado. Assim, o processo da taxa de juro de curto prazo é escolhido de forma que a curva de juro relativa à data corrente obtida pelo modelo seja igual à observada no mercado. Os principais modelos de não-arbitragem são os de Ho e Lee (1986) e Hull e White (1990).

As dinâmicas de Vasicek e CIR foram os primeiros modelos de taxa de juro apresentados na literatura para a classe das dinâmicas de equilíbrio⁴. Apesar do sucesso desses modelos, eles produzem uma estrutura a termo endógena, não compatível com a estrutura a termo verificada no mercado, uma vez que o reduzido número de parâmetros não é suficiente para reproduzir a estrutura a termo corrente.

Na década de 1980, os modelos de equilíbrio passaram a ser transformados em modelos de não-arbitragem. A estratégia dessa transformação era incluir parâmetros dependentes do tempo na tendência do processo. Hull e White (1990) utilizaram essa metodologia no desenvolvimento dos modelos Vasicek Estendido e CIR Estendido, nos quais a tendência da taxa de juro é uma função do tempo.

Em 1992, Heath et al. (1992), doravante HJM, desenvolveram um sistema, para a modelagem da estrutura a termo, genérico e promissor, no qual a principal variável de estado é a taxa *forward* instantânea, ao invés da taxa de juro de curto prazo. A importância desse modelo está no fato de que qualquer modelo de estrutura a termo exógeno pode ser derivado a partir do sistema HJM, isto é, todo modelo de não-arbitragem é um modelo HJM⁵.

Além disso, a derivação de modelos de taxa de juro a partir do sistema HJM não fornece apenas uma compreensão maior dos parâmetros do modelo, como também fornece dinâmicas que são automaticamente ajustadas à estrutura a termo corrente⁶.

2.3 Modelos para Opções de Taxa de Juros no Brasil

A precificação das opções de taxas de juros no Brasil, isto é, opções de IDI, vem sendo abordada recentemente pela literatura. Uma razão para isso se encontra no fato desses papéis apresentarem características peculiares que o diferenciam substancialmente das opções sobre taxa de juro típicas dos principais mercados do mundo.

⁴Brigo e Mercurio (2001) afirmam que a ampla aceitação dos modelos de equilíbrio foi em grande parte ocasionado pela possibilidade de avaliação analítica de alguns instrumentos financeiros, tais como títulos de renda fixa e opções europeias de taxa de juros.

⁵Chiarella e Kwon (1999) mostraram que os modelos CIR Estendido, Black-Karasinski e Hull-White com dois fatores são casos especiais do modelo HJM.

⁶O maior problema dos modelos HJM se resume no fato de que ele supõe que o processo da taxa de juro de curto prazo possa não ser um processo Markoviano. Para uma discussão mais detalhada dessa questão ver Ritchken e Sankarabramanian (1995).

O contrato futuro de taxa de juros padrão dos EUA e da Europa possuem como ativo objeto um título de renda fixa que vencerá em um determinado número de meses depois do vencimento do futuro. Por questões de *hedge* e combinação de posições, as opções sobre taxa de juros padrão destes mercados são próximas aos contratos futuros, ou seja, o ativo objeto também é um título de renda fixa vencendo meses depois da maturidade da opção.

Pela instabilidade macroeconômica e financeira que o Brasil sofreu no último quartel do século XX, o mercado de juros negocia títulos com prazos mais curtos que os norte-americanos e europeus. Com isso, o contrato futuro de taxa de juro criado no país, o Futuro DI, é distinto do padrão verificado nos mercados mais influentes. O Futuro DI se comporta, na verdade, como um título de renda fixa no mercado **à vista**, sujeito a “marcações” a mercado diárias (ajustes diários), e não exatamente como um contrato **futuro** de juros Neto e Pereira (1999). Assim como no caso das opções dos mercados americano e europeu, a opção sobre taxa de juro brasileira, a opção de IDI, também reflete a forma do contrato futuro. Esse contrato negocia a taxa de juros acumulada entre a data corrente e sua própria data de vencimento, de maneira análoga ao que faz o futuro, ao invés de possuir um título de renda fixa como ativo objeto.

A peculiaridade das opções sobre taxa de juros no Brasil torna seu preço, bem como os fatores que o afetam, distinto daqueles ligados às opções sobre juros tradicionais. Por esse motivo a literatura tem abordado esse tema em especial, devido à carência do desenvolvimento de critérios próprios de análise e avaliação.

Os primeiros a abordar esse problema foram Neto e Pereira (1999). Eles desenvolveram uma fórmula fechada para o cálculo do prêmio das opções sobre IDI com base no modelo de Vasicek (1977). Derivada a fórmula analítica, propuseram dois métodos para estimação dos parâmetros volatilidade e velocidade de reversão à média da taxa de juro de curto prazo (implícitos no modelo): método endógeno (estimação incondicional) e exógeno (estimação condicional).

Apesar da alternativa de valoração das opções sobre IDI sugerida pelos autores, não o verificaram empiricamente, devido à baixa liquidez dos contratos, uma vez que foram introduzidos na Bolsa a partir de 1997⁷. Posteriormente, Cassetari e Neto (2001) mostraram que o modelo proposto por Neto e Pereira (1999) possui suposições que não condizem com o mercado brasileiro. Dessa forma, com base em uma análise do comportamento da taxa de juro de curto prazo brasileira, os autores evidenciaram que os prêmios das opções sobre IDI, de maneira geral, são mais sensíveis à volatilidade dessa taxa,

⁷ Outra crítica ao modelo de Neto e Pereira (1999) se baseia no fato de que, de acordo com a metodologia teórica desenvolvida, o modelo adota um processo de Ornstein-Uhlenbeck, que permite a existência de taxas de juros negativas.

assim como aos parâmetros de reversão a média e a taxa de juros de longo prazo, porém, em menor grau.

Gluckstern et al. (2002) aplicaram o modelo de Hull e White (1990), com base em árvores trinomiais, para precificar as opções sobre IDI, no período de 1997 a 2000. Seus resultados foram comparados com os obtidos por meio do modelo de Black (1976), e indicaram a superioridade do modelo de Hull-White.

Em seguida, Almeida et al. (2003) calibram o modelo proposto por Neto e Pereira (1999), de forma a avaliar as oscilações dos parâmetros implícitos, velocidade de reversão à média e volatilidade da taxa de juro de curto prazo, para o período de 2000 a 2001. Os resultados mostraram que a calibração se torna mais robusta em períodos de estabilidade econômico-financeira. Porém, em períodos com grande volatilidade ou pós-crise, o mesmo modelo gera parâmetros instáveis.

Barbachan e Ornelas (2003) derivaram uma fórmula analítica fechada para o apreamento das opções de IDI, com base no modelo de Cox et al. (1985), na tentativa de superar as limitações do trabalho de Neto e Pereira (1999), por se tratar agora de um modelo que não admite taxas de juro negativas. Os autores testaram seu modelo apenas para uma série de opções de compra de IDI (série AB17 de abril de 2003), ou seja, uma amostra com apenas oito observações. Os resultados foram comparados com os obtidos pelo modelo de Neto e Pereira (1999), e mostraram que as fórmulas derivadas pelos autores obtiveram preços teóricos mais próximos aos observados no mercado.

No modelo de Barbachan e Ornelas (2003) são quatro os parâmetros a serem estimados: velocidade de reversão à média e volatilidade da taxa de juros de curto prazo, assim como a taxa de juros de longo prazo e o preço do mercado do risco. Os autores estimaram a volatilidade com base no desvio padrão e a taxa de juros de longo prazo com base nos dados históricos do Certificado de Depósitos Interfinanceiros (CDI), enquanto que os demais parâmetros foram arbitrados.

Seguindo o mesmo caminho de Gluckstern et al. (2002), Junior et al. (2003) testaram um modelo de precificação para as opções de IDI, baseado no modelo de Hull e White (1990), para os dados observados no mercado para o período de 1999 a 2002. Para isso realizaram três testes: calibração (verificou-se o quanto a volatilidade implícita obtida das opções explica seus preços); avaliação de operações *delta-hedge* para verificar possibilidades de arbitragem, e seleção da melhor estimativa para o parâmetro de reversão a média que resultasse em uma volatilidade implícita de menor variância. Os resultados indicaram que a volatilidade implícita desses contratos chega a ser superior a 100%, o que indica inconsistência do modelo. Com relação ao teste de posições *delta*, verificou-se que o modelo é aceitável, gerando níveis baixos de erro ao se computar a operação. E, finalmente, com

relação ao parâmetro de reversão a média, a melhor estimativa foi zero, que gerou a menor variância da volatilidade implícita.

Almeida e Vicente (2003) estimaram por máxima verossimilhança um modelo Gaussiano multi-fator dinâmico para a estrutura a termo da taxa de juros brasileira. O objetivo do trabalho foi verificar a performance dessa estimação para precificar e realizar operações de *hedge* para as opções de IDI. Os resultados subestimaram os preços das opções, mas apresentou bons indicadores para as operações *delta*. Num segundo estágio, os autores inseriram os preços das opções no modelo dinâmico.

Como os próprios autores afirmaram, essa metodologia envolve uma capacidade computacional muito pesada, o que pode dificultar a aplicação do modelo na prática. Além disso, sua metodologia faz a assertiva de que o movimento do ativo objeto das opções de IDI está estritamente relacionado ao movimento das taxas implícitas dos contratos de Futuro DI. Essa pode ser uma afirmação muito forte, uma vez que existe o elemento expectativas dos agentes desse mercado influenciando o movimento da estrutura a termo.

Por fim, Barbedo et al. (2009) implementaram o modelo HJM para a precificação das opções de IDI. Os resultados evidenciaram que o modelo HJM consistentemente subprecifica os papéis, e verificaram uma relação direta entre a maturidade desses contratos e o erro de precificação em contraposição com uma relação negativa de acordo com o grau de *moneyness*.

Apesar das diversas metodologias utilizadas para se obter o prêmio das opções de IDI no mercado brasileiro, os resultados ainda indicam que os preços teóricos consistentemente se distanciam dos preços de fechamento verificados, demandando ainda propostas à essa questão, na tentativa de incrementar a liquidez neste segmento de contratos derivativos no país.

2.4 Redes Neurais na Precificação de Opções

Muitos estudos com redes neurais aplicadas às finanças, em especial na precificação de instrumentos derivativos, têm sido desenvolvidos na literatura recentemente, devido sua alta capacidade em se modelar relações de tipo entrada-saída. Hutchinson et al. (1994), por exemplo, avaliaram uma abordagem de redes neurais para o apreçamento e aplicação da operação *delta-hedge* nas opções sobre futuros do índice S&P 500, no período de 1987 a 1991. Seus resultados mostraram que as redes neurais foram capazes de aprender o funcionamento da fórmula de Black e Scholes (1973), com base em uma amostra para ajuste dos parâmetros composta por preços diários das opções no período de dois anos. Sobre tudo, os autores indicaram a possibilidade de apreçar e formar operações *delta-hedge* com sucesso para

opções não utilizadas na amostra apresentada à rede. Ainda, sugeriram que o apreçamento por tal método pode ser mais eficiente quando a dinâmica do ativo objeto da opção é desconhecida.

O estudo de Qi e Maddala (1996), por sua vez, aplicou um modelo de rede neural multicamadas para o apreçamento de *calls* sobre o índice S&P 500, no período que compreende dezembro de 1994 a janeiro de 1995. Os resultados apresentaram a superioridade do método de inteligência artificial em comparação à fórmula de Black-Scholes (BS). Com base na análise dos pesos da rede neural, confirmaram que quanto menor o preço de exercício ou maior o preço do ativo objeto ou, ainda, quanto mais distante o prazo para o vencimento, maior será o preço da *call*.

Outra pesquisa que se observa o uso de redes neurais para opções é o de Lajbcygier et al. (1996). Os autores utilizaram dois modelos para apreçar opções sobre o futuro do índice representativo das ações ordinárias comercializadas no mercado de ações australiano. O primeiro deles teve duas variáveis de entrada (quociente entre o preço *spot* e o preço de exercício, e o prazo de vencimento do papel). Já o segundo utilizou as mesmas variáveis, adicionando a taxa de juros e a volatilidade (estimada por meio da volatilidade histórica com uma janela de 60 dias), na tentativa de comparar com o modelo utilizado por Hutchinson et al. (1994). Os resultados mostraram que o modelo com quatro variáveis de entrada superou o modelo de duas variáveis, assim como a metodologia de Black-Scholes.

Podemos citar ainda os trabalhos de Geigle e Aronson (1999) e Malliaris e Salchenberger (1996). Estes primeiros examinaram o desempenho das redes neurais para a avaliação de opções americanas sobre futuro do índice S&P 500 e confirmaram a superioridade do modelo se comparado ao de BS. Já Malliaris e Salchenberger (1996) desenvolveram um modelo de rede neural para a previsão da volatilidade futura das opções sobre futuro do índice S&P 100, com base na volatilidade passada e outros fatores que afetavam os preços das opções comercializadas neste mercado. A rede foi estruturada com mais 13 fatores de entrada e a estimativa da volatilidade permitiria aos agentes estabelecer posições estratégicas na antecipação de mudanças nos movimentos do mercado. Os resultados da rede neural foram comparados com a volatilidade implícita das opções, obtidas pelo modelo de Black-Scholes. De acordo com os seus resultados, a estimação da volatilidade pelo modelo de redes neurais foi superior às modelagens de volatilidade histórica e implícita.

O mercado de opções norte-americano é o mais avaliado pela literatura sobre redes neurais, como se pode verificar nos trabalhos de Kelly (1994), Kitamura e Ebisuda (1998), Garcia e Gençay (2000), ?, ?, Ghaziri et al. (2001), e Andreou et al. (2008). Esses trabalhos aplicaram modelos de redes neurais para precificação de opções sobre o índice S&P 500, e seus resultados têm apontado superioridade

desses modelos em comparação com os da fórmula analítica de Black-Scholes. Para o Reino Unido, Niranjani (1996), Freitas et al. (2000), Healy et al. (2002) e Bennel e Sutcliffe (2004) apontaram melhor performance do método de rede neural para precificar as opções sobre o índice FTSE 100, em contraposição aos modelos tradicionais de valoração.

Winne et al. (2001), por exemplo, implementaram uma RNA para precificar as opções sobre o índice CAC 40, comercializadas no mercado de derivativos francês. Os resultados do método aplicado foram mais acurados que os do modelo binomial para essas opções. Anders et al. (1998), com base nas opções de compra de estilo europeu para o índice DAX, confirmaram a superioridade do modelo de rede neural sobre o modelo de Black-Scholes para avaliação dessas opções no mercado alemão, assim como Ormoneit (1999) e Krause (1996). Yao et al. (2000), ainda, utilizaram uma RNA para precificar as opções sobre futuros do índice Nikkei 225, que são de estilo americano, e verificaram seu melhor desempenho comparado com as equações analíticas de BS. Por fim, Amilon (2001) comparou a performance de um modelo de rede neural com o método de Black-Scholes para avaliar as opções de compra sobre o índice OMX, principal contrato de opções transacionado no mercado da Suécia. Baseado tanto em volatilidade histórica como implícita, o modelo de rede neural foi, em geral, superior.

Para o mercado brasileiro, Freitas e Souza (2002) avaliaram o modelo de redes neurais para precificar opções de compra européias sobre a ação da Telebrás PN, no período de janeiro de 1995 a dezembro de 1999. A rede neural aplicada pelos autores apresentou resultados superiores em comparação ao modelo de Black-Scholes, especialmente quando as *calls* estavam *in-the-money*. Maciel et al. (2009) também apresentaram um modelo de RNA multicamadas para avaliar as opções de taxa cambial R\$/US\$ comercializadas na BM&FBOVESPA, e mostraram que o modelo avaliado superou os resultados obtidos pela fórmula analítica de Black (1976), utilizada para precificação dessas opções na Bolsa brasileira. Entretanto, ainda não existe trabalho na literatura que apresente um estudo para valoração de opções sobre taxa de juros no Brasil com base em modelos de inteligência artificial, o que confirma a relevância deste trabalho.

Capítulo 3

Modelos de Taxa de Juros

Os modelos indiretos para precificação de opções de taxa de juros serão descritos neste capítulo. Além disso, serão apresentadas as características das opções de IDI, bem como os modelos derivados para sua valoração.

3.1 Precificação Livre de Arbitragem

A avaliação do preço de uma opção de compra de um título pré-fixado e outros tipos de opções sobre taxas de juros se baseia em metodologias que utilizam a precificação livre de arbitragem (*"arbitrage-free pricing"*), também conhecida como avaliação neutra ao risco. Assim, considera-se como dados os preços de um conjunto de ativos primários, assim como sua evolução estocástica, para então precificar um conjunto de ativos secundários Neto e Pereira (1999).

Os ativos secundários são precificados por meio da construção de uma carteira composta por ativos primários, dinamicamente rebalanceada no tempo, de tal modo que o fluxo de caixa e o valor dessa carteira replicam o fluxo de caixa e o valor do ativo secundário. Dessa forma, as opções são apreçadas num mundo neutro ao risco.

Como afirmado no capítulo anterior, os modelos de estrutura a termo da taxa de juro podem ser de abordagem direta ou indireta. Os modelos de abordagem direta partem da especificação direta do processo estocástico dos preços dos títulos primários, em vez de derivarem este processo de conceitos mais primitivos, como taxa de juro de curto prazo ou a relação de preferência dos agentes econômicos. Já os modelos de abordagem indireta partem da especificação do processo estocástico seguido pela taxa de juro de curto prazo ou pela taxa *forward*.

Por ser de interesse deste trabalho e além de ser aplicada pela maior parte das pesquisas sobre a estrutura a termo da taxa de juro, descreveremos apenas os modelos indiretos¹. Esses modelos têm dependido da incorporação dos principais fatos estilizados que procuram descrever a dinâmica evolutiva da ETTJ. Usualmente, eles se baseiam num processo estocástico para a taxa de juros de curto prazo num mundo neutro ao risco. De acordo com a consideração das variáveis (ou fatores) estocásticas que compõem o modelo, eles são chamados de modelos de n fatores. Os modelos estudados neste trabalho consideram como fator estocástico apenas a taxa de juros de curto prazo, por isso são conhecidos como modelos de um fator.

3.2 Modelos Indiretos

Os modelos indiretos apresentam duas abordagens distintas: modelos indiretos endógenos (ou modelos de equilíbrio geral) e modelos indiretos exógenos (ou modelos de não-arbitragem). Nos modelos de equilíbrio geral, os parâmetros estimados não necessariamente geram preços de títulos adequados à estrutura a termo vigente no mercado real. Enquanto que os modelos de não-arbitragem permitem que a estimação obtida gere uma curva de juros automaticamente ajustada à curva de juros do mercado Neto e Pereira (1999).

3.2.1 Modelos Indiretos Endógenos

Modelo de Merton (1973)

Em seu artigo clássico, *Theory of Rational Option Pricing*, em uma nota de rodapé, Merton introduz o primeiro modelo sobre a estrutura a termo onde os preços dos títulos são obtidos em função do processo estocástico seguido pela taxa de juros de curto prazo. Dessa forma, a equação diferencial estocástica descrita pela taxa de juros de curto prazo é:

$$dr(t) = \theta dt - \sigma dW(t) \quad (3.1)$$

em que $r(t)$ representa a taxa de juros de curto prazo, θ a taxa de juros de longo prazo, σ a volatilidade da taxa de juros de curto prazo e $W(t)$ é um processo de Wiener unidimensional.

