

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

BRENO DE OLIVEIRA ARANTES

**PREVISÃO DA ESTRUTURA A TERMO DA TAXA DE JUROS
BRASILEIRA USANDO REDES NEURAIAS ARTIFICIAIS**

Porto Alegre

2013

BRENO DE OLIVEIRA ARANTES

**PREVISÃO DA ESTRUTURA A TERMO DA TAXA DE JUROS
BRASILEIRA USANDO REDES NEURAIS ARTIFICIAIS**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Orientador: Prof. Dr. Marcelo Savino Portugal.

Co-orientador: Prof. Dr. João Fróis Caldeira.

Porto Alegre

2013

CIP - Catalogação na Publicação

Arantes, Breno de Oliveira

Previsão da estrutura a termo da taxa de juros brasileira usando redes neurais artificiais. / Breno de Oliveira Arantes. -- 2013.

39 f.

Orientador: Marcelo Savino Portugal.

Coorientador: João Fróis Caldeira.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2013.

1. Finanças. 2. Métodos Quantitativos. 3. Estrutura a Termo da Taxa de Juros. 4. Redes Neurais Artificiais. I. Portugal, Marcelo Savino, orient. II. Caldeira, João Fróis, coorient. III. Título.

BRENO DE OLIVEIRA ARANTES

**PREVISÃO DA ESTRUTURA A TERMO DA TAXA DE JUROS
BRASILEIRA USANDO REDES NEURAIAS ARTIFICIAIS**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia, com ênfase em Economia Aplicada.

Aprovada em: Porto Alegre, 02 de maio de 2013.

BANCA EXAMINADORA:

Prof. Dr. Gilberto de Oliveira Kloeckner
Universidade Federal do Rio Grande do Sul – UFRGS

Prof. Dr. Guilherme Ribeiro de Mâcedo
Universidade Federal do Rio Grande do Sul – UFRGS

Prof. Dr. Márcio Poletti Laurini
Universidade de São Paulo – USP

RESUMO

Avaliamos as previsões fora da amostra da curva de juros geradas por modelos de redes neurais artificiais e as comparamos com os modelos tradicionalmente usados para este fim. A curva de juros foi segmentada em três regiões distintas e para cada uma dessas regiões e horizontes de previsão, foram estimados duas classes de modelos de redes neurais (modelos dinâmicos e modelos estáticos), totalizando 198 modelos estimados. Afim de verificar a significância estatística das previsões dos modelos de RNA em relação aos outros modelos, foi realizado o teste Diebold-Mariano. Os resultados mostram que em média, os modelos estimados através de RNA conseguiram superar as previsões realizadas pelo *random walk* em todos os horizontes de previsão, sendo essas previsões em torno de 2% e 5% melhores para os horizontes de 1 dia e 1 semana a frente, e de 12% e 7% melhores para os horizontes mais longos, de 1 mês e 3 meses. Além disso, apresentou previsões em torno de 15% e 10% melhores que o modelo de NS para os horizontes de 1 mês e 3 meses a frente. Concluimos que os modelos de redes neurais são capazes de realizar previsões superiores para todos os horizontes testados, principalmente para região de curto prazo da curva, com destaque especial para as previsões com horizontes de 1 dia e 1 semana a frente.

Palavras-chave: Estrutura a termo da taxa de juros. Previsão. Redes neurais artificiais.

ABSTRACT

We evaluate the out of sample forecasts of the yield curve generated by artificial neural network models and compare them with the models traditionally used for this purpose. The yield curve was segmented into three distinct regions and for each region and forecast horizons, we estimated two classes of neural network models (dynamic models and static models), totaling 198 models estimated. In order to check the statistical significance of the model predictions of RNA compared to other models, was performed the Diebold-Mariano's test. The results show that on average, the models estimated using RNA overcame the predictions made by the random walk at all forecast horizons, and these forecasts around 2% and 5% better for horizons of 1 day and 1 week forward and 12% and 7% better for longer horizons, 1 month and 3 months. Moreover, forecasts showed around 15% and 10% better than the NS model for horizons of 1 month and 3 months ahead. We conclude that the neural network models are capable of superior forecasts for all horizons tested, especially for short-term region of the curve, with particular attention to the forecasts with horizons of 1 day and 1 week ahead.

Keywords: Term structure of interest rates. Forecast. Artificial neural networks.

SUMÁRIO

1	INTRODUÇÃO	6
2	ESTRUTURA A TERMO DA TAXA DE JUROS.....	10
2.1	Titulo zero-cupom	10
2.2	Taxa de juros	10
2.3	Função de desconto, Curva Foward e Curva de juros	11
3	REDES NEURAIIS ARTIFICIAIS.....	14
3.1	Redes Multicamadas	15
3.2	Treinamento de Redes Neurais	16
3.3	Aplicação de RNA à previsão da estrutura a termo da taxa de juros.	19
3.3.1	Classe de modelos e variáveis de entrada.....	19
3.3.2	Número de camadas ocultas e neurônios.....	20
3.3.3	Escolha dos modelos	21
3.3.4	Treinamento e Função de ativação	21
4	MODELOS COMPETIDORES.....	23
4.1	O modelo de Diebold e Li	23
4.2	O modelo de Svensson	24
4.3	O modelo AR, VAR e Random Walk	24
5	DADOS, MÉTODO DE COMPARAÇÃO E ANÁLISE DE RESULTADOS	26
5.1	Dados	26
5.2	Método de Comparação.....	28
5.3	Análise de Resultados.....	29
6	CONCLUSÕES.....	36
	REFERÊNCIAS	37

1 INTRODUÇÃO

Estrutura a termo da taxa de juros é uma relação entre o yield e a maturidade de um título zero-cupom e fornece uma medida do retorno que o investidor pode esperar dos diferentes períodos de investimento em um mercado de renda fixa. A modelagem e previsão da estrutura a termo da taxa de juros é extremamente importante para economia e finanças. Produzir previsões dela é crucial para precificação de ativos e derivativos financeiros, seleção de carteiras, gestão de risco e condução da política monetária.

Todos esses desafios essenciais em economia financeira envolvem as taxas de juros e a evolução dinâmica da curva de juros, como resultado a modelagem e estimação da estrutura a termo tem recebido considerável atenção de pesquisadores desde os anos sessenta.

Durante muito tempo a literatura a respeito do assunto se concentrou no debate da Hipótese da Expectativa da estrutura a termo da taxa de juros (ETTJ) formulado por Fisher (1896). Nesta, as taxas de juros dos títulos de diferentes maturidades são relacionadas com as expectativas dos agentes do mercado sobre a evolução da taxa de juros no futuro. De acordo com esta hipótese, as taxas da curva de juros podem fornecer previsões ótimas para a evolução da taxa à vista, portanto, nenhum modelo alternativo poderia produzir previsões superiores. No entanto, estudos empíricos rejeitaram a Hipótese da Expectativa (HE), pois foram encontrados resultados que apontam que os prêmios de risco são variantes ao longo do tempo (CALDEIRA; FURLANI, 2011) - apesar que a HE funciona bem para determinados trechos da curva de juros.

Posteriormente, a literatura se concentrou no conceito de equilíbrio. Nessa abordagem, as taxas de juros das diversas maturidades são funções afins (constante mais linear) de um pequeno número de fatores latentes (fatores não observáveis), que são extraídos por meio de um painel de taxas de juros no tempo e maturidades. Além do mais, os modelos “padrão” dessa classe impõem restrições nas equações de forma a excluir oportunidades de arbitragem. Os modelos de equilíbrio fornecem ajustes exatos para a curva de juros observada (LONGSTAFF; SCHWARTZ, 1992). O modelo pioneiro de Vasicek (1977) foi seguido por outros, como Brennan e Schwartz (1979) e Cox e Ingersoll (1985).

Os modelos de equilíbrio normalmente utilizam a classe dos modelos afins da ETTJ. Nesse tipo de abordagem, as taxas de juros das diversas maturidades são funções afins (constante mais linear) de um pequeno número de fatores latentes (fatores não observáveis), que são extraídos por meio de um painel de taxas de juros no tempo e maturidades. Além do

mais, os modelos “padrão” dessa classe impõem restrições nas equações de forma a excluir oportunidades de arbitragem.

A outra classe de modelo tem seu foco na existência de uma característica essencial para a precificação de derivativos de taxas de juros: ausência de condições de arbitragem. Como o próprio nome já indica, essa classe é conhecida como modelos de não-arbitragem. Para tanto, os parâmetros dessa classe são ajustados de modo a coincidir o preço do ativo subjacente do modelo com o seu preço de mercado – calibração do modelo para fins de precificação, não permitindo a realização de previsões. Os principais destaques dessa classe são os modelos de Ho e Lee (1986), Hull e White (1990) e Black et al. (1990).

Os modelos afins possuem propriedades dinâmicas, desta forma, pode ser testada a qualidade de suas previsões. Duffee (2002) testa o desempenho dos modelos “padrões” em realizar previsões para a taxa de juros americana - por padrão o autor considera diversos modelos, que incluem a generalização multifatorial de Vasicek (1977) e Cox e Ingersoll (1985), com também, o modelo afim “completo” de Dai e Singleton (2002). O resultado encontrado foi que além de falharem ao fazer as previsões, o modo como estas ocorriam era bastante preocupante. Os modelos produziam erros de previsões das taxas que são negativamente relacionadas com a inclinação da curva de juros, ou seja, falham em replicar a relação empírica encontrada entre os retornos esperados e a inclinação da curva.

Devido ao fato do baixo desempenho dos modelos afins, Diebold e Li (2006) adotam uma abordagem que não utiliza nem o arcabouço da não-arbitragem, que se foca no ajuste em cross-section das taxas, e nem o arcabouço dos modelos afins, que se foca na dinâmica ao longo do tempo, mas dão pouca atenção ao ajuste em cross-section. Os autores utilizam uma variação do modelo de fatores de Nelson (1987) para modelar toda a curva de juros, período por período, através de três fatores. Além do mais, os mesmos, apontam que esses três parâmetros podem ser interpretados como fatores (latentes) correspondentes ao nível, inclinação e curvatura da curva de juros. Os resultados obtidos com o modelo foram promissores, principalmente para longos horizontes.

De Pooter et al. (2010) compara empiricamente diversos modelos de fatores e diferentes formas de estimação, através da comparação do poder de previsão dos modelos fora da amostra. Outro trabalho que deve ser destacado, apesar de não realizar previsões, é o de Koopman et al. (2010), no qual os autores consideram os parâmetros do modelo variantes no tempo.

Existem também os modelos de previsão que são baseados na interação entre a curva de juros e variáveis macroeconômicas; a abordagem usual é estender os modelos de fatores

latentes existentes através da adição de fatores macroeconômicos observáveis. De Pooter et al. (2010) estenderam a abordagem de Nelson-Siegel para incluir fatores macroeconômicos observáveis e Diebold e Li (2006), também com base em Nelson-Siegel, encontram ganhos preditivos ao se adicionar variáveis macroeconômicas no modelo. Temos também, Ang e Piazzesi (2003), que desenvolveram um modelo afim de fatores latentes, acrescidos de variáveis macroeconômicas e Mönch (2008), que analisou a capacidade preditiva dos componentes principais da curva de juros envolvendo dados macroeconômicos de taxas de juros.

