



NOVA

IMS

Information
Management
School

MEGI

Mestrado em Estatística e Gestão de Informação

Master Program in Statistics and Information Management

Previsão da Estrutura Temporal da Taxa de Juro na Zona Euro

Aproximação paramétrica e por métodos de
aprendizagem automática

Inês Regina Portela Costa Garcia

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Estatística e Gestão de
Informação

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

PREVISÃO DA ESTRUTURA TEMPORAL DA TAXA DE JURO NA ZONA EURO

por

Inês Regina Portela Costa Garcia

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Estatística e Gestão de Informação, Especialização em Análise e Gestão de Risco

Orientador: Professor Doutor Jorge Miguel Ventura Bravo

Coorientador: Professor Doutor Mauro Castelli

novembro 2015

AGRADECIMENTOS

Em primeiro lugar gostaria de agradecer ao Professor Doutor Jorge Miguel Ventura Bravo e ao Professor Doutor Mauro Castelli, orientador e coorientador deste trabalho, que contribuíram com os seus ensinamentos e sugestões para que este estudo fosse concluído com sucesso. É óbvio que qualquer erro ou omissão são da minha inteira responsabilidade.

Aos meus pais, José e Lídia, e ao meu irmão, Flávio, agradeço todo o apoio e incentivo que me deram para abraçar este difícil projeto, proporcionando-me todas as condições ao seu alcance para que a sua realização fosse possível. Obrigada pelo apoio incondicional.

Não posso deixar de agradecer a todos os meus amigos que me ajudaram a ganhar motivação e que me acompanharam ao longo deste trabalho entendendo as fases de maior pressão e ausência. E deixo ainda um agradecimento especial aos que leram este estudo. Obrigada pela paciência, pelas críticas, pelos conselhos e pela vossa disponibilidade.

RESUMO

A estrutura temporal das taxas de juro, também conhecida por *yield curve* ou curva de rendimentos define a relação entre as taxas de juros e o prazo de vencimento (ou maturidades) dos investimentos feitos. Assim, o desenvolvimento de modelos que possibilitem a obtenção de previsões precisas sobre a estrutura temporal das taxas de juro e que permitam estudar a dinâmica da evolução das taxas de juro é de crucial importância em diversas áreas de financiamento.

Neste estudo investigou-se a performance de diferentes métodos de previsão para obter a estrutura temporal das taxas de juro da Zona Euro, considerando o período entre 2009 e 2015. Em termos mais específicos, foi analisada a capacidade preditiva do modelo de Nelson-Siegel & Svensson assumindo que os parâmetros resultantes da estimação da especificação paramétrica podem ser modelizados através de métodos de séries temporais univariados (modelos ARIMA, *Random walk*) e multivariados (modelos VAR) e Redes Neurais Artificiais (RNA) individuais e conjuntas.

Os resultados deste estudo mostram que (i) as RNA com a previsão dos parâmetros em simultâneo exibem os valores de erro mais baixos para as maturidades de curto e médio prazo (3 meses a 5 anos); (ii) As RNAs individuais são melhores para prever as taxas de juro nas maturidades compreendidas entre os 7 e os 10 anos, e que (iii) para as maturidades de longo e muito longo prazo (15 e 30 anos respetivamente) deverá ser escolhido o modelo VAR(1). Estes resultados são robustos e consistentes para todos os horizontes de previsão analisados (1,2 e 3 meses). Contudo, no período analisado nenhum dos modelos testados apresenta valores de erro inferiores aos obtidos com o modelo *Random Walk*.

PALAVRAS-CHAVE

Taxas de Juro, Zona Euro, Modelo Nelson-Siegel & Svensson, Redes Neurais Artificiais, Previsão

ABSTRACT

The yield curve defines the relationship between interest rates and the maturity of the investments made. Thus develop models that are able to produce an accurate prediction about the term structure of interest rates and that allow us to study the dynamics of changes in interest has a crucial importance in various areas of finance.

This study focus on the predictive power of the yield curves the Eurozone, comparing the performance of several forecasting methodologies using data from 2009 to 2015. Taking into account the factors set by the specifications of Nelson-Siegel & Svensson were tested: Univariate models (ARIMA models, Random Walk) and multivariate (VAR) and Artificial Neural Networks (with an ANN that predicts each parameter individually and an ANN that forecast all the parameter simultaneously).

The results of this study show that (i) RNA that forecast all the parameter simultaneously produce more accurate forecasts for the short-medium maturities previsions (3-5 years), (ii) ANN that predicts each parameter individually has better results with yields with maturities between 7 and 10 years, and that (iii) VAR(1) are better for the long and very long term maturities (15 and 30 years respectively). These conclusions held in all the time horizons forecasts. Although none of the developed models has lower error than the Random Walk.

KEYWORDS

Yield curve, Euro-Area, Nelson-Siegel & Svensson Model, Neural Networks, Forecast

ÍNDICE

1. Introdução	1
2. Revisão da Literatura	5
2.1. Modelação da <i>Yield curve</i>	5
2.2. Modelos de Previsão Paramétricos	11
2.3. Modelos de Aprendizagem Automática	14
2.4. Estudos Anteriores Sobre a Temática da Previsão da ETTJ	17
3. Metodologia	19
3.1. Dados	19
3.1.1. Perfis de Volatilidade e de Correlação para os Choques de Taxas de Juro..	22
3.1.2. Dinâmica dos fatores latentes da função de Nelson-Siegel & Svensson	23
3.1.3. Decomposição dos Choques de Taxa de Juro	24
3.2. Previsão	26
3.2.1. Modelos Univariados ARIMA (p,d,q)	28
3.2.2. Modelos VAR	33
3.2.3. Modelos de Aprendizagem Automática	36
3.2.4. <i>Random walk</i>	42
3.3. Critérios de Desempenho dos Modelos	43
4. Resultados e Discussão	44
5. Conclusões	49
5.1. Limitações e Recomendações para Trabalhos Futuros	50
6. Bibliografia	51
7. Anexos	55

ÍNDICE DE FIGURAS

Figura 1: Exemplo Ilustrativo da <i>Yield curve</i> na Zona Euro (taxas <i>spot</i>)	6
Figura 2: Representação Rede Neuronal <i>Feed-Forward</i>	15
Figura 3: Evolução das Taxas de Juro na Zona Euro entre 2004 e 2015	20
Figura 4 : Três Fatores Principais da ETTJ	25
Figura 5: Gráficos ACF e PACF do parâmetro β_0	29
Figura 6: Análise dos resíduos pelo Ljung-Box Test do parâmetro β_0	32
Figura 7: Rede Neuronal β_0	41
Figura 8: Gráficos ACF e PACF do parâmetro β_1	55
Figura 9: Gráficos ACF e PACF do parâmetro β_2	55
Figura 10: Gráficos ACF e PACF do parâmetro β_3	56
Figura 11: Análise dos resíduos do parâmetro β_1	57
Figura 12: Análise dos resíduos do parâmetro β_2	57
Figura 13: Análise dos resíduos do parâmetro β_3	58
Figura 14: Rede Neuronal Conjunta.....	61
Figura 15: Rede Neuronal β_1	62
Figura 16: Rede Neuronal β_2	62
Figura 17: Rede Neuronal β_3	63

ÍNDICE DE TABELAS

Tabela 1: Matriz de Correlação entre as variações da Taxa de Juro	22
Tabela 2: Principais Indicadores Estatísticos	23
Tabela 3: Tendências Estatísticas.....	23
Tabela 4: Matriz de Correlação Entre os Choques sobre o Nível, Inclinação e Curvatura da ETTJ.....	24
Tabela 5: Importância Relativa dos Vários Fatores para cada Prazo.....	24
Tabela 6: Exemplo Ilustrativo de aplicação do método <i>recursive expanding window forecasts</i>	26
Tabela 7 : Resumo dos resultados dos testes de raízes unitárias.....	30
Tabela 8: Estatísticas resumo do modelo univariado ótimo selecionado para cada parâmetro	30
Tabela 9: Resultados do Teste Ljung-Box	31
Tabela 10: Parâmetros obtidos com os modelos univariados ARIMA(p,d,q).....	33
Tabela 11: <i>Lags</i> testados para os modelos VAR	34
Tabela 12: Resultados da Estimação do modelo VAR (período junho 2009-dezembro2013)	35
Tabela 13: Quadro Resumo das estatísticas de teste dos resíduos dos modelos VAR	36
Tabela 14: Configurações testadas à RNA conjunta	40
Tabela 15: Configurações final da RNA individual	41
Tabela 16: Resultados da previsão para diferentes horizontes temporais	45
Tabela 17: Configurações testadas à RNA para o parâmetro β_0	59
Tabela 18: Configurações testadas à RNA para o parâmetro β_1	59
Tabela 19: Configurações testadas à RNA para o parâmetro β_2	60
Tabela 20: Configurações testadas à RNA para o parâmetro β_3	60

LISTA DE SIGLAS E ABREVIATURAS

ACP	Análise de Componentes Principais
ADF	Augmented Dickey–Fuller
AIC	AKAIKE Information Criterion
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ANN	Artificial Neural Network
BCE	Banco Central Europeu
ETTJ	Estrutura Temporal das Taxas de Juro
MA	Moving Average
MAE	Mean Absolute Error
ME	Mean Error
NS	Nelson-Siegel
NSS	Nelson-Siegel & Svensson
PP	Philip-Perrow
RPROP+	Resilient BackPropagation
RMSE	Root-Mean-Square Error
RNA	Redes Neuronalis Artificias
RW	<i>Random walk</i>
VAR	Vetor Auto-Regressivo
VECM	Vetor Error Correction Model

1. INTRODUÇÃO

“Life must be understood looking backwards, but must be lived looking forward.”

(McNelis, 2005)

Atualmente, e não apenas na área da economia e das finanças, o sucesso da implementação de qualquer estratégia depende da qualidade das previsões efetuadas, ou seja, da extração de conhecimento a partir da informação já existente, de forma a criar um *input* útil para o futuro. Ao garantir que as previsões são fidedignas, as instituições financeiras poderão tomar decisões mais conscientes e de menor risco ao longo do tempo.

A estrutura temporal das taxas de juro (doravante abreviada como ETTJ), também conhecida por “*yield curve*” ou curva de rendimentos, descreve a relação entre a taxa de rendibilidade (*yield*) implícita no valor de obrigações e outros instrumentos de dívida e taxa de juro transacionados, como função das suas diferentes maturidades.

O conceito de *yield curve* para a Zona Euro tornou-se relevante com a criação do BCE e com a introdução do Euro a 1 de janeiro de 1999 (Berk e Bergeijk, 2000). Apesar de os membros da Zona Euro continuarem a ter as suas *yield curves* nacionais, os *spreads* entre as diferentes *yields* diminuíram consideravelmente em virtude da política monetária única e da eliminação da incerteza da taxa de câmbio (Pederson e Wormstrup, 2001).