¹ Para estudo dos principais modelos diretos, verificar os trabalhos de Ball e Tóuross (1983) e Munnik (1992).

Neste modelo, o processo estocástico suposto para a taxa de juros de curto prazo permite que ela assuma valores negativos, por isso, conforme afirma Merton (1973), trata-se de uma suposição irrealista. Além disso, se $\theta > 0$, a taxa tende a crescer ilimitadamente com o passar do tempo. Com isso, o modelo deve ser considerado como um exemplo didático, servindo como instrumento utilizado no desenvolvimento de modelos mais realistas.

Toda a incerteza da economia está contida no processo de Wiener unidimensional $W(t)$, sendo a única fonte de incerteza capaz de afetar o movimento do preço de um título referenciado à taxa de juro de curto prazo.

No modelo de Merton (1973), a estrutura a termo é igual a soma da taxa de juros de curto prazo $r(t)$ com uma função quadrática côncava do prazo para o vencimento do título referenciado. O termo estocástico $r(t)$ possui distribuição normal, fazendo assim com que a estrutura a termo também seja normalmente distribuída e toda sua evolução seja condicionada à evolução de $r(t)$.

Modelo de Black (1976)

Posteriormente, o modelo de Black (1976) foi desenvolvido para apreçar opções sobre futuros, e é bastante aplicado para avaliação de opções sobre títulos pré-fixados. Trata-se de uma variação do modelo de Merton (1973), em que o ativo objeto se trata de um contrato futuro, ao invés de um ativo à vista. A equação diferencial estocástica nesse modelo pode ser descrita como:

$$dF(t) = \sigma dW(t) \quad (3.2)$$

onde $F(t)$ é o preço futuro do ativo.

Em opções sobre taxa de juros, o preço do contrato futuro é substituído pelo preço a termo (*forward*) do título². Nesse modelo supõe-se que o preço do ativo objeto segue uma distribuição log-normal e não se pressupõe a existência de reversão à média. Além disso, assume taxas de juros nominais negativas e volatilidade constante para o ativo objeto que, no caso de títulos pré-fixados, apresenta-se mais irreal, pois quanto maior for o tempo até o vencimento, maior será a volatilidade³.

Modelo de Vasicek (1977)

²Essa suposição é questionável, mas compensa integralmente a hipótese de que as taxas de juro são constantes para efeito de desconto. Para mais detalhes, ver Hull (2003).

³Apenas no caso de a vida da opção ser muito curta em relação à vida do título-objeto é que se poderia considerar uma volatilidade constante ao longo da vida da opção.

Vasicek (1977) seguiu o mesmo caminho de Merton (1973) com modelos “gaussianos” sobre a estrutura a termo em tempo contínuo, porém, envolvendo uma álgebra mais complexa. Nesse caso, o processo estocástico seguido pela taxa de juros de curto prazo é dado por:

$$dr(t) = \alpha(\theta - r(t)) - \sigma dW(t) \quad (3.3)$$

onde o parâmetro α representa a velocidade de reversão a média da taxa de juros de curto prazo.

Vasicek (1977) foi o primeiro a propor o comportamento de reversão à média, fato estilizado do comportamento das taxas de juros de curto prazo. Isso significa que se a taxa de curto prazo está acima de uma média dita de longo prazo, a tendência dela é de queda. Se estiver abaixo da média esperada de longo prazo, a tendência é de alta.

O modelo de Vasicek procura ser mais realista que o de Merton, uma vez que agora a taxa de juros de curto prazo não mais possui uma tendência de crescimento ilimitada. Sempre que $r(t)$ estiver acima de θ , o termo $\alpha(\theta - r(t))$ será negativo e “ $\alpha(\theta - r(t))$ ” “puxará” a taxa para baixo Neto e Pereira (1999). Porém, quando $r(t)$ estiver abaixo de θ , o oposto se verifica.

Apesar de solucionado este problema encontrado em Merton, o modelo de Vasicek ainda continua a permitir taxas de juros negativas, uma vez que a distribuição condicional de $r(t)$ é normal⁴. Além disso, a volatilidade da taxa de juros de curto prazo é considerada constante, o que não acontece com dados reais.

A partir da equação diferencial estocástica, Vasicek obteve uma expressão analítica para o preço de um título pré-fixado. Jamshidian (1989), posteriormente, desenvolveu uma fórmula de precificação de opções européias sobre títulos pré-fixados usando o modelo de difusão de Vasicek.

Modelo de Cox, Ingersoll e Ross (CIR)(1985)

O modelo de Cox et al. (1985), ou modelo CIR, permite a solução do problema do *crescimento ilimitado da taxa de juros*, verificado em Merton, como a questão da *possibilidade de taxas de juros negativas*, presente em Merton e Vasicek. Em contrapartida, segundo Neto e Pereira (1999), a tratabilidade analítica dos dois modelos anteriores é perdida, pela sua complexidade.

No modelo CIR, a equação diferencial estocástica que descreve o movimento da taxa de juros de curto prazo é:

⁴Entretanto, assim como o modelo de Merton, ele apresenta grande tratabilidade analítica.

$$dr(t) = \alpha(\theta - r(t)) - \sigma\sqrt{r(t)}dW(t) \quad (3.4)$$

Assim como em Vasicek, as taxas de juros tendem a gravitar em torno de θ , caracterizando a *taxa de juros de curto prazo esperada no longo prazo*.

Pode-se demonstrar, Feller (1951), que se $2\alpha\theta \geq \sigma^2$ e $r(t) > 0$, então $r(t)$ nunca será negativo. Como a volatilidade da taxa de juros de curto prazo é $\sigma\sqrt{r(t)}$, sempre que $r(t)$ se aproximar de zero, sua raiz quadrada será um número muito pequeno, tornando a volatilidade desprezível. Dessa forma, a magnitude do termo " $\alpha(\theta - r(t))$ " será dominante, fazendo com que $r(t)$ se afaste da origem e seja sempre positivo⁵.

Outros Modelos

Podemos citar, ainda, o modelo de Chan et al. (1992), ou modelo CKLS. Ele possui características tais como reversão à média, e permite que a volatilidade da taxa de juros dependa de uma potência γ da própria taxa de juros. A equação diferencial estocástica para o modelo CKLS é:

$$dr(t) = \alpha(\theta - r(t)) - \sigma r(t)^\gamma dW(t) \quad (3.5)$$

Pode-se perceber que os modelos de Vasicek e CIR são casos particulares do modelo CKLS. Podemos citar, ainda, outros três modelos que são casos particulares do CKLS: Dothan (1978), Brennan e Schwartz (1982) e Constantinides e Ingersoll (1984)⁶

3.2.2 Modelos Indiretos Exógenos

Os modelos de equilíbrio geral apresentados apresentam uma limitação em comum: não se ajustam automaticamente à estrutura a termo da taxa de juros observada no mercado. Essa característica decorre do fato de que os parâmetros estimados com base em preços passados não irão gerar necessariamente uma estrutura a termo ajustada ao mercado real. Com isso, surgiram os modelos indiretos exógenos, ou de não-arbitragem. Eles foram desenvolvidos com o intuito de se gerar uma estrutura a termo que seja consistente com a estrutura a termo atual do mercado.

⁵Para mais detalhes desses modelos, ver Munnik (1992).

⁶Os modelos de Dothan (1978) e Constantinides e Ingersoll (1984) não possuem reversão à média, mas a volatilidade aumenta com o nível da taxa de juros. O modelo de Brennan e Schwartz (1982) já possui a reversão à média, e se distingue do modelo CIR apenas pelo $\gamma = 1$, ao invés de 0,5.

Modelo de Ho e Lee (1986) (Merton Extendido)

Originalmente desenvolvido em tempo discreto, na forma de uma árvore binomial, o modelo de Ho e Lee (1986) foi transformado em um modelo em tempo contínuo bem definido conforme o intervalo de tempo entre cada instante da árvore tende a zero, dessa forma, a equação diferencial estocástica seguida pela taxa de juros de curto prazo é:

$$dr(t) = \frac{\partial \theta(t)}{\partial t} dt - \sigma dW^Q(t) \quad (3.6)$$

em que $\theta(t)$ é uma função determinística, e $W^Q(t)$ é um processo de Wiener unidimensional, com espaço de probabilidade da medida martingal equivalente Q .

A função $\theta(t)$ é escolhida de tal forma que a estrutura a termo gerada pelo modelo, na data inicial, seja exatamente a mesma à estrutura a termo observada no mercado. Conforme já foi evidenciado, esta é a característica que distingue os modelos indiretos exógenos dos endógenos⁷. O modelo de Ho-Lee foi o precursor dos modelos exógenos, tendo grande importância conceitual e teórica, dando partida a uma linha de pesquisa original e promissora.

Este modelo permite a derivação de fórmulas fechadas para o preço de opções sobre títulos. Uma de suas limitações é não permitir a reversão à média. Entretanto, o grande avanço deste modelo foi permitir que o parâmetro θ , ou chamado *drift*, varie com o tempo. Na realidade, o modelo Ho-Lee é basicamente o modelo de ? com *drift* variável no tempo.

Modelo de Hull e White (1990) (Vasicek Extendido)

Hull e White (1990) propuseram um modelo na busca de superar alguns problemas verificados no modelo de Ho-Lee, da mesma forma com que o modelo de Vasicek fez com o modelo de Merton⁸. Com isso, o processo estocástico seguido pela taxa de juros de curto prazo é definido por:

$$dr(t) = (\theta(t) - \alpha r(t)) dt + \sigma dW^Q(t) \quad (3.7)$$

Assim como no modelo de Ho-Lee, a função $\theta(t)$ é escolhida de forma que a estrutura a termo gerada pelo modelo, na data inicial, seja exatamente igual à estrutura a termo verificada no mercado⁹.

⁷Para ver de que forma a função $\theta(t)$ é determinada, ver Neto e Pereira (1999).

⁸O trabalho de Hull e White (1990) pode ser caracterizado como "Ho e Lee com reversão à média" ou como "Vasicek com estrutura a termo inicial exógena".

⁹Em geral, esse procedimento é realizado por intermédio da taxa *forward*.

Podemos dizer que o modelo de Hull-White é o modelo de Vasicek com uma taxa de juros de longo prazo, $r(t)$, que varia com o tempo. Ou, ainda, trata-se de um modelo de Ho-Lee com a introdução da reversão à média, ou seja, se o parâmetro (α) for zero, o modelo de Hull-White se transforma no de Ho-Lee.

A propriedade de normalidade da taxa de juros de curto prazo indica uma distribuição log-normal para os títulos. Dessa forma, pode-se derivar fórmulas fechadas para os prêmios de opções européias¹⁰.

Modelo de Heath, Jarrow e Morton (1992)

O mais geral dos modelos precificação por não-arbitragem é o de Heath et al. (1992), doravante HJM. Os autores desenvolveram uma metodologia geral que engloba grande parte dos modelos sobre a estrutura a termo da taxa de juros¹¹. O modelo HJM não parte de uma equação diferencial para a taxa de juros, mas de uma equação diferencial estocástica para o preço de um título pré-fixado:

$$dP(t, T) = r(t) P(t, T) dt + \sigma P(t, T) W^Q(t) \quad (3.8)$$

onde $P(t, T)$ representa o preço de título *zero cupom* negociado em t e com maturidade em T , sendo $t < T$.

No caso geral, o modelo HJM gera uma árvore que é não-recombinante, isto é, um movimento para cima seguido de um movimento para baixo não leva ao mesmo estado de movimento para baixo seguido de um movimento para cima¹².

O processo da taxa de juro de curto prazo do modelo HJM é não-Markov. Dessa forma, para se obter o comportamento estocástico de $r(t)$ em um determinado período pequeno no futuro deve-se saber não só o valor de r no começo do período, mas também o caminho descrito pela taxa até chegar naquele valor Barbachan e Ornelas (2003)¹³.

Modelo de Black, Derman e Toy (BDT) (1990)

O modelo de Black et al. (1990), BDT em diante, também pode ser considerado um caso particular

¹⁰No mesmo artigo, Hull e White (1990) também estenderam o modelo CIR para se ajustar à estrutura a termo inicial, de maneira análoga ao que fizeram com o Vasicek.

¹¹Neste caso, os modelos de Ho-Lee e Hull-White podem ser obtidos como casos particulares de um modelo HJM. Para este tópico, ver Neto e Pereira (1999).

¹²Essa característica torna o modelo custoso em termos computacionais.

¹³Nos modelos de Ho e Lee e Hull e White, casos particulares do modelo HJM, pode-se obter um movimento Markov para r , economizando esforço computacional.

do modelo HJM, porém, tem-se distribuição log-normal para a taxa de juros de curto prazo, assim como o fenômeno de reversão à média. A equação diferencial estocástica que descreve o processo no modelo BDT é:

$$d \ln(r(t)) = (\theta(t) - \alpha \ln(r(t))) dt + \sigma dW^Q(t) \quad (3.9)$$

Ao se utilizar uma distribuição log-normal para a taxa de juros de curto prazo, o problema da negatividade da taxa é solucionado.

A partir do modelo BDT não é possível chegar a uma fórmula fechada para o preço de uma opção européia. Pois, para se chegar ao prêmio da opção, é necessária a geração de uma árvore binomial das taxas de juros *spot*, do presente para o futuro. Nesse caso, sendo r um processo de Markov, a árvore gerada é recombinante (ao contrário do modelo HJM em geral), o que reduz o número de nós da árvore, economizando o esforço computacional. Desenhando a árvore de taxas *spot* do futuro para o presente, constrói-se, então, as árvores dos preços dos títulos e das opções utilizando avaliação neutra ao risco.

3.3 Precificação das Opções de IDI no Brasil

Os modelos apresentados anteriormente são base da teoria de avaliação de títulos contingentes em mercados completos e livres de arbitragem, e, por meio deles se deriva uma solução analítica simples para o cálculo do preço de uma opção sobre taxa de juros. Deve-se perceber, entretanto, que a literatura desenvolveu fórmulas analíticas fechadas para o preço de opções de taxa de juros com base nas características dos instrumentos transacionados no mercado norte-americano e europeu. Como as opções de taxa de juros típicas desses mercados são substancialmente diferentes da opção de IDI, esses modelos devem ser adequados à realização brasileira.

Os contratos de opções de taxa de juros padrão dos EUA e Europa possuem como ativo objeto um título de renda fixa que vencerá em determinado número de meses depois do vencimento da opção. As opções de IDI, contudo, ao invés de ter como ativo objeto um título de renda fixa com vencimento superior ao seu, elas negociam a taxa de juros acumulada entre a data corrente e a sua própria data de vencimento. Portanto, a forma específica das opções sobre taxa de juro brasileira torna seu preço, bem como os fatores que o afetam, distinto daqueles ligados às opções sobre juros tradicionais.

Uma opção de IDI envolve as mesmas variáveis financeiras de um contrato de *swap* de taxa de juros, mas possui um padrão de comportamento distinto. O índice DI (IDI) foi definido pela

BM&FBOVESPA como sendo o valor de 100.000 corrigido pela *taza média de depósitos interfinanceiros e 1 dia*, o CDI, entre uma determinada data inicial e a data corrente. Em notação matemática, temos:

$$IDI_t = 100.000 \cdot \prod_{j=t_0}^{t-1} (1 + CDI_j) \quad (3.10)$$

em que IDI_t indica o valor do IDI na data t e CDI_t o valor do CDI na data t , convertido para percentual ao dia.

Uma opção de compra sobre IDI permite ao seu titular o direito de “comprar”, na data de maturidade da opção (T), o IDI ao preço de exercício da opção (X). Dessa forma, o valor de liquidação do contrato é:

$$VL_c = \max \{0, IDI_t - X\} \quad (3.11)$$

Analogamente, uma opção de venda permite ao seu titular o direito de “vender” o IDI ao preço de exercício da opção. Assim, o valor de liquidação será:

$$VL_c = \max \{0, X - IDI_t\} \quad (3.12)$$

As opções de compra (*calls*) e de venda (*puts*) são ambas de estilo europeu.

Dessa forma, as subseções seguintes apresentam os modelos de valoração utilizados para a precificação das opções sobre IDI no Brasil.

3.3.1 Modelo de Black para as Opções de IDI

Utilizado pela BM&FBOVESPA, o modelo de Black (1976) destina-se a opções sobre futuros, e é bastante utilizado para a avaliação de opções sobre títulos pré-fixados. Este modelo de Black é uma variação do modelo de Black e Scholes (1973), com o ativo objeto sendo um contrato futuro ao invés de um ativo à vista.

Para utilizar o modelo de Black em opções sobre taxa de juros usa-se o preço a termo do título no lugar do preço do contrato futuro. Assim como o modelo de BS, supondo que o ativo-objeto segue uma distribuição log-normal. Nesse caso, não se pressupõe a existência de reversão à média.

Assim como a maioria dos modelos, este considera a volatilidade do ativo-objeto como sendo constante ao longo da vida da opção. Entretanto, no caso dos títulos pré-fixados, quanto maior for

o tempo até o vencimento, maior será a volatilidade. No caso da vida da opção ser muito curta em relação à vida do título-objeto é que poderíamos considerar uma volatilidade praticamente constante durante a vida da opção. Além disso, vale acrescentar que o modelo de Black (1976) assume taxas de juros nominais negativas.

Para a opção de IDI, a Bolsa utiliza o modelo de Black. Neste caso, o ativo-objeto é representado pelo valor do IDI corrigido pela taxa de juros esperada (DI Futuro) até a data do vencimento da opção, ou seja, cria-se um preço a termo do IDI. O problema da volatilidade decrescente é geralmente amenizado por meio da utilização de uma volatilidade “média”. A fórmula analítica para uma opção de compra sobre IDI na data t é dada por:

$$c_t = IDI_t N(d_1) - XP(t, T) N(d_2) \quad (3.13)$$

Em que:

$$d_1 = \frac{\ln \left(IDI_t / XP(t, T) \right) + \frac{\sigma^2 (T-t)^3}{6}}{\sigma \sqrt{(T-t)^3/3}} \quad (3.14)$$

$$d_2 = \frac{\ln \left(IDI_t / XP(t, T) \right) - \frac{\sigma^2 (T-t)^3}{6}}{\sigma \sqrt{(T-t)^3/3}} \quad (3.15)$$

onde $P(t, T)$ é o preço em t de um título pré-fixado com vencimento em T ; IDI_t é o valor do IDI em t ; e σ é a volatilidade de um título pré-fixado de curto prazo.

Este é o modelo mais utilizado no mercado brasileiro para a valoração das opções de IDI no Brasil, apesar de suas inconsistências teóricas. Por isso, ele será aplicado neste trabalho como um dos modelos que representam o *benchmark* para os métodos de precificação.