Outra vertente de modelos são aqueles que se utilizam de dados macroeconômicos para prever a curva de juros, no entanto, não através da abordagem dos modelos de fatores. Como exemplos, temos o trabalho de Evans e Marshall (1998), que se utilizam do VAR estrutural e Hördahl et al. (2006).

Como dito anteriormente, os modelos baseados na abordagem de Nelson-Siegel, apesar de possuírem boas propriedades estatísticas, não criam restrições para que não existam oportunidades de arbitragem. Esta condição implica que a ETTJ não deve permitir operações entre títulos de diferentes maturidades que resultem em lucros livres de risco. Nesse sentido, Christensen et al. (2007) derivam um modelo afim que mantém a estrutura de fatores de Nelson-Siegel mas que possui a qualidade desejada de não permitir a existência arbitragem. Essa abordagem originou uma nova classe de modelos afins da ETTJ, conhecidos pela sigla AFDNS. Entretanto, o incremento da previsão advindo da imposição de ausência de arbitragem não é consenso.

É importante destacar que, nos últimos 20 anos a literatura relacionada à previsão de séries financeiras tem registrado importantes avanços, principalmente no que diz respeito ao desenvolvimento de metodologias que permitam melhor entender os padrões presentes nos dados financeiros. Essas novas metodologias estão permitindo identificar e prever sistemas complexos, em que se considera a existência de relações dinâmicas não lineares nos dados; situação essa, impossível de ser replicada pelos modelos lineares tradicionais.

No Brasil, vários trabalhos se dedicaram a realizar previsões da ETTJ. Alguns destes propõem alguma expansão no modelo de fatores de Diebold-LI [Caldeira et al. (2010); Laurini e Hotta (2010); Almeida et al. (2009); de Rezende et al. (2008)], outros se preocupam em fazer a previsão sem se utilizar da abordagem dos modelos de fatores [Leite et al. (2009); Lima et al. (2006); Vereda et al. (2008); Audrino e Medeiros (2011)] ou então, apenas em testar a capacidade preditiva de vários tipos de modelos (VICENTE; TABAK, 2008).

O objetivo desse trabalho é avaliar as previsões fora da amostra da curva de juros utilizando o arcabouço de redes neurais artificiais (RNA). Os modelos baseados na estrutura de redes neurais artificiais, apesar de serem vistos como modelos “black box” - dado que às vezes é difícil atribuir significado econômico para as relações estimadas - oferecem uma classe bem geral de modelos não lineares, que tem sido bastante utilizados como ferramentas de previsão em diversas áreas e também em finanças [Zhang et al. (1998); Nakamura (2005); Moshiri e Cameron (2000); Choudhary e Haider (2012); Tkacz (2001)]. No entanto, existem relativamente poucas aplicações de redes neurais artificiais para previsão da curva de juros - dois exemplos são Abid e Ben Salah (2002) e Täppinen (1998). Além disso, White (1992) mostra que as RNA são capazes de gerar boas aproximações de relações lineares (sem nenhuma camada oculta) e não lineares (com pelo menos uma camada de neurônios oculta) entre variáveis. As RNA “aprendem” as relações entre variáveis através de um processo de tentativa e erro. Em termos mais precisos, White (1981) mostrou que há uma convergência assintótica dos parâmetros de um modelo estimado por mínimos quadrados não lineares e os pesos ótimos “aprendidos” por uma RNA.

Nesse trabalho, foram estimadas duas classes de redes neurais (dinâmica e estática) para cada uma das regiões da curva de juros - seccionamos a curva em três regiões distintas, pois consideramos importante levar em consideração o fato de que a dinâmica das taxas depende da região que esta se encontra - totalizando 192 modelos estimados. Para verificar a significância estatística das previsões dos modelos de RNA em relação aos outros modelos, foi realizado o teste Diebold-Mariano. Os resultados apontam que os modelos de redes neurais artificiais são capazes de realizar previsões superiores para todos os horizontes testados e principalmente para região de curto prazo da curva, com destaque especial para as previsões com horizontes de 1 dia e 1 semana.

O trabalho é dividido em 5 seções além desta introdução, onde será apresentado uma breve explanação sobre conceitos relativos à estrutura a termo da taxa de juros (seção 2) e sobre redes neurais artificiais e seu uso neste trabalho (seção 3), os modelos tradicionalmente utilizados para realizar previsões da curva de juros (seção 4), como foi obtida a base de dados utilizada e os resultados alcançados com essa pesquisa (seção 5) e por fim, apresentaremos nossas conclusões (seção 6).

2 ESTRUTURA A TERMO DA TAXA DE JUROS

Nesta seção são apresentados alguns conceitos e definições relacionados com a estrutura a termo da taxa de juros, bem como as relações existentes entre as diferentes taxas.

2.1 Título zero-cupom

Títulos são instrumentos que podem ser usados por governos ou empresas privadas para financiar suas atividades; são instrumentos de dívida, os quais, diferentemente de ações, não implicam propriedade da firma por parte do comprador. Basicamente, o comprador de um título empresta dinheiro ao emissor, por um intervalo de tempo que termina na maturidade do título. Na maturidade, o emissor paga ao detentor do título uma quantidade de dinheiro correspondente ao valor de face, também chamado o valor do par, do título. Além disso, podem ser feitos pagamentos periódicos, chamados cupons.

Os títulos zero-cupom (zero-coupon bonds) são títulos que prometem apenas um único pagamento em uma data futura, conhecida como data de vencimento do título (ou maturidade). Uma importante destaque, é que estes títulos são tipicamente caracterizados por curtas maturidades e que estes são fundamentais para precificação de outros títulos.

2.2 Taxa de juros

A taxa interna de retorno, ou simplesmente yield to maturity, no tempo t de um título com maturidade em τ é definida implicitamente como a taxa $y_y(\tau)$ que desconta seu fluxo de pagamentos obtendo o preço P :

$$P_t(\tau) = \sum_{i=1}^n C_i e^{-(t_i-t)y_t(\tau)} \quad (1)$$

O lado direito da expressão é o valor presente, descontado à taxa $y_y(\tau)$, dos pagamentos de restantes para o detentor do título. Para um título zero-cupom, a expressão reduz-se a:

$$P_t(\tau) = e^{-y_t(\tau)\tau} \quad (2)$$

O preço $P_t(\tau)$ de um título zero-cupom é esperado crescer gradualmente com o tempo até a maturidade τ , quando alcança seu máximo. O aumento no preço do título ao longo do período em que ele é mantido representa o retorno para o seu detentor. O verdadeiro aumento no preço, uma vez que é determinado pelas forças do mercado, pode não ser estável, e oscilar ao longo do tempo. Assim, é útil ter alguma medida do aumento futuro no preço do título que está implícito em $P_t(\tau)$. O yield to maturity (taxa interna de retorno) $y_t(\tau)$ no tempo t , para um título com maturidade τ , pode ser definido, dado $P_t(\tau)$ como a taxa estável, à qual o preço do título deveria crescer, para que o título valha seu valor máximo, no vencimento.

2.3 Função de desconto, Curva Forward e Curva de juros

A estrutura a termo das taxas de juros é representada por um conjunto de pontos no espaço taxa de juros spot versus prazo, no qual cada ponto corresponde a uma taxa de juros $y_t(\tau)$ associada a uma maturidade τ , obtida com base em algum título negociado no mercado, (VARGA, 2009).

Em qualquer ponto do tempo t , existirá uma coleção de títulos sem cupom que diferem uns dos outros apenas pela maturidade. No entanto, em um dado momento do tempo pode não existir título disponível para todas as maturidades desejadas, já que no mercado financeiro não são negociados títulos para todas as maturidades possíveis.

A função de desconto é uma das construções mais fundamentais que descreve a estrutura a termo da taxa de juros, geralmente a partir da qual as demais curvas são derivadas. Seja $P_t(\tau)$, o preço de um título zero-cupom, no tempo t , e que paga \$1, na maturidade τ . Por suposição, todo título zero-cupom é livre de default e tem preços estritamente positivos.

Assim, a função de desconto é definida por:

$$P_t(\tau) = e^{-y_t(\tau)\tau} \quad (3)$$

A taxa de juros $y_t(\tau)$, à qual o título é descontado, é a taxa interna de retorno do título zero-cupom, no tempo t , e com maturidade τ , expressa por:

$$y_t(\tau) = \frac{-\ln(P_t(\tau))}{\tau} \quad (4)$$

A taxa forward, no tempo t , aplicada para o intervalo de tempo entre τ_1 e τ_2 , é definida por:

$$f_t(\tau_1, \tau_2) = \frac{1}{\tau_1 - \tau_2} \int y(x) dx \quad (5)$$

O mesmo argumento se aplica para taxas forwards para k -períodos. A taxa forward pode ser interpretada como a taxa marginal de retorno para manter um título por um instante adicional. O limite da expressão (2) quando τ_1 se aproxima de τ_2 , denotada por $f_t(\tau)$, é a taxa forward instantânea:

$$f_t(\tau) = -\frac{P_t'(\tau)}{P_t(\tau)} \quad (6)$$

A curva de taxa forward instantânea $f_t(\tau)$, fornece a taxa de decaimento da função de desconto $P_t(\tau)$ em cada ponto τ . A curva de juros é a média da taxa de decaimento para o intervalo de 0 a τ , expressa por:

$$y_t(\tau) = \frac{1}{\tau} \int_0^\tau f_t(x) dx \quad (7)$$

A função $f_t(\tau)$, das taxas forward, descreve a taxa de retorno (instantânea) de um investimento mantido por um intervalo de tempo muito curto. A curva de taxas forwards instantâneas é uma construção teórica muito importante, apesar de seu valor para uma única maturidade τ ser de pouco interesse prático, devido ao elevado custo de transação para se fazer um contrato entre dois pontos no futuro se esses dois pontos forem muito próximos um do outro. Apenas a média de $f_t(\tau)$ para um intervalo de tempo futuro é de interesse prático.

Em qualquer ponto do tempo t , haverá um conjunto de títulos com diferentes maturidades, τ , e diferentes fluxos de pagamentos, que podem ser usados para estimar as curvas de juros, curvas de desconto e curvas forwards, que não são observáveis na prática. Existem algumas abordagens para construção de curvas de juros. McCulloch (1971, 1975) e Vasicek e Fong (1977) constroem curvas de juros a partir de curvas de desconto suaves estimadas, convertendo-as para taxas nas maturidades relevantes. A primeira abordagem faz uma interpolação da função de desconto com cubic splines e a segunda sugere que o emprego

de splines exponenciais para ajustar a função de desconto, que eliminaria o problema da divergência para maturidades mais longas.

Também foram empregados modelos estatísticos não lineares para estimar a estrutura a termo da taxa de juros, ao invés da função de desconto, como em Nelson e Siegel (1987), Svensson (1994) entre outros. Esses modelos se mostraram bastante úteis para análise e precificação de títulos de renda fixa, com destaque para o primeiro trabalho, que recebeu uma nova interpretação em Diebold e Li (2006), onde os fatores de curto, médio e longo prazos passaram a ser interpretados como fatores de nível, inclinação e curvatura.