A compreensão da dinâmica da *yield curve* ou estrutura temporal de taxas de juro é importante em muitos domínios, incluindo na avaliação de ativos financeiros e seus derivados, na gestão do risco financeiro, na escolha e gestão de carteiras de investimentos, na estruturação da dívida pública, na condução da política monetária, na obtenção de indicadores sobre a evolução esperada do ciclo económico ou da inflação, nas expectativas dos agentes económicos ou na valorização de bens (Berk, 1998).

Por esse motivo, os investigadores têm-se dedicado ao desenvolvimento de um conjunto numeroso e variado de técnicas e modelos de estimação e previsão da *yield curve*. Entre as abordagens mais populares encontram-se os modelos de equilíbrio e os modelos de não-arbitragem.

Os modelos de não-arbitragem são desenhados de modo a que a dinâmica de toda a ETTJ seja automaticamente consistente com dados observados no mercado para as taxas de juro e, nalguns casos, para a sua estrutura por prazos de volatilidade (veja-se, p.e., Hull e White, 1990; Heath et al., 1992).

Os modelos de equilíbrio para a ETTJ assumem que os valores das obrigações isentas de risco de incumprimento e de várias maturidades podem ser explicados convenientemente pelos processos estocásticos seguidos, ao longo do tempo, por um conjunto reduzido de variáveis. A especificação do processo estocástico pode ser feita de forma exógena ou determinada endogenamente a partir de modelos de equilíbrio geral para a ETTJ, assumindo, neste último caso, hipóteses sobre as preferências dos investidores (veja-se, p.e., Vasicek, 1977, Cox, Ingersoll & Ross (1985) e Duffie e Kan, 1996).

Não obstante a sua capacidade para ajustar adequadamente as taxas de juro observadas no mercado, ambas as abordagens têm-se revelado incapazes de proporcionar um bom desempenho preditivo, uma vez que a previsão não constitui, regra geral, o seu principal objetivo. Com efeito, os modelos de não-arbitragem são usados preferencialmente para efeitos de calibragem da *yield curve* e *pricing* de instrumentos financeiros e não para a obtenção de previsões.

Para este fim, Diebold e Li (2006) propuseram a utilização da função paramétrica de Nelson-Siegel para decompor a evolução de toda a *yield curve* mediante um vetor de parâmetros com dinâmica representada por processos estocásticos.

O modelo dinâmico de Nelson-Siegel ganhou popularidade entre os profissionais do mercado financeiro e os bancos centrais. Esta popularidade deve-se não apenas à sua parcimónia e facilidade na obtenção de estimativas dos parâmetros mas também, apesar da sua estrutura simples, à sua capacidade de representar as propriedades empíricas da ETTJ historicamente observadas e aos resultados encorajadores obtidos em exercícios de previsão quando comparados com os resultados obtidos por modelos alternativos.

Para além da boa capacidade de ajustamento do modelo, os fatores dados pela forma exponencial do modelo de Nelson-Siegel oferecem uma interpretação intuitiva para a dinâmica das taxas de juro, representando genericamente a evolução do nível, da inclinação e da curvatura da *yield curve*. Estes resultados abriram caminho para o estudo do comportamento das séries temporais dos fatores, bem como, da sua relação com variáveis macroeconómicas relevantes. Estes trabalhos acrescentaram à especificação de Nelson e Siegel uma fundamentação económica de que carecia inicialmente (Gilli, Große, & Schumann, 2010).

O principal objetivo desta investigação é avaliar a capacidade preditiva da estrutura temporal de taxas de juro da Zona Euro considerando diferentes métodos de previsão. Em termos mais específicos, é analisada a capacidade preditiva da versão dinâmica do modelo de Nelson e Siegel e sua extensão desenvolvida por Svensson (1994) usada inicialmente por Diebold e Li (2006). Para tal, assume-se que a dinâmica temporal dos fatores latentes (parâmetros) resultantes da estimação da especificação de Nelson-Siegel & Svensson (NSS) pode ser especificada através de métodos de previsão séries temporais univariados (modelos ARIMA, *Random walk*) e multivariados (modelos VAR) e Redes Neurais Artificiais (RNA).

A base de dados considerada neste estudo consiste na *yield curve* das obrigações da Zona Euro com notação de *rating* AAA e nas estimativas dos parâmetros da função de NSS publicadas pelo Banco Central Europeu para todos os dias úteis no período compreendido entre 6 de Setembro de 2004 e 25 de maio de 2015.

A modelação das séries temporais dos quatro parâmetros da especificação de NSS ($\beta_0, \beta_1, \beta_2, \beta_3$) foi feita usando processos estocástico univariados ou multivariados, com ou sem possibilidade de interação entre os parâmetros, permite a obtenção de previsões para os fatores latentes. Estas podem, por sua vez, ser incorporadas na forma funcional do modelo de NSS permitindo a derivação de projeções para as taxas de juros de todas as maturidades. O contraste dos valores projetados com os valores observados para as taxas de juro permite o cômputo de métricas de avaliação do poder preditivo dos modelos, quer no caso em que estes são individualmente considerados, quer no caso

em que se afere a sua performance relativa, e a realização de inferência estatística quanto aos resultados.

À semelhança de outros estudos, os resultados obtidos nesta dissertação são comparados com o desempenho do modelo *Random Walk (RW)* que é utilizado como *benchmark* natural aos modelos propostos.

A inclusão das redes neuronais artificiais neste estudo justifica-se pela crescente importância que esta técnica tem assumido em diversas áreas ou domínios científicos, e pelas vantagens que apresenta relativamente a outros métodos de modelação e previsão (Bodyanskiy e Popov, 2006).

Uma rede neuronal artificial é um modelo matemático que tem como base as redes neuronais biológicas e consiste num conjunto de neurónios artificiais ligados entre si e que são capazes de processar informação, reconhecer padrões e aprender (Kohonen, 1987). Resumidamente, é um modelo inspirado na estrutura neuronal de organismos inteligentes e que adquire conhecimento através da experiência.

Entre as principais vantagens desta técnica relativamente aos métodos estatísticos convencionais destaca-se, por exemplo, a sua capacidade de aprendizagem a partir de um conjunto de dados e a aptidão para efetuar previsões a partir de conjuntos de dados com *missing values* (valores omissos) (Hill, O'Connor e Remus, 1996). Em vários estudos comparativos tem sido demonstrado que os modelos obtidos usando redes neuronais apresentam uma maior capacidade de previsão relativamente a outros (Täppinen, 1998).

Esta investigação insere-se na literatura académica e dos mercados financeiros dedicada à modelação da dinâmica da *yield curve* e à temática da derivação de projeções de longo prazo da estrutura por prazos das taxas de juro. Estudos similares incluem, entre outros, Diebold e Li (2006), Matsumura et al. (2011), Luo et al. (2012) e Vela (2013).

Pela sua proximidade com o presente trabalho, destaca-se o trabalho de Vela (2013), onde se elabora um estudo da capacidade preditiva da ETTJ de vários países da América Latina, tendo em conta os fatores latentes da especificação de NSS. O seu trabalho consistiu num estudo comparativo de várias metodologias de previsão, entre as quais modelos autorregressivos, modelos vetoriais autorregressivos, redes neuronais artificiais independentes para cada fator e redes neuronais artificiais para todos os fatores de NSS. O estudo concluiu que a aproximação realizada através de redes neuronais apresenta previsões para o mês seguinte mais fidedignas na maioria das curvas estudadas. No entanto, o autor afirma que não existe evidência suficiente que suporte as redes neuronais como sendo o melhor método de previsão, constituindo assim mais uma alternativa para prever a ETTJ.

Os resultados empíricos desta investigação permitiram concluir que, entre os modelos testados, não existe um que apresente uma performance superior em todas as maturidades e para todos os horizontes temporais estudados. No entanto, de acordo com os resultados obtidos, verifica-se que: (i) as RNA com a previsão dos parâmetros em simultâneo exibem os valores de erro mais baixos para as maturidades de curto e médio prazo (3 meses a 5 anos); (ii) As RNAs individuais são melhores para prever as taxas de juro nas maturidades compreendidas entre os 7 e os 10 anos, e que (iii) para as maturidades de longo e muito-longo prazo (15 e 30 anos) deverá ser escolhido o modelo VAR(1).

Refira-se, contudo, que no período analisado nenhum dos modelos testados apresenta valores de erro inferiores aos obtidos com o modelo *Random walk*. Este resultado pode estar relacionado com o facto de no período em análise a *yield curve* ter exibido uma volatilidade significativa e ter tomado valores atípicos, incluindo taxas de juro negativas nalgumas maturidades, dificultando deste modo a previsão, e daí que um modelo com uma estrutura mais simples tenha conseguido obter valores de previsões relativamente próximos dos valores observados.

Este estudo contribui para a literatura sobre este tema de duas maneiras principais. Em primeiro lugar, o estudo compara o desempenho em termos preditivos de um conjunto de métodos alternativos de previsão da *yield curve* da Zona Euro, replicando em parte exercícios similares realizados noutros países. Em segundo lugar, o estudo avalia o desempenho preditivo dos modelos num contexto e num período temporal marcados por uma extrema volatilidade nos mercados de taxa de juro da Zona Euro, na sequência da crise financeira de 2008, da crise económica global que se lhe sucedeu, da crise da dívida pública e da implementação de planos de resgate financeiro em vários países da Zona Euro (Portugal, Irlanda, Grécia), da mudança na condução da política monetária pelo BCE (por exemplo, *quantitative easing program*) e da instabilidade política em países como Portugal, Grécia ou Espanha. Neste contexto, as capacidades de previsão desses modelos são testados dentro de um período em que existiram mudanças substanciais na *yield curve*, contrastando assim com exercícios similares realizados em contextos de maior estabilidade nos mercados de dívida.

O remanescente desta dissertação está organizado da seguinte forma. No capítulo 2 faz-se um breve enquadramento do tema em estudo e uma revisão da literatura científica mais relevante sobre o tema, fornecendo uma maior contextualização do problema. Posteriormente, descrevem-se os modelos de NS e NSS, com maior destaque para este último uma vez que é aquele que servirá de base ao estudo empírico. No Capítulo 3 apresentam-se os dados das curvas de rendimento da Zona Euro e uma breve análise descritiva dos dados recolhidos, bem como uma análise de componentes principais para compreender o contributo de cada parâmetro para a dinâmica da *yield curve*. Na última secção desde capítulo, é feita uma previsão através dos métodos paramétricos e dos métodos de aprendizagem automática. No Capítulo 4 são apresentados os principais resultados das diferentes abordagens e é feita uma comparação das *yield curves* resultantes da amostra. Finalmente, no Capítulo 5 apresentam-se as conclusões retiradas do estudo.