3.3.2 Modelo de Vasicek para as Opções de IDI

Vasicek (1977) foi o primeiro a propor o comportamento de reversão à média, fato estilizado das taxas de juros de curto prazo. Isso significa que se a taxa de curto prazo está acima de uma média dita de longo prazo, a tendência dela é de queda. Se estiver abaixo da média esperada de longo prazo, a tendência é de alta. Com base na equação 3.3 para o processo estocástico de difusão da taxa de juros de curto prazo, Neto e Pereira (1999) propuseram uma fórmula analítica fechada para precificação das opções de IDI. Segundo os autores, o prêmio de uma opção de compra de IDI (c_t) é dado por:

$$c_t = IDI_t N(h) - XP(t, T) N(h - k) \quad (3.16)$$

Sendo:

$$h = \frac{\ln\left(\frac{IDI_t}{P(t, T)X}\right) + \frac{k^2}{2}}{k} \quad (3.17)$$

$$k^2 = \sigma_t^2 \frac{(4e^{-\alpha\tau} - e^{2\alpha\tau} + 2\alpha\tau - 3)}{2\alpha^3} \quad (3.18)$$

em que $N(\cdot)$ denota a função de distribuição normal padrão acumulada; IDI_t indica o valor do IDI em t ; X é o preço de exercício da opção; $P(t, T)$ é o preço, em t , de um título de renda fixa que paga 1 unidade monetária em T ¹⁴; α é o parâmetro de reversão à média da taxa de juros de curto prazo; $\tau = T - t$, ou seja, o prazo de maturidade opção; e σ_t^2 indica a volatilidade da taxa de juros no curto prazo.

3.3.3 Modelo CIR para as Opções de IDI

O modelo descrito por Vasicek (1977) apresenta uma limitação ao permitir que a taxa de juros assuma valores negativos. Entretanto, no modelo CIR, a volatilidade da taxa de juros de curto prazo é $\sigma_t\sqrt{r(t)}$, assim, sempre que $r(t)$ se aproximar de zero, sua raiz quadrada será um número muito pequeno, tornando a volatilidade desprezível, então, a magnitude do termo " $\sigma_t\sqrt{r(t)}$ " será dominante, fazendo com que $r(t)$ se afaste da origem e seja sempre positivo. Com base nesse processo de difusão para a taxa de juros de curto prazo, Barbachan e Ornelas (2003) construíram uma forma analítica fechada para o cálculo do prêmio de uma opção de compra de IDI (c_t). Nesse caso, temos que:

$$c_t = \psi IDI_t \left[1 - \chi^2 \left(\ln\left(\frac{X}{IDI_{t-1}}\right); \eta, \rho \right) \right] - X \psi 3^{\frac{-\eta}{2}} e^{-\mu} \left[1 - \chi^2 \left(\ln\left(\frac{X}{IDI_{t-1}}\right) / 3; \eta, \mu \right) \right] / (1 + CIDI_t) \quad (3.19)$$

Em que:

$$\rho = \frac{4CIDI_t}{\sigma_t^2} \left(\ln\left(\frac{1 - e^{\theta^* \tau}}{1 - e^{-\theta^*}}\right) \right) \quad (3.20)$$

¹⁴Naturalmente, $P(T, T) = 1$.

$$\eta = \frac{4\theta^*\alpha^*\Delta t}{\sigma_2^t} \quad (3.21)$$

$$\psi = \frac{\sigma_2^t}{4\theta^*} + \frac{\sigma_2^t}{4\Delta t e^{\theta^*}} (e^\tau - 1) \quad (3.22)$$

sendo $\theta^* = (\theta - \lambda)$; $\alpha^* = \left(\frac{\alpha\theta}{\theta - \lambda}\right)$; e $\mu = \frac{\rho\psi}{2+\psi}$, em que α é o parâmetro da velocidade de reversão à média da taxa de juros; θ a taxa de juros de longo prazo; σ^2 a volatilidade da taxa de juros de curto prazo; $\tau = T - t$; λ o preço de mercado do risco; $\Delta t = T - t - \xi^{15}$; χ^2 denota uma distribuição chi-quadrado não central.

Dessa forma, esses modelos (Black, Vasicek e CIR) compõem o *benchmark* deste trabalho para obtenção dos prêmios das opções de compra de IDI no período considerado.

¹⁵Segundo Barbachan e Ornelas (2003) adota-se $\xi = \frac{1}{252}$, que corresponde a um dia útil dividido pelo total de dias úteis durante o período de um ano regular.

Capítulo 4

Redes Neurais Artificiais

O objetivo deste capítulo é apresentar um panorama geral dos modelos de redes neurais artificiais, abordando suas características, processos de treinamento, assim como suas arquiteturas a serem utilizadas neste trabalho (*feedforward*) e recorrentes.

4.1 Introdução

Os métodos baseados em redes neurais foram estimulados na tentativa de se definir um modelo computacional para a forma de processamento de informação do cérebro humano. Sua origem remonta o trabalho de Warren McCulloch e Walter Pitts de 1943 intitulado “*A Logical Calculus of the Ideas Immanent in Nervous Activity*”, em que se apresentou o chamado neurônio booleano de MacCulloch Haykin (1999). O neurônio artificial de McCulloch foi o primeiro modelo computacional para o neurônio biológico. Ele tinha sua saída calculada com base em suas variáveis de entrada, gerando uma resposta determinada pela soma ponderada das entradas, sendo esses fatores de ponderação denominados de “ganhos sinápticos”. A Figura 4.1 apresenta um modelo não-linear de um neurônio artificial Haykin (1999).

A saída do neurônio pode ser descrita da seguinte forma:

$$y_i = \varphi \left(\sum_{j=1}^m w_{ij} x_j \right) \quad (4.1)$$

onde $x_1, x_2, \dots, x_m$ são os sinais de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio i ; u_i é a saída do combinador linear devido aos sinais de entrada; $\varphi(\cdot)$ é a função de ativação; e y_i é o

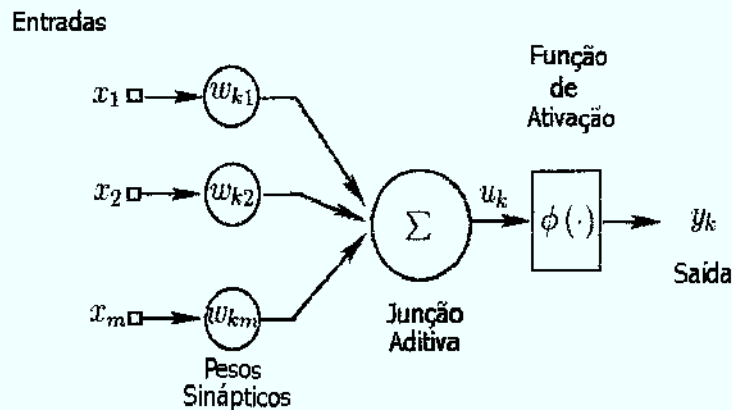


Fig. 4.1: Modelo não-linear de um neurônio artificial

sinal de saída do neurônio.

Com base nas idéias de McCulloch, as redes neurais artificiais, semelhantes às que conhecemos hoje, apareceram pela primeira vez na década de 1950. Nessa década, Frank Rosenblatt criou o chamado *perceptron*, uma rede de múltiplos neurônios. Nesta rede, os neurônios estão dispostos em diversas camadas. Os neurônios que recebem as entradas constituem a primeira camada (camada de entrada). Os neurônios que recebem as saídas dos neurônios localizados na camada de entrada formam a segunda camada, e assim sucessivamente, até a última camada (camada de saída). Uma grande vantagem em relação ao modelo de McCulloch e Pitts é que o *perceptron* de Rosenblatt pode lidar com estímulos de valores contínuos, e não apenas binários.

As maiores contribuições ao longo da década de 1970 foram referentes a topologias e utilizações de redes neurais. No entanto, o maior impulso para o estabelecimento das redes neurais viria apenas em 1986. David Rumelhart desenvolveu um método para treinar redes neurais multicamadas chamado *algoritmo de retropropagação* e resolveu o problema pendente desde o fim da década de 1950, no tocante à estimação de pesos de redes neurais, se tornando um dos métodos mais utilizado para treinamento de perceptrons multicamadas. Desde então, muitos trabalhos foram feitos para melhorar o algoritmo de retropropagação, tais como evitar sobre-ajuste dos dados, acelerar a velocidade do algoritmo e evitar mínimos locais Haykin (1999).

Em geral, as redes neurais podem executar diferentes tarefas: classificação ou reconhecimento, associação ou memorização, codificação, predição e simulação. Há dois tipos de aprendizagem de redes neurais: aprendizagens supervisionada e não-supervisionada. Aprendizagem supervisionada é

utilizada quando há um valor alvo (desejável) associado com cada entrada do conjunto de treinamento (amostra utilizada para ajuste dos pesos). A saída da rede é comparada com o valor alvo e sua diferença é utilizada para alterar os pesos. Aprendizagem não-supervisionada ocorre quando não há valores alvo para a saída da rede. Nesse caso, a rede aprende através de correlações entre os padrões das variáveis de entrada Haykin (1999). Como o objetivo deste trabalho é predição, apenas a aprendizagem supervisionada será utilizada.

4.2 Rede Perceptron Multi-Camadas

A rede perceptron multi-camadas (*Multilayer perceptrons - MLP*) se trata de uma generalização do perceptron proposto por Roseblatt, sendo uma das mais conhecidas e aplicadas arquiteturas de redes neurais.

A topologia deste modelo consiste de uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída. A camada de entrada, cujos neurônios pertencentes a esta camada são denominados unidades de entrada, o qual difunde as entradas para as camadas seguintes sem nenhuma modificação. As camadas intermediárias transmitem informações através das conexões entre as camadas de entrada e saída. Nessa camada estão os chamados neurônios intermediários. Por fim, na camada de saída, cujos neurônios são chamados de unidades de saídas, transmite-se a resposta da rede neural à entrada aplicada na camada de entrada. A figura 4.2 apresenta uma estrutura de rede neural multi-camadas com duas camadas intermediárias.

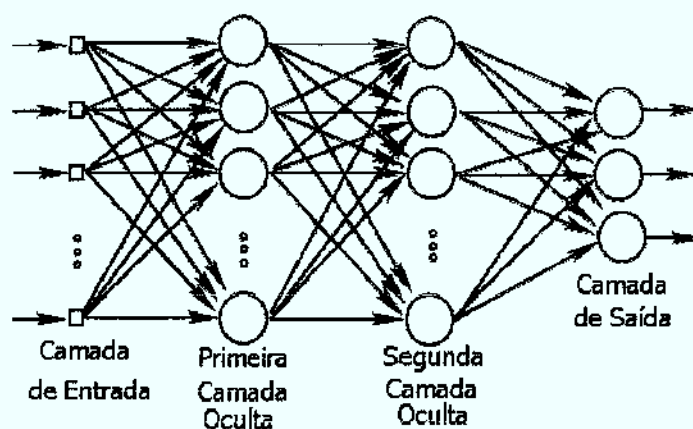


Fig. 4.2: Topologia multi-layer perceptrons com duas camadas ocultas

Os neurônios estão conectados por arcos e a cada arco está associado um peso. A comunicação entre os neurônios é realizada através das interconexões ponderadas. Porém, entre unidades que estão na mesma camada a comunicação entre elas não se verifica.

As redes tipo MLP se distiguem quanto as seguintes características: número de camadas e neurônios intermediários na rede, tipo de conexão determinada pelas sinapses e função de ativação.

O treinamento do modelo MLP é supervisionado, isto é, um “professor” indica a resposta desejada para o padrão de entrada apresentado à rede durante seu aprendizado. Define-se então como sinal de erro a diferença entre a resposta desejada e a resposta observada. Este sinal, por consequência, é a base para o ajuste dos parâmetros da rede (pesos e limiares). A adaptação da rede se baseia em iterações sucessivas para ajuste dos pesos, o qual pode ser expresso na forma de um método de aprendizagem.

Para o treinamento das redes de tipo MLP, o método de aprendizagem mais utilizado é o algoritmo de *retropropagação do erro*. Esse método compõe-se por duas fases. Na primeira fase, conhecida como fase *forward*, as entradas são apresentadas e propagadas através da rede, camada por camada, calculando a saída de cada neurônio. Nesta fase, os pesos são fixos e a saída é calculada e comparada com a saída desejada, resultando em um erro para cada unidade. A segunda fase, *backward*, caracteriza-se pela propagação do erro da camada de saída para a camada de entrada, em que os pesos agora são ajustados de acordo com a regra de correção do erro, o que explica a denominação “retro-propagação do erro” Ballini (2000).

4.3 Redes Neurais *Feedforward*

Um rede neural de topologia MLP, com ligações unidirecionais entre os neurônios (na direção das entradas para as saídas), é chamada de rede neural progressiva ou *feedforward*. A rede neural é um conjunto de unidades computacionais, ou nós computacionais, interconectadas e organizadas em camadas. As ligações entre os neurônios. As ligações entre neurônios, ou entre entradas e neurônios, têm, cada uma, um peso sináptico correspondente, denotado por w_{ij} (conexão do j -ésimo nó, ou j -ésima entrada, para o i -ésimo nó). A saída do j -ésimo neurônio, ou a j -ésima entrada, é multiplicada por w_{ij} e é uma entrada para o i -ésimo nó.

Cada neurônio tem uma função não-linear correspondente $\varphi(\cdot)$, pois ela determina a saída do neurônio. Essa função deve ser não-linear diferenciável e limitada. Dentre as mais utilizadas, podemos citar a função sigmóide, cujo gráfico tem a forma de s , e é a mais comum na construção de redes neurais artificiais. Ela é definida como uma função estritamente crescente que exibe um balanceamento

adequado entre comportamento linear e não-linear. Um exemplo de função sigmóide é a *função logística*, definida por:

$$\varphi(u) = \frac{1}{1 + e^{(-au)}} \quad (4.2)$$

onde a é o parâmetro de inclinação da função sigmóide.

Além disso, quando é desejável que a função se estenda de -1 a $+1$ na sua imagem, assumindo uma forma anti-simétrica em relação à origem, utiliza-se a função tangente hiperbólica, definida como:

$$\varphi(u) = \tanh(u) \quad (4.3)$$

A figura 4.3 apresenta exemplos de funções de ativação logística e tangente hiperbólica.

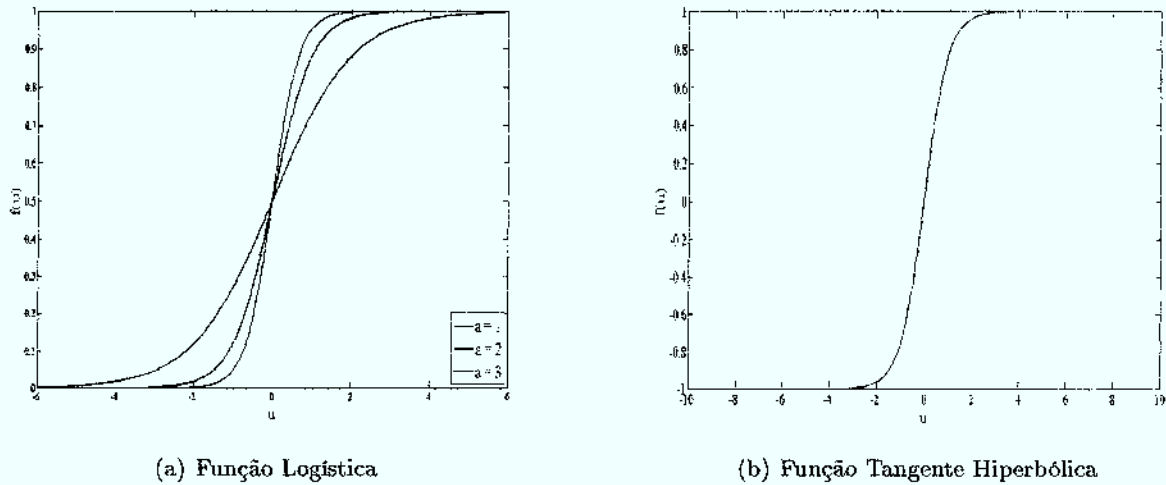


Fig. 4.3: Exemplos de Funções de Ativação

A saída do j -ésimo neurônio da camada intermediária é dada por $v_{pj} = \varphi \left(\sum_{i=1}^N w_{ji} x_{pi} \right)$, onde N é o número total de entradas, w_{ji} é o peso que relaciona a entrada i ao neurônio j , x_{pi} é o valor da p -ésima observação ("pattern") da i -ésima entrada.

O k -ésimo neurônio dos O neurônios da camada de saída tem como saída $\hat{y}_{pk} = r \left(\sum_{j=1}^Q W_{kj} v_{pj} \right)$ onde Q é o número de neurônios da camada anterior, W_{kj} é o peso que relaciona a saída do j -ésimo neurônio da camada intermediária do neurônio k da camada de saída e r é uma função de ativação. A fórmula geral para as saídas de uma rede *feedforward* com uma camada intermediária e uma camada de saída fica, então, para $k = 1, 2, 3, \dots, O$:

$$\hat{y}_{pk} = \sum_{j=1}^Q W_{kj} \varphi \left(\sum_{i=1}^N W_{ji} v_{pi} \right) \quad (4.4)$$

4.4 Redes Neurais Recorrentes

A estrutura de uma rede neural recorrente tem por base uma rede *feedforward* com algumas modificações, notadamente a introdução de realimentação, o que amplia sua potencialidade de modelagem de dados temporais ou espaciais. Uma rede *feedforward*, portanto, pode ser vista como um caso particular de uma rede recorrente. As realimentações consistem em saídas de neurônios de determinada camada serem reintroduzidas como entradas de neurônios de camadas anteriores ou da própria. Um neurônio pode ser realimentado por sua própria saída. Essas possibilidades fazem com que a arquitetura de uma rede recorrente possa tomar diversas formas.

No caso de tempo discreto, como utilizado aqui, as realimentações são acompanhadas de defasagens de uma ou mais unidades do tempo.

Com essa topologia de rede, ampliam-se as possibilidades de modelagem de estruturas de autodependência de dados em caso de séries temporais ou espaciais, não se limitando apenas às estruturas auto-regressivas padrão passíveis de representação por redes *feedforward*.

Para o caso de predição de séries temporais, duas das redes recorrentes mais utilizadas são as do tipo Elman e as do tipo Jordan Mandic e Chambers (2001). As saídas da camada intermediária (Elman) ou da camada de saída (Jordan) são defasadas e reintroduzidas como entradas da rede. As figuras 4.4 e 4.4 mostram tipos particulares dessas redes, com uma camada intermediária.

4.5 Algoritmo de Retro-Propagação do Erro

O algoritmo de aprendizagem por retropropagação do erro ("back-propagation") é um método de estimação dos pesos sinápticos de uma rede multi-camadas. Ele foi desenvolvido por Rumelhart et al. (1986). O objetivo do algoritmo nada mais é do que minimizar a soma de quadrados de erros.