3 REDES NEURAIAS ARTIFICIAIS

As redes neurais artificiais originaram-se na tentativa de imitar o funcionamento do cérebro humano, com o objetivo de criar um modelo que possuísse suas características, ou seja, que possuísse um sistema de processamento distribuído não paralelo composto por unidades de processamento simples (neurônios). Como no cérebro, o conhecimento é adquirido pela rede através de um processo de aprendizagem.

O conhecimento da rede é armazenado em suas conexões através dos pesos, o que mostra quão relevantes são as conexões no processo de aprendizagem de uma RNA. Uma rede treinada com um número baixo de neurônios (poucas conexões) pode não aproveitar o máximo de seu potencial. No entanto, uma rede treinada com um número excessivo de neurônios (muitas conexões) pode adaptar-se aos ruídos e prejudicar sua capacidade de generalização. Nesse último caso, é dito que a rede “decorou” os dados¹.

Do ponto de vista econométrico, uma grande vantagem das RNA sobre os modelos tradicionais está no fato das redes neurais serem capazes de aproximar qualquer função sem se preocupar com a especificação do modelo e, segundo Cao e Tay (2003), serem capazes de aprender relações complexas (não lineares) mesmo com poucos dados. Por conta da primeira característica, é dita uma técnica não paramétrica. Assim, dada uma amostra de observações geradas pelo modelo $y_{it} = g(x_{it}) + \varepsilon_{it}$, uma RNA permite construir uma aproximação da função desconhecida $g(\cdot)$, com a forma:

$$f(x_{it}, \theta) = w_{00} + \sum_{j=1}^h w_{0j} \Psi(\tilde{x}_{it}, w_{1j}) \quad (8)$$

onde:

$$w \equiv (w_{00}, w_{01}, \dots, w_{0h}, w_{11}, \dots, w_{ih})$$

$$\square x_{it} = (1, x_{it}), \text{ matriz com as variáveis explicativas.}$$

$\Psi(\cdot)$ = função de ativação.

De acordo com Fausett (1994), uma RNA pode ser caracterizada por três elementos essenciais: função de ativação, arquitetura da rede e aprendizado (treinamento). A função de ativação (ou de transferência) produz a saída de um neurônio, utilizando como argumento a

¹ Uma técnica que pode ser usada para minimizar essa chance, é dividir os dados em dois conjuntos (um para treinamento e o outro para validação) e parar o treinamento quando os erros do conjunto de validação começarem a aumentar.

soma ponderada de suas entradas. A arquitetura da rede² refere-se a maneira com a qual os neurônios estão estruturados (número de neurônios de entrada, número de neurônios em cada camada e o número de camadas ocultas). E o aprendizado refere-se ao conjunto de regras (algoritmo) que será utilizado para a que rede “aprenda”.

Existem muitos tipos diferentes de redes neurais. Do ponto de vista estrutural, as RNA podem ser classificadas como redes neurais estáticas ou como redes neurais dinâmicas (HAYKIN, 2001). As redes estáticas são caracterizadas por não apresentar elementos de realimentação e não apresentar “*delays*”, dessa forma, a saída da rede é calculada diretamente das entradas apresentadas, através de conexões alimentadas adiante (*feed foward*). Já nas redes dinâmicas, a saída pode não depender apenas da entrada corrente da rede, mas sim dos valores presentes e passados das entradas, saídas e do próprio estado da rede. As redes dinâmicas podem ser divididas em duas categorias: aquelas que possuem apenas conexões feedforward, e aquelas que possuem conexões com *feedback*.

3.1 Redes Multicamadas

As redes com múltiplas camadas podem ser vistas com uma generalização da rede com apenas uma camada. Essas redes contêm três níveis - camada de entrada, camada(s) oculta(s) e camada de saída - onde cada neurônio de uma determinada camada conecta-se com os neurônios da camada seguinte, até que a informação recebida (das camadas anteriores) seja enviada ao ambiente externo pelo(s) neurônio(s) da camada de saída. Não necessariamente todas as camadas devem estar diretamente conectadas e nem existir conexões entre todos os seus neurônios; as configurações criadas podem ser muito complexas e o seu uso depende geralmente do problema com o qual se está interessado em resolver.

O poder das RNA em tratar com problemas complexos está na inclusão de camadas de neurônios intermediárias (camada oculta) entre as entradas da rede e a camada de saída. Teoricamente, uma camada oculta é suficiente para aproximar qualquer função não linear, desde que esta possua um número suficiente de neurônios³ (HORNIK et al, 1989).

Formalmente, o produto de um neurônio j na camada de saída de uma rede alimentada adiante com uma camada oculta é:

$$y_j^o = \rho \left(\sum_{i=1}^m w_{ji}^o y_i^h + b_j^o \right) \quad (9)$$

² Apesar desta ser fundamental para o sucesso da aplicação, a sua escolha é realizada através de um processo empírico, o que não garante a certeza de obtenção de um bom resultado.

³ As redes com uma única camada oculta são muito utilizadas em problemas de previsão.

Onde ρ é a função de ativação (normalmente linear) e w_{ji}^o e b_j^o são os pesos a estimar (no qual o sobrescrito se refere à camada de saída). E a saída de um neurônio na camada oculta:

$$y_i^h = \varphi \left(\sum_{p=1}^k w_{ip}^h x_p + b_i^h \right) \quad (10)$$

Onde φ é a função de ativação dos neurônios intermediários (normalmente não linear) e w_{ip}^h e b_i^h são os pesos a serem estimados. E x_p ($p=1..k$) são os sinais de entrada. Dessa forma, tem-se que:

$$y_j^o = \rho \left(\sum_{i=1}^m w_{ji}^o \left(\varphi \left(\sum_{p=1}^k w_{ip}^h x_p + b_i^h \right) \right) + b_j^o \right) \quad (11)$$

3.2 Treinamento de Redes Neurais

O problema de otimização dos pesos pode ser visto como um problema de busca em um espaço: dada uma função de perda e a complexidade da rede (quantidade de conexões), a solução é encontrada através da busca de um ponto no qual a perda seja mínima. Para encontrar esse ponto, é necessário percorrer a função com algum algoritmo de busca; sendo comum que vários pontos de mínimos locais sejam encontrados.

O algoritmo mais conhecido para treinamento de redes de múltiplas camadas é o algoritmo de retropropagação. Esse algoritmo é uma generalização da regra de aprendizado Widrow-Hoff para redes com múltiplas camadas e para funções de transferência não lineares diferenciáveis. O algoritmo de retropropagação é um algoritmo estático e supervisionado, já que o vetor entrada e o correspondente vetor alvo são usados para treinar a rede até que esta consiga aproximar a função, associando cada vetor de entrada com um específico vetor de saída.

O termo retropropagação se refere a maneira pelo qual o gradiente da rede multicamadas é computado. Basicamente, a aprendizagem pela retropropagação consiste em duas fases.

Na primeira fase, um conjunto de sinais é submetido à camada de entrada e seu efeito é propagado adiante na rede, camada por camada, até que um resultado seja fornecido na camada de saída - sendo que nessa fase não há mudança no valor dos pesos sinápticos. Já na segunda fase, ocorre o aprendizado por meio de ajuste de pesos para a correção dos erros. O erro é calculado ao final da primeira fase pela subtração do resultado fornecido pela rede com

o resultado conhecido, e então, é propagado para trás no sentido inverso ao da propagação. Fases essas que acontecerão até que um critério de parada seja alcançado.

As redes dinâmicas podem ser treinadas usando os mesmos algoritmos (baseados no gradiente) utilizados para treinar as redes estáticas, no entanto, a performance dos algoritmos em redes dinâmicas podem ser um pouco diferentes, e o gradiente deve ser computado de uma maneira mais complexa. Espera-se que a retropropagação dinâmica leve um tempo maior para treinar, e além disso, a superfície de erro para essas redes tende a ser mais complexa do que para as redes estáticas. Existe uma chance maior do treinamento ficar preso em um mínimo local, sugerindo, que a rede deve ser treinada diversas vezes até ser encontrado um resultado ótimo.

Assim, considerando um modelo de aproximação $\hat{g}(\cdot, w)$, a distância euclidiana entre a função e sua aproximação $dist(g(\cdot), \hat{g}(\cdot, w))$ – que é função apenas do vetor de parâmetros w – produz a seguinte expressão (erro quadrado médio):

$$E(w) = \frac{1}{N} \sum_{i=1}^N (g(x) - \hat{g}(x, w))^2 \quad (12)$$

Onde o funcional $E: \mathfrak{R}^p \rightarrow \mathfrak{R}$ é denominado superfície de erro do problema de aproximação. Como o termo funcional corresponde a toda função $f: X \subset \mathfrak{R}^p \rightarrow \mathfrak{R}$, o problema de minimizar $E(w)$ torna-se um problema de minimização funcional. Com isso, dada a superfície de erro, o problema de aproximação passa a ser um problema de otimização cuja solução é :

$$W^* = \underset{w \in \mathfrak{R}^p}{\operatorname{argmin}} E(w) \quad (13)$$

Se considerarmos que o problema pode ter múltiplas camadas de saída e consideramos como o funcional sendo o a soma dos erros quadráticos, obtemos a seguinte expressão para o funcional dos erros:

$$E(w) = \sum_{i=1}^N \sum_{j=1}^m (g_{ij}(x) - \hat{g}_{ij}(x, w))^2 = \sum_{k=1}^q \eta^2 \quad (14)$$

Onde N é o número de amostras, l é o número de unidades intermediárias, r o erro residual, m o número de saídas e q o produto Nxm .

Seja \mathbf{J} o jacobiano do funcional E dado pela equação 14. Esta matriz pode ser escrita como:

$$\mathbf{J} \equiv \begin{pmatrix} \nabla_{r_1}^T \\ \vdots \\ \nabla_{r_q}^T \end{pmatrix} \quad (15)$$

Onde r é o erro residual.

Diferenciando a equação 3.14 obtemos:

$$\nabla J = 2J^T r = 2 \sum_{k=1}^q r_k \nabla r_k \quad (16)$$

$$\nabla^2 J = 2J^T J + \sum_{k=1}^q r_k \nabla^2 r_k \quad (17)$$

Quando os erros residuais são suficiente pequenos, a matriz Hessiana (eq. 17) pode ser aproximada da seguinte forma:

$$\nabla^2 J \approx 2J^T J \quad (18)$$

Essa aproximação normalmente é válida para um mínimo de J para a maioria dos propósitos, e é base para o método de quase-Newton (HAGAN, 1994). A lei de atualização fica na forma:

$$\Delta w = [J^T J]^{-1} J^T r \quad (19)$$

Neste trabalho usou-se o algoritmo de treinamento de Levenberg-Marquardt, o qual substitui a regra de mudança dos pesos do algoritmo de retropropagação (regra delta) e consiste em uma modificação do método de Newton que utiliza-se de uma aproximação da matriz Hessiana⁴ (segundas derivadas). A utilização desse algoritmo pode reduzir significativamente o tempo de treinamento e por isso, é um substituto natural ao algoritmo de retropropagação.