2. REVISÃO DA LITERATURA

Com o intuito de contextualizar e fundamentar a importância do tema escolhido, este capítulo aprofunda algumas temáticas relevantes para a compreensão do tema proposto. Este capítulo está dividido em quatro subcapítulos. No primeiro descrevem-se os diversos modelos paramétricos existentes para realizar modelações, fazendo também referência aos modelos de equilíbrio e aos modelos de não-arbitragem. No segundo, faz-se uma descrição dos modelos de previsão paramétricos, com particular destaque para os testados no presente estudo. No terceiro, faz-se uma breve descrição dos modelos de aprendizagem automática e dos tipos de redes neuronais existentes. Por último, são apresentados alguns dos estudos similares já elaborados sobre a temática da previsão da ETTJ.

2.1. MODELAÇÃO DA *YIELD CURVE*

A taxa de juro é um instrumento financeiro de grande relevância para a política monetária. Assim, o acompanhamento da sua evolução e da do mercado obrigacionista é também importante para a componente de supervisão do sistema financeiro. Naturalmente, para além das autoridades monetárias, também para os participantes nos mercados financeiros, emitentes de dívida e investidores, a análise da ETTJ assume importância para a definição das suas estratégias (Campbell, 1995).

A representação gráfica da ETTJ procura traduzir visualmente o comportamento das taxas de juro para diversas maturidades. A curva pode assumir múltiplas configurações (Normal ou Crescente, Invertida ou Decrescente, Estática ou Constante, *Steep Yield Curve*, etc.), correspondendo a diferentes contextos quanto ao *market price* do risco de liquidez ou de crédito, à preferência por determinadas maturidades dos investidores, à preferência por liquidez, às expectativas dos agentes económicos quanto ao crescimento, ao equilíbrio no mercado de fundos. As principais teorias explicativas da ETTJ que procuram fundamentar as diferentes configurações possíveis da ETTJ são a Teoria das expectativas Puras ou racionais, a Teoria da Segmentação dos Mercados, a Teoria dos Prémios de Risco (ou de Liquidez) e a Teoria do Habitat Preferido.¹

A *yield curve* pode ser definida de forma equivalente através das taxas de juro *spot*, taxas de juro *forward* ou através dos fatores de desconto. A taxa de juro *spot* é a taxa de juro em vigor no momento zero (hoje), para uma aplicação com início imediato e com um único *cash-flow* no momento futuro t . A taxa de juro *forward* é uma taxa de juro fixada no presente para um empréstimo a concretizar numa determinada data futura. A característica principal desta taxa resulta de o investimento ser realizado no futuro e não hoje (Bravo, 2001).

Na Figura 1 está representada, a título ilustrativo, a *yield curve* estimada e publicada pelo BCE no dia 24 de outubro de 2015 para as diferentes maturidades entre os 0 e os 30 anos (taxas *spot*).

¹ Para uma revisão mais rigorosa e detalhada das teorias explicativas da ETTJ recomenda-se a leitura de Shiller & McCulloch (1990) e Fonseca (1999).

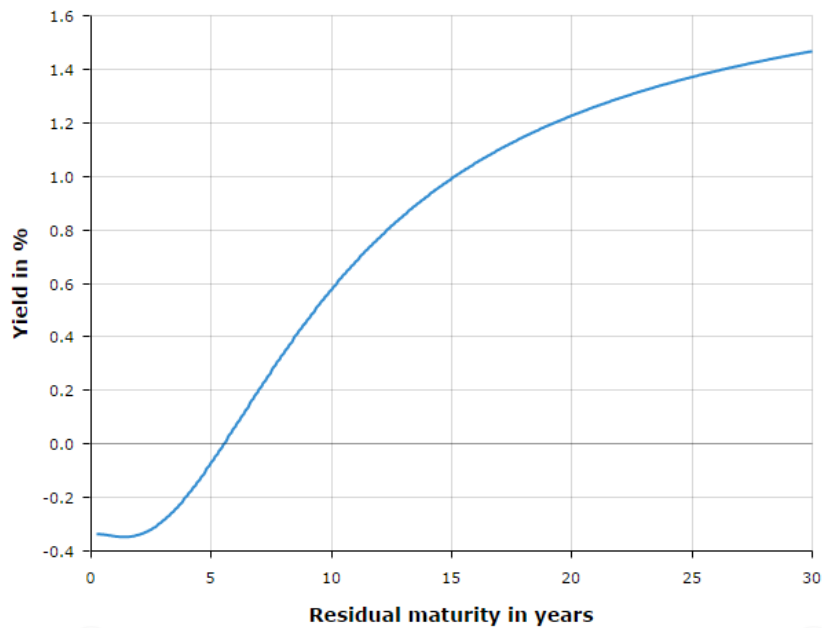


Figura 1: Exemplo Ilustrativo da *Yield curve* na Zona Euro (taxas *spot*)
Fonte: BCE

A relação entre as taxas de juro *spot* e *forward* e a respetiva maturidade observadas nos mercados monetários interbancários e de dívida pública serve de *benchmark* para a determinação das taxas de juro a aplicar em operações de crédito à economia e ao consumo e é um *input* fundamental no *pricing* (definição de preços) de instrumentos financeiros (D Bolder & Stréliski, 1999).

A estrutura temporal das taxas de juro não é diretamente observável, e não pode ser estimada para todas as maturidades com base nas cotações observadas no mercado para obrigações de cupão zero, uma vez que, estas não são normalmente suficientes para abranger todo o espectro temporal desejado. Assim sendo, torna-se necessário proceder à sua estimação (Cox et al., 1985) através de uma combinação de instrumentos financeiros dos mercados monetário, de dívida e de derivados de taxa de taxas de juro (obrigações de taxa fixa, contratos de futuros, *swaps*).

Ao longo do tempo, foram sendo desenvolvidos diversos métodos (diretos e indiretos) de estimação da estrutura temporal das taxas de juro com distintas propriedades, vantagens e desvantagens, pelo que a escolha do modelo a ser utilizado para estimar as taxas de juro dependerá, não só do objetivo do estudo, mas também da ponderação de prós e contras feita pelos autores (Martellini et al., 2001; Ioannides, 2003).

Na literatura académica e dos mercados financeiros são inúmeras as investigações dedicadas à modelação da dinâmica das taxas de juro e à temática da derivação de projeções de longo prazo da *yield curve* (Monteiro, 2005). Os métodos de estimação da *yield curve* são normalmente agrupados em três grandes categorias: os modelos de equilíbrio, os modelos de não-arbitragem e os modelos paramétricos.

O princípio subjacente aos modelos paramétricos, também conhecidos como modelos baseados em funções ou modelos estatísticos, é o da especificação de uma única função definida para todo o espectro de maturidades considerado na análise e a estimação dos parâmetros através do método dos mínimos quadrados ponderados ou de métodos de máxima verosimilhança (BIS, 2005).

O modelo desenvolvido por Nelson & Siegel (1987) tenta estimar a relação entre os preços teóricos e os preços observados das obrigações ajustando, para um determinado ponto no tempo, a função de desconto. Os autores propuseram ajustar a ETTJ usando uma função suave e flexível capaz de representar com um pequeno número de parâmetros uma grande variedade de formas típicas que a ETTJ pode ter ao longo do tempo.

O modelo de Nelson e Siegel apresenta a seguinte expressão matemática para as taxas de juro *forward*:

$$f(t) = \beta_0 + \beta_1 e^{-\frac{t}{\tau}} + \beta_2 \frac{t}{\tau} e^{-\frac{t}{\tau}} \quad [1]$$

onde t representa a maturidade da taxa de juro, e β_0 , β_1 , β_2 e τ são parâmetros a estimar. β_0 representa o nível de taxas de juro, β_1 a inclinação e β_2 a curvatura da *yield curve*.

A equação (1) implica as seguintes expressões para a taxa *spot* $y(t)$:

$$y(t) = \beta_0 + \beta_1 \frac{\tau}{t} (1 - e^{-t/\tau}) + \beta_2 \frac{\tau}{t} \left[1 - e^{-t/\tau} \left(1 + \frac{t}{\tau} \right) \right] \quad [2]$$

e para a função de desconto, $d(t)$:

$$d(t) = \exp[-i(t)t] = \exp \left\{ -\beta_0 t + \beta_1 \tau (1 - e^{-t/\tau}) + \beta_2 \tau \left[1 - e^{-t/\tau} \left(1 + \frac{t}{\tau} \right) \right] \right\} \quad [3]$$

respetivamente.

A equação (2) acomoda a maioria das configurações que a ETTJ pode tomar. Cada um dos três parâmetros do modelo define um determinado comportamento para a taxa de juro *forward* instantânea (Bravo, 2001):

- O parâmetro β_0 , obrigatoriamente positivo, corresponde ao valor assintótico das taxas *forward* (e *spot*), isto é, corresponde ao nível para o qual as taxas de juro convergem à medida que a maturidade aumenta. Alterações neste parâmetro traduzem deslocamentos paralelos da ETTJ.
- O parâmetro β_1 determina se a inclinação da ETTJ é positiva (no caso em que $\beta_1 < 0$), ou negativa (no caso em que $\beta_1 > 0$).
- O parâmetro β_2 é responsável pela curvatura da ETTJ.

Bolder (2006) explicou que a maior desvantagem da utilização desta abordagem é não existir uma fundamentação teórica para o modelo, o que poderá causar previsões pouco fidedignas. Outras limitações na capacidade de previsão do modelo NS prendem-se com a fraca capacidade de adaptação aos dados quando comparado com os modelos polinomiais e quando os dados utilizados para construir o modelo não são regulares, o que pode levar a um modelo que se desvie bastante da realidade das taxas observadas.

Apesar do modelo de NS ser capaz de produzir uma grande variedade de formas da ETTJ, patenteia algumas dificuldades para a ajustar corretamente ao segmento de curto prazo da curva, especialmente em períodos de grande volatilidade em que a configuração da *yield curve* exibe mais do que uma alteração de curvatura. Para ultrapassar este problema e permitir um melhor ajuste da ETTJ, Svensson (1994) sugere a inclusão de parâmetros adicionais na especificação funcional para permitir uma maior flexibilidade (maior flexibilidade neste caso significa uma maior capacidade de ajustamento).

Em termos formais, Svensson (1994) e Bliss (1996) estenderam o modelo NS, incluindo termos adicionais β_3 e θ que definem a curvatura (côncava ou convexa) da *yield curve*.

A equação base do modelo de Svensson é a seguinte:

$$f(t) = \beta_0 + \beta_1 e^{-\frac{t}{\tau}} + \beta_2 \frac{t}{\tau} e^{-\frac{t}{\tau}} + \beta_3 \frac{t}{\theta} e^{-\frac{t}{\theta}} \quad [4]$$

onde os coeficientes β_3 e θ têm uma interpretação similar ao coeficiente β_2 e τ presentes na equação de Nelson & Siegel.