Para efetuar essa aprendizagem, alguma medida de performance é necessária. A mais utilizada no caso de respostas gaussianas é a soma de quadrados de erros:

$$E = \frac{1}{2} \sum_p \sum_k \left(y_{pk} - \hat{y}_{pk} \right)^2 \quad (4.5)$$

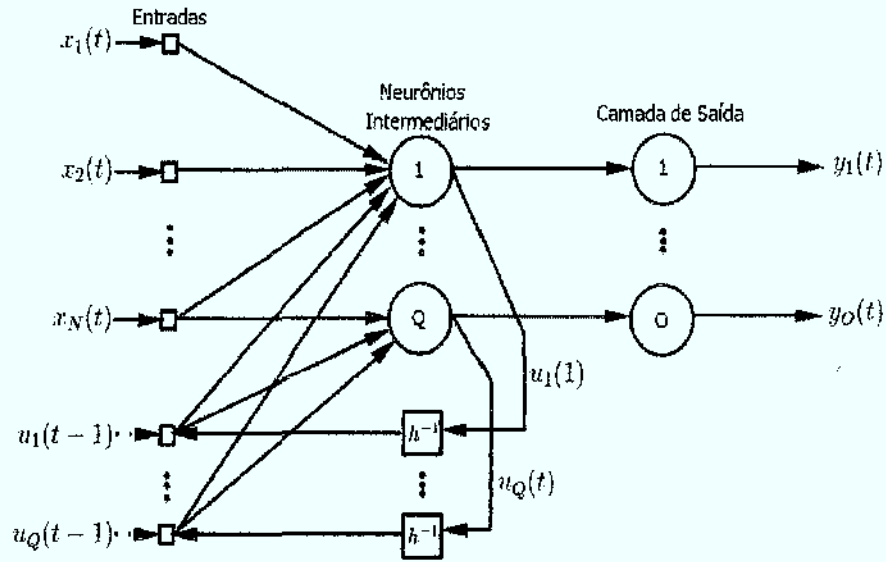


Fig. 4.4: Rede Neural Recorrente de Elman

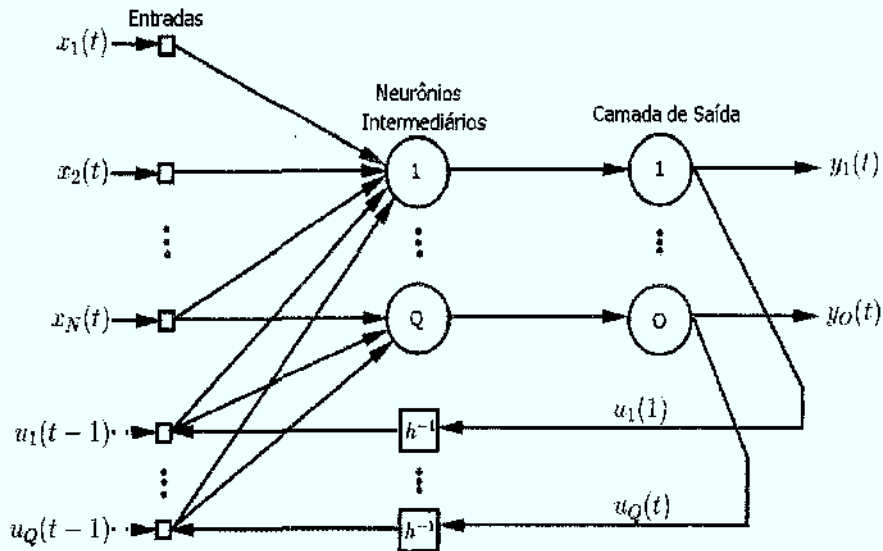


Fig. 4.5: Rede Neural Recorrente de Jordan

onde p é referente às observações, o índice k é referente aos nós de saída, y é a resposta desejada (alvo) e $\hat{y}$ é a saída da rede (valor ajustado). Os pesos da rede são ajustados de forma a minimizar esse erro.

A cada passo do algoritmo, ou seja, quando a observação p é processada, o objetivo é minimizar:

$$E_p = \frac{1}{2} \sum_k \left(y_{pk} - \hat{y}_{pk} \right)^2 \quad (4.6)$$

o que implica que $E = \sum_p E_p$.

Consideremos, para ilustração, uma rede com duas camadas (intermediária e saída). O objetivo é encontrar os pesos w_{ji} e W_{kj} que minimizem a função E . Para isso, utiliza-se o procedimento de minimização por gradiente descendente, percorrendo a rede da saída para a entrada, onde a correção dos pesos é definida pela regra delta Haykin (1999):

$$\Delta w_{kj} = -\eta \frac{\partial E_p}{\partial W_{kj}} \quad (4.7)$$

onde η é a taxa de aprendizagem. O sinal negativo da equação 4.7 é necessário para alterar o valor dos pesos na direção de uma redução no valor de E . Substituindo a equação 4.4 na equação 4.6, temos que:

$$E_p = \frac{1}{2} \sum_{k=1}^O \left(y_{pk} - \left(\sum_{j=1}^M W_{kj} \phi \left(\sum_{i=1}^M w_{ji} x_{pi} \right) \right) \right)^2 \quad (4.8)$$

Pela regra da cadeia, temos:

$$\frac{\partial E_p}{\partial W_{kj}} = \frac{\partial E_p}{\partial \hat{y}_{pk}} \frac{\partial \hat{y}_{pk}}{\partial f_{pk}} \frac{\partial f_{pk}}{\partial W_{kj}} \quad (4.9)$$

Podemos ver que:

$$\frac{\partial E_p}{\partial \hat{y}_{pk}} = - \left(y_{pk} - \hat{y}_{pk} \right) \quad (4.10)$$

$$\frac{\partial \hat{y}_{pk}}{\partial f_{pk}} = 1 \quad (4.11)$$

$$\frac{\partial f_{pk}}{\partial W_{kj}} = v_{pj} \quad (4.12)$$

Substituindo em 4.7, obtemos:

$$\Delta W_{kj} = \eta \left(y_{pk} - \hat{y}_{pk} \right) v_{pj} \quad (4.13)$$

Com isso, a fórmula para atualização dos pesos da camada de saída fica:

$$W_{kj}^{p+1} = W_{kj}^p + \Delta W_{kj} \quad (4.14)$$

Agora, calculamos, para os pesos da camada intermediária:

$$\Delta w_{ji} = -\eta \frac{\partial E_p}{\partial w_{ji}} \quad (4.15)$$

Pela regra da cadeia, obtemos

$$\frac{\partial E_p}{\partial w_{ji}} = \sum_{k=1}^O \frac{\partial E_p}{\partial \hat{y}_{pk}} \frac{\partial \hat{y}_{pk}}{\partial f_{pk}} \frac{\partial f_{pk}}{\partial v_{pj}} \frac{\partial v_{pj}}{\partial h_{pj}} \frac{\partial h_{pj}}{\partial w_{ij}} \quad (4.16)$$

Mas

$$\frac{\partial E_p}{\partial \hat{y}_{pk}} = -\left(y_{pk} - \hat{y}_{pk}\right) \quad (4.17)$$

$$\frac{\partial \hat{y}_{pk}}{\partial f_{pk}} = 1 \quad (4.18)$$

$$\frac{\partial f_{pk}}{\partial v_{pj}} = W_{kj} \quad (4.19)$$

$$\frac{\partial v_{pj}}{\partial h_{pj}} = \phi'(h_{pj}) \quad (4.20)$$

$$\frac{\partial h_{pj}}{\partial w_{ij}} = x_{pi} \quad (4.21)$$

Peas equações 4.15 e 4.16,

$$\Delta w_{ji} = \eta \sum_{k=1}^O \left(y_{pk} - \hat{y}_{pk}\right) W_{kj} \phi'(h_{pj}) x_{pi} \quad (4.22)$$

onde a parte interna da expressão acima (entre η e x_{pi}), por analogia com a equação 4.13 (onde $y_{pk} - \hat{y}_{pk}$ é o erro de saída) é o erro da camada intermediária, obtido por retropropagação do erro de saída.

4.5.1 Taxa de Aprendizagem

A taxa de aprendizagem η é uma constante de proporcionalidade, que assume valores no intervalo $(0, 1)$, e que determina a natureza de adaptação dos pesos.

Um alto valor de η resultará numa rápida aprendizagem da rede, decorrendo em elevadas alterações nos pesos. Todavia, o risco de instabilidade da rede aumenta, podendo resultar na saturação dos neurônios e oscilação ao redor da solução Ballini (2000).

Em contrapartida, valores pequenos de η indicam que as mudanças nos pesos sinápticos serão menores de uma iteração para outra e as trajetórias no espaço definido pelos pesos serão muito suaves. Porém, esse benefício é obtido à um alto custo no processo de treinamento, pois ele se torna mais lento por requerer muito mais iterações até a convergência Ballini (2000).

4.5.2 Modos de Treinamento

Na aplicação prática do algoritmo de retropropagação, o modo de treinamento pode variar de acordo com a maneira com que o conjunto de treinamento (amostra utilizada para a estimação dos pesos) é apresentado. Uma apresentação de todo o conjunto de treinamento durante o processo de aprendizagem é chamado de época e os pesos são alterados depois que o algoritmo processa uma observação ou todo o conjunto de treinamento. Assim, o algoritmo de retropropagação pode ser implementado de duas formas Haykin (1999):

- Modo sequencial: também referido como modo *on-line*. A atualização dos pesos é feita após a apresentação de cada uma das observações do conjunto de treinamento (forma que o algoritmo de retropropagação é apresentado nesta seção). Quando o conjunto de treinamento inteiro é processado, o treinamento é recommçado e o processo continua, até que seja interrompido através de um critério de parada.
- Modo por batelada (*batch*): a atualização dos pesos é feita após a apresentação de todas as observações do conjunto de treinamento, que constituem uma época. A cada passo do algoritmo, a função E (equação 4.5) é minimizada, ao invés de apenas E_p (equação 4.6). O conjunto de treinamento é apresentado ao algoritmo quantas vezes forem necessárias para a estimação dos pesos (o treinamento é interrompido através de um critério de parada).

Neste trabalho, o modo de treinamento utilizado será o *batch*.

4.5.3 Critério de Parada e Validação

Os métodos de otimização para encontrar os pesos são iterativos. Isso significa que, em um número grande de iterações, o valor da função E pode ser tão pequeno quanto se queira. Isso normalmente representa um bom ajuste dentro da amostra, mas em geral o modelo não conseguirá ajustar valores fora do plano amostral, pois neste caso o modelo estaria incluindo o erro em sua predição, o que não interessa. Para evitar esse inconveniente, usa-se em redes neurais, entre outras, a técnica do “*earlystopping*”. O conjunto de dados é separado em uma amostra de treinamento, que será utilizada para estimar os pesos, e em uma amostra de validação. Quando os pesos são estimados pela amostra de treinamento, ao mesmo tempo avaliam-se os erros nas amostras, ou seja, verifica-se se o valor de E continua pequeno mesmo fora da amostra de treinamento. Desta forma, tomam-se os pesos da iteração que corresponde ao valor mínimo de E na amostra de validação (mas com os pesos calculados baseados na amostra de treinamento), evitando-se assim o que se chama de sobre-ajuste. Pode ser mostrado que este procedimento corresponde a um método de regularização Bishop (1995).

É comum separar o conjunto de dados em mais uma amostra, chamada de amostra de teste, para eventuais cálculos de performance preditiva e comparações entre diferentes modelos.

Neste trabalho, utilizaremos a técnica estatística chamada de validação cruzada Stone (1974), no qual os dados serão divididos em três sub-amostras: treinamento, validação e teste. Em cada amostra será avaliado o erro quadrático médio correspondente. Assim, com essa técnica podemos avaliar o modelo em um conjunto diferente do usado para estimar os parâmetros da rede, podendo, assim, verificar o desempenho de vários modelos e escolher o que apresentar melhor desempenho. Após encontrada a melhor arquitetura, a rede é novamente treinada com todo o conjunto treinamento. Então, o conjunto teste é utilizado para se testar a capacidade de generalização da rede. Estes resultados auxiliam na escolha da melhor topologia, em termos de aprendizagem e generalização Ballini (2000).

Capítulo 5

Metodologia

Neste capítulo será descrita a metodologia utilizada para a estimação dos prêmios das opções de compra de IDI. Descreveremos a amostra de dados, assim como a estimação dos parâmetros dos modelos de Black, Vasicek e CIR. Além disso, as topologias dos modelos de redes neurais *feedforward* e recorrentes. Por fim, os métodos para avaliação dos resultados, compostos por métricas usuais de erro e testes estatísticos para comparação de modelos competitivos de predição.

5.1 Amostra

Os dados deste trabalho correspondem aos preços de fechamento das opções européias de compra de IDI, transacionadas na BM&FBOVESPA, para diferentes prazos e preços de exercício, que se iniciam em 2 de janeiro de 2003, com fim em 31 de julho de 2007, totalizando 1.638 dias de negociação e 12.801 observações¹.

Temos associado, para cada contrato de opção, o preço de exercício, o preço do contrato DI Futuro de mesmo prazo da opção, o número de contratos negociados, o volume financeiro e prazo para o vencimento em dias úteis². Nossa amostra se compõe dos contratos mais líquidos negociados na Bolsa, assim, avaliamos os papéis com mais de 800 contratos negociados.

Ainda, para o mesmo período, coletamos os dados da média anualizada diária do CDI (Certificado de Depósitos Interfinanceiros) de um dia, uma vez que essa taxa é a referência dos derivativos de juros no Brasil³.

¹A avaliação das opções de venda (puts) não foi considerada, devido sua baixa liquidez na Bolsa.

²Os dados foram fornecidos pela BM&FBOVESPA.

³O CDI é divulgado diariamente pela ANBID (Associação Nacional dos Bancos de Investimento) e os dados foram coletados em: (<http://www.cetip.com.br/> - Acesso em: 23/07/2009).

A figura 5.1 ilustra o comportamento do CDI para o período avaliado, bem como a evolução da taxa básica de juros na economia brasileira, a taxa Selic. A formação do CDI é baseada na taxa Selic, uma vez que as relações de empréstimos inter-bancários diferem pouco da taxa básica, garantindo o *spread* bancário.

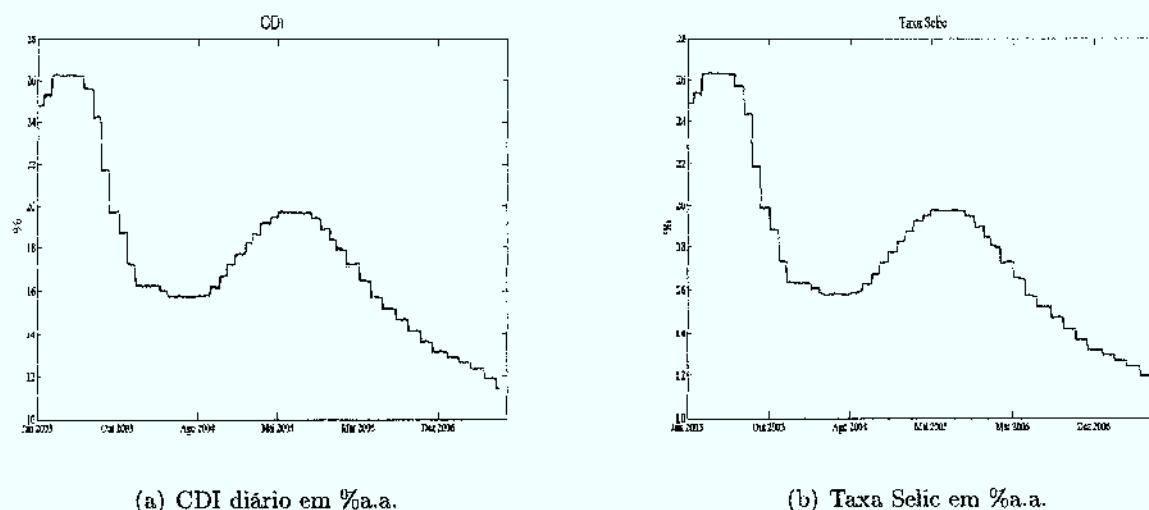


Fig. 5.1: Evolução das Taxas CDI e Selic

Pode-se perceber uma política de redução do juros iniciada em 2003, interrompida no segundo semestre de 2004. A partir de setembro de 2004, o Bacen, Banco Central do Brasil, reverteu a tendência da política monetária e deu início a um ciclo de alta da taxa de juros (que havia ficado inalterada entre abril e agosto), elevando a taxa Selic de 16,25% em setembro para 17,25% em novembro. A postura mais restritiva da política monetária esteve associada à propagação de choques - como, por exemplo, a alta do preço do petróleo no mercado internacional - e à aceleração do crescimento da demanda, provocando a deterioração das expectativas inflacionárias tanto para esse ano quanto para o próximo.

Em 2005, retorna-se um movimento de baixa nas taxas de juros, norteadas pelo objetivo principal de promover a convergência da inflação para a meta de 5,1%, ajustada pelo Banco Central, a meta original de 2005 foi fixada, em 2003, pelo Conselho Monetário Nacional (CMN), em 4,5%, mesmo valor estabelecido também para 2006, embora com intervalo de tolerância menor. Foi esse objetivo que guiou as decisões do Comitê de Política Monetária (Copom) nas suas reuniões mensais para definir a taxa Selic, levando em conta o comportamento corrente e previsto da demanda agregada, eventuais choques de preços e sua propagação pelo sistema econômico e a evolução das expectativas em relação à inflação futura.

5.2 Estimação dos Modelos

Nesta seção serão descritos os métodos para obtenção dos prêmios das opções de acordo com os modelos avaliados: modelo de Black, Vasicek, CIR, e de redes neurais *feedforward* e recorrentes.

5.2.1 Modelo de Black

No modelo de Black (1976), utilizado pela BM&FBOVESPA para a precificação das opções sobre IDI, temos apenas que obter o parâmetro referente à volatilidade da taxa de juros de curto prazo, uma vez que temos os valores do IDI (atualizados de acordo com o CDI), o preço de exercício de cada opção, e o valor de $P(t, T)$, que indica o preço em t de um título pré-fixado com vencimento em T , já são observados.

Em todos os modelos temos incorporados o valor de $P(t, T)$. Neste caso, ele será obtido de acordo com as taxas da curva *zero cupom* para títulos pré-fixados, fornecidas pela ANBIMA (Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais).

Como no modelo de Black (1976), a volatilidade (σ), que representa a variabilidade de um título pré-fixado de prazo unitário, será obtida por meio da estimação de um processo GARCH(p, q), considerando as taxas *zero cupom* para obtenção dos preços dos títulos pré-fixados.

5.2.2 Modelo de Vasicek

Para o cálculo dos prêmios das opções de compra de IDI, segundo o modelo proposto por Neto e Pereira (1999), baseados no processo de Vasicek (1977) para o comportamento da taxa de juros de curto prazo, são dois os parâmetros não observáveis: a taxa de reversão à média (α) e a volatilidade (σ^2) da taxa de juros de curto prazo.

Neto e Pereira (1999) afirmam que são duas as formas para obtenção desses parâmetros. Sendo elas:

- Método Endógeno: nele utiliza-se o próprio modelo para a construção de uma função de verossimilhança que permite a estimação dos parâmetros. A partir desses parâmetros obtém-se os preços das opções.
- Método Exógeno: nesse método, a estimação dos parâmetros é obtida implicitamente, por meio da obtenção de k (equação 3.18). Isso é possível, uma vez que, segundo os autores, k é a volatilidade da taxa de juros acumulada ao longo de (t, T) . Neste caso, Neto e Pereira (1999)

proporam um modelo para a estimação condicional de k denominada *GARCH-D*, onde a letra D provém de *duration*.

Neste trabalho, utilizaremos o método de estimação exógeno, i.e., os parâmetros não observáveis serão estimados por meio de um modelo GARCH-D.

O método de estimação exógeno se refere ao problema de estimação da volatilidade. Para isso, deve-se notar que para cada série de opções sobre IDI existe um contrato futuro de DI com mesma data de vencimento. Cada um à sua maneira, os dois contratos negociam a taxa de juros acumulada entre a data corrente e a data de vencimento. Dessa forma, a volatilidade do *yield to maturity* do preço de ajuste do contrato futuro de DI pode ser usada como uma proxy prospectiva (ao invés de retrospectiva) da volatilidade do rendimento. O termo *prospectiva* é utilizado para enfatizar o fato de que o *yield to maturity* envolve a expectativa sobre o CDI futuro e, nesse sentido, é uma variável que “olha para frente ao invés de olhar para trás”.