Para o ajuste de curvas, a fórmula básica desse algoritmo é:

⁴ A matriz Hessiana é atualizada a cada iteração do algoritmo, sendo que essa atualização é baseada no gradiente da função.

$$\Delta w = -(J^T J + \lambda I)^{-1} J^T e \quad (20)$$

Onde J é a matriz Jacobiana do erro da rede em relação aos pesos, λ é uma constante que controla a velocidade do ajuste, I é matriz identidade, Δw é um vetor de mudança nos pesos e e é um vetor com os erros.⁵

3.3 Aplicação de RNA à previsão da estrutura a termo da taxa de juros.

Nessa seção, destaca-se a metodologia adotada no trabalho para se realizar as estimativas e avaliar os modelos estimados. É importante destacar que estudos recentes apontam vantagens em se estimar e prever as taxas de juros considerando-se a região em que a maturidade prevista encontra-se (ALMEIDA ET AL, 2012). Baseando-se nessas evidências, a curva de juros foi dividida em três regiões, uma de Curto Prazo (compreendendo as maturidades 1, 3, 4, 6 e 9 meses), uma de Médio Prazo (compreendendo as maturidades de 12, 15, 18, 21 e 24 meses) e outra de Longo Prazo (compreendendo as maturidades de 27, 30, 36, 42 e 48 meses), e então, os modelos foram estimados em cada uma dessas regiões.

3.3.1 Classe de modelos e variáveis de entrada.

Os modelos estimados podem ser divididos em duas grandes classes: modelos estáticos e modelos dinâmicos.

Para primeira classe, seguiu-se em parte o trabalho de Abid & Salah (2002), no qual as informações relativas às mudanças futuras nas taxas de juros estão contidas tanto no nível das taxas quanto nos spreads entre as taxas (ver figura 1). Ou seja, para cada região da curva de juros, o vetor de entrada contém os spreads e os níveis, e o vetor de saída, a mudança futura⁶:

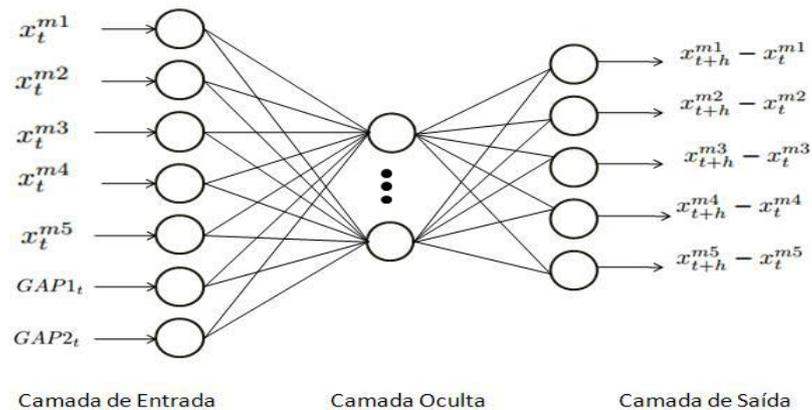
$$x_{t+h}^m - x_t^m = f(GAP_t, x_t) \quad (21)$$

Onde h representa o horizonte de previsão e m a maturidade.

⁵ $J^T J$ é considerado como a aproximação da matriz Hessiana dos erros.

⁶ Neste trabalho, os GAPS dependem de qual região da curva se está interessado. Para a região de curto prazo temos $GAP1_t = x_t^{m=3} - x_t^{m=1}$ e $GAP2_t = x_t^{m=9} - x_t^{m=6}$; para a região de médio prazo $GAP1_t = x_t^{m=15} - x_t^{m=12}$ e $GAP2_t = x_t^{m=24} - x_t^{m=21}$ e para a região de longo prazo $GAP1_t = x_t^{m=30} - x_t^{m=27}$ e $GAP2_t = x_t^{m=48} - x_t^{m=42}$.

Figura 1: Arquitetura Rede Neural Estática.



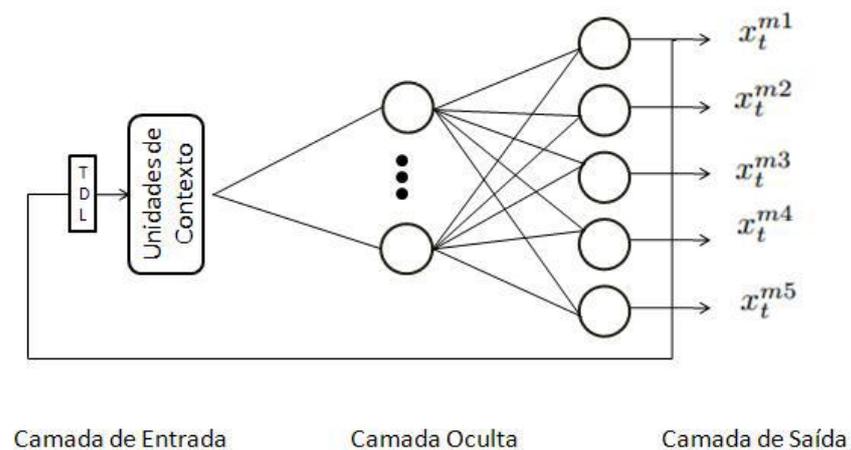
Fonte: elaboração própria (2013).

Para a segunda classe de modelo, considerou-se que os valores passados das taxas contêm as informações necessárias para se conhecer o seu valor futuro (ver figura 2) . Dessa forma, para cada parte da curva de juros, o vetor de entrada contém os níveis no passado, e o vetor de saída, os seus níveis correntes:

$$x_t^m = f(x_{t-1}^m, \dots, x_{t-i}^m) \quad (22)$$

No qual i representa a ordem da defasagem e m a maturidade.

Figura 2: Arquitetura Rede Neural Dinâmica.



Fonte: elaboração própria (2013).

3.3.2 Número de camadas ocultas e neurônios.

Como dito anteriormente, uma camada oculta é suficiente para aproximar qualquer função não linear, desde que esta possua um número suficiente de neurônios (HORNIK EL AL, 1989), por isso, nessa pesquisa escolheu-se trabalhar com redes com apenas uma camada oculta, procurando encontrar os melhores modelos, variando-se o número de neurônios nessa camada.

Apesar de existirem “regras de bolso” para a escolha do número de neurônios, não existe uma forma conclusiva para se realizar essa tarefa, o que torna a situação em uma questão puramente empírica. Foram estimados modelos com 1, 3, 5 e 7 neurônios na camada oculta pra os modelos estáticos, e com 4, 3, 2 e 1 neurônios para os modelos dinâmicos. Para este último, considerou-se também modelos com 1, 2 e 3 defasagens temporais.

Portanto, foram testadas 48 configurações para os modelos estáticos (4 horizontes de previsão, 3 regiões da curva, 4 configurações de neurônios na camada oculta) e 144 configurações para os modelos dinâmicos (4 horizontes de previsão, 3 regiões da curva, 4 configurações de neurônios na camada oculta e 3 defasagens temporais).

3.3.3 Escolha dos modelos

A avaliação da capacidade preditiva dos modelos é realizada dividindo-se a amostra em duas partes. Uma parte é usada para fazer a estimação dos modelos, com dados de janeiro de 2007 até meados de março de 2010, representando 80% da amostra total. A segunda parte, os últimos 200 dados da amostra, são reservados para avaliar a performance das previsões, com período de março de 2010 até janeiro de 2011 . São realizadas previsões para horizontes de um dia, um mês e três meses à frente para ambas as classes de modelo – esses horizontes forem escolhidos por serem bastante utilizados na literatura. Para o caso dos modelos estáticos, as previsões realizadas são de um passo à frente. Cada uma das configurações foi estimada 10 vezes, com reinícios aleatórios nos pesos iniciais, e os que apresentaram menores erros quadráticos médios fora da amostra foram selecionados para serem apresentados⁷.

3.3.4 Treinamento e Função de ativação

As redes foram treinadas com o algoritmo de Levenberg-Marquardt, mostrado anteriormente. Para evitar a possibilidade de overfitting, reservou-se 20% dos dados da amostra de treinamento para validação, parando o treinamento quando o erro quadrado médio

⁷ O reinício aleatório dos pesos é recomendável, já que o treinamento de uma RNA é bastante sensível aos pesos iniciais.

nessa amostra deixasse de cair ou começasse a aumentar. Além do mais, o algoritmo foi configurado para parar as interações quando não fosse mais possível aumentar o valor da função em pelo menos $1.0e^{-10}$.

Foi utilizada a função sigmoide para os neurônios da camada intermediária e a função linear (ou identidade) para o neurônio de saída.

4 MODELOS COMPETIDORES

Nesta seção, serão apresentados os modelos comumente usados para estimar e prever a estrutura a termo da taxa de juros. Estarão aqui presente o modelo de Diebold e Li, o modelo de Svensson, o autorregressivo, o vetor-autorregressivo e o random walk.

4.1 O modelo de Diebold e Li

Diebold & Li (2006) reinterpretam o modelo exponencial de Nelson & Siegel (1987), considerando uma forma paramétrica para a evolução da curva de juros ao longo do tempo; além do mais, os autores consideram que a “classificação” nível, inclinação e curvatura é apropriada para esses fatores. A curva de juros spot correspondente é:

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3,t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) + \varepsilon_t^\tau \quad (23)$$

A forma da curva de juros é determinada pelos três fatores e pelos factor loadings associados a eles. O parâmetro λ governa a taxa de decaimento exponencial, pequenos valores de λ estão associados a um decaimento suave, e ajusta melhor as maturidades longas. Os autores propõem fixar λ em um valor a priori, e estimar a equação (23) por meio de mínimos quadrados ordinários para cada t, de onde serão obtidas as séries de tempo dos fatores.

Ao colocarem o modelo de Nelson-Siegel em um contexto dinâmico, Diebold & Li (2006) mostram também que o modelo é capaz de replicar os principais fatos empíricos da estrutura a termo da taxa de juros ao longo do tempo: a curva média é positivamente inclinada e côncava, a dinâmica das taxas são altamente persistentes, com as taxas de maturidade mais longas sendo mais persistentes do que as maturidades de curto prazo, e que a volatilidade das taxas de juros é decrescente para maturidades longas.

Diebold & Li (2006) modelam os fatores de duas maneiras, primeiramente assumem um processo univariado AR(1) e também por um processo VAR(1). A previsão realizada tem a forma:

$$\hat{y}_{t+h/\tau}(\tau) = \hat{\beta}_{1,t+h/\tau} + \hat{\beta}_{2,t+h/\tau} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \hat{\beta}_{3,t+h/\tau} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \quad (24)$$

Onde, para a primeira especificação

$$\hat{\beta}_{i,t+h/\tau} = \hat{c}_i + \hat{y}_i \hat{\beta}_{i,t}, i = 1,2,3 \quad (25)$$

E para a segunda especificação

$$\hat{\beta}_{t+h/\tau} = \hat{c} + \hat{\Gamma} \hat{\beta}_t \quad (26)$$

Caso os fatores não sejam altamente correlacionados, é esperado pouco, ou nenhum, aumento na capacidade de previsão ao se passar para o modelo multivariado.

4.2 O modelo de Svensson

Embora o modelo de Diebold & Li(2006) seja capaz de capturar grande parte dos formatos da curva de juros, ele não é capaz de gerar todos os formatos que a estrutura a termo assume durante o tempo. Como forma de “corrigir” tal problema, muitas outras especificações mais flexíveis tem sido sugeridas. As extensões do modelo de Nelson-Siegel aumentam a flexibilidade do modelo através da adição de mais fatores, mais parâmetros de decaimento, ou uma combinação de ambos.