Considerando a relação entre as taxas *spot* e as taxas *forward*:

$$y(t) = \frac{1}{t} \int_t^t f(s) ds \quad [5]$$

$$f(t) = y(t) + t \frac{dy(t)}{dt} \quad [6]$$

obtém-se a seguinte forma funcional para as taxas de juro *spot*²

$$y(t) = \beta_0 + \beta_1 \left[\frac{1 - e^{-\frac{t}{\tau_1}}}{\frac{t}{\tau_1}} \right] + \beta_2 \left[\frac{1 - e^{-\frac{t}{\tau_1}}}{\frac{t}{\tau_1}} - e^{-\frac{t}{\tau_1}} \right] + \beta_3 \left[\frac{1 - e^{-\frac{t}{\tau_2}}}{\frac{t}{\tau_2}} - e^{-\frac{t}{\tau_2}} \right] \quad [7]$$

onde t representa o tempo para a maturidade e $\beta_i \tau_j$ são os parâmetros a estimar.

² Para mais informações consultar o documento técnico do BCE: BIS. (2005). Zero-Coupon *yield curves*: technical documentation. *BIS Papers*, No.25.

O modelo NSS é atualmente o modelo mais usado pelos bancos centrais.

Outros autores avaliaram a capacidade de previsão de formas complementares da *yield curve* que incluem mais fatores. Por exemplo, Cziráky (2007) analisou a inclusão do quarto termo de Svensson na equação NS, enquanto De Pooter (2007) avaliou uma versão ajustada do modelo NSS, do modelo Bliss (1997) e do modelo de Bjork & Christensen (1999). Além disso, a partir dos modelos listados, De Rezende & Ferreira (2011) propuseram uma equação com cinco fatores. Quase todos os autores verificaram uma melhor adaptação dos modelos que apresentam fatores extra pois contribui para uma maior flexibilidade do modelo. No entanto, modelos com mais fatores não implicam necessariamente um bom desempenho fora do período da amostra, devido ao risco de *overfitting*.

Em relação às restantes abordagens paramétricas já estudadas para prever o comportamento das taxas de juro, Dolan (1999) classificou os modelos de curva de rendimento em três tipos diferentes: Estocásticos, de Componentes Principais e Fundamental. Os modelos que apresentaram melhor performance são aqueles que têm implícitos fatores latentes que determinam a estrutura por prazo das taxas de juro, sendo estes os modelos de componentes principais. Dolan foi o primeiro a ligar estas componentes principais à forma funcional da ETTJ proposta por Nelson & Siegel (1987), determinando que os três fatores que compõem o seu modelo são o nível, a inclinação e a curvatura da curva de rendimentos.

Mais tarde, Diebold e Li (2006) propuseram a utilização da função paramétrica de Nelson-Siegel para decompor a evolução de toda a *yield curve* mediante um vetor de parâmetros com dinâmica representada por processos estocásticos. Os autores introduzem um procedimento de estimação em dois passos: na primeira etapa, as equações de medida são tratadas como um modelo *cross-section* e é utilizado o método de mínimos quadrados para estimar os parâmetros para cada período separadamente. No segundo passo, as séries temporais são especificadas e ajustadas para os fatores.

Outros autores, como De Pooter (2007), De Rezende e Ferreira (2011) ou Yu e Zivot (2011), estudaram a capacidade de previsão de formas funcionais alternativas à Nelson & Siegel, como sejam o modelo de Svensson (1994) ou a especificação proposta por Bjork & Christensen (1999).

Contrariamente a alguns dos modelos determinísticos analisados anteriormente, para os quais a natureza estocástica das taxas de juro era ou arbitrariamente pré-determinada, ou condicionada pela especificação paramétrica adotada para a ETTJ, não garantindo a consistência com as condições de equilíbrio nos mercados, nos modelos de equilíbrio a incerteza subjacente à evolução futura das taxas de juro exige que o seu comportamento seja modelizado em termos de uma variável aleatória.

Os modelos de equilíbrio para a ETTJ assumem que os valores das obrigações isentas de risco de incumprimento e de várias maturidades podem ser explicados convenientemente pelos processos estocásticos seguidos, ao longo do tempo, por um conjunto reduzido de variáveis. A especificação do processo estocástico pode ser feita de forma exógena ou determinada endogenamente a partir de modelos de equilíbrio geral para a ETTJ, assumindo, neste último caso, hipóteses sobre as preferências dos investidores. Entre os principais expoentes desta categoria de modelos encontram-se as abordagens de Vasicek (1977), Cox, Ingersoll e Ross (1985) ou Duffie e Kan (1996).

A especificação convencional dos modelos de equilíbrio apenas define um número de pontos na ETTJ igual ao número de variáveis usadas. Deste modo, e porque o objetivo é caracterizar toda a ETTJ, os preços de obrigações com maturidade diferente devem satisfazer um conjunto de relações de consistência interna de forma a eliminar oportunidades de arbitragem no mercado. Definidas estas relações, o cálculo dos preços de equilíbrio é conseguido a partir da resolução de uma determinada equação diferencial parcial (Bravo, 2001).

Cox, Ingersoll e Ross (1985) (CIR) desenvolveram um modelo de equilíbrio geral no qual a ETTJ, a sua dinâmica, e a forma do preço de mercado são determinados endogenamente. Neste modelo, variáveis como as expectativas, a aversão ao risco, as alternativas de investimento e as preferências dos consumidores relativamente ao momento de efetuar o consumo assumem um papel preponderante na determinação do preço das obrigações. O modelo de CIR relaciona a variabilidade da taxa de juro com o seu próprio nível permitindo, deste modo, que a volatilidade exiba um comportamento heterocedástico.

O principal inconveniente dos modelos de equilíbrio diz respeito à sua incapacidade de refletir, de forma adequada e automática, a estrutura de taxas observada em cada momento nos mercados. Nestes modelos a ETTJ é não um *input*, mas um *output*. Os modelos de não-arbitragem são desenhados, pelo contrário, de modo a que a dinâmica de toda a ETTJ seja imediatamente consistente com dados de mercado para as taxas de juro.

Os modelos de Ho e Lee (1986) e Hull e White (1990) tentam aumentar os graus de liberdade do processo estocástico seguido pela taxa de juro de curto prazo admitindo que os parâmetros são função do tempo. Outros modelos usados para assegurar a consistência são os modelos de Heath, Jarrow e Morton (1990; 1992) (HJM), que consideram um vetor de variáveis de estado compreendendo toda a ETTJ, impondo em seguida restrições à tendência do processo estocástico por forma a garantir que o modelo não oferece oportunidades de arbitragem.

Os modelos de não-arbitragem são bastante utilizados para na avaliação de derivados financeiros de taxa de juro, dos mais simples aos mais complexos. Contudo, por terem uma natureza *cross-section*, ou seja, limitada ao ponto no tempo em que os dados são analisados e sem considerar a dinâmica observada na evolução da estrutura a termo, não podem ser utilizados para previsão da curva de juros.

2.2. MODELOS DE PREVISÃO PARAMÉTRICOS

Nesta secção descrevemos, de forma sumária, os principais modelos paramétricos de previsão usados em estudos de avaliação da capacidade preditiva da *yield curve*, em particular os modelos univariados ARIMA e os modelos vetoriais autorregressivos (VAR).

Começando pelos modelos univariados, Diebold & Li (2006) descrevem os modelos AR(1) (autorregressivos univariados) como sendo os modelos mais simples e que podem ser usados como ponto de referência para as previsões da *yield curve*.

Diz-se que são modelos univariados porque não admitem a possibilidade de existirem relações entre a variável de interesse e outras variáveis económicas. Assim, o comportamento da *yield curve* é explicado apenas com base nos valores anteriores da série temporal. Portanto, num processo autorregressivo, o valor presente da série temporal Y_t é expresso linearmente em termos dos valores passados da série e da perturbação aleatória ε_t registada no instante t . A ordem deste processo depende do valor mais antigo sobre o qual a regressão tem lugar. Num processo autorregressivo de ordem p , o modelo pode escrever-se como:³

$$Y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad [8]$$

onde:

- Y_t : é a função a prever e que depende dos seus desfasamentos y_{t-p} ;
- $\phi_1, \phi_2, \dots, \phi_p$ são constantes reais (parâmetros) a estimar;
- ε_t : termo aleatório *white noise* (ruído branco).

Neste estudo não são usados exclusivamente modelos AR, mas sim modelos ARIMA (Modelos autorregressivos integrados e de média móvel), sendo estes uma generalização dos modelos ARMA, que incorporam modelos AR e modelos de média móvel (MA).

Num processo de média móvel, o valor presente da série expressa-se em função dos valores presente e passados das perturbações aleatórias, que formam uma série de ruído branco. A ordem deste processo depende do valor mais antigo da série de ruído branco considerado. Para um processo de média móvel de ordem q , a série exprime-se como:

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad [9]$$

onde:

- Y_t : é a função a prever e que depende dos seus desfasamentos y_{t-p} ;
- $\theta_1, \theta_2, \dots, \theta_q$ são os parâmetros a estimar.

Isto é, um modelo de média móvel constrói-se como uma regressão linear do valor presente da série sobre as perturbações aleatórias de um ou mais valores anteriores da série. Admite-se que estas perturbações são geradas por uma mesma distribuição, habitualmente normal, de média e desvio padrão constantes. Este modelo difere do AR(p) na medida em que cada perturbação aleatória se propaga para os valores futuros da série (Makridakis, 1997).

³ Para uma descrição mais detalhada dos modelos ARIMA veja-se, por exemplo, Hamilton (1994).

Box e Jenkins popularizaram uma técnica que combina as características dos modelos autorregressivos AR e de média móvel MA. Embora estes modelos fossem já conhecidos e investigados, a contribuição de Box e Jenkins consistiu no desenvolvimento de um processo eficaz e sistemático de identificação e estimação de modelos que pudessem incorporar em simultâneo as ambas características (Box & Jenkins, 1994).

Assim, os modelos ARMA de ordem (p,q) representam-se por:

$$Y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad [10]$$

O modelo ARIMA é uma generalização do modelo autorregressivo de média móvel (ARMA). A representação ARIMA (p,d,q) refere-se, respetivamente, às ordens de auto-regressão, de integração e de média móvel, i.e., p denota o número de termos autorregressivos, d representa é o número de diferenças não sazonais necessário para obter a estacionaridade da série, e q simboliza o número de termos da média móvel. No caso particular em que d = 0, temos o modelo ARMA(p, q), e no caso particular de também q ser igual a zero, temos os modelo AR(p).

Os modelos ARIMA são ajustados a séries temporais quer para compreender melhor os dados quer para realizar previsões. Ao contrário dos modelos ARMA, estes modelos podem ser usados para prever séries temporais não estacionárias, uma vez que permitem incorporar uma componente de séries integradas, ou seja, se a série for não estacionária mas for transformada numa série estacionária após n diferenciações (Zhang, 2003). A designação “integrado” explica-se pelo facto de se reconstruir a série original a partir da série diferenciada, por uma operação de integração ou soma recursiva.