Sabe-se que a volatilidade do PU (ou de seu *yield*), preço unitário, tende a decrescer conforme diminui o prazo para o vencimento do contrato, devido ao assim chamado *efeito duração*.

Em vista disso, Neto e Pereira (1999) propuseram um modelo para a estimação da volatilidade do *yield to maturity* que incorpora a tendência cíclica da volatilidade das séries quando estas são classificadas de acordo com a ordem para o vencimento. Em essência, trata-se da inclusão do efeito duração num modelo GARCH(p,q).

Sejam $R(t, T_n)$ a taxa de juros implícita no preço de ajuste do contrato futuro de DI de n -ésimo vencimento, isto é, $R(t, T_n) = \frac{100.000}{PU(t, T_n)} - 1$ e $r(t)$ o CDI observado na data t . O GARCH(p,q)-D é composto pelas três relações abaixo:

$$R(t+1, T_n) = R(t, T_n) * +\sigma(t+1, T_n) \varepsilon_n(t+1) \quad (5.1)$$

$$\sigma(t+1, T_n)^2 = d(\beta_1, \beta_2, t+1, T_n) + \beta_3 + \sum_{i=1}^p \varphi_i \sigma(t+1-i, T_n)^2 \quad (5.2)$$

$$+ \sum_{i=1}^q \delta_i [R(t+1-i, T_n) - R(t-i, T_n) *]^2$$

$$\varepsilon(t+1) I(0, 1) \quad (5.3)$$

onde $R(t, T_n)^* = \frac{1+R(t, T_n)}{1+r(t)} - 1$ se $t+1$ não for o primeiro dia útil do mês e $R(t, T_n)^* = \frac{1+R(t, T_{n+1})}{1+r(t)} - 1$ caso contrário.

A função $d(\cdot)$ é determinista e tem como objetivo capturar a tendência cíclica de decrescimento da volatilidade conforme diminui o prazo para o vencimento do contrato. Uma possível forma para $d(\cdot)$ é a própria fórmula de k^2 . Outras formas possíveis incluem: $d(t, T_n) = \beta_1 (T_n - t)^{\beta_2}$, onde a curvatura da função é determinada por β_2 ; $d(t, T_n) = \beta_1 (T_n - t)$, forma linear e mais parcimoniosa em termos de quantidade de parâmetros, que será utilizada neste trabalho.

Dessa forma, para obtenção do prêmio das opções de IDI seguindo o modelo de Vasicek (1977), os parâmetros não observáveis serão estimado implicitamente pela estimação da volatilidade com base num processo GARCH(p,q)-D.

5.2.3 Modelo CIR

No modelo proposto por Barbachan e Ornelas (2003) temos quatro parâmetros para serem estimados: velocidade de reversão à média e volatilidade da taxa de juros de curto prazo, assim como o preço de mercado do risco e a taxa de juros de longo prazo. Na estimação desses parâmetros utilizaremos a mesma metodologia proposta por Barbachan e Ornelas (2003), exceto para a obtenção da volatilidade.

O parâmetro que representa a velocidade de reversão à média da taxa de juros será arbitrada em 0,006, uma vez que esse valor é uma média aproximada que a literatura apresenta, obtendo-a implicitamente por meio de inúmeros modelos que descrevem a estrutura a termo da taxa de juros no mercado brasileiro, segundo Barbachan e Ornelas (2003).

O preço de mercado do risco também será arbitrado, sendo seu valor igual a -1 , conforme sugerem os autores. O valor negativo significa que o risco é suposto indesejável, já o valor unitário nos diz que para cada unidade adicional de risco, o investidor exige uma unidade a mais de retorno.

Com base nesses parâmetros arbitrados, a volatilidade da taxa de juros de curto prazo e a taxa de juros esperada de longo prazo serão estimados com base em dados históricos.

O parâmetro que representa a volatilidade será obtido por meio de um processo GARCH(p,q)-D, uma vez que permite o cálculo da variância levando em consideração o efeito da *duration*, que melhor se adequa à *yield curve*. Esse método nos permite tornar o modelo o mais real possível, respeitando as características de contratos relacionados à níveis da taxa de juros. Barbachan e Ornelas (2003) estimam a volatilidade com base no desvio padrão das variações absolutas do CDI. Entretanto, dessa forma o modelo se limita no sentido de não levar em consideração os fatos estilizados que são observados

em contratos de taxa de juros⁴.

Por fim, conforme sugerem Barbachan e Ornelas (2003), a taxa de juros esperada de longo prazo será estimada por meio da média móvel do CDI *over*, com base em uma janela amostral com período de 3 anos.

5.2.4 Modelos de Redes Neurais

Neste trabalho, foram avaliados três tipos distintos de redes neurais: um modelo de rede *feedforward*, e os modelos de rede recorrentes de Elman e de Jordan. Elas se distinguem com relação a propagação dos sinais na rede, sendo a rede de tipo *feedforward* é caracterizada pela progressão dos sinais da camada de entrada em sentido da camada de saída, enquanto que as redes recorrentes de Elman e Jordan apresentam recorrência na propagação dos sinais dos neurônios intermediários para as camadas anteriores e dos neurônios da camada de saída para a camada de entrada, respectivamente.

Na literatura sobre redes neurais, não existe uma técnica que especifique qual é a estrutura ideal, ou seja, qual o número de camadas intermediárias e número de neurônios nesta camada que resulte nos melhores resultados. Portanto, o processo envolve escolhas empíricas que variam de acordo com a especificidade dos dados e o objetivo de estimação e previsão Kaastra e Boyd (1995).

Os três tipos de rede serão treinadas com o algoritmo de retro-propagação do erro e a seleção da estrutura da rede será determinada de acordo com o processo estatístico de validação cruzada. Os conjuntos treinamento, validação e teste possuem um total de 70%, 20% e 10% do total da amostra, respectivamente, que correspondem aos seguintes períodos: de 02/01/2003 a 02/02/2006 (70% da amostra), de 03/02/2006 a 08/12/2006 (20% da amostra), e de 11/12/2006 a 31/07/2007 (10% do total da amostra, com 1.270 observações).

A partir da delimitação dos conjuntos treinamento, validação e teste, o próximo passo é determinar a topologia da rede, em termos de número de neurônios de entrada, número de camadas e neurônios ocultos, assim como a caracterização da camada de saída.

Para determinação do número de entradas, utilizaremos as variáveis que afetam o prêmio da opção de compra de IDI e avaliaremos os resultados em termos do erro quadrático médio (*EQM*), definido por:

⁴Deve-se perceber que ao arbitrar os parâmetros da velocidade de reversão à média e preço do mercado do risco também incorremos em limitações. Entretanto, como o objetivo do trabalho é avaliar os modelos de redes neurais, por isso seguimos os procedimentos propostos por Barbachan e Ornelas (2003).

$$EQM = \frac{1}{N} \sum_{i=1}^N (c_i - \hat{c}_i)^2 \quad (5.4)$$

sendo c_i e $\hat{c}_i$ os prêmios das opções observados no mercado e obtidos pelo modelo, respectivamente, e N o tamanho da amostra.

Dessa forma, estimaremos, inicialmente, um modelo que considere como variáveis de entrada: valor do IDI, atualizado de acordo com o CDI; preço de exercício da opção (X); maturidade ou tempo de vencimento da opção ($\tau = T - t$); o preço de um título pré-fixado em t com vencimento em T ($P(t, T)$); a taxa de juros estimada de longo prazo, de acordo com a média móvel do CDI *over* com um janela amostral de 3 anos (θ); e a volatilidade da taxa de juros de curto prazo (σ^2), obtida por um modelo GARCH(p,q)-D, conforme proposto por Neto e Pereira (1999), uma vez que se mostra como uma alternativa mais realista para a obtenção da variância de taxas de juros, ao levar em conta o efeito duração.

Com base nessas variáveis, obteremos o *EQM* do modelo. Assim, retiramos da estrutura da rede as variáveis de entrada uma a uma e controlaremos o comportamento da medida de erro, para então obter o modelo que considere apenas as variáveis que permitam a obtenção do menor erro quadrático, deixando de lado as variáveis que não sejam significativas para a obtenção do valor do prêmio da opção.

Determinada as variáveis de entrada para os três modelos de rede, a obtenção do melhor modelo passará para a determinação do número de camadas e neurônios intermediários, uma vez que a camada de saída é representada por um único neurônio, que fornece o valor do prêmio da opção de compra de IDI. Para isso, foram utilizados os critérios de informação e qualidade de ajuste de Akaike (AIC), Hannan-Quinn (HQC), Schwarz (BIC) e Shibata (SIC), todos baseados no erro quadrático médio⁵, definidos como:

$$AIC = -2 \left(\frac{\ln(EQM)}{N} \right) + 2 \left(\frac{K}{N} \right) \quad (5.5)$$

$$HQC = -2 \left(\frac{\ln(EQM)}{N} \right) + 2 \left(\frac{K \ln(\ln(N))}{N} \right) \quad (5.6)$$

$$BIC = -2 \left(\frac{\ln(EQM)}{N} \right) + 2 \left(\frac{K \ln(K)}{N} \right) \quad (5.7)$$

⁵Para detalhes dos testes ver: Akaike (1974), Hannan e Quinn (1979), Schwarz (1978) e Shibata (1976).

$$SIC = -2 \left(\frac{\ln(EQM)}{N} \right) + 2 \left(\frac{N + 2K}{N} \right) \quad (5.8)$$

onde K representa o número de camadas intermediárias ou o número de neurônios das camadas intermediárias.

Esses critérios apresentam um compromisso entre a minimização do erro quadrático médio e a complexidade da rede, pois penalizam os modelos que não são parcimoniosos.

Neste trabalho, utilizamos a aprendizagem *batch* ou chamada de aprendizagem padrão a padrão, em que cada par de entrada-saída é apresentado à rede para a atualização dos pesos. Ao final de uma época, o erro quadrático médio é calculado e se o valor for menor que o nível de erro estipulado, o processo de aprendizagem é encerrado. Caso contrário, inicia-se uma nova época. Com isso, o número de épocas (iterações) foi estipulado em 500 para os três modelos de rede neural. O nível de erro foi estipulado em torno de 10^{-4} , uma vez que são níveis de erro aceitáveis para um modelo de estimação e previsão de séries temporais financeiras.

Com relação a taxa de aprendizagem, selecionamos os seguintes valores: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 e 0.9. Assim, a taxa selecionada será aquela que apresentar maior capacidade de generalização da rede, isto é, menor erro nas amostras de validação e teste.

Dessa forma, para os três diferentes modelos de rede neural foram obtidas as topologias que apresentaram os melhores resultados para a obtenção dos prêmios das opções de compra de IDI no período selecionado.

5.3 Métricas de Erro

Para avaliação dos modelos empregados neste trabalho, comparamos os prêmios das opções de compra de IDI teóricos (i.e., obtidos pelos modelos) em relação aos prêmios observados no mercado. Para tanto, foram utilizadas as métricas tradicionais de erro de predição: Erro Percentual Médio (MPE - *Mean Percentage Error*), Erro Percentual Máximo (MPE - *Maximum Percentage Error*), Raiz do Erro Quadrático Médio (RMSE - *Root Mean Squared Error*), Erro Absoluto Médio (MAE - *Maximum Absolute Error*), Erro Médio Absoluto Percentual (MAPE - *Maximum Absolute Percentage Error*), e o coeficiente de desigualdade de Theil (TIC), definidos como:

$$MEP = \frac{100}{N} \sum_{i=1}^N \frac{|c_i - \hat{c}_i|}{c_i} \quad (5.9)$$

$$MPE = \max_{i=1, \dots, N} \left\{ 100 \frac{|c_i - \hat{c}_i|}{c_i} \right\} \quad (5.10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (c_i - \hat{c}_i)^2} \quad (5.11)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |c_i - \hat{c}_i| \quad (5.12)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{c_i - \hat{c}_i}{c_i} \right| \quad (5.13)$$

$$TIC = \frac{\sqrt{\sum_{t=1}^N \frac{(\hat{c}_t - c_t)^2}{N}}}{\sqrt{\sum_{t=1}^N \frac{\hat{c}_t^2}{N}} + \sqrt{\sum_{t=1}^N \frac{c_t^2}{N}}} \quad (5.14)$$

Para as métricas usuais de error, seu menor valor indica o melhor modelo, enquanto que o coeficiente de desigualdade de Theil varia entre zero e um, em que zero indica o ajuste perfeito.

Os resultados dos modelos ainda foram avaliados de acordo com o grau de *moneyness* (M) das opções, sendo este definido pela relação entre o valor presente do preço de exercício da opção e o preço à vista do ativo objeto:

$$M = \frac{IDI_t}{X \cdot e^{-rT}} \quad (5.15)$$

em que r indica a taxa implícita associada ao contrato de Futuro DI para cada opção.

Neste caso, as opções da amostra teste foram divididas conforme o grau de *moneyness*: *out-of-the-money* ($M \leq 1 - \alpha\%$), *at-the-money* ($1 - \alpha\% < M < 1 + \alpha\%$) e *in-the-money* ($M > 1 + \alpha\%$), considerando $\alpha = 5\%$.

Além disso, calculou-se o coeficiente de determinação, R^2 , da regressão linear dos preços das opções observadas no mercado sobre os preços teóricos:

$$c_i = \phi_1 + \phi_2 \cdot \hat{c}_i + \zeta_t \quad (5.16)$$

no qual ϕ_1 e ϕ_2 são os parâmetros linear e angular da regressão, respectivamente, e ζ_t um ruído

branco.

Os resultados empíricos na literatura se baseiam extensivamente em medidas padrões de acurácia, como as métricas de erro apresentadas acima, e não levam em consideração a superioridade de um modelo a outro em termos de significância estatística. Assim, neste trabalho foram aplicados testes paramétricos e não-paramétricos para verificar se existe diferença estatisticamente significativa entre os resultados dos modelos avaliados. Esses testes são apresentados a seguir.

5.4 Testes Estatísticos

Nesta seção serão descritos dois testes paramétricos (AGS e MGN) e um teste não-paramétrico (SIGN) para avaliação dos modelos.

Teste AGS

O teste paramétrico aqui descrito se baseia no trabalho de Ashley et al. (1980), denominado teste AGS. O teste AGS permite a avaliação da existência de significância estatística entre a diferença do erro quadrático médio (*EQM*) entre modelos de previsão.

Seja $\Delta e_i^{B,RNA}$ a diferença entre os erros de previsão obtidos pelos modelos convencionais de precificação (Black, Vasicek ou CIR), tomados como *benchmark*, e de Redes Neurais, e_i^B e e_i^{RNA} , respectivamente. Assim, temos que:

$$\Delta e_i^{B,RNA} = e_i^B - e_i^{RNA} = (c_i^B - \hat{c}_i^B) - (c_i^{RNA} - \hat{c}_i^{RNA}) \quad (5.17)$$

Scndo $i = 1, \dots, N$. Define-se, ainda, $S_i^{B,RNA}$ a soma dos erros de previsão e μ_S a média amostral de $S_i^{B,RNA}$. O teste pode ser descrito segundo a equação abaixo:

$$\Delta e_i^{B,RNA} = \beta_1 + \beta_2 (S_i^{B,RNA} - \mu_S) + v_i \quad (5.18)$$

Na qual v_i indica um processo ruído branco.

O teste AGS mostra que β_1 representa a diferença entre os erros quadráticos médios de previsão obtidos pelos modelos e β_2 é proporcional à diferença entre as variâncias dos erros de previsão dos modelos. O teste de significância tem como hipótese nula $H_0 : \beta_1 = \beta_2 = 0$, enquanto que a hipótese alternativa consiste em $H_1 : \beta_1 > 0$ e/ou $\beta_2 > 0$. A estatística para este teste, obtida por meio

da estimação da equação 5.18, tem distribuição F com 2 e $(N - 2)$ graus de liberdade, assumindo normalidade para os erros. A rejeição da hipótese nula indica que os modelos baseados em Redes Neurais superam os métodos tomados como *benchmark*.

Neste caso, enfatizaremos as comparações entre as metodologias tradicionais com as propostas neste trabalho, baseadas nos modelos de redes neurais. Além disso, iremos comparar a acurácia entre os próprios *benchmarks* e de redes neurais para auferição do melhor método para as fórmulas analíticas e para as metodologias de inteligência artificial.

Teste MGN

Nesta subseção é descrito outro teste paramétrico para igualdade entre acurácia de previsão. Trata-se do teste de Morgan-Granger-Newbold (teste MGN), baseado inicialmente no trabalho de Granger e Newbold (1977). A utilização desse teste é recomendada para avaliar modelos de predição e é amplamente aplicado quando se relaxa a assunção de não existência de correlação entre os erros dos modelos. O valor da estatística para esse teste pode ser calculado como:

$$MGN = \frac{\hat{\rho}_{s\Delta}}{(1 - \hat{\rho}_{s\Delta})^{1/2}} (N - 1)^{1/2} \quad (5.19)$$

Em que $\hat{\rho}_{s\Delta}$ é o coeficiente de correlação estimado entre $S_i^{B,RNA}$ e $\Delta e_i^{B,RNA}$, e N representa o número de observações. A estatística para o teste MGN tem distribuição t com $(N - 1)$ graus de liberdade. Para este teste, se as previsões são igualmente acuradas então a correlação entre $S_i^{B,RNA}$ e $\Delta e_i^{B,RNA}$ será zero (hipótese nula).

Teste SIGN

O teste SIGN (*Significance Test*), derivado do trabalho de Lehmann (1988), é uma avaliação não-paramétrica para modelos de predição que não requer suposições como normalidade e não-correlação serial dos erros. A estatística para esse teste pode ser calculada como:

$$SIGN = \left(\psi(N) - \frac{N}{2} \right) \cdot 2(N)^{1/2} \quad (5.20)$$

Sendo que o termo ψ_t denota o número de vezes em que, neste caso, os erros dos modelos tradicionais superaram os erros derivados dos modelos de Redes Neurais. A hipótese nula é $H_0 : \psi(N) = (N/2)$,

contra a hipótese alternativa de que $H_1 : \psi(N) > (N/2)$. Assume-se que N é o número de observações previstas por ambos os modelos e, se os erros de previsão dos modelos do *benchmark* são da mesma proporção que os de inteligência artificial em exatamente $(N/2)$ observações, de acordo com o teste SIGN, a acurácia dos modelos pode ser considerada equivalente, sem diferença estatisticamente significativa. Entretanto, se a proporção de erros do modelo analíticos superam os dos modelos de Redes Neurais numa proporção maior que $(N/2)$, o teste indica que os modelos tradicionais são menos acurados que os modelos de Redes Neurais. A estatística para o teste SIGN tem distribuição gaussiana com média zero e variância unitária. Assim, se a estatística SIGN é significativamente elevada, a hipótese nula de equivalência de predição pode ser rejeitada em favor da hipótese alternativa.

Capítulo 6

Resultados e Discussão

Neste capítulo serão descritos os resultados obtidos pelos modelos tradicionais de precificação e pelos modelos de redes neurais avaliados. Serão apresentadas as topologias selecionadas para os modelos de redes neurais, seguido pela comparação dos preços teóricos obtidos com os preços de fechamento das opções de compra de IDI transacionadas na BM&FBOVESPA para o período avaliado, em termos das métricas tradicionais de erro, como para os testes estatísticos de performance.