Svensson (1994) propõe o aumento da flexibilidade do modelo NS através da inclusão de um quarto componente exponencial, uma segunda curvatura, que no entanto, apresenta um parâmetro λ_i diferente. O modelo que se ajusta às curvas forward é:

$$y_i(\tau) = \beta_{1,t} + \beta_{2,t} \left(\frac{1 - e^{-\lambda_{1,t}\tau}}{\lambda_{1,t}\tau} \right) + \beta_{3,t} \left(\frac{1 - e^{-\lambda_{1,t}\tau}}{\lambda_{1,t}\tau} - e^{-\lambda_{1,t}\tau} \right) + \beta_{4,t} \left(\frac{1 - e^{-\lambda_{2,t}\tau}}{\lambda_{2,t}\tau} - e^{-\lambda_{2,t}\tau} \right) + \varepsilon_t^\tau \quad (27)$$

O modelo SV poderia assim, teoricamente, melhor se ajustar aos diversos formatos das curvas de juros. A previsão realizada tem a forma:

$$\hat{y}_{t+h/\tau}(\tau) = \hat{\beta}_{1,t+h/\tau} + \hat{\beta}_{2,t+h/\tau} \left(\frac{1 - e^{-\lambda_{1,t}\tau}}{\lambda_{1,t}\tau} \right) + \hat{\beta}_{3,t+h/\tau} \left(\frac{1 - e^{-\lambda_{1,t}\tau}}{\lambda_{1,t}\tau} - e^{-\lambda_{1,t}\tau} \right) + \hat{\beta}_{4,t+h/\tau} \left(\frac{1 - e^{-\lambda_{2,t}\tau}}{\lambda_{2,t}\tau} - e^{-\lambda_{2,t}\tau} \right) \quad (28)$$

Onde, para uma especificação de uma dinâmica univariada AR(1) temos:

$$\hat{\beta}_{i,t+h/\tau} = \hat{c}_i + \hat{y}_i \hat{\beta}_{i,t}, i = 1,2,3 \quad (29)$$

E segundo uma especificação VAR(1)

$$\hat{\beta}_{t+h/\tau} = \hat{c} + \hat{\Gamma} \hat{\beta}_t \quad (30)$$

4.3 O modelo AR, VAR e Random Walk

Esses modelos mais simples, caracterizam-se por aplicarem-se diretamente nos vértices da curva de juros. São eles:

Na literatura de previsão da ETTJ, o modelo de random walk é utilizado como *benchmark* natural aos modelos propostos. A hipótese de que as taxas de juros seguem esse processo estocástico implica que a previsão para uma determinada taxa de juros é o seu próprio valor defasado. Desta forma, a previsão da ETTJ como um todo, é o valor defasado das taxas em cada um dos vértices da curva. A formulação deste modelo é:

$$\hat{y}_{t+h}(\tau) = y_t(\tau) + \varepsilon_t^\tau \quad (31)$$

Desde o trabalho de Box & Jenkins (1970), modelos autoregressivos e de médias móveis são estimados para realizar previsões, os quais têm se mostrado bastante satisfatórios em várias aplicações nas mais diversas áreas. Em virtude da sua simplicidade e da qualidade de suas previsões, esses modelos são usualmente a base de comparação (*benchmark*) para outros modelos. No presente trabalho, utilizaremos um modelo autoregressivo de ordem 1, por ser bastante utilizado na literatura para previsão da curva de juros:

$$\hat{y}_{t+h}(\tau) = \hat{c} + \hat{\theta}y_t(\tau) + \varepsilon_t^\tau \quad (32)$$

Existe a possibilidade de outras variáveis conterem informações adicionais sobre a variável de interesse - no nosso caso, se a curva de juros como um todo contém informação adicional a respeito da taxa para uma maturidade específica. Por essa razão, é interessante investigar se modelos multivariados de séries temporais podem gerar previsões mais eficientes do que os modelos univariados. A sua formulação é:

$$\hat{y}_{t+h}(\tau) = \hat{C} + \hat{\Gamma}PC_t + \varepsilon_t \quad (33)$$

onde $PC_t = [pc_{t,1}, \dots, pc_{t,n}]$ são os primeiros n primeiros componentes principais da matriz de covariância das taxas de juros.

5 DADOS, MÉTODO DE COMPARAÇÃO E ANÁLISE DE RESULTADOS

O contrato futuro de depósito interbancário (DI Futuro) com maturidade é um contrato futuro do qual o ativo básico é taxa de juros acumulada diariamente (DI), capitalizada entre o momento da negociação t , e τ . O valor do contrato é dado pelo seu valor no vencimento, R\$ 100.000,00 descontado pela taxa de juros acumulada, negociada entre o vendedor e o comprador do contrato.

Ao comprar um contrato de DI Futuro ao preço DI no tempo t e mantê-lo até a maturidade τ , o ganho ou perda é dado por:

$$100 \left(\frac{\prod_{i=1}^{\zeta(t,\tau)} (1 + y_i)^{1/252}}{(1 + DI)^{\zeta(t,\tau)/252}} - 1 \right) \quad (34)$$

Onde y_i denota a taxa DI, $(i-1)$ dias após o dia da negociação. A função $\zeta(t, \tau)$ representa o número de dias entre t e τ .

O contrato de DI é muito similar ao zero-cupom Bond, exceto pelo fato de pagar os ajustes de margem diariamente. A cada dia o fluxo de caixa é a diferença entre o preço de ajuste (liquidação) do dia corrente e o preço de ajuste do dia anterior, corrigido pela taxa DI do dia anterior.

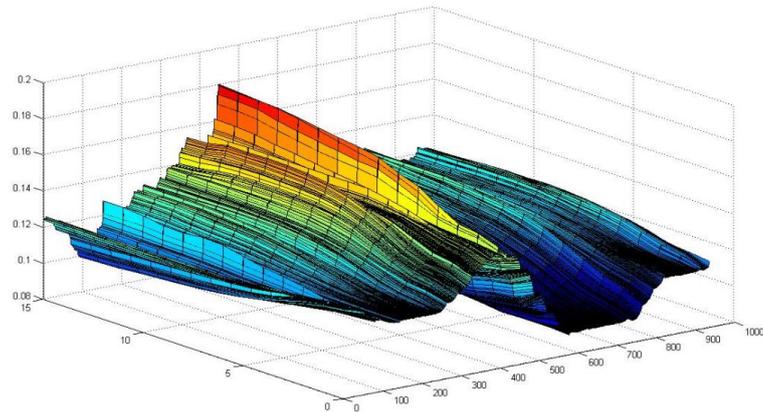
Os contratos de DI Futuro são negociados na BM&F, que determina o número de vencimentos com contratos autorizados. Geralmente, para cada dia, existem em torno de 20 vencimentos com contratos autorizados, mas nem todos apresentam liquidez. Normalmente por volta de 10 vencimentos apresentam contratos com maior liquidez. Existem contratos com vencimentos mensais para os meses que iniciam cada trimestre, janeiro, abril, julho e outubro. Além disso, existem contratos com vencimentos para os quatro meses subsequentes ao mês corrente. A data de vencimento é o primeiro dia útil do mês de vencimento do contrato.

5.1 Dados

Os dados empregados nesse artigo consistem de observações diárias das taxas dos contratos de DI Futuro, preços de fechamento. Como na prática não são observados diariamente contratos com todas as maturidades utilizadas, a partir das taxas observadas diariamente para as maturidades disponíveis, os dados foram convertidos para maturidades fixas de 1, 3, 4, 6, 9, 12, 15, 18, 21, 24, 27, 30, 36, 42 e 48 meses, por meio de interpolações

pelo método exponencial splines. Os dados foram observados para o período de Janeiro de 2007 a Janeiro de 2011 (986 dados diários), e representam os contratos de DI mais líquidos negociados para o período analisado. Na figura 2, mostra-se a evolução das curvas de juros no tempo.

Figura 3: Estrutura a termo da taxa de juros



Fonte: elaboração própria.

São utilizados apenas os dados de ajuste dos contratos de DI Futuro, não incluindo taxas SWAP. Conforme critério de apuração da BM&F, o dado de fechamento das taxas de SWAP PRÉ DI são obtidos a partir dos dados do ajuste dos contratos de DI Futuro negociado na BM&F, não correspondendo a dados de negócios efetivamente executados sob a modalidade SWAP. Logo, como os dados de SWAP são obtidos através do próprio dado do DI Futuro ou através de uma interpolação deste, consideramos que ao utilizar apenas o próprio dado de DI Futuro o modelo estará livre de distorções decorrentes da utilização das taxas divulgadas de SWAP como se fossem informações de negócios efetivamente executados. A tabela 1 apresenta algumas estatísticas descritivas da curva de juros. Percebe-se que a média incondicional das taxas aumenta com o aumento da maturidade e o desvio-padrão parece exibir um evolução irregular através das diferentes maturidades. Os coeficientes de autocorrelação decaem muito devagar e são consistentemente positivos, mostrando evidências de que as taxas de juros são não estacionárias.⁸

⁸ O teste Dickey Fuller Aumentado confirma que as taxas de juros são estacionárias na primeira diferença.

Tabela 1: Estatísticas descritivas

maturidade	Média	Dev.Pad	Mínimo	Máximo	ACF1	ACF5	ACF21
1	0,1100	0,0157	0,0851	0,1413	0,9988	0,9927	0,9540
3	0,1104	0,0156	0,0859	0,1452	0,9990	0,9938	0,9562
4	0,1109	0,0156	0,0858	0,1469	0,9989	0,9937	0,9554
6	0,1119	0,0157	0,0861	0,1532	0,9988	0,9930	0,9506
9	0,1138	0,0157	0,0873	0,1604	0,9983	0,9908	0,9394
12	0,1158	0,0155	0,0899	0,1640	0,9978	0,9888	0,9292
15	0,1178	0,0153	0,0935	0,1691	0,9973	0,9863	0,9179
18	0,1193	0,0148	0,0955	0,1712	0,9967	0,9838	0,9050
21	0,1206	0,0144	0,0979	0,1726	0,9961	0,9813	0,8911
24	0,1216	0,0140	0,1006	0,1744	0,9956	0,9789	0,8780
27	0,1223	0,0138	0,1026	0,1762	0,9952	0,9766	0,8649
30	0,1228	0,0136	0,1018	0,1778	0,9947	0,9745	0,8530
36	0,1236	0,0131	0,1009	0,1783	0,9939	0,9707	0,8343
42	0,1242	0,0126	0,1003	0,1793	0,9931	0,9675	0,8210
48	0,1246	0,0126	0,0997	0,1800	0,9926	0,9656	0,8102

Fonte: elaboração própria.

5.2 Método de Comparação

Como forma de apresentarmos a diferença entre as previsões das taxas geradas pelos modelos estimados e os seus valores verdadeiros, utilizamos a raiz do erro quadrado médio (RMSE). Utilizamos essa medida porque é uma forma simples e direta de verificar o quão bem os modelos se adequam à realidade. A equação 25 apresenta a sua forma de cálculo.

$$RMSE = \sqrt{\frac{\sum_{j=1}^m e_{t+j}^2}{m}} \quad (35)$$

onde e é o erro de previsão calculado para a observação $t+j$ e m é o número de observações.