De um modo sucinto, uma vez que numa série temporal os dados são registados sequencialmente ao longo do tempo, espera-se que esta apresente autocorrelação (*serial correlation*) no tempo. Assim, a metodologia de Box-Jenkins visa captar o comportamento da correlação parcial da série ou autocorrelação entre os valores observados da série temporal, para posteriormente realizar previsões.

Os modelos vetoriais autorregressivos (VAR) surgiram nos anos 80 como uma generalização dos modelos AR(1), e têm em consideração a existência de relações de interdependência entre as variáveis e permitem avaliar o impacto dinâmico das perturbações aleatórias sobre o sistema de variáveis (Sims, 1980). Este tipo de modelos são particularmente úteis e eficientes na previsão do comportamento futuro de séries temporais interrelacionadas. Nestes modelos todas as variáveis em estudo são endógenas e são modelos de regressão definidos por sistemas de equações que podem ser vistos como uma combinação de séries temporais de sistemas de equações simultâneas (Caiado, 2002).

O modelo VAR trata todas as relações lineares existentes entre as variáveis endógenas e os valores passados das mesmas sem impor restrições quanto à sua independência e dependência, permitindo ainda a inclusão de variáveis exógenas ao modelo. É um modelo mais flexível que os modelos univariados, que possibilita testar a direção de causalidade e previsões de melhor qualidade do que nos modelos estruturais tradicionais (Sims, 1980).

Assim, um modelo VAR (p) pode ser descrito genericamente como:

$$Y_t = A_0 + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t \quad [11]$$

Onde

- Y_t é um vetor coluna $Y_t = (Y_1, \dots, Y_k)'$ com k variáveis;
- A_0 é um vetor de termos independentes;
- A_i é uma matriz de coeficientes (k x k) com $i=1, \dots, p$;
- u_t é um vetor de perturbações aleatórias não correlacionadas com os seus próprios valores passados e não correlacionadas com nenhuma das variáveis do segundo membro, embora possam estar contemporaneamente correlacionadas, e com matriz de covariâncias não singular.

A abordagem VAR tem em consideração a existência de relações de interdependência entre as variáveis e permitem avaliar o impacto dinâmico das perturbações aleatórias sobre o sistema de variáveis, o que torna estes modelos particularmente úteis e eficientes na previsão do comportamento futuro de séries temporais interrelacionadas (Vereda et al. 2014).

Ao tratarem todas as variáveis simetricamente sem impor qualquer restrição quanto à independência e dependência entre elas, estes modelos permitem ainda descrever cada uma das variáveis endógenas no sistema como uma função dos valores desfasados de todas as variáveis endógenas (Ramos, 2011).

No entanto, apesar das vantagens reconhecidas nos modelos VAR, estes evidenciam algumas limitações. Além de outros inconvenientes, os modelos VAR requerem que todas as componentes do modelo sejam estacionárias. Caso tal não se verifique, é necessário adotar procedimentos alternativos com vista à correção do erro – MCE, surgindo assim os modelos VECM (*Vetor Error Correction Model*), nos quais é dado especial ênfase às relações de longo prazo. Com efeito, os modelos VECM distinguem-se dos modelos VAR pelo facto de apresentarem como variáveis determinísticas os MCE (sob a forma de vetores) correspondentes às relações de cointegração entre variáveis (Félix e Nunes, 2003).

2.3. MODELOS DE APRENDIZAGEM AUTOMÁTICA

Os métodos de *machine learning*, ou de aprendizagem automática, nos quais se incluem as redes neuronais artificiais, são ferramentas que possuem a capacidade de se adaptar e de aprender a partir de um conjunto de exemplos de dados, à semelhança do que acontece com os neurónios no cérebro humano. Podem ser usados para encontrar padrões em conjuntos de dados, classificar objetos numa determinada classe ou em tarefas de previsão quantitativa (*regression tasks*) (Haykin, 1999).

Assim como acontece nos neurónios no cérebro humano também as RNA utilizam a informação passada para conseguirem dar resposta às novas situações. Como tal, podem ser descritas na seguinte função:

$$\hat{\beta} = f(\beta_{t-1}, \beta_{t-2}, \dots, \beta_{t-p}, W) + \varepsilon_t \quad [12]$$

onde $f(\cdot)$ é a função que determina a estrutura e conexão entre as diferentes camadas e os respetivos pesos (W) adicionando a componente de erro existente em todos os modelos de previsão (Zhang et al. , 1998).

O objetivo das RNA consiste na compreensão e conseguinte aplicação das bases do sistema biológico, de forma a otimizar a resolução de problemas complexos (Basheer & Hajmeer, 2000). As redes neuronais oferecem boas soluções para este tipo de problemas, uma vez que conseguem explorar estruturas complexas e encontrar interações lineares e não lineares e, ainda, descobrir padrões. Atualmente as RNA são mais utilizadas em problemas de classificação, análise de séries temporais, predição e *clustering*, entre outros (Cui et al., 2006).

Falar em algoritmos de aprendizagem automática conduz, normalmente, à referência de dois paradigmas: aprendizagem supervisionada e não supervisionada (Stimpson & Cummings, 2014). Na aprendizagem supervisionada são apresentados dois conjuntos de dados, o conjunto de *input* e o conjunto de *output* esperado (Winandy, Borges Filho, & Bento, 2007). Na aprendizagem não supervisionada apenas é apresentado um conjunto de *input* (Alpaydin, 2004).

A aprendizagem supervisionada está diretamente relacionada com a previsão enquanto a aprendizagem não supervisionada se relaciona mais com a descoberta de padrões num conjunto de dados (Stimpson & Cummings, 2014). No presente estudo utilizar-se-á uma aprendizagem supervisionada.

Existem diversos tipos de redes neuronais. Contudo, no presente estudo serão aplicadas redes neuronais de multicamada. A arquitetura destas redes está organizada em 3 camadas: a camada de *input*, "*input layer*", (constituída por neurónios de *input*, com tantos neurónios quantas as variáveis de *input* do modelo), uma ou várias camadas intermédias, "*hidden layer*", cada um delas com um certo número de neurónios intermédios e uma camada de *output* que pode ter um ou mais neurónios consoante o número de *targets* que se pretende prever com o modelo (Kaastra & Boyd, 1996).

Do ponto de vista estrutural, as RNA podem ser classificadas como redes neuronais estáticas ou como redes neuronais dinâmicas (Haykin, 1999). As redes estáticas são caracterizadas por não apresentarem elementos de realimentação e não apresentarem desfasamentos (*delays*). Dessa

forma, a resposta (*output* ou previsão) da rede é calculada diretamente através das entradas apresentadas, *input*, através de conexões alimentadas adiante (*feed forward*). Já nas redes dinâmicas, a resposta pode não depender apenas da entrada corrente da rede, mas sim dos valores atuais e anteriores de *input* e *output* e do próprio estado da rede. As redes dinâmicas podem ser divididas em duas categorias, tal como afirmam Tissot, Camargo e Pozo (2012):

- *Feed-forward*: são estruturas de redes neuronais cujas conexões entre neurónios de diferentes camadas obedecem à direção *input-output*, não havendo assim conexões entre neurónios pertencentes à mesma camada nem às camadas anteriores.
- *Feed-back*: cada camada pode conter conexões entre os neurónios da mesma camada, de camadas anteriores ou camadas posteriores. Não existe portanto um sentido único para o fluxo de dados entre neurónios ou entre camadas.

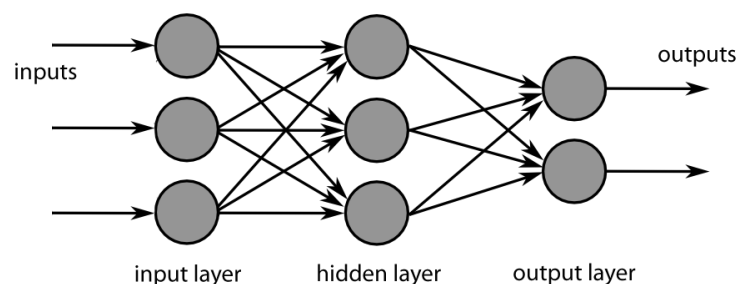


Figura 2: Representação Rede Neuronal *Feed-Forward*
Fonte: (Täppinen, 1998)

As RNA usadas nesta dissertação são chamadas de *feed-forward*, o que significa que a informação apenas é direcionada da camada de *input* e através da *hidden layer* chegará ao *output*. Não existem ciclos nem *loops* neste tipo de estruturas de redes neuronais.

O “conhecimento” de cada rede é armazenado nas suas conexões através dos pesos, que mostram quão relevantes são as conexões no processo de aprendizagem de uma RNA. Uma rede treinada com um número baixo de neurónios (poucas conexões) pode não aproveitar o máximo de seu potencial. No entanto, uma rede treinada com um número excessivo de neurónios (muitas conexões) pode adaptar-se aos ruídos e prejudicar a sua capacidade de generalização, dando origem a problemas de *overfitting* do modelo aos dados de treino (Arantes et al., 2014). Assim, os autores, quando preveem aa ETTJ brasileira usando redes neuronais, alertam para a necessidade de uma correta configuração das redes neuronais a utilizar aquando das previsões.

Do ponto de vista econométrico, uma das grandes vantagens das RNA sobre os modelos tradicionais está no facto das redes neuronais serem capazes de aproximar qualquer função sem haver necessidade de especificar o modelo e serem capazes de aprender relações complexas (não lineares) mesmo com poucos dados. Por conta da primeira característica, esta é dita uma técnica não paramétrica (Moshiri & Cameron, 2000).

Allende, Moraga e Salas (2002) realizaram um estudo comparativo entre a capacidade preditiva das RNA e dos modelos estatísticos habitualmente usados, como é o caso dos modelos ARIMA. Os autores concluem que as RNA têm ganhado cada vez mais importância como ferramenta de análise de dados e previsão de séries temporais e apresentam como uma das vantagens das RNA o facto de

ser uma metodologia de aproximação independente do modelo formado pelos dados, não sendo necessário efetuar nenhuma análise prévia aos dados nem ter em atenção os pressupostos dos modelos. Assim, a abordagem RNA é sugerida como sendo um método competitivo e fiável para a análise e previsão de modelos, não só pelo facto de não ter de seguir as especificações de um modelo mas também pela qualidade das previsões feitas (Allende et al., 2002).

No entanto, a utilização de RNA também apresenta algumas desvantagens, sendo que aquela que se destaca é o facto de as RNA poderem convergir para um ponto ótimo local, podendo não chegar a encontrar o ótimo global, o que fará com que a solução encontrada não seja a que minimiza o erro das previsões (distância entre as previsões e os valores observados). Para ultrapassar esta desvantagem alguns autores mencionam outras formas de treino dos dados, como por exemplo McNelis (2005), que sugere a utilização de Algoritmos Genéticos ou do *Simulated Annealing*.