6.1 Topologias das Redes Neurais

Foram estimados, neste trabalho, três tipos de redes neurais artificiais: um modelo de tipo *feedforward* ou progressiva, e dois modelos de redes recorrentes (rede de Elman e rede de Jordan).

Inicialmente, os modelos de redes neurais foram avaliados considerando como entradas todas as variáveis que “afetam” o preço de uma opção de compra de IDI, ou seja: valor do IDI, atualizado de acordo com o CDI (IDI_t); preço de exercício da opção (X); maturidade ou tempo de vencimento da opção ($\tau = T - t$); o preço de um título pré-fixado em t com vencimento em T ($P(t, T)$); a taxa de juros estimada de longo prazo, de acordo com a média móvel do CDI-*over* com um janela amostral de 3 anos (θ); e a volatilidade da taxa de juros de curto prazo (σ^2), obtida por um modelo GARCH(p,q)-D. Com base nessas entradas, determinou-se os parâmetros das redes, para então avaliá-las levando em consideração apenas as entradas que incrementam a performance da rede neural.

Rede Neural *Feedforward*

A tabela 6.1 apresenta os valores obtidos para os critérios de informação adotados para a escolha

do número de camadas e neurônios intermediários para a rede neural progressiva.

Tab. 6.1: Critérios de informação para determinação da estrutura da rede neural *feedforward* para o apreçamento das opções de compra de IDI

1 Camada Interm.									
Nº Neurônios	AIC	HQC	BIC	SIC					
1	-4,762	-4,638	-4,559	-4,698					
2	-4,791	-4,512	-4,487	-4,723					
3	-4,271	-4,762	-4,701	-4,681					
4	-4,162	-4,472	-4,672	-4,592					
5	-4,661	-4,726	-4,529	-4,615					
6	-4,762	-4,721	-4,552	-4,583					
2 Camadas Interm.									
1º Camada Interm.		2º Camada Interm.							
Nº Neurônios	AIC	HQC	BIC	SIC	Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,762	-4,563	-4,472	-4,611	1	-4,625	-4,827	-4,598	-4,550
2	-4,276	-4,819	-2,623	-4,726	2*	-4,872	-4,876	-4,766	-4,578
3*	-4,965	-4,369	-4,877	-4,904	3	-4,532	-4,719	-4,574	-4,372
4	-4,623	-4,704	-4,621	-4,625	4	-4,666	-4,800	-4,801	-4,726
5	-4,575	-4,593	-4,476	-4,263	5	-4,627	-4,799	-4,563	-4,492
6	-4,670	-4,632	-4,701	-4,732	6	-4,637	-4,352	-4,362	-4,532
3 Camadas Interm.									
1º Camada Interm.		2º Camada Interm.							
Nº Neurônios	AIC	HQC	BIC	SIC	Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,472	-4,625	-4,532	-4,449	1	-4,652	-4,293	-4,574	-4,532
2	-4,726	-4,623	-4,263	-4,392	2	-4,327	-4,672	-4,623	-4,329
3	-4,700	-4,627	-4,632	-4,442	3	-4,524	-4,333	-4,660	-4,281
4	-4,271	-4,124	-4,382	-4,527	4	-4,372	-4,221	-4,632	-4,701
5	-4,362	-4,213	-4,392	-4,271	5	-4,555	-4,621	-4,699	-4,612
6	-4,251	-4,625	-4,222	-4,173	6	-4,273	-4,615	-4,542	-4,251
3º Camada Interm.									
Nº Neurônios	AIC	HQC	BIC	SIC					
1	-4,621	-4,392	-4,271	-4,600					
2	-4,382	-4,427	-4,558	-4,699					
3	-4,371	-4,712	-4,329	-4,134					
4	-4,293	-4,222	-4,709	-4,612					
5	-4,371	-4,721	-4,251	-4,261					
6	-4,299	-4,198	-4,283	-4,653					

O sobrescrito (*) indica a melhor estrutura, referente ao menor valor dos critérios de informação.

Deve-se perceber que, na tabela 6.1, o limite de camadas intermediárias apresentadas é três, enquanto que o número de neurônios das camadas intermediárias se estendem até seis. Isso se deve ao fato de que, para maiores números de camadas e neurônios intermediários, os critérios de informação indicaram piora no desempenho da rede, uma vez que o aumento da complexidade da rede não resultou numa melhora da previsão dos prêmios das opções, indicando que as redes de topologia mais complexa acabam perdendo a parcimônia.

Os resultados da tabela 6.1 mostram que o modelo de rede progressiva que apresentou melhor desempenho possui 2 camadas intermediárias, sendo a primeira composta por três neurônios e a segunda por dois neurônios, segundo os critérios de informação adotados (valores em negrito). Além disso, de acordo com os mesmos critérios utilizados, as entradas que indicaram a melhor performance da rede

foram: valor do IDI, atualizado de acordo com o CDI (IDI_t); preço de exercício da opção (X); maturidade ou tempo de vencimento da opção ($\tau = T - t$); a taxa de juros estimada de longo prazo, de acordo com a média móvel do CDI-*over* (θ); e a volatilidade da taxa de juros de curto prazo (σ^2).

A figura 6.1 apresenta uma representação da rede neural progressiva para a obtenção do prêmio das opções de compra de IDI.

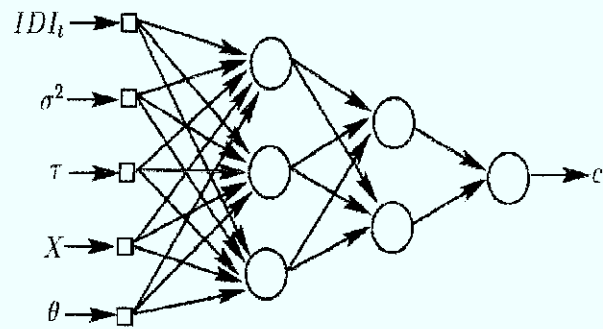


Fig. 6.1: Topologia da rede neural progressiva para o prêmio das opções de compra de IDI

Determinada a topologia da rede neural progressiva, o próximo passo da estimação se deu na seleção da taxa de aprendizagem (η) que permitiu a melhor capacidade de generalização da rede. Assim, com base no valor do erro quadrático médio (EQM) obtido para diferentes taxas de aprendizagem, selecionamos a que resultou no melhor modelo. A tabela 6.2 associa o valor do EQM às diferentes taxas de aprendizagem, com base na amostra teste.

Tab. 6.2: EQM para as diferentes taxas de aprendizagem com base na amostra teste para a rede neural *feedforward*

	η								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
EQM	132,651	139,872	133,762	129,872	129,999	122,345	154,342	165,872	172,392

Por meio da tabela 6.2 pode-se perceber que o menor valor para o EQM corresponde a uma taxa de aprendizagem $\eta = 0,6$. Pode-se perceber que para as taxas maiores que 0,6 a rede perde a capacidade de generalização, enquanto que para valores menores que 0,6 os erros associados são próximos, entretanto, deve-se ressaltar que, neste caso, o processo de treinamento se torna mais custoso, em termos computacionais, decorrendo num maior tempo para o treinamento da rede.

Com isso, com base na topologia obtida e com uma taxa de aprendizagem igual a 0.6, selecionamos o modelo de rede neural progressiva que apresentou os melhores resultados, sendo que a taxa de

convergência ao erro estipulado (10^{-4}) foi obtida com base em 76 épocas.

Rede Neural Recorrente de Elman

Na tabela 6.3 podemos verificar os valores dos critérios de informação associados ao número de camadas e neurônios ocultos para o modelo de rede recorrente de Elman para a valoração das opções de compra de IDI.

Tab. 6.3: Critérios de informação para determinação da estrutura da rede neural recorrente de Elman para o apreçamento das opções de compra de IDI

1 Camada Interm.				
Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,762	-4,652	-4,632	-4,475
2	-4,342	-4,583	-4,598	-4,273
3	-4,382	-4,753	-4,287	-4,333
4	-4,573	-4,349	-4,440	-4,672
5	-4,276	-4,234	-4,328	-4,762
6	-4,376	-4,625	-4,271	-4,372
2 Camadas Interm.				
1º Camada Interm.				
Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,765	-4,371	-4,382	-4,349
2	-4,376	-4,372	-4,291	-4,330
3	-4,494	-4,550	-4,318	-4,392
4	-4,103	-4,293	-4,029	-4,099
5	-4,726	-4,701	-4,698	-4,376
6	-4,387	-4,729	-4,376	-4,372
2º Camada Interm.				
Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,381	-4,731	-4,574	-4,392
2	-4,311	-4,092	-4,382	-4,303
3	-4,392	-4,398	-4,400	-4,381
4	-4,276	-4,362	-4,361	-4,311
5	-3,982	-4,039	-4,382	-4,387
6	-4,493	-4,302	-4,790	-4,311
3 Camadas Interm.				
1º Camada Interm.				
Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,376	-4,762	-4,809	-4,376
2	-4,309	-4,693	-4,583	-4,601
3	-4,372	-4,387	-4,766	-4,898
4*	-4,880	-4,981	-4,872	-4,762
5	-4,365	-4,576	-4,602	-4,723
6	-4,371	-4,447	-4,762	-4,809
2º Camada Interm.				
Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,382	-4,387	-4,310	-4,722
2	-4,362	-4,672	-4,800	-4,372
3	-4,376	-4,772	-4,808	-4,626
4	-4,653	-4,762	-4,376	-4,387
5*	-4,981	-4,899	-4,870	-4,910
6	-4,365	-4,687	-4,761	-4,632
3º Camada Interm.				
Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,376	-4,487	-4,583	-4,372
2	-4,389	-4,287	-4,723	-4,372
3*	-4,762	-4,810	-4,928	-4,891
4	-4,876	-4,365	-4,376	-4,780
5	-4,371	-4,376	-4,493	-4,582
6				

O sobrescrito (*) indica a melhor estrutura, referente ao menor valor dos critérios de informação.

Com base nos critérios de parcimônia, tabela 6.3, o modelo de rede neural recorrente de Elman que apresentou o melhor desempenho possui três camadas escondidas. A primeira é composta por quatro neurônios, enquanto que a segunda e a terceira possuem cinco e três neurônios ocultos, respectivamente.

Determinado parte da topologia, verificou-se a composição da camada de entrada, de acordo com as variáveis que influenciam o preço das opções de IDI. Assim, com base nos critérios de informação, a

rede recorrente de Elman têm como entrada as seguintes variáveis: valor do IDI, atualizado de acordo com o CDI (IDI_t); preço de exercício da opção (X); maturidade ou tempo de vencimento da opção ($\tau = T - t$); e a volatilidade da taxa de juros de curto prazo (σ^2).

Podemos perceber que, diferente do modelo de rede neural *feedforward*, a rede de Elman desconsiderou o parâmetro que representa a taxa de juros esperada de longo, uma vez que sua inserção no modelo não resulta em ganhos de performance.

A figura 6.1 elucida a topologia da rede recorrente de Elman que apresentou a melhor performance para a obtenção dos prêmios das opções de compra de IDI. Neste caso, a estrutura da rede recorrente de Elman se mostrou mais complexa, em termos do número de neurônios e camadas ocultas, se comparado com o modelo de rede progressiva ou *feedforward*.

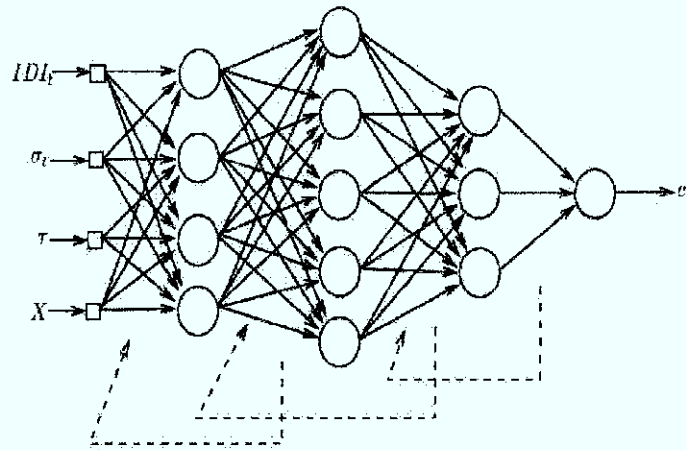


Fig. 6.2: Topologia da rede neural recorrente de Elman para o prêmio das opções de compra de IDI

Na tabela 6.4 podemos observar a variação no valor do *EQM* associado às diferentes taxas de aprendizagem. Para a rede recorrente de Elman, o valor de η que resultou no melhor estimador para os prêmios das opções de compra de IDI foi de 0.4. Essa taxa de aprendizagem foi capaz de atingir o nível de erro estipulado (10^{-4}) com base em 102 épocas.

Tab. 6.4: *EQM* para as diferentes taxas de aprendizagem com base na amostra teste para a rede neural recorrente de Elman

	η								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
<i>EQM</i>	102,324	100,982	90,298	70,023	98,763	102,971	104,128	115,298	101,982

Podemos perceber que a taxa de aprendizagem associada ao modelo da rede recorrente de Elman foi menor que a obtida para o modelo *feedforward*, o que explica em parte o maior número de épocas associado ao modelo recorrente, uma vez que, quanto menor a taxa, maior será o tempo para o treinamento da rede, de forma a atingir o limiar de convergência determinado.

Rede Neural Recorrente de Jordan

A estimação do modelo de rede neural recorrente de Jordan, dentre os modelos de rede avaliados, apresentou a estrutura mais complexa, em termos de neurônios e camadas intermediárias. Na tabela 6.5 são apresentados os valores dos critérios de informação para determinação da topologia dessa rede.

Tab. 6.5: Critérios de informação para determinação da estrutura da rede neural recorrente de Jordan para o apreçamento das opções de compra de IDI

1 Camada Interm.									
Nº Neurônios	AIC	HQC	BIC	SIC					
1	-4,625	-4,493	-4,672	-4,376					
2	-4,762	-4,665	-4,376	-4,762					
3	-4,376	-4,293	-4,653	-4,372					
4	-4,312	-4,423	-4,762	-4,473					
5	-4,387	-4,233	-4,387	-4,592					
6	-4,584	-4,685	-4,677	-4,584					
2 Camadas Interm.									
1º Camada Interm.		2º Camada Interm.							
Nº Neurônios	AIC	HQC	BIC	SIC	Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,447	-4,376	-4,788	-4,574	1	-4,376	-4,372	-4,389	-4,311
2	-4,122	-4,733	-4,500	-4,329	2	-4,394	-4,449	-4,355	-4,485
3	-4,388	-4,785	-4,653	-4,444	3	-4,493	-4,374	-4,503	-4,573
4	-4,756	-4,723	-4,364	-4,521	4	-4,376	-4,387	-4,312	-4,376
5	-4,239	-4,392	-4,403	-4,376	5	-4,594	-4,376	-4,754	-4,423
6	-4,376	-4,320	-4,493	-4,362	6	-4,732	-4,766	-4,320	-3,983
3 Camadas Interm.									
1º Camada Interm.		2º Camada Interm.							
Nº Neurônios	AIC	HQC	BIC	SIC	Nº Neurônios	AIC	HQC	BIC	SIC
1	-4,762	-4,449	-4,311	-4,345	1	-4,448	-4,372	-4,365	-4,367
2	-4,376	-4,598	-4,764	-4,653	2	-4,362	-4,387	-4,366	-4,213
3	-4,387	-4,322	-4,459	-4,563	3	-4,365	-4,376	-4,376	-4,345
4	-4,362	-4,487	-4,322	-4,722	4	-3,983	-4,034	-4,763	-4,387
5*	-4,901	-4,801	-4,918	-4,872	5	-4,762	-4,376	-4,442	-4,371
6	-4,376	-4,635	-4,376	-4,448	6*	-4,932	-4,891	-4,888	-4,872
3º Camada Interm.									
Nº Neurônios	AIC	HQC	BIC	SIC					
1	-4,313	-4,111	-4,243	-4,382					
2	-4,387	-4,762	-4,387	-4,673					
3	-4,566	-4,763	-4,829	-4,387					
4*	-4,856	-4,900	-4,652	-4,891					
5	-4,322	-4,803	-4,821	-4,332					
6	-4,263	-4,092	-4,387	-4,222					

O sobrescrito (*) indica a melhor estrutura, referente ao menor valor dos critérios de informação.

Dessa forma, por meio da tabela 6.5, verificou-se que o modelo de rede que aprendeu os melhores resultados, de acordo com os critérios adotados, apresenta três camadas intermediárias. A primeira e

a segunda camadas são formadas por cinco e seis neurônios ocultos, respectivamente, enquanto que a terceira camada apresenta quatro neurônios.

Assim como no modelo de rede recorrente de Elman, as variáveis que compõem a camada de entrada da rede recorrente de Jordan são: valor do IDI, atualizado de acordo com o CDI (IDI_t); preço de exercício da opção (X); maturidade ou tempo de vencimento da opção ($\tau = T - t$); e a volatilidade da taxa de juros de curto prazo (σ^2).

Na figura 6.1 temos a estrutura que caracteriza a rede recorrente de Jordan que apresentou a melhor performance para a obtenção dos prêmios das opções de compra de IDI.

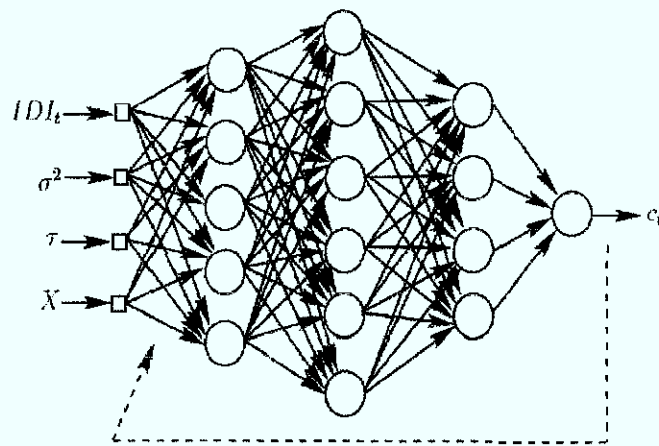


Fig. 6.3: Topologia da rede neural recorrente de Jordan para o prêmio das opções de compra de IDI

Por fim, a tabela 6.6 nos mostra o valor da taxa de aprendizagem associada ao treinamento da estrutura que apresentou mais baixo EQM . Neste caso, temos que tal taxa é $\eta = 0.7$, com base em 61 iterações, indicando uma alta convergência no processo de treinamento para o erro estipulado (10^{-4}).

Tab. 6.6: EQM para as diferentes taxas de aprendizagem com base na amostra teste para a rede neural recorrente de Jordan

	η								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
EQM	63,392	66,293	65,763	62,387	64,983	65,876	59,653	64,875	72,387

6.2 Comparação das Métricas de Erro

Para confrontar o desempenho dos modelos baseados em redes neurais com os métodos de Black, Vasicek e CIR, foram comparados os preços teóricos das opções obtidos por meio de cada modelo com os

preços observados no mercado. Nessa comparação, os dados utilizados pertenceram ao grupo de teste da estruturação dos modelos de redes neurais, uma vez que o conjunto treinamento e validação foram utilizados para determinar a estrutura mais adequada associada a cada topologia de rede, estimar os pesos sinápticos e aprender o processo gerador dos preços de compra das opções de dólar. Dessa forma, o conjunto teste, ou conjunto de previsão, permitiu avaliar a capacidade de aprendizado e generalização da rede para precificar as opções com dados de entrada nunca antes apresentados à rede. O conjunto teste foi composto por negociações que se iniciam em 11/12/2006 e vão até 31/07/2007, totalizando 1.270 observações.