Para confirmar se as diferenças entre as previsões realizadas através dos diversos modelos são estatisticamente significantes, foi aplicado o teste de Diebold Mariano (ver Diebold e Mariano, 1995). Comparamos aos pares as previsões geradas pelos modelos de redes neurais, autorregressivo, vetor-autorregressivo, *random walk* e pelos modelos de Diebold-Li e Svensson. Seja $\{d_i\}_{i=1}^n$ uma função das diferenças dos quadrados dos erros de previsão produzidas por dois modelos, temos:

$$d_i = (\hat{y}_{t+h,i}(\tau) - y_{t+h}(\tau))^2 - (\hat{y}_{t+h,j}(\tau) - y_{t+h}(\tau))^2 \quad (36)$$

onde i e j representam os dois modelos que serão comparados e as variáveis \hat{y}_{t+h} são as previsões h passos à frente no tempo t destes modelos. A forma como Diebold Mariano propõem verificar a diferença entre as previsões é dado por:

$$DM = \frac{\frac{1}{n} \sum_{i=1}^n d_i}{\sqrt{\frac{\hat{\delta}}{n}}} \xrightarrow{d} N(0,1) \quad (37)$$

onde $\hat{\delta}$ é a estimativa da matriz de covariância de longo prazo de d_i - a correlação serial dos erros de previsão foi controlada (ver Newey & West, 1987).

5.3 Análise de Resultados

Na última seção foi explicada a forma como os modelos de RNA foram estimados e como se daria a escolha dos modelos que seriam usados para comparação com os modelos usualmente utilizados para prever a curva de juros.

Observando-se a tabela 2, constata-se que a classe dos modelos de redes neurais que melhor preveram a curva de juros está relacionado com o horizonte de previsão escolhido: os modelos pertencentes à classe estática se saíram melhor para os horizontes de previsão de curto prazo (1 dia e 1 semana), e os modelos da classe dinâmica, apresentaram melhores resultados para os horizontes de médio prazo (1 e 3 meses). Além disso, parece existir relação entre o número de neurônios na camada oculta, a classe que pertence o modelo, e a região da curva de juros: modelos (dinâmicos) com 3 e 4 neurônios se saíram melhor na região de curto e médio prazo, e modelos (estáticos) com 3 neurônios adaptaram-se melhor na região de longo prazo.

Tabela 2: Modelos com menor RMSE fora da amostra por região e horizonte de previsão

Modelos horizonte previsão	Modelos dinâmicos			Modelos estáticos		
	CP	MP	LP	CP	MP	LP
1 dia	AR2NN4	AR1NN4	AR1NN3	NN5*	NN3*	NN3*
5 dias	AR2NN4	AR2NN4	AR2NN3	NN3*	NN5*	NN3*
21 dias	AR1NN4	AR3NN3*	AR1NN4*	NN5*	NN7	NN5
63 dias	AR2NN4*	AR3NN3*	AR1NN3*	NN7	NN7	NN3

Nota: Essa tabela apresenta os modelos de redes neurais artificiais que melhor previram a curva de juros para o período que se estende de março/2010 a janeiro/2011. Na primeira parte da tabela encontram-se os modelos dinâmicos e na segunda, os estáticos. Para ambos os casos, o número após NN significa a quantidade de neurônios na camada oculta e para os modelos dinâmicos, o número após AR significa a quantidade de defasagens dinâmicas; por exemplo, AR2NN4, significa um modelo autoregressivo 2 com 4 neurônios na camada oculta - ver seção 3.3 para mais detalhes. O destaque (*) significa que o modelo apresentou menor RMSE fora da amostra. Legenda: CP = região de curto prazo da curva de juros (maturidades 1, 3, 4, 6 e 9

meses); MP= região de médio prazo da curva de juros (maturidades 12, 15, 18, 21, e 24 meses); LP= região de longo prazo da curva (maturidades 27, 30, 36, 42 e 48 meses);
 Fonte: elaboração própria.

A tabela 3 apresenta a raiz do erro quadrado médio (RMSE) das previsões de cada um dos modelos testados, para os quatro horizontes de previsão. Percebe-se pela sua análise, que em média, apenas os modelos estimados através de RNA conseguiram superar as previsões realizadas pelo *random walk* em todos os horizontes de previsão, sendo essas previsões em torno de 2% e 5% melhores para os horizontes de 1 dia e 1 semana à frente, e de 12% e 7% melhores para os horizontes mais longos, de 1 mês e 3 meses. Além disso, apresentou previsões em torno de 15% e 10% melhores que o modelo de DNS para os horizontes de 1 mês e 3 meses à frente. Um ponto a ser destacado, é que para o caso dos horizontes de previsão mais curtos, os modelos auto-regressivos (AR1 e VAR1) apresentaram desempenho médio muito próximo aos gerados pelas previsões do random walk, situação esta, não mantida para os horizontes mais longos. Já para o modelo de DNS, temos a situação oposta, pois o modelo apresentou melhores resultados para os horizontes de previsão mais longos do que para os horizontes mais curtos.

Tabela 3: RMSE das previsões fora da amostra

horizonte previsão	RW	RNA	AR1	VAR1	NS	SV
1 dia	0,0616	0,0605	0,0624	0,0655	0,1325	0,1092
5 dias	0,1593	0,1504	0,1630	0,1614	0,1988	0,1862
21 dias	0,3813	0,3369	0,4029	0,4017	0,3939	0,3889
63 dias	0,6381	0,5969	0,7843	0,8069	0,6522	0,6499

Nota: Essa tabela apresenta a raiz do erro quadrado médio das previsões fora da amostra (em base points) de todos os modelos estimados para os quatro horizontes de previsão. O período das previsões se estende de março/2010 a janeiro/2011. Legenda: RNA = modelo de redes neurais artificiais; AR1= modelo autoregressivo(1); VAR1= vetor-autoregressivo(1); NS= modelo de Nelson e Siegel estimado pelo método de dois passos; SV= modelo de Svensson ; RW= random walk.

Fonte: elaboração própria.

Uma descrição mais detalhada dos desempenhos dos modelos por horizonte de previsão e maturidade pode ser vista nas Tabelas 4 e 5, em que na primeira apresenta-se o desempenho geral (RMSE por maturidade) e na segunda o desempenho relativo ao *random walk* de cada modelo. Constata-se que os modelos apresentaram desempenhos diferentes quando se considera a região da curva e o horizonte de previsão.

Tabela 4: RMSE das previsões fora da amostra por maturidade

1 dia frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	0,035	0,035	0,041	0,046	0,051	0,056	0,058	0,062	0,064	0,067	0,070	0,075	0,080	0,081	0,086
RW	0,039	0,037	0,043	0,048	0,052	0,057	0,059	0,062	0,065	0,068	0,071	0,076	0,081	0,082	0,086
NS 2P	0,152	0,050	0,074	0,118	0,121	0,100	0,110	0,135	0,152	0,157	0,154	0,145	0,147	0,172	0,200
SV	0,046	0,062	0,069	0,075	0,084	0,097	0,108	0,118	0,123	0,129	0,137	0,140	0,145	0,151	0,157
AR1	0,040	0,038	0,044	0,049	0,054	0,058	0,060	0,063	0,065	0,069	0,071	0,076	0,081	0,082	0,087
VAR1	0,041	0,042	0,046	0,060	0,060	0,059	0,063	0,067	0,070	0,070	0,073	0,078	0,084	0,083	0,087
5 dias frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	0,071	0,085	0,112	0,139	0,151	0,150	0,152	0,155	0,159	0,167	0,174	0,179	0,183	0,188	0,191
RW	0,109	0,111	0,126	0,147	0,158	0,159	0,160	0,162	0,164	0,171	0,175	0,180	0,185	0,190	0,193
NS 2P	0,196	0,105	0,122	0,161	0,168	0,170	0,188	0,213	0,227	0,230	0,229	0,225	0,228	0,248	0,271
SV	0,106	0,123	0,145	0,166	0,165	0,171	0,180	0,193	0,200	0,205	0,214	0,220	0,227	0,235	0,242
AR1	0,121	0,123	0,136	0,154	0,163	0,163	0,163	0,164	0,165	0,171	0,175	0,180	0,185	0,190	0,193
VAR1	0,115	0,115	0,130	0,158	0,161	0,158	0,156	0,163	0,166	0,169	0,176	0,183	0,187	0,190	0,194
21 dias frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	0,296	0,300	0,315	0,362	0,400	0,373	0,350	0,338	0,329	0,330	0,335	0,334	0,325	0,333	0,334
RW	0,360	0,347	0,367	0,404	0,424	0,399	0,390	0,385	0,376	0,379	0,382	0,384	0,378	0,375	0,369
NS 2P	0,435	0,334	0,342	0,369	0,378	0,387	0,404	0,425	0,427	0,420	0,413	0,403	0,386	0,390	0,396
SV	0,354	0,356	0,386	0,418	0,409	0,392	0,388	0,396	0,392	0,389	0,392	0,392	0,387	0,391	0,390
AR1	0,441	0,412	0,424	0,445	0,451	0,417	0,402	0,393	0,383	0,384	0,384	0,385	0,379	0,375	0,369
VAR1	0,434	0,403	0,416	0,446	0,444	0,418	0,399	0,393	0,384	0,382	0,388	0,389	0,380	0,376	0,373
63 dias frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	0,712	0,647	0,618	0,570	0,538	0,524	0,467	0,461	0,481	0,506	0,692	0,716	0,706	0,669	0,646
RW	0,699	0,685	0,696	0,710	0,713	0,711	0,694	0,676	0,637	0,598	0,586	0,582	0,543	0,528	0,513
NS 2P	0,784	0,662	0,640	0,617	0,620	0,661	0,705	0,735	0,727	0,687	0,654	0,633	0,575	0,550	0,534
SV	0,702	0,686	0,701	0,714	0,700	0,692	0,689	0,689	0,671	0,633	0,609	0,603	0,571	0,556	0,533
AR1	0,951	0,945	0,949	0,955	0,933	0,899	0,857	0,816	0,754	0,691	0,666	0,652	0,590	0,564	0,542
VAR1	0,906	0,921	0,937	0,966	0,975	0,958	0,924	0,876	0,812	0,742	0,702	0,677	0,609	0,562	0,536

Nota: Essa tabela apresenta a raiz do erro quadrado médio das previsões fora da amostra (em base points) de todos os modelos estimados - por maturidade - para os quatro horizontes de previsão. O período das previsões se estende de março/2010 a janeiro/2011. Legenda: RNA = modelo de redes neurais artificiais; AR1= modelo autoregressivo (1); VAR1= vetor-autoregressivo(1); NS= modelo de nelson e siegel estimado pelo método de dois passos, SV= modelo de Svensson; RW= random walk.

Fonte: elaboração própria.