Os Algoritmos Genéticos partem de soluções geradas aleatoriamente e tentam encontrar soluções com uma qualidade melhor e de forma iterativa. O *Simulated Annealing* tem a capacidade de continuar a procura por uma solução melhor em vez de terminar a pesquisa assim que é encontrada uma solução, de acordo com os critérios especificados no início da resolução do problema. No entanto, estes métodos também não garantem que a solução encontrada seja um ótimo global. Não existe uma solução para eliminar este problema, apenas existem estratégias para o minimizar.

A utilização de redes neuronais para previsão da taxa de juro será uma mais valia na medida em que as redes neuronais, além de outras propriedades, exibem segundo Warner e Misra (1996), a capacidade de adaptação e aprendizagem a novas situações, i.e., os modelos de redes neuronais podem lidar com dados imprecisos e situações não totalmente definidas. Por outro lado, estes modelos exibem capacidade de generalização, i.e., após a fase de aprendizagem são capazes de generalizar e dar resposta a situações que não estão presentes nos dados já conhecidos usados na aprendizagem.

2.4. ESTUDOS ANTERIORES SOBRE A TEMÁTICA DA PREVISÃO DA ETTJ

Nesta secção sintetizam-se os principais resultados de estudos realizados sobre a mesma temática desta dissertação.

Aljinović e Poklepović (2013) seguem uma abordagem semelhante à realizada neste estudo e utilizam dados da *yield curve* na Croácia para o período entre 2011 e 2012. Com base nesta informação, desenvolveram uma rede neuronal e o modelo de NS para a estimação e previsão da *yield curve*. O estudo concluiu que as redes neuronais apresentam previsões mais precisas para a *yield curve* em comparação com os métodos de estimação habituais.

Tabak et al. (2012) comparam a capacidade preditiva da *yield curve* na Zona Euro usando o modelo Functional Signal plus Noise com correção ao erro (FSN-ECM) desenvolvido por Bowsher and Meeks (2008), o modelo dinâmico de Nelson–Siegel apresentado por Diebold and Li (2006) (DN) e o modelo *Random walk*. Os autores utilizam dados de janeiro de 1999 a abril de 2011, o que significa que a sua amostra engloba os anos da crise internacional (2007 e 2008) onde as *yield curves* apresentam valores bastante diferentes dos valores observados habitualmente. O horizonte temporal escolhido foi de 1, 3 e 6 meses. Os resultados do estudo indicam que o modelo FSN e o modelo DN prevê melhor do que o modelo *Random walk* apesar de a diferença ser muito ténue. Comparando os modelos FNS e o modelo DN, concluem que o modelo FSN faz previsões mais precisas a curto prazo.

Luo, Han, e Zhang (2012) elaboraram o primeiro estudo de previsão para a ETTJ na China, usando os seguintes modelos para a modelação da *yield curve*: Nelson–Siegel (1987), Björk and Christensen (1999) (DBC), Svensson (1994) e Diebold and Li (2006). Para a previsão da *yield curve* testaram os seguintes métodos: AR(1) linear e não linear, VAR(1), GARCH linear e não linear. Os autores concluíram que diferentes especificações devem ser usadas com diferentes métodos de previsão. O modelo DL deverá ser utilizado com o processo AR(1) não linear, o modelo DBC com o processo VAR e o DS com o AR(1) linear. Verificou-se ainda que os modelos GARCH linear e não linear não são muito bem sucedidos na previsão da *yield curve*.

Täppinen (1998) estudou o potencial da utilização de redes neuronais na previsão da taxa de juro para os EUA comparando os resultados obtidos com os gerados mediante modelos de regressão linear. O autor salientou que, apesar de as RNA apresentarem de um modo geral previsões mais robustas, um dos problemas apresentados pela utilização de redes neuronais é o facto de estas serem sensíveis à arquitetura de estimação selecionada, ou seja, a utilização de diferentes parâmetros na mesma rede provoca resultados muito distintos. Assim, dependendo dos parâmetros usados para inicializar a rede as previsões poderão ser mais ou menos robustas.

Arantes et al. (2014) realizaram um estudo para a previsão da ETTJ Brasileira usando redes neuronais artificiais, uma rede estática e uma rede dinâmica. No final, a performance dos métodos de aprendizagem automática é comparada com os modelos de Diebold & Li (2006), com o modelo de Svensson (1994), o modelo autorregressivo, o modelo vetorial autorregressivo e o *Random walk*. Os autores concluem que o conjunto de modelos de redes neuronais com melhor capacidade de previsão está diretamente relacionado com o horizonte temporal de previsão escolhido: os modelos pertencentes à classe estática são melhores 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. Os autores ressaltam que os modelos de redes neuronais artificiais serão uma

alternativa promissora para previsão da ETTJ, uma vez que apenas os modelos estimados através de RNA conseguiram superar as previsões realizadas pelo *Random walk* em todos os horizontes de previsão. Analisando os horizontes de previsão a mais curto prazo, os modelos autorregressivos (AR(1) e VAR(1)) apresentaram um desempenho médio muito próximo dos gerados pelas previsões do *Random walk*, situação que não se mantém para os horizontes de previsão a mais longo prazo.

3. METODOLOGIA

O seguinte capítulo está dividido em 4 partes: inicialmente é feita uma descrição dos dados utilizados, bem como uma análise dos perfis de volatilidade e correlação para os choques na taxa de juro. Em complemento, é estudada a dinâmica dos choques nas taxas de juro e ainda uma decomposição dos choques de taxas de juro. Posteriormente, descrevem-se os métodos de previsão usados e toda a metodologia seguida. Por último, apresentam-se os critérios de desempenho dos modelos.

3.1. DADOS

O BCE publica diariamente as taxas de juro estimadas para a Zona Euro para ativos com determinada classe de risco de crédito. A *yield* é estimada com base no modelo de Svensson. São estes os dados utilizados nesta dissertação para a previsão da ETTJ.

Os dados recolhidos têm como frequência os dias úteis de negociação, ou seja, todos os feriados e fins de semana foram excluídos da análise. Estes dados dizem respeito à *yield curve* das obrigações com notação de *rating* AAA, ou seja, obrigações com o menor risco de incumprimento possível, com maturidades entre os 3 meses e os 30 anos e são referente ao período entre 6 de setembro de 2004 e 25 de maio de 2015.

À semelhança de Diebold e Li (2006), são usados dados mensais na análise. Na estimação da *yield curve* pelo BCE, apenas as obrigações que respeitam os seguintes critérios são recolhidas:

- Obrigações em euros e referentes aos países da Zona Euro;
- Obrigações com montantes de pelo menos 5 biliões €;
- Obrigações de cupão fixo com uma maturidade finita e obrigações de cupão-zero, incluindo STRIPS (*Separate Trading of Registered Interest and Principal Securities*);
- Obrigações negociadas pelo governo central com um máximo *bid-ask spread per quote* de três pontos-base. Os preços/*yields* são os apresentados no fecho do mercado na data de referência.

Às obrigações selecionadas é ainda aplicado um mecanismo para remoção dos *outliers*: as obrigações são removidas se se desviarem mais do que o dobro do desvio-padrão da média da *yield* para uma mesma maturidade.

Após a recolha de dados, é feita uma análise de estatísticas descritivas, onde são apresentadas de seguida uma análise gráfica das taxas de juro; uma análise de correlação dos parâmetros e uma análise das componentes principais à taxa de juro, com o intuito de obter alguma intuição sobre o comportamento das taxas de juro neste período e de justificar a utilização dos modelos assentes em fatores latentes.

Na figura 3 é representada a evolução diária das taxas de juro desde 6 de Setembro 2004 até 25 de maio de 2015 na Zona Euro, para as diferentes maturidades selecionadas (3 meses a 30 anos):

Evolução da Yield Curve na Zona Euro

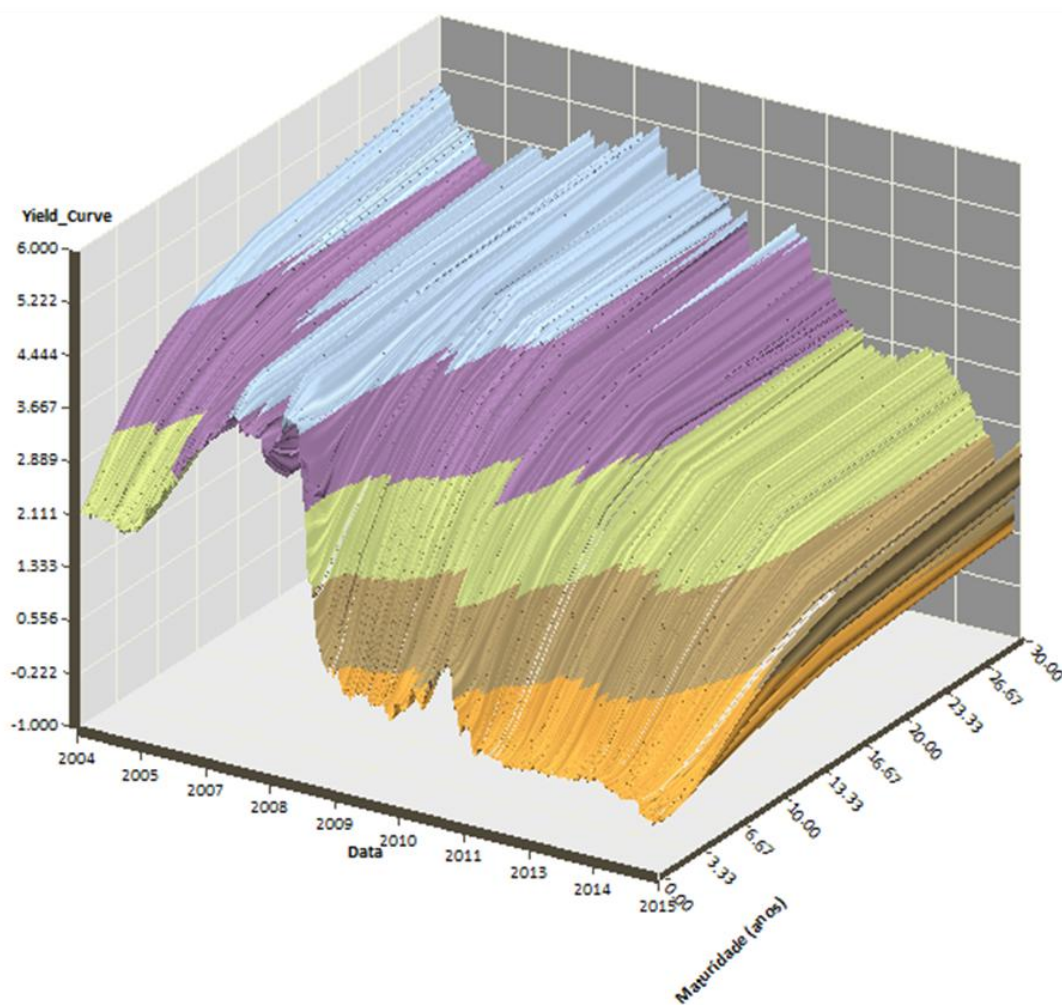


Figura 3: Evolução das Taxas de Juro na Zona Euro entre 2004 e 2015

Fonte: Elaboração própria

Nota: O gráfico foi obtido com os valores observados da *spot rate* entre dia 6 de setembro de 2004 e o dia 25 de maio de 2015. As maturidades estão expressas em anos.