Na tabela 6.7 são apresentados os valores das métricas de erro obtidos para os modelos de precificação abordados neste trabalho, sendo eles: Black, Vasicek, CIR, Rede Neural *Feedforward* (*RNF*), Rede Neural Recorrente de Elman (*RNRE*) e Rede Neural Recorrente de Jordan (*RNRJ*). Além disso, são indicados os valores do coeficiente de desigualdade de Theil e do coeficiente de determinação (R^2) da regressão dos preços teóricos sobre os prêmios observados no mercado.

Tab. 6.7: Aderência dos preços calculados pelos modelos tradicionais e de Redes Neurais aos preços de fechamento das *calls* de IDI européias negociadas na BM&FBOVESPA entre 11/12/2006 e 31/07/2007. Amostra com 1.270 observações

Modelos	Métricas						
	MEP	MPE	RMSE	MAE	MAPE	TIC	R^2
Black	87,98%	96,15%	44,823	123,434	34,589	0,737	0,553
Vasicek	80,12%	94,54%	33,265	99,872	33,291	0,698	0,622
CIR	74,65%	90,20%	24,983	93,652	29,382	0,507	0,663
RNF	59,87%	74,11%	12,873	58,763	17,843	0,432	0,753
RNRE	46,98%	66,82%	8,283	45,214	14,028	0,312	0,910
RNRJ	51,98%	70,99%	7,873	41,862	15,386	0,375	0,892

Por meio da avaliação das métricas de erros (tabela 6.7), é possível verificar que os modelos baseados em redes neurais apresentaram preços teóricos mais próximos aos preços de mercado, explicando os baixos valores dos erros obtidos, em comparação aos modelos de fórmula analítica fechada.

Dentre os modelos tomados como *benchmark*, i.e., Black, Vasicek e CIR, podemos inferir que a metodologia mais adequada para a valoração das opções de compra de IDI é o modelo proposto por Barbachan e Ornelas (2003), pois ele indicou baixos erros, em relação aos modelos de Black e CIR, que revelaram inadequação ao tipo de opção transacionado na Bolsa. Dentre as metodologias de redes neurais, o modelo de rede recorrente de Elman foi o que apresentou os menores valores das métricas de erro, o que indica ser o método mais adequado quando o objetivo é a obtenção dos prêmios das opções de compra de IDI. Deve-se ressaltar, sobretudo, que a rede *feedforward* foi a que apresentou os

maiores erros, levando em consideração os modelos recorrentes, o que nos sugere que a dependência temporal das variáveis que determinam o preço da opção de IDI é relevante, pois apenas nos métodos de rede recorrente se permite capturar tal característica.

O coeficiente de desigualdade de Theil (TIC) mostrou um ajuste superior dos modelos de rede neural aos dados das opções, pois apresentou um valor bem mais próximo de zero em comparação aos modelos convencionais. Adicionalmente, o coeficiente de determinação da regressão dos preços de mercado sobre os preços teóricos confirmou as análises expostas acima, dado o maior valor do (R^2) quando utilizados os prêmios gerados pelo modelos de inteligência artificial. Nesta última análise, cabe notar que os preços do mercado não foram explicados fundamentalmente pelos preços teóricos, o que leva a concluir que outros fatores estariam determinando os prêmios dessas opções. Ainda, esses indicadores confirmam a superioridade do modelo de rede neural recorrente de Elman sobre todos os demais modelos avaliados.

De forma complementar as análises, a tabela 6.8 apresenta os valores das métricas de erro para os conjuntos de treinamento e validação. Os resultados confirmaram o baixo desempenho dos modelos analíticos em toda a amostra dos dados. Os erros destes modelos se apresentaram relativamente iguais nos três conjuntos de dados, uma vez que não faz diferença a distinção entre elas para a aplicação das fórmulas analíticas. Entretanto, como era de se esperar, nos conjuntos treinamento e validação, para os métodos de redes neurais, os erros foram maiores, se comparados aos próprios resultados do conjunto teste das redes, visto que foram os dados apresentados à rede e que a estruturaram de forma a convergir para o nível de erro estipulado, de acordo com o processo de aprendizagem supervisionado.

Foram avaliados, ainda, os resultados dos modelos de acordo com o grau de *moneyness* (M) das opções. Verificou-se que 51% das opções da amostra teste estavam dentro do dinheiro, enquanto 29% e 20% estão no dinheiro e fora do dinheiro, respectivamente. A tabela 6.9 apresenta os resultados das métricas de erro para o conjunto da amostra teste de acordo com o grau de *moneyness* das opções. Tanto os modelos tradicionais quanto os baseados em redes neurais foram mais acurados no apuração das opções *in-the-money*. Os maiores níveis de erros foram verificados nas opções *out-of-the-money*, para os modelos de fórmulas analíticas fechadas. Nota-se, de forma geral que, nos diferentes graus de *moneyness*, os resultados dos modelos baseados em inteligência artificial foram superiores em comparação aos modelos tomados como *benchmark*.

Os resultados indicados na tabela 6.9 evidenciam a superioridade dos modelos de redes neurais propostos para a precificação das opções de compra de IDI, de acordo com os baixos valores obtidos

Tab. 6.8: Aderência dos preços calculados pelos modelos tradicionais e de Redes Neurais aos preços de fechamento das *calls* de IDI européias negociadas na BM&FBOVESPA entre 02/01/2003 e 08/12/2006. Amostra com 11.531 observações

Treinamento							
Modelos	MEP	MPE	RMSE	MAE	MAPE	TIC	R^2
Black	83,32%	94,72%	46,823	120,342	33,324	0,754	0,523
Vasicek	79,32%	93,55%	32,532	101,432	35,534	0,632	0,642
CIR	76,12%	91,13%	25,312	92,312	31,332	0,542	0,654
RNF	60,32%	79,21%	17,313	66,992	21,532	0,532	0,634
RNRE	48,32%	73,53%	12,543	55,214	19,133	0,412	0,790
RNRJ	60,12%	78,12%	11,531	59,421	21,542	0,475	0,722
Validação							
Modelos	MEP	MPE	RMSE	MAE	MAPE	TIC	R^2
Black	88,21%	93,32%	43,421	121,232	35,312	0,722	0,516
Vasicek	81,43%	90,23%	34,555	96,822	35,003	0,621	0,692
CIR	76,21%	88,53%	25,321	91,872	31,321	0,532	0,611
RNF	63,32%	78,12%	14,111	64,231	22,100	0,501	0,699
RNRE	50,71%	71,19%	12,432	53,092	18,321	0,429	0,872
RNRJ	63,22%	77,28%	11,532	46,523	18,993	0,452	0,781

Tab. 6.9: Aderência dos preços calculados pelos modelos tradicionais e de Redes Neurais aos preços de fechamento das *calls* de IDI européias negociadas na BM&FBOVESPA entre 11/12/2006 e 31/07/2007 de acordo com o grau de *moneyness*. Amostra com 1.270 observações

Out-of-the-Money							
Modelos	MEP	MPE	RMSE	MAE	MAPE	TIC	R^2
Black	92,32%	102,23%	49,287	138,233	41,872	0,797	0,493
Vasicek	86,21%	104,32%	39,422	105,643	39,092	0,788	0,503
CIR	79,83%	94,82%	31,625	105,123	31,726	0,639	0,537
RNF	61,92%	78,72%	15,762	62,872	19,826	0,411	0,700
RNRE	48,83%	67,21	9,993	47,321	16,862	0,392	0,876
RNRJ	54,32%	75,42%	8,982	49,421	17,421	0,399	0,802
At-the-Money							
Modelos	MEP	MPE	RMSE	MAE	MAPE	TIC	R^2
Black	86,32%	95,21%	45,421	122,532	33,312	0,772	0,582
Vasicek	81,11%	95,82%	31,312	100,321	34,321	0,701	0,611
CIR	75,23%	91,33%	22,872	94,421	31,312	0,531	0,692
RNF	61,32%	76,73%	11,321	59,982	14,321	0,491	0,777
RNRE	48,11%	68,93%	7,983	48,937	12,399	0,374	0,876
RNRJ	54,92%	68,99%	9,32	45,422	13,532	0,366	0,800
In-the-Money							
Modelos	MEP	MPE	RMSE	MAE	MAPE	TIC	R^2
Black	86,99%	95,21%	42,321	121,312	33,873	0,732	0,521
Vasicek	82,32%	96,98%	32,111	102,321	31,321	0,672	0,601
CIR	75,32%	90,89%	26,321	91,321	30,532	0,532	0,688
RNF	60,32%	75,93%	13,424	56,532	14,421	0,483	0,799
RNRE	43,83%	64,42%	9,222	43,824	16,384	0,298	0,921
RNRJ	53,32%	68,54%	6,999	42,321	13,313	0,388	0,876

das métricas de erro de acordo com os graus de *moneyness*. Deve-se perceber, sobretudo, que o modelo de precificação baseado no modelo de Black (1976) apresentou os piores resultados. Os prêmios das

opções obtidos pelos modelos que compõem o *benchmark* indicam a inconsistência de tais metodologias para o mercado de opções de IDI.

Para ilustrar os resultados, selecionamos as opções de compra de IDI no conjunto da amostra teste com um dos *strikes* mais líquido, isto é, as opções de compra com preço de exercício igual a 220.000 pontos do IDI, sendo o valor de cada ponto dado em reais estabelecido pela Bolsa, e com vencimento fevereiro de 2008. As figuras 6.2 a 6.2 apresentam a evolução dos prêmios obtidos pelos modelos para as opções de compra de IDI em relação aos prêmios de fechamento observados no mercado.

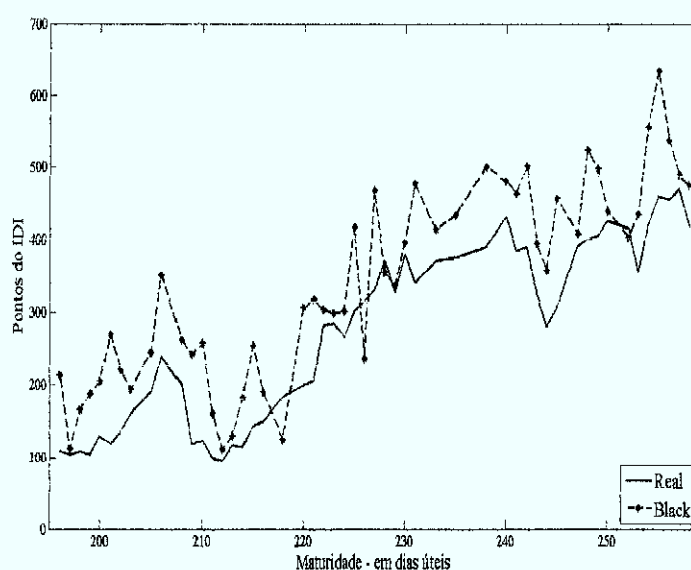


Fig. 6.4: Evolução dos prêmios de mercado e do modelo de Black

É evidente a superioridade dos modelos de redes neurais, uma vez que conseguem acompanhar a evolução do prêmio dessas opções de forma mais acurada que os modelos baseados nos processos de Black, Vasicek e CIR, para o comportamento da taxa de juros de curto prazo. Além disso, mesmo com o aumento do prazo de maturidade dessas opções, os resultados dos modelos de redes neurais não se afastam dos prêmios observados no mercado. Porém, a diferença de performance entre os modelos de inteligência artificial ficam mais claros quando se observam as métricas de erro.

É possível observar o bom desempenho obtido pelos modelos de redes neurais para a avaliação das opções estudadas. Constata-se que, para as opções com vencimentos mais curtos, os modelos teóricos apresentaram resultados similares, porém, os modelos convencionais resultaram em erros de apreçamento maiores de acordo com a elevação da maturidade da opção.

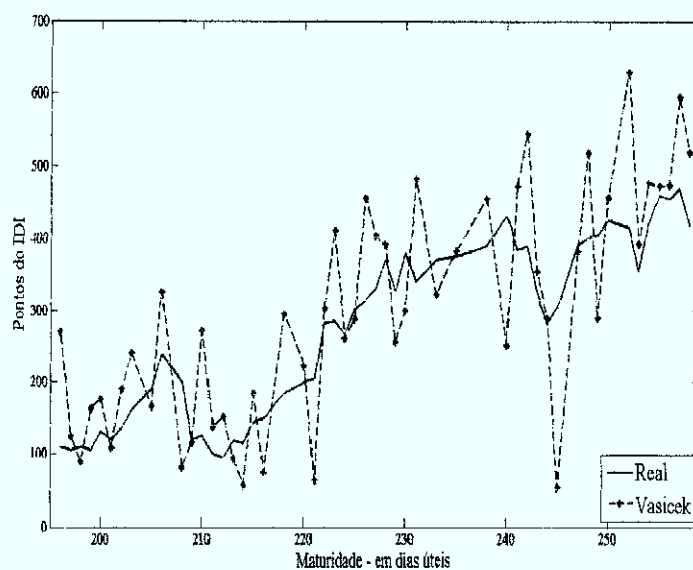


Fig. 6.5: Evolução dos prêmios de mercado e do modelo de Vasicek

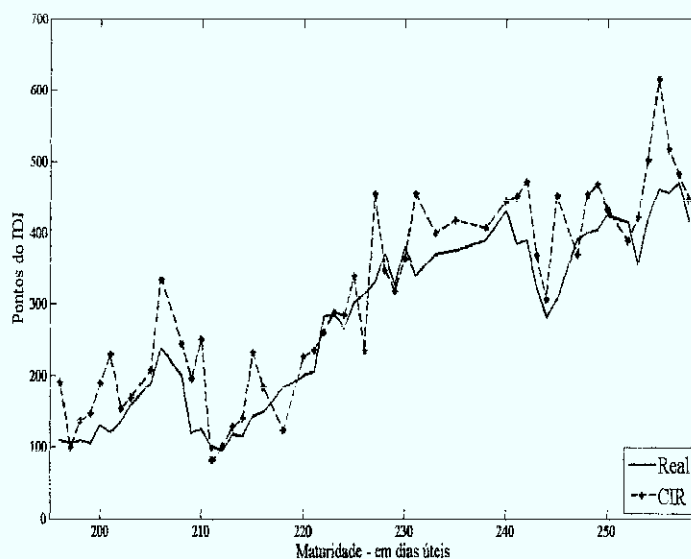


Fig. 6.6: Evolução dos prêmios de mercado e do modelo de CIR

6.3 Avaliação dos Testes Estatísticos

Apesar dos resultados favoráveis aos modelos de inteligência artificial, como descritos na seção anterior, é necessário verificar se a diferença entre os erros obtidos pelos modelos é estatisticamente

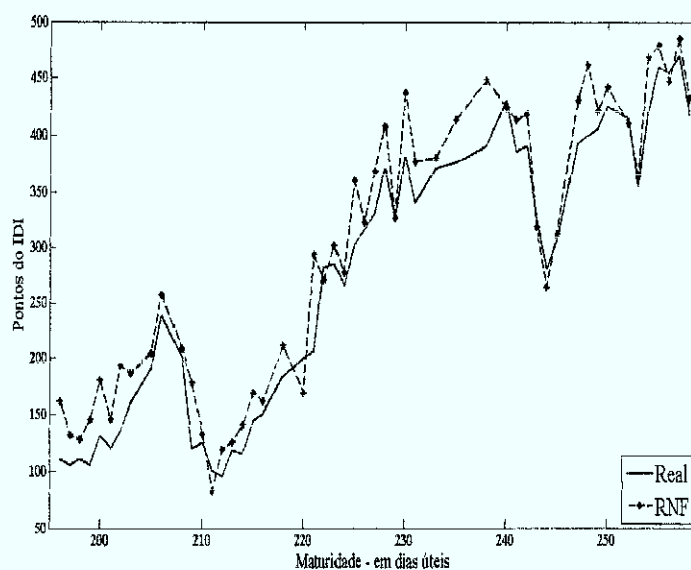


Fig. 6.7: Evolução dos prêmios de mercado e da RNF

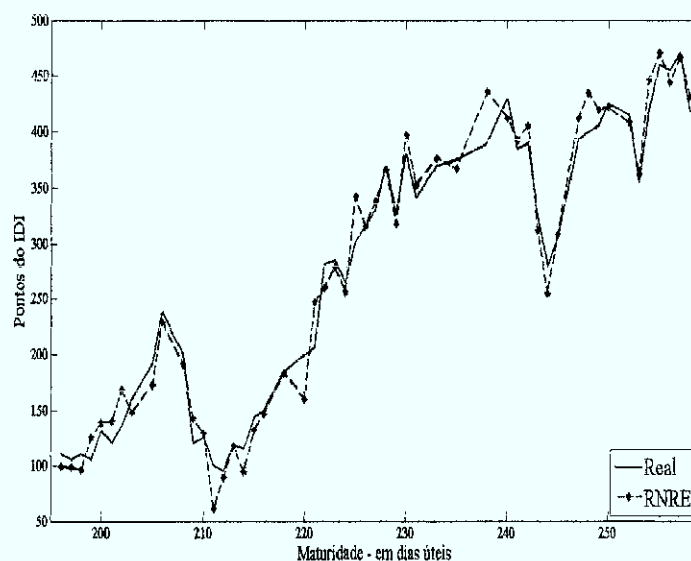


Fig. 6.8: Evolução dos prêmios de mercado e da RNRE

significativa. A tabela 6.10 elucida essa análise por meio dos resultados dos testes AGS, MGN e SIGN.

Os resultados do teste AGS, que examina os erros de previsão em termos do erro quadrático médio dos modelos avaliados, indicaram que os modelos de redes neurais apresentaram superioridade significativa na precificação das opções de compra de IDI no Brasil, no período em questão. Na

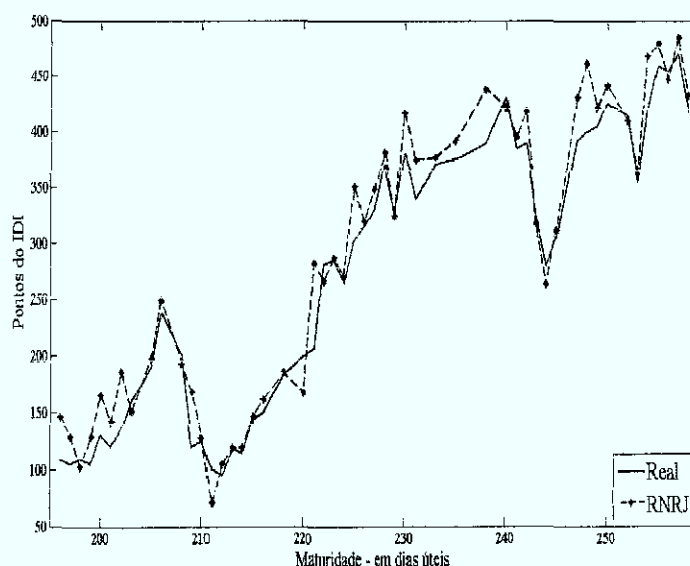


Fig. 6.9: Evolução dos prêmios de mercado e da RNRJ

comparação entre os modelos tradicionais, pode-se perceber que o modelo de CIR foi o que apresentou melhor desempenho. Ao se comparar os modelos de redes neurais entre si, confirmou-se o verificado com base nas métricas de erro, isto é, o modelo de rede recorrente de Elman superou os demais em termos de capacidade para obtenção do processo gerador dos prêmios das opções de compra de IDI.