Tabela 5: RMSE relativo das previsões fora da amostra por maturidade

1 dia frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	0,902	0,949	0,947	0,958	0,972	0,987	0,991	0,991	0,989	0,983	0,997	0,998	0,996	0,996	0,996
NS 2P	3,874	1,372	1,708	2,478	2,298	1,770	1,865	2,165	2,346	2,294	2,187	1,916	1,829	2,115	2,322
SV	1,171	1,685	1,585	1,570	1,592	1,710	1,839	1,890	1,899	1,893	1,934	1,851	1,802	1,849	1,816
AR1	1,021	1,031	1,023	1,018	1,025	1,023	1,017	1,010	1,010	1,007	1,007	1,003	1,002	1,005	1,003
VAR1	1,048	1,151	1,060	1,249	1,145	1,047	1,062	1,078	1,077	1,019	1,036	1,035	1,037	1,020	1,007
5 dias frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	0,650	0,761	0,888	0,943	0,957	0,943	0,949	0,957	0,970	0,979	0,998	0,996	0,991	0,992	0,990
NS 2P	1,808	0,946	0,970	1,095	1,065	1,068	1,173	1,316	1,383	1,346	1,313	1,250	1,231	1,311	1,402
SV	0,972	1,106	1,153	1,131	1,048	1,076	1,126	1,191	1,214	1,202	1,223	1,222	1,225	1,242	1,253
AR1	1,110	1,102	1,081	1,050	1,035	1,023	1,015	1,010	1,007	1,004	1,002	1,001	1,000	1,000	0,999
VAR1	1,062	1,030	1,034	1,078	1,023	0,993	0,976	1,002	1,009	0,988	1,007	1,018	1,010	1,001	1,006
21 dias frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	0,823	0,865	0,858	0,897	0,944	0,936	0,896	0,879	0,873	0,871	0,875	0,869	0,860	0,888	0,906
NS 2P	1,210	0,964	0,933	0,915	0,893	0,969	1,036	1,106	1,135	1,106	1,080	1,048	1,019	1,041	1,072
SV	0,985	1,028	1,053	1,036	0,964	0,984	0,995	1,029	1,042	1,025	1,024	1,021	1,023	1,044	1,056
AR1	1,225	1,189	1,156	1,102	1,065	1,046	1,030	1,022	1,017	1,011	1,004	1,003	1,001	1,000	0,999
VAR1	1,208	1,163	1,134	1,106	1,048	1,046	1,022	1,021	1,022	1,007	1,015	1,011	1,005	1,003	1,011
63 dias frente															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
RNA	1,019	0,944	0,888	0,802	0,755	0,737	0,673	0,682	0,755	0,847	1,181	1,229	1,302	1,266	1,259
NS 2P	1,122	0,967	0,920	0,869	0,869	0,929	1,016	1,087	1,141	1,149	1,116	1,087	1,059	1,041	1,039
SV	1,005	1,001	1,006	1,006	0,982	0,972	0,992	1,019	1,053	1,059	1,040	1,035	1,052	1,053	1,039
AR1	1,361	1,379	1,364	1,344	1,308	1,264	1,235	1,206	1,184	1,156	1,136	1,120	1,088	1,069	1,056
VAR1	1,297	1,345	1,346	1,360	1,367	1,347	1,331	1,295	1,274	1,241	1,198	1,163	1,123	1,064	1,044

Nota: Essa tabela apresenta a raiz do erro quadrado médio em relação ao random walk das previsões fora da amostra de todos os modelos estimados - por maturidade - para os quatro horizontes de previsão. O período das previsões se estende de março/2010 a janeiro/2011. Legenda: RNA = modelo de redes neurais artificiais; AR1= modelo autoregressivo (1); VAR1= vetor-autoregressivo(1); NS= modelo de Nelson e Siegel estimado pelo método de dois passos; SV= modelo de Svensson.

Fonte: elaboração própria.

No caso dos modelos de RNA, verifica-se que este apresentou, de uma maneira geral, um desempenho superior na região de curto prazo - considerando-se o horizonte de 1 dia, por exemplo, seu desempenho foi aproximadamente 5% melhor do que o *random walk* nessa região, mas foi praticamente nulo quando considerado a região de longo prazo. O interessante notar aqui, é que estes resultados independem do trecho da curva de juros considerado, indicando que os modelos de RNA são capazes de obter resultados interessantes para as taxas

de juros de curto e médio prazo, mas não tão interessantes assim quando se considera horizontes de previsão mais longos. Tais fatos podem ser melhor verificados na Tabela 6.

Tabela 6: RMSE relativo das previsões fora da amostra por região da curva

Modelos horizonte previsão	RNA			AR1			VAR1			NS			SV		
	CP	MP	LP												
1 dia	0,946	0,988	0,997	1,024	1,013	1,004	1,131	1,057	1,027	2,346	2,088	2,074	1,521	1,846	1,850
5 dias	0,840	0,959	0,993	1,075	1,012	1,000	1,045	0,994	1,008	1,177	1,257	1,301	1,082	1,162	1,233
21 dias	0,877	0,891	0,880	1,147	1,025	1,001	1,132	1,024	1,009	0,983	1,070	1,052	1,013	1,015	1,034
63 dias	0,881	0,739	1,247	1,351	1,209	1,094	1,343	1,298	1,118	0,949	1,065	1,118	1,000	1,019	1,044

Nota: Essa tabela apresenta o erro quadrado médio em relação ao random walk das previsões fora da amostra, por região da curva de juros de todos os modelos estimados e para todos os horizontes de tempo. O período das previsões se estende de março/2010 a janeiro/2011. Legenda: RNA = modelo de redes neurais artificiais; AR1= modelo autoregressivo (1); VAR1= vetor-autoregressivo(1); NS= modelo de Nelson e Siegel estimado pelo método de dois passos; SV= modelo de Svensson; CP = região de curto prazo da curva de juros(maturidades 1, 3, 4, 6 e 9 meses); MP= região de médio prazo da curva de juros (maturidades 12, 15, 18, 21, e 24 meses); LP= região de longo prazo da curva (maturidades 27, 30, 36, 42 e 48 meses).

Fonte: elaboração própria.

Tabela 7: Teste Diebold-Mariano ModelosxRW

AR1xRW															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	0,818	2,147	3,781	3,659	2,846	2,409	2,108	1,594	1,544	1,276	1,655	1,006	0,729	1,553	1,194
5 dias frente	5,495	3,649	3,083	2,303	1,982	1,683	1,478	1,226	1,057	0,926	0,637	0,306	-0,011	-0,089	-0,572
21 dias frente	2,480	2,122	1,925	1,637	1,388	1,154	0,964	0,821	0,732	0,544	0,403	0,295	0,123	0,019	-0,117
63 dias frente	1,788	2,119	2,283	2,709	3,200	2,642	2,069	1,722	1,544	1,452	1,344	1,238	1,168	1,147	1,204
VAR1xRW															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	1,380	1,560	1,086	5,200	3,041	1,086	1,582	2,948	2,819	0,881	1,937	1,771	1,535	1,383	0,494
5 dias frente	2,553	0,721	1,089	2,425	0,551	-0,289	-0,908	0,117	0,354	-0,778	0,633	0,970	0,371	0,070	0,645
21 dias frente	2,463	2,110	2,096	1,982	1,097	1,156	0,737	1,174	0,946	0,139	0,884	1,312	0,397	0,645	0,930
63 dias frente	1,720	2,189	2,404	2,693	3,170	3,223	2,870	2,501	2,362	2,040	1,935	1,852	1,278	0,815	0,619
NSxRW															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	16,583	3,007	5,637	7,116	5,164	2,928	3,441	5,231	5,958	5,897	5,571	4,341	3,639	4,443	5,063
5 dias frente	6,752	-0,866	-0,326	0,788	0,609	1,005	2,245	2,883	2,788	2,594	2,531	2,171	1,897	2,191	2,475
21 dias frente	2,949	-3,207	-1,727	-1,250	-1,747	-2,071	1,021	1,597	1,377	1,046	0,954	0,797	0,710	1,119	1,410
63 dias frente	1,564	-1,371	-1,605	-2,015	-2,801	-1,840	0,668	2,594	2,241	1,700	1,806	2,382	3,294	1,284	0,464
SVxRW															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	3,985	7,690	8,755	5,241	3,032	2,833	2,853	3,090	3,328	3,401	3,681	3,625	3,652	3,989	4,086
5 dias frente	-0,963	2,743	3,934	2,792	1,141	1,263	1,515	2,003	1,980	1,789	1,889	1,892	1,925	2,095	2,203
21 dias frente	-1,584	1,286	1,708	1,478	-2,066	-2,114	-0,463	1,801	1,027	0,450	0,582	0,605	0,740	1,405	1,609
63 dias frente	0,002	-0,216	0,642	0,526	-1,987	-1,000	-0,396	1,615	1,619	1,102	1,293	2,744	3,608	2,884	1,429

Nota: Essa tabela apresenta a estatística de Diebold-Mariano para os erros de previsão fora da amostra dos modelos estimados e do random walk para os quatro horizontes de previsão - comparando os modelos aos pares.

O período das previsões se estende de março/2010 a janeiro/2011. Valores negativos indicam superioridade das previsões do primeiro método do par. Legenda: RNA = modelo de redes neurais artificiais; AR1= modelo autoregressivo (1); VAR1= vetor-autoregressivo(1); NS= modelo de Nelson e Siegel estimado pelo método de dois passos; SV= modelo de Svensson ; RW= random walk.

Fonte: elaboração própria.