Conforme se observa através do gráfico, as *spot rate* apresentam algumas alterações ao longo do tempo, sendo que em 2008 foi o ano em que registaram valores mais baixos e a partir dessa data apresentam valores negativos. Tal é explicado, em parte, pela resposta do BCE aos sucessivos eventos que ocorreram após agosto de 2007.

Como referido, existiram dois contextos de crise bastante diferentes: o primeiro com origem nos mercados financeiros dos EUA e na sequência da falência do banco Lehman Brothers e da crise de *subprime*, e o segundo na sequência da instabilidade dos mercados de obrigações soberanas de alguns países da Zona Euro.

Com efeito, em resposta à denominada crise do mercado imobiliário *subprime*, o BCE começou por tomar um conjunto de medidas denominadas de *credit enhanced support*, destinadas a tranquilizar os mercados em relação aos riscos de liquidez do sector financeiro apoiando os bancos da Zona Euro, num contexto em que os efeitos da crise económica e financeira já se faziam sentir de forma mais

intensa. Esta estratégia não inviabilizou que o BCE mantivesse a orientação da política monetária no sentido do cumprimento do objetivo de estabilidade de preços, enquanto as medidas de gestão de liquidez se destinaram a suportar o regular funcionamento dos mercados monetários.

A utilização de Operações de Mercado Aberto e das Facilidades Permanentes de Liquidez aumentou a oferta de liquidez na parte inicial do período de manutenção e reduziu na parte final, e a descida continuada das taxas de juro de referência do BCE durante a crise tiveram reflexos naturais no comportamento da *yield curve* da Zona Euro neste período.

No início de 2010, em resposta à denominada crise das dívidas soberanas, e perante a súbita intensificação das tensões nos mercados financeiros com o aparecimento de pressões nos mercados de dívida de longo prazo, com os *spreads* e as volatilidades a apresentarem valores recorde e com a deterioração do funcionamento do mercado monetário, o BCE reintroduziu uma série de medidas especiais de apoio ao sistema financeiro.

No contexto da crise das dívidas soberanas e em resultado das consecutivas descidas de notação de *rating* das obrigações soberanas de alguns países, o BCE avançou com a suspensão do limiar mínimo da notação de crédito nos requisitos de elegibilidade, nomeadamente para os instrumentos de dívida transacionáveis emitidos ou garantidos pelo governo grego (maio de 2010), pelo governo irlandês (março de 2011) e pelo governo português (julho de 2011). Em outubro de 2011, o Conselho do BCE lançou um novo programa de aquisição de *covered bonds* de forma a injetar liquidez na economia e a pressionar a descida das taxas de juro.

Como referem Estrella & Trubin (2006), a política monetária e as expectativas dos investidores influenciam a inclinação da *yield curve*. Habitualmente uma política monetária mais rígida leva a uma subida das taxas de juro a curto-prazo, o que conseqüentemente levará a uma redução da inflação. Quando essas pressões diminuem espera-se que os governos pratiquem uma política mais flexível, com taxas de juro menores (Cieslak & Povala, 2013).

Ao analisar o gráfico da série temporal é possível observar que existe uma quebra estrutural na série e que os valores observados em 2008 conferiam uma grande volatilidade à estrutura temporal das taxas de juro, apresentados valores muito atípicos que são justificáveis devido à falência do Lehman Brothers e ao início da crise financeira internacional que obrigou à intervenção do BCE nos mercados financeiros.

Tal como estudaram Bech & Lengwiler (2012), podem observar-se diversas fases nos gráficos das *yield curves*: após uma fase "normal" que termina em meados de 2004, observa-se um período que é caracterizado por uma notável ausência de volatilidade na *yield*. Esta fase de "moderação" termina com o início da crise financeira em agosto de 2007. A primeira parte da crise, que pode ser vista como uma "crise de liquidez" caracteriza-se por problemas de turbulência do mercado monetário e de liquidez. Deste modo, observa-se enorme volatilidade no espectro de maturidades de curto e médio da *yield curve*. Este padrão altera-se abruptamente em setembro de 2008 com a falência da Lehman Brothers onde se passa a observar-se uma grande volatilidade nas maturidades a longo prazo.

Assim, para obter modelos com melhor capacidade de previsão em face desta alteração estrutural nos mercados de taxa de juro, optou-se por restringir o período amostral para a construção do

modelo e foram excluídas do período de estimação dos modelos todas as observações anteriores a junho de 2009.

Existem modelos paramétricos capazes de prever séries temporais que apresentam este tipo de volatilidade como é o caso do Modelo GARCH (*General Autoregressive Conditional Heteroscedasticity*) (Bollerslev, 1986). No entanto, o objetivo deste estudo prende-se apenas com a comparação dos modelos ARIMA, VAR e RNA.

Uma vez que a amostra de dados recolhida não era significativamente elevada, na presença de *missing values* serão utilizados os dados da observação imediatamente anterior.

3.1.1. Perfis de Volatilidade e de Correlação para os Choques de Taxas de Juro

Para analisar a dinâmica das taxas de juro observada neste período foi elaborada uma análise às correlações entre as variações das taxas de juro para as diferentes maturidades. A Tabela 1 representa os coeficientes de correlação linear entre as variações mensais nas estimativas diretas de taxa de juro *spot* obtidas a partir da especificação de Nelson-Siegel & Svensson.

Tabela 1: Matriz de Correlação entre as variações da Taxa de Juro

	0.25	0.5	1	2	3	5	7	10
0.25	1							
0.5	0.919394	1						
1	0.826751	0.937664	1					
2	0.684909	0.782235	0.942533	1				
3	0.58476	0.701302	0.880594	0.979519	1			
5	0.414177	0.555296	0.72418	0.85209	0.934969	1		
7	0.289225	0.422333	0.570627	0.703351	0.814632	0.962707	1	
10	0.187321	0.293508	0.419428	0.545407	0.661684	0.854943	0.958049	1

Nota: Os valores indicados no quadro correspondem aos coeficientes de correlação linear entre as variações mensais nas estimativas diretas de taxa de juro spot obtidas a partir da especificação de Nelson-Siegel & Svensson. Ao todo foram consideradas 2744 observações diárias respeitantes ao período setembro 2004 a maio 2015. Todos os prazos estão expressos em anos.

A leitura e interpretação destes valores permite-nos deduzir que:

- O facto de os coeficientes serem elevados sugere a existência de fatores comuns na explicação dos choques de taxa de juro;
- O padrão de correlação imperfeito mostra que a influência desses fatores não é uniforme ao longo de toda a ETTJ, pelo que modelos multifator podem ser necessários para explicar a dinâmica das taxas de juro;
- Existe uma correlação bastante forte em segmentos próximos da ETTJ – por exemplo, para as taxas de 2 e 3 anos, relativamente com prazos de 3 meses e 10 anos. Este padrão de correlação é natural, uma vez que é de esperar que uma obrigação com 3 meses para o vencimento apresente um comportamento bastante mais próximo de uma obrigação com 6 meses de vida do que de uma emissão com 10 anos de vida.

3.1.2. Dinâmica dos fatores latentes da função de Nelson-Siegel & Svensson

Por forma a desenvolver alguma intuição sobre o tipo de fatores que comandam a dinâmica dos fatores latentes da função de Nelson-Siegel & Svensson, efetuamos nesta secção uma análise das estatísticas descritivas que resumiam o comportamento dos parâmetros que descrevem as taxas de juro.

Tabela 2: Principais Indicadores Estatísticos

	β_0	β_1	β_2	β_3	τ_1	τ_2
Média	3.0108	-1.6577	10.0768	-8.50889	4.4236	2.8759
Mediana	3.3113	-1.6644	7.9477	-4.10417	4.2642	2.6097
Moda	0.01	-2.6577	0.0773	-1.20535	0.25	2.9215
Desvio-padrão	1.5872	1.0329	9.8690	8.4693	3.2172	1.5054
Variância da amostra	2.5192	1.0669	97.3980	71.7289	10.3501	2.2663
Mínimo	0.01	-5.0352	-2.6984	-36.4651	0.25	0.25
Máximo	5.5596	2.3234	36.8629	-0.09215	14.5967	9.3758

Tabela 3: Tendências Estatísticas

		Percentagem do Tempo												
		Valor			$\beta_1 > 0$		$\beta_1 < 0$		$\beta_1 > 0$			$\beta_1 < 0$		
		Positivo	Negativo	Nulo	$\beta_2 > 0$	$\beta_2 < 0$	$\beta_2 > 0$	$\beta_2 < 0$	$\beta_3 > 0$	$\beta_3 < 0$	$\beta_3 = 0$	$\beta_3 > 0$	$\beta_3 < 0$	$\beta_3 = 0$
β_1		5.6	94.4	0.0	5.6	0.0	86.0	8.4	0.0	5.6	0.0	0.0	94.4	0.0
β_2		91.6	8.4	0.0										
β_3		0.0	100.0	0.0										

Nota: Os indicadores estatísticos reportados foram calculados com base numa amostra total de 2744 observações diárias, compreendidas entre setembro 2004 e maio 2015. Todos os valores estão em percentagem exceto os parâmetros τ_1 e τ_2 que são medidos em anos.

De acordo com a informação apresentada nas tabelas acima, é possível verificar que em 94,4% das observações o coeficiente β_1 é negativo, ou seja, na maioria dos casos a ETTJ apresenta uma inclinação positiva e o coeficiente β_2 é positivo 91,6% das observações, ou seja, a ETTJ assume uma forma côncava. O cruzamento destes dados permite-nos deduzir que a ETTJ exhibe predominantemente (em 86% dos casos) uma tendência crescente e uma configuração côncava. O coeficiente β_3 apresenta sempre valores negativos.

A seguinte tabela representa a correlação histórica dos principais tipos de choque de taxa de juro, onde se pode verificar que existe uma correlação negativa entre o nível de taxas de juro e os indicadores de inclinação e curvatura da ETTJ.

Tabela 4: Matriz de Correlação Entre os Choques sobre o Nível, Inclinação e Curvatura da ETTJ

	BETA 0	BETA 1	BETA 2	BETA 3
BETA 0	1			
BETA 1	-0.924	1		
BETA 2	-0.569	0.489	1	
BETA 3	-0.095	0.101	-0.665	1

Nota: Os valores indicados na tabela correspondem aos coeficientes de correlação linear entre as variações mensais nas estimativas dos parâmetros β_0 (nível), β_1 (inclinação), β_2 e β_3 (curvatura) da equação Nelson-Siegel & Svensson, foram consideradas 2744 observações referentes ao período setembro 2004 e maio 2015.

3.1.3. Decomposição dos Choques de Taxa de Juro

Para verificar qual o número de fatores relevantes sobre a sua importância relativa e sobre a forma como cada um deles contribui para a explicação dos movimentos nas taxas para os diferentes prazos, foi feita uma Análise de Componentes Principais (ACP). O objetivo da ACP é o de explicar o comportamento de um conjunto alargado de variáveis observáveis (choques de taxa de juro) usando um número reduzido de variáveis implícitas não observáveis (fatores), independentes umas das outras. A ACP pretende reduzir a dimensionalidade do problema, com pouca perda de informação, tornando-a mais compreensível.

Os principais resultados da ACP efetuada com base em 8 maturidades representativas – 3 e 6 meses, 1, 2, 3, 5, 7 e 10 anos – são reportados na Tabela 5:

Tabela 5: Importância Relativa dos Vários Fatores para cada Prazo

	Fator 1	Fator 2	Fator 3	Fatores (1, 2, 3)
3M	50.27%	38.63%	7.57%	96.46%
6M	66.58%	27.59%	2.75%	96.92%
1A	85.28%	12.27%	0.80%	98.35%
2A	91.26%	1.23%	6.66%	99.15%
3A	93.15%	0.49%	5.94%	99.58%
5A	85.25%	13.09%	0.46%	98.80%
7A	69.61%	28.66%	1.30%	99.57%
10A	50.80%	39.89%	8.01%	98.70%
% do Total	74.02%	20.23%	4.19%	98.44%

Notas: Todos os valores se encontram em percentagem. As maturidades de referência estão expressas em meses (M) e Anos (A). A análise foi conduzida sobre a matriz de correlação apresentada na tabela 4, e para o período de setembro 2004 a maio 2015

Após a análise da tabela acima apresentada, podemos verificar que o primeiro fator é o mais importante na medida em que explica, para o conjunto dos prazos, 74% da variância total. Os fatores 2 e 3 têm menos importância que o primeiro mas contribuem, ainda, para explicar 24,42% da volatilidade total. É ainda significativo o facto dos três primeiros fatores serem responsáveis por 98,44% da variância dos choques de taxa de juro.

Para termos alguma intuição sobre a interpretação a dar aos três principais fatores, representamos no gráfico seguinte os seus coeficientes (*factor loadings*):

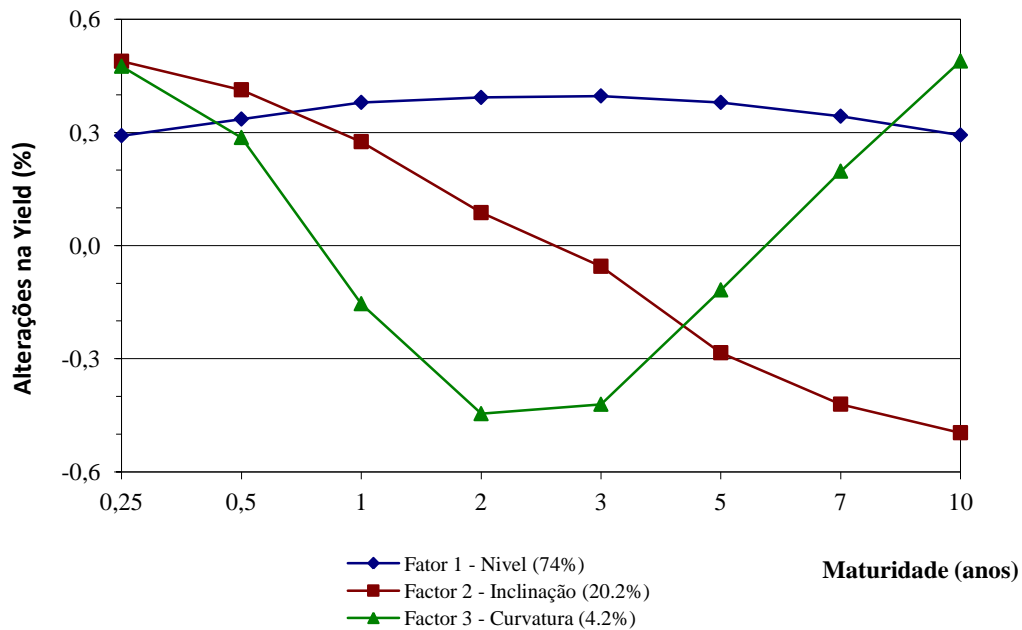


Figura 4 : Três Fatores Principais da ETTJ

Constata-se que as variações de taxas de juro induzidas por um choque sobre o primeiro fator são aproximadamente iguais para todas as maturidades. Assim, pode-se interpretar o primeiro fator como sendo fator de nível, representando essencialmente choques paralelos sobre a ETTJ.

O gráfico mostra ainda que um choque sobre o fator 2 diminui as taxas de juro para todas as maturidades, isto é, alterações neste fator induzem rotações na ETTJ. Assumimos então que este fator é responsável pelas alterações na inclinação da curva de rendimentos.

O terceiro fator exerce um efeito positivo sobre as taxas situadas nos extremos da ETTJ e um efeito negativo sobre as taxas intermédias (1 a 3 anos). Podemos associar este fator às modificações na curvatura da ETTJ.

Em síntese, os resultados proporcionados pela ACP permitem-nos concluir que a quase totalidade da volatilidade das taxas de juro é explicada por apenas três fatores: choques sobre o nível, choques sobre a inclinação, e choques sobre a curvatura da ETTJ. Reconhecer que estes três parâmetros têm esta interpretação ajuda a implementar estratégias de investimento e a obter mais retorno (Fabozzi, Martellini, & Priaulet, 2005).

3.2. PREVISÃO

Para analisar o poder preditivo dos vários modelos foi adotado uma técnica de previsão recursiva com janela de observação crescente (*recursive expanding window forecast*). Segundo esta metodologia, um período de observação inicial $t = 1, \dots, T$ é usado para estimar os modelos e obter previsões da variável de interesse para o horizonte temporal h com início em T . No período (mês) seguinte, o período de observação é incrementado numa unidade, os modelos são re-estimados, e as previsões para o horizonte temporal h com início em $T + 1$ são derivadas. Em termos esquemáticos:

$$\begin{aligned}
 & [1, \dots, t, \dots, t+h] \\
 & [1, \dots, t+1, \dots, t+h+1] \\
 & \quad \quad \quad \text{M} \\
 & [1, \dots, t+N, \dots, t+h+N]
 \end{aligned}$$

Apresenta-se um exemplo ilustrativo do funcionamento do método *recursive expanding window forecasts* utilizado nesta dissertação para o horizonte temporal de previsão a 1 mês.

Tabela 6: Exemplo Ilustrativo de aplicação do método *recursive expanding window forecasts*

Período de observação												Forecast		
dez	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	jan	fev

O método *recursive expanding window* contrasta com o método *Rolling (moving window) forecasts*. Neste último, um período de observação inicial $t = 1, \dots, T$ é usado para estimar os modelos e obter previsões da variável de interesse para o horizonte temporal h com início em T . A janela de observação é então deslocada uma observação para a frente, os modelos são re-estimados usando os dados referentes ao período $t = 2, \dots, T + 1$, e as previsões para o horizonte temporal h com início em $T + 1$ são derivadas. Em termos esquemáticos:

$$\begin{aligned}
 & [1, \dots, t, \dots, t+h] \\
 & [2, \dots, t+1, \dots, t+h+1] \\
 & \quad \quad \quad \text{M} \\
 & [N, \dots, t+N, \dots, t+h+N]
 \end{aligned}$$

A amostra utilizada para realizar as previsões compreende as observações mensais dos parâmetros da especificação de NSS e da *yield curve* entre junho de 2009 e maio de 2015. A opção pela não consideração das observações anteriores a esta data na estimação dos modelos prende-se com a verificação de uma quebra estrutural nas séries após a crise financeira de 2008, que reduzia substancialmente a capacidade preditiva do tipo de modelos em estudo.

O procedimento de obtenção de previsões para a *yield curve* desenvolve-se em duas etapas:

- Numa primeira fase, são elaboradas previsões para os quatro fatores latentes da especificação de NSS com recurso aos métodos univariados (*Random walk*, ARIMA, RNA) e multivariados (VAR, RNA) descritos anteriormente.
- Numa segunda etapa, as previsões dos parâmetros são incorporadas na especificação (7) de modo a obter previsões das taxas de juro para um conjunto selecionado de maturidades. A comparação entre os valores projetados das taxas de juro e os correspondentes valores observados permite a realização de inferência sobre desempenho preditivos dos vários modelos testados.

Dada a reduzida dimensão da amostra recolhida, as previsões foram elaboradas apenas para um horizonte temporal de 1, 2 e 3 meses, horizontes que comparam com os escolhidos em estudos similares realizados no passado. No total, foram elaboradas 12 previsões com recurso à técnica de previsão recursiva acima descrita.

As maturidades selecionadas para avaliar a capacidade preditiva dos modelos foram de 0.25, 0.5, 1, 2, 3, 5, 7, 10, 15 e 30 anos, uma vez que estas são as maturidades mais utilizadas neste tipo de estudos. Este leque de maturidades permite fazer uma análise da performance dos modelos no segmento de curto prazo da curva (3 meses a 1 ano), no segmento de médio prazo (2 anos a 7 anos) e segmento longo prazo (10 e 15 anos) e muito longo prazo (30 anos).

Refira-se ainda que, à semelhança do realizado em estudos análogos, na previsão da *yield curve* se assumiu que os parâmetros τ_1 e τ_2 se mantêm inalterados face à última observação conhecida. Assim, para estas variáveis foram considerados os valores de τ_1 e τ_2 observados no último dia do período usado para prever.

Existem várias ferramentas que auxiliam no processo de identificação e previsão dos modelos de séries temporais. Nesta dissertação foi utilizado o programa estatístico R, quer para a construção dos métodos paramétricos quer para as RNA.

3.2.1. Modelos Univariados ARIMA (p,d,q)

Antecedendo a implementação dos procedimentos práticos deste projeto, é necessário referir que nestes modelos com séries temporais admite-se que as variáveis assumem valores gerados por meio de um processo estocástico. Tendo como objetivo desenvolver e utilizar modelos capazes de prever a ETTJ, um dos pressupostos a ser validado é a estacionaridade das séries temporais, isto é, deve verificar-se que as médias e variâncias das séries temporais são constantes ao longo do tempo e que as covariâncias entre os valores desfasados da série dependem apenas do valor do desfasamento, isto é, da “distância” temporal entre elas (Ramos, 2011).

Tal como referido anteriormente e confirmado depois de uma extensa análise das séries temporais, verificou-se que para realizar previsões mais fidedignas e que se ajustem melhor aos dados recolhidos