Os mesmos resultados se verificaram de acordo com o teste MGN, principalmente na comparação do modelo de Vasicek com o modelo CIR, pois neste caso, os testes indicaram que ambos modelos podem ser considerados equivalentes em relação à acurácia para se obter o prêmio das opções avaliadas. Entretanto, ainda os modelos de redes neurais performaram os métodos convencionais, com destaque para as redes recorrentes de Elman e de Jordan.

Por fim, o teste não-paramétrico (SIGN) rejeitou a equivalência dos modelos de inteligência artificial em comparação com os modelos analíticos, com destaque para as redes recorrentes.

Dessa forma, podemos elucidar nossas conclusões da seguinte forma:

- Os testes estatísticos indicaram que os modelos avaliados baseados em redes neurais obtiveram os prêmios para as opções de compra de IDI mais próximos dos observados no mercado, em comparação às metodologias tradicionais, baseadas em fórmulas analíticas fechadas.
- Dentre os modelos que compõem o *benchmark* deste trabalho, os testes paramétricos e não-paramétrico auferiram que os métodos de Vasicek e CIR performaram o modelo de Black, entre-

Tab. 6.10: Testes AGS, MGN e SIGN para avaliação dos modelos de precificação para as opções de compra de IDI européias negociadas na BM&FBOVESPA entre 11/12/2006 e 31/07/2007, amostra com 1.270 observações.

Modelos	Teste AGS ($F_{2,1268}$) [*]	Teste MGN (t_{1269}) [*]	Teste SIGN $N(0,1)$ [*]
Black Vs. Vasicek	64,872 (0,0000)	16,324 (0,0000)	-4,623 (0,0001)
Black Vs. CIR	67,724 (0,0000)	17,922 (0,0000)	-4,762 (0,0001)
CIR Vs. Vasicek	14,762 (0,2687)	3,255 (0,4402)	-1,982 (0,2791)
Black Vs. RNF	75,762 (0,0000)	19,3752 (0,0002)	-8,982 (0,0001)
Black Vs. RNRE	107,987 (0,0000)	25,782 (0,0001)	-7,287 (0,0000)
Black Vs. RNRJ	122,238 (0,0000)	31,972 (0,0000)	-8,982 (0,0000)
Vasicek Vs. RNF	97,362 (0,0000)	28,364 (0,0000)	-7,983 (0,0000)
Vasicek Vs. RNRE	98,872 (0,0000)	34,762 (0,0000)	-8,761 (0,0000)
Vasicek Vs. RNRJ	105,762 (0,0000)	33,999 (0,0000)	-9,872 (0,0000)
CIR Vs. RNF	89,293 (0,0000)	28,333 (0,0000)	-8,763 (0,0000)
CIR Vs. RNRE	102,762 (0,0000)	42,819 (0,0000)	-9,623 (0,0000)
CIR Vs. RNRJ	96,534 (0,0000)	36,829 (0,0000)	-8,777 (0,0000)
RNF Vs. RNRE	77,837 (0,0001)	19,557 (0,0002)	-8,889 (0,0000)
RNF Vs. RNRJ	113,762 (0,0000)	26,983 (0,0000)	-11,659 (0,0000)
RNRE Vs RNRJ	12,653 (0,2318)	4,762 (0,3209)	-1,876 (0,2583)

(*) Os respectivos p-valores estão representados entre parêntesis.

tanto, quando comparados entre si, os testes indicaram equivalência para os modelos de Vasicek e CIR.

- As redes neurais recorrentes de Jordan e Elman podem ser consideradas mais acuradas, quando comparadas com a rede neural progressiva, segundo os testes, e ainda, são consideradas equivalentes quando avaliadas conjuntamente. Mas mesmo assim, a rede *feedforward* superou os modelos tradicionais na valoração das *calls* de IDI.

Capítulo 7

Conclusões

A precificação de opções, em geral, é de grande importância para os agentes de mercado, pois esses papéis são capazes de refletir a informação sobre o futuro de seus ativos objeto, diminuindo, assim, o grau de incerteza presente na economia. Entretanto, os preços desses ativos muitas vezes são obtidos por modelos que não incorporam características reais dos ativos subjacentes, fazendo com que seus resultados difiram dos preços de mercado. Dessa forma, alternativas baseadas em inteligência artificial, como as redes neurais, têm se mostrado promissoras na área de precificação de derivativos, sobretudo opções, demonstrada sua capacidade em se modelar padrões não-lineares de tipo entrada-saída.

No mercado de opções brasileiro, não existe um modelo teórico consistente, baseado em fórmulas fechadas, para se precificar as opções sobre o Índice de Depósitos Interfinanceiros de Um Dia (IDI), uma vez que a Bolsa utiliza o modelo de Black (1976), que é adequado para opções sobre futuros, não sobre taxa de juros. As opções de IDI são peculiares do nosso mercado e fazem com que seja necessária a adequação dos modelos tradicionais norte-americanos e europeus para o caso brasileiro. Com isso, Neto e Pereira (1999) e Barbachan e Ornelas (2003) propuseram modelos analíticos fechados para a precificação das opções de IDI, com base, respectivamente, nos modelos de Vasicek (1977) e Cox et al. (1985). Entretanto, seus resultados ainda evidenciam que os preços teóricos não são capazes de explicar os preços das opções observados no mercado.

Dessa forma, este trabalho avaliou modelos de redes neurais progressivas (*feedforward*) e recorrentes de tipo Elman e Jordan para se obter o prêmio das opções de compra de IDI, no período de janeiro de 2003 a julho de 2007. Esse modelo se baseou nas principais variáveis que afetam o preço da opção de IDI: valor do IDI, maturidade do contrato, preço de exercício, taxa de juros de longo prazo e volatilidade da taxa de juros de curto prazo. Neste caso, a volatilidade foi estimada com base em um

modelo GARCH(p,q) que leva em consideração o efeito duration.

Avaliamos, então, os modelo de Black, Vasicek e CIR, assim como os métodos de redes neurais para a valoração das opções de IDI. Os preços teóricos dos modelos foram comparados com os preços de fechamento dessas opções, observados no mercado. Para tanto, utilizamos métricas erro convencionais e testes estatísticos paramétricos e não-paramétricos para modelos competitivos de predição.

Nossos resultados inferiram a superioridade do modelo de rede neural em relação aos demais modelos, devido aos baixos erros apresentados. Além disso, para as opções *out-of-the-money*, o modelo de rede neural recorrente de Elman apresentou os melhores resultados em termos comparativos, uma vez que para as opções *at-the-money* e *in-the-money* os modelos tradicionais não apresentaram erros tão elevados quanto para as opções fora do dinheiro. Dentre os modelos convencionais, o modelo proposto por Barbachan e Ornelas (2003) apresentou os menores erros. Enquanto que para os modelos baseados em inteligência artificial a rede recorrente de Elman se destacou, no sentido de melhor acurácia para a obtenção dos prêmios das opções de compra de IDI no período avaliado.

Para validar os resultados, os testes estatísticos AGS, MGN e SIGN indicaram que a acurácia de todos modelos de redes neurais foi superior a dos modelos tradicionais, tornados como *benchmark*. Dentre os modelos de rede, os mesmos testes indicaram que as redes recorrentes podem ser consideradas equivalentes para a precificação das opções de IDI, porém, superam, em termos de acurácia, o modelo de rede progressiva. Resultado semelhante se verificou com os modelos convencionais, sendo que os modelos de Vasicek e CIR, de acordo com os testes estatísticos, podem também ser considerados equivalentes, mas superiores quando comparados com o modelo utilizado pela Bolsa (Black).

Com isso, verificou-se a inconsistência do modelo utilizado pela Bolsa, assim como dos modelos de Neto e Pereira (1999) e Barbachan e Ornelas (2003), para a precificação das opções de compra de IDI no período considerado. Entretanto, os modelos de redes neurais mostraram um grande potencial para o mesmo fim, resultando em estruturas que permitem captar o processo gerador dos prêmios desses papéis, o que se mostra como uma alternativa para se tentar ampliar a liquidez nesse mercado que, talvez, por deficiências teóricas, apresentam baixo volume de negociações.

7.1 Trabalhos Futuros

Os resultados apresentados indicam que os modelos de redes neurais se mostram como ferramentas que fornecem um bom desempenho quando o objetivo é avaliar o prêmio de uma opção de IDI no mercado brasileiro. Como continuidade, podemos focar no estudo de novas estruturas de rede, capazes

de incorporar mais características que representam os fatos estilizados do mercado de juros no Brasil. São diversos os modelos alternativos, como, por exemplo, redes neurais bayesianas, uma vez que permite uma maior tratabilidade estatística aos dados.

Além disso, pode-se avaliar outros modelos relacionados à inteligência computacional, tal como os modelos baseados em inferência *fuzzy*, algoritmos genéticos e programação genética, uma vez que esse campo de estudo têm mostrado resultados estimulantes quando associados à chamada engenharia financeira.

Por fim, tais modelos, como os apresentados neste trabalho, podem ser aplicados à outros segmentos do mercado de opções brasileiro, o que permite avaliar sua tratabilidade nos mais diferentes tipos de opções, permitindo uma maior generalização dos resultados.

Referências Bibliográficas

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6), 716–723.
- Almeida, C. I. R., Vicente, J. V., 2003. Pricing and hedging brazilian fixed income options. In: *Anais do VI Encontro da Sociedade Brasileira de Finanças*. Vitória, Brasil.
- Almeida, L. A., Yoshino, J. A., Schimer, P. P., 2003. Derivativos de renda fixa no brasil: modelo Hull-White. *Política e Planejamento Econômico* 33 (2), 299–333.
- Amilon, H., 2001. A neural network versus black-scholes: A comparison of pricing and hedging performances. Working Paper: Lund University 1.
- Anders, U., Korn, O., Schmitt, C., 1998. Improving the pricing of options: A neural network approach. *Journal of Forecasting* 17 (5–6), 369–388.
- Andreou, P. C., Charalambous, C., Martzoukos, S. H., 2008. Pricing and trading european options by combining artificial neural networks and parametric models with implied parameters. *European Journal of Operational Research* 185 (3), 1415–1433.
- Ashley, R., Granger, J. W. C., Schmalense, R., 1980. Advertising and aggregate consumption: an analysis of causality. *Econometrica* 48 (5), 1149–1167.
- Ball, A. C., Touros, W. N., 1983. Bond price dynamics and options. *Journal of Financial and Quantitative Analysis* 18, 517–531.
- Ballini, R., 2000. Análise e previsão de vazões utilizando modelos de séries temporais, redes neurais e redes neurais nebulosas. Ph.D. thesis, Faculdade de Engenharia Elétrica e de Computação - Universidade Estadual de Campinas.

- Barbachan, J. S. F., Ornelas, J. R. H., 2003. Apreçamento de opções de IDI usando o modelo CIR. *Estudos Econômicos* 33 (2), 287–323.
- Barbedo, C. H. S., Vicente, J. V. M., Lion, O. M. B., 2009. Pricing asian interest rate options with a Three-Factor HJM model. Working paper do Banco Central.
- Bennel, J., Sutcliffe, C., 2004. Black-Scholes versus artificial neural networks in pricing FTSE-100 options. *Intelligent Systems in Accounting, Finance and Management* 12 (4), 243–260.
- Bishop, C. M., 1995. *Neural Networks for Pattern Recognition*.
- Black, F., 1976. The pricing of commodity contracts. *Journal of Financial Economics* 3, 167–179.
- Black, F., Derman, E., Toy, W., 1990. A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal* 46, 33–39.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81 (3), 637–654.
- Brennan, M., Schwartz, E., 1982. An equilibrium model of bond prices and a test of market efficiency. *Journal of Financial and Quantitative Analysis* 17, 301–329.
- Brigo, D., Mercurio, F., 2001. *Interest rate models theory and practice*, 2nd Edition. Springer.
- Callen, J. L., Kwan, C. C. Y., Yip, P. C. Y., Yuan, Y., 1996. Neural network forecasting of quarterly accounting earnings. *Journal of Forecasting* 4 (12), 475–482.
- Cassetari, A., Neto, L. F., 2001. A precificação de derivativos de taxa de juro no brasil. *Revista BM&F* 145, 35–47.
- Chan, C. K., Karolyi, G. A., Longstaff, F. A., Sanders, A. B., 1992. The volatility of Japanese interest rates: A comparison of alternative term structure models. In: *Pacific Basin Capital Markets Research*. Elsevier.
- Chiarella, C., Kwon, O., 1999. Forward rate dependent markovian transformations of the heath-jarrow-morton term structure model. *Research paper series - University of Technology (Sydney)* (5).
- Constantinides, G. M., Ingersoll, J. E., 1984. Optimal bond trading with personal taxes. *Journal of Financial Economics* 13 (3), 299–335.

- Cox, J. C., Ingersoll, J. E., Ross, S. A., 1985. A theory of the term structure of interest rates. *Econometrica* 55, 385–407.
- Dothan, U. L., 1978. On the term structure of interest rates. *Journal of Financial Economics* 6, 59–69.
- Feller, W., 1951. Two singular diffusion problems. *Annals of Mathematics* 54, 173–182.
- Freitas, J. F. G. D., Niranjana, M., Gee, A. H., Doucet, A., 2000. Sequential monte carlo methods on train neural network models. *Neural Computation* 12 (4), 955–993.
- Freitas, O. S., Souza, A. A., 2002. Utilização de redes neurais na precificação de opções. *Resenha BM&F* 150, 63–73.
- Garcia, M. G. P., Barcinski, A., Julho 1998. Capital flows to Brazil in the nineties: Macroeconomic aspects and the effectiveness of capital controls. *The Quarterly Review of Economics and Finance* 38 (3), 319–357.
- Garcia, R., Gençay, R., 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics* 94 (1–2), 93–115.
- Geigle, S. D., Aronson, E. J., 1999. An artificial neural network approach to the valuation of options and forecasting of volatility. *Journal of Computational Intelligence in Finance* 7 (6), 19–25.
- Ghaziri, H., Elfakhani, S., Assi, J., 2001. Neural networks approach to pricing options. *Neural Network World* 10 (1–2), 271–277.
- Gluckstern, M. C., Francisco, G., Jr, W. E., 2002. Aplicação do modelo de hull-white à precificação de opções de IDI. In: *Anais do II Encontro da Sociedade Brasileira de Finanças*. Rio de Janeiro, Brasil.
- Granger, J. W. C., Newbold, P., 1977. *Forecasting economic time series*. New York: Academic Press.
- Hannan, E. J., Quinn, B. G., 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society* 41 (2), 190–195.
- Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation*, 2nd Edition. Prentice Hall, New Jersey.
- Healy, J., Dixon, M., Read, B., Cai, F. F., 2002. A data-centric approach to understanding the pricing of financial options. *European Physical Journal B* 27 (2), 219–227.

- Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: a new methodology. *Econometrica* 60, 77–105.
- Ho, T. S. Y., Lee, S. B., 1986. Term structure movements and pricing interest rate contingent claims. *Journal of Finance* 41 (5), 1011–1029.
- Hull, J., White, A., 1990. Pricing interest rate derivative securities. *Review of Financial Studies* 3, 573–592.
- Hull, J. C., 2003. *Options, Futures and Other Derivatives*, 5th Edition. Hardcover.
- Hutchinson, J. M., Lo, A., Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance* 49 (3), 851–889.
- Jamshidian, F., 1989. An exact bond option formula. *Journal of Finance* 44, 205–209.
- Junior, A. F., Greco, F., Lauro, C., Francisco, G., Rosenfeld, R., Oliveira, R., 2003. Application of hull-white model to brazilian idi options. In: *Anais do III Encontro da Sociedade Brasileira de Finanças*. Florianópolis, Brasil.
- Kaastra, I., Boyd, M., 1995. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10 (3), 215–236.
- Kelly, L. D., 1994. Valuing and hedging american put options using neural networks. Working Paper: University of California 2.
- Kitamura, T., Ebisuda, S., 1998. Pricing options using neural networks. Working Paper: Carnegie Mellon University 11.
- Krause, J., 1996. Option pricing with neural networks. In: *Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing*. Verlag Mainz - Aachen, pp. 2206–2210.
- Lajbcygier, P., Boek, C., Palaniswami, M., Flitman, F., 1996. Comparing conventional and artificial neural network models for the pricing of options on futures. *Neurovest Journal* 4 (5), 16–24.
- Lehmann, E. L., 1988. *Nonparametric statistical methods based on ranks*. New Jersey: Prentice Hall.
- Maciel, L. S., Ballini, R., Silveira, R. L. F., 2009. Apreçamento de opções de dólar no Brasil: uma avaliação dos modelos de redes neurais. In: *XXXVII Encontro Nacional de Economia - ANPEC*. Foz do Iguaçu - Paraná.

- Malliaris, M., Salchenberger, L., 1996. A neural network model for estimating option prices. *Journal of Applied Intelligence* 3 (3), 193–206.
- Mandic, D. P., Chambers, J. A., 2001. *Recurrent neural networks for prediction: Architectures, learning algorithms and stability*. Wiley.
- Merton, R., 1973. The theory of rational option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.
- Munnik, J., 1992. *The Valuation of Interest Rate Derivative Securities*. NY: Routledge.
- Neto, C. A. V., Pereira, P. V., 1999. Derivação de uma fórmula para o cálculo do preço livre de arbitragem das opções sobre o índice de depósitos financeiros de um dia - IDI. Working Paper - IBMEC 2.
- Niranjan, M., 1996. Sequential tracking in pricing financial options using model based and neural network approaches. In: Mozer, M. C., Petsche, T. (Eds.), *Advances in Neural Information Processing Systems*. Vol. 9. MIT Press - Cambridge, pp. 960–966.
- Ormonet, D., 1999. A regularization approach to continuous learning with an application to financial derivatives pricing. *Neural Networks* 12 (10), 1405–1412.
- Qi, M., Maddala, G. S., 1996. Option pricing using artificial neural networks: the case of s&p 500 index call options. In: *Proceedings of 3rd International Conference On Neural Networks in the Capital Markets*. Londres, Reino Unido.
- Rendleman, R. J., Bartter, B. J., 1980. The pricing of options on debt securities. *Journal of Financial and Quantitative Analysis* 15 (1), 11–24.
- Ritchken, P., Sankarasubramanian, L., 1995. Volatility structures of forward rates and the dynamics of the term structure. *Mathematical Finance* 5 (1), 55–72.
- Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning internal representations by error propagation. In: Rumelhart, D. E., McClelland, J. L. (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press.
- Schwarz, G. E., 1978. Estimating the dimension of a model. *Annals of Statistics* 6 (2), 461–464.

- Shapiro, A., 2003. Capital market applications of neural networks, fuzzy logic and genetic algorithms. In: Proceedings of the 13th International AFIR Colloquium. Maastricht, Holanda.
- Shibata, R., 1976. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* 63 (1), 117–126.
- Stone, M., 1974. Cross-Validation choice assessment of statistical prediction. *Journal of the Royal Statistical Society B* 36, 111–133.
- Vasicek, O. A., 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
- Wuine, R. D., Francois-Heude, A., Meurisse, B., 2001. Market microstructure and option pricing: A neural network approach. Working Paper: Facultés Universitaires Catholiques de Mons 1.
- Yao, J., Li, Y., Tan, L., 2000. Option price forecasting using neural networks. *Omega* 28 (4), 455–466.