Tabela 8: Teste Diebold-Mariano RNAXModelos

RNAxRW															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	-3,705	-3,342	-2,037	-1,886	-2,224	-0,945	-0,897	-0,698	-0,892	-1,979	-1,027	-1,024	-1,158	-1,378	-0,946
5 dias frente	-1,567	-1,829	-1,905	-1,538	-0,592	-2,128	-1,534	-1,090	-0,706	-0,476	-0,072	-0,151	-0,322	-0,288	-0,345
21 dias frente	-1,590	-1,901	-1,970	-1,566	-0,639	-1,686	-1,751	-1,322	-1,069	-0,989	-2,393	-1,857	-1,730	-1,020	-0,814
63 dias frente	0,170	-0,414	-0,743	-1,147	-1,296	-4,520	-4,305	-3,005	-2,206	-1,325	2,189	2,244	2,147	2,388	2,218
RNAxAR1															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	-5,585	-4,521	-2,814	-2,282	-2,858	-1,853	-1,840	-1,222	-1,390	-2,231	-2,125	-1,560	-1,409	-2,234	-1,461
5 dias frente	-2,964	-2,310	-1,393	-0,890	-1,218	-2,186	-1,614	-1,156	-0,777	-0,526	-0,147	-0,193	-0,344	-0,299	-0,327
21 dias frente	-2,056	-2,134	-2,000	-1,595	-0,899	-2,472	-2,043	-1,363	-1,081	-0,982	-2,304	-1,784	-1,654	-0,981	-0,787
63 dias frente	-1,262	-1,541	-1,680	-1,924	-2,052	-4,050	-2,926	-2,284	-1,969	-2,003	0,501	1,302	2,254	2,044	1,754
RNAxVAR1															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	-3,854	-2,171	-2,670	-5,280	-3,747	-1,695	-1,981	-2,792	-2,853	-1,719	-2,052	-1,793	-1,675	-1,571	-0,719
5 dias frente	-3,019	-2,150	-1,343	-1,287	-1,018	-1,212	-0,652	-1,128	-0,774	-0,184	-0,276	-0,487	-0,351	-0,222	-0,562
21 dias frente	-2,001	-2,196	-2,061	-1,731	-0,754	-2,515	-1,958	-1,408	-1,159	-0,948	-2,582	-1,915	-1,655	-1,006	-0,804
63 dias frente	-1,127	-1,515	-1,698	-1,948	-2,130	-4,267	-3,601	-2,903	-2,621	-2,512	-0,164	0,687	1,584	1,661	1,490
RNAxNS															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	-16,615	-3,498	-6,295	-7,285	-5,250	-2,964	-3,464	-5,236	-5,954	-5,931	-5,580	-4,349	-3,654	-4,455	-5,081
5 dias frente	-7,312	-1,604	-0,744	-1,189	-1,222	-1,922	-2,442	-2,710	-2,569	-2,354	-2,807	-2,299	-1,899	-2,121	-2,392
21 dias frente	-2,322	-1,542	-1,015	-0,338	0,908	-1,059	-2,371	-1,757	-1,454	-1,186	-1,868	-1,635	-1,608	-1,344	-1,366
63 dias frente	-0,557	-0,179	-0,272	-0,558	-0,720	-2,280	-6,436	-3,851	-2,743	-2,367	0,888	1,777	2,023	1,997	1,576
RNAxSV															
maturidade	1	3	4	6	9	12	15	18	21	24	27	30	36	42	48
1 dia frente	-5,084	-8,448	-9,191	-5,153	-3,144	-2,870	-2,874	-3,107	-3,350	-3,444	-3,691	-3,631	-3,667	-4,002	-4,099
5 dias frente	-2,741	-2,663	-1,960	-1,409	-1,512	-2,178	-2,016	-2,220	-2,014	-1,787	-1,974	-1,927	-1,946	-2,086	-2,223
21 dias frente	-1,586	-2,011	-1,873	-1,615	-0,225	-1,771	-2,041	-1,561	-1,237	-0,959	-1,958	-1,670	-1,589	-1,211	-1,059
63 dias frente	0,164	-0,414	-0,799	-1,173	-1,175	-3,276	-5,501	-3,400	-2,324	-1,738	1,940	2,117	2,021	2,015	1,850

Nota: Essa tabela apresenta a estatística de Diebold-Mariano para os erros de previsão fora da amostra dos modelos de redes neurais artificiais e dos modelos concorrentes para os quatro horizontes de previsão - comparando os modelos aos pares. O período das previsões se estende de março/2010 a janeiro/2011. Valores negativos indicam superioridade das previsões do primeiro método do par. Legenda: RNA = modelo de redes neurais artificiais; AR1= modelo autoregressivo (1); VAR1= vetor-autoregressivo(1); NS= modelo de Nelson e Siegel estimado pelo método de dois passos; SV= modelo de Svensson

Fonte: elaboração própria.

Para testar a significância estatística da diferença das previsões, foi aplicado o teste de Diebold & Mariano (1995). Apresentam-se as estatísticas de teste para uma função de perda

quadrática, após controlada a correlação serial dos erros de previsão (ver Newey and West, 1987) - tabela 7 e tabela 8. Valores negativos indicam superioridade do primeiro método do par e valores absolutos maiores do que 1.96 e 1.66 indicam rejeição da hipótese nula (diferença entre a média dos erros quadrados é nula) a 95% e 90% de confiança respectivamente. Percebe-se a capacidade dos modelos de RNA de gerarem previsões da estrutura a termo em todos os horizontes de previsão e praticamente para as três regiões da curva. Destaca-se que muitos dos resultados são significativos aos níveis de 90% e 95% de confiança, com destaque especial para as previsões de um dia e uma semana à frente na região de curto prazo da curva.

6 CONCLUSÕES

Neste trabalho nos propusemos testar a capacidade preditiva dos modelos de redes neurais frente aos modelos tradicionalmente utilizados para prever a estrutura a termo da taxa de juros. Para tanto, utilizamos duas configurações de redes (estática e dinâmica) e levamos em consideração que a dinâmica das taxas dependem em qual das 3 regiões da curva que estas se encontram. Como forma de avaliar a capacidade preditiva dos modelos, foram utilizadas as últimas 200 observações de um painel composto por 986 observações diárias de 15 maturidades da curva de juros brasileira. Os resultados encontrados foram animadores, colocando os modelos de redes neurais artificiais como uma alternativa promissora para previsão da estrutura a termo da curva de juros.

Em média, os modelos estimados através de RNA conseguiram superar as previsões realizadas pelo *random walk* em todos os horizontes de previsão, sendo essas previsões em torno de 2% e 5% melhores para os horizontes de 1 dia e 1 semana à frente, e de 12% e 7% melhores para os horizontes mais longos, de 1 mês e 3 meses. Além disso, apresentou previsões em torno de 15% e 10% melhores que o modelo de NS para os horizontes de 1 mês e 3 meses à frente.

Neste sentido, o trabalho contribui com a literatura ao apontar que é possível encontrar modelos de redes neurais artificiais com boas habilidades preditivas, tanto em relação as previsões do *random walk* quanto dos modelos tradicionalmente utilizados para este fim, ratificando a capacidade dos modelos não-lineares em relação aos modelos lineares. Ademais, verificou-se que os modelos de redes neurais artificiais são capazes de realizar boas previsões para todos os horizontes testados, principalmente para região de curto prazo da curva, com destaque especial para as previsões com horizontes de 1 dia e 1 semana. Uma extensão natural em relação a este trabalho é procurar verificar se esta capacidade de previsão se mantém no tempo e se a utilização do arcabouço bayesiano para a determinação dos pesos sinápticos é capaz de melhorar as previsões.

REFERÊNCIAS

- ABID, F; BEN SALAH, M. Estimating term structure of interest rates: neural network vs one factor parametric models. *University of Sfax, Business and Economics Working Paper*, Tunisia, 2002.
- ALMEIDA, C. E VICENTE, J.V. Movimentos da estrutura a termo e critérios de minimização do erro de previsão em um modelo paramétrico exponencial. *Revista Brasileira de Economia*, 2008
- ALMEIDA, C. et al. Does curvature enhance forecasting? *International Journal of Theoretical and Applied Finance*, 2009.
- ALMEIDA, C.; SIMONSEN, A.; VICENTE, J. Forecasting bond yields with segmented term structure models. *Central Bank of Brazil, Brasília-DF, Working Paper 288*, 2012.
- ANG, A.; PIAZZESI, M. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, 2003.
- AUDRINO, F.; MEDEIROS, M. C. Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. *Journal of Applied Econometrics*, 2011.
- CALDEIRA, J.; VALLE MOURA, G.; SAVINO PORTUGAL, M. Efficient yield curve estimation and forecasting in brazil. *Revista Economia*, 2010.
- CALDEIRA, J. F.; LAURINI, M. P.; PORTUGAL, M. S. Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility. *Brazilian Review of Econometrics*, 2010.
- CALDEIRA, J.F.; FURLANI, L.C. Break-even inflation rate and the risk premium: an alternative approach to the VAR models in forecasting the CPI. *Economics Buletin*, 2011.
- CAO, L.J.; TAY, F. E. Support vector machine with adaptive parameters in financial time series forecasting. *Neural Networks, IEEE Transactions on*, 2003.
- CHOUDHARY, M. A.; HAIDER, A. Neural network models for inflation forecasting: an appraisal. *Applied Economics*, 2012.
- CHRISTENSEN, J., DIEBOLD, F.; RUDEBUSCH, G. *An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model*. *The Econometrics Journal*, 2009.
- CORONEO, L., NYHOLM, K., VIDOVA-KOLEVA, R. *How arbitrage-free is the Nelson-Siegel model?* *Journal of Empirical Finance*, 2011.
- COX, J. C.; INGERSOLL, E. A theory of the term structure of interest rates. *Econometrica*, 1985.
- DAI, Q.; SINGLETON, K. J. Specification analysis of affine term structure models. *The Journal of Finance*, 2002.

- DE POOTER, M.; RAVAZZOLO, F.; VAN DIJK, D. Term structure forecasting using macro factors and forecast combination. *FRB International Finance Discussion Paper*, 2010.
- De JONG, F. Time series and cross section information in affine term structure models. *Journal of Business and Economic Statistics*, 2000.
- DIEBOLD, F. X.; LI, C. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 2006.
- DUFFEE, G. R. Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 2002.
- FAUSETT, L. V. *Fundamentals of neural networks: architectures, algorithms, and applications*. Prentice-Hall Englewood Cliffs, NJ, 1994.
- HAYKIN, S. S. *Redes neurais-principios e prática*. Porto Alegre: Bookman, 2001.
- HEATH, D.; JARROW, R.; MORTON, A. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 1992.
- HÖRDAHL, P.; TRISTANI, O.; VESTIN, D. A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics*, 2006.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural networks*, 1989.
- HO, T. S.; LEE, S.-B. Term structure movements and pricing interest rate contingent claims. *The Journal of Finance*, 1986.
- HULL, J.; WHITE, A. Pricing interest-rate-derivative securities. *Review of financial studies*, 1990.
- LAURINI, M. P.; HOTTA, L. K. Bayesian extensions to diebold-li term structure model. *International Review of Financial Analysis*, 2010.
- LONGSTAFF, F. A.; SCHWARTZ, E. S.. *Interest rate volatility and the term structure: A two-factor general equilibrium model*. *Journal of Finance*. 1992.
- MÖNCH, E. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, 2008.
- MOSHIRI, S.; CAMERON, N. Neural network versus econometric models in forecasting inflation. *Journal of Forecast*, 2000.
- NAKAMURA, E. Inflation forecasting using a neural network. *Economics Letters*, 2005.
- NELSON, CHARLES R.; SIEGEL, A. F. Parsimonious modeling of yield curves. *Journal of Business*, 1987.

NEWKEY, W. K.; WEST, K. D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 1987.

REZENDE, R. B.; FERREIRA, M. S. Modeling and forecasting the yield curve by an extended Nelson-Siegel class of models: A quantile autoregression approach. *Journal of Forecasting*, 2011.

SVENSSON, L. E. O. *Estimating Forward Interest Rates with the Extended Nelson- Siegel Method*. Quarterly Review Sveriges Riksbank 3: 13 – 26, 1995.

TÄPPINEN, J. *Interest rate forecasting with neural networks*, Discussion Papers 170, Government Institute for Economic Research Finland (VATT), 1998.

TKACZ, G. Neural network forecasting of canadian gdp growth. *International Journal of Forecasting*, 2001.

VASICEK, O. A.; FONG, H. G. Term structure modeling using exponential splines. *The Journal of Finance*, 1977.

VASICEK, O. An equilibrium characterization of the term structure. *Journal of financial economics*, 1977.

VEREDA, L., LOPES, H. FUKUDA, R. Estimating var models for the term structure of interest rates. *Insurance: Mathematics and Economics*, 42 (2): 548–559, 2008.

VICENTE, J.; TABAK, B. M. Forecasting bond yields in the brazilian fixed income market. *International Journal of Forecasting*, 2008.

WHITE, H. *Artificial neural networks: approximation and learning theory*. Blackwell Publishers, Inc., 1992.

ZHANG, G.; EDDY PATUWO, B.; Y HU, M. Forecasting with artificial neural networks: The state of the art. *International journal of forecasting*, 1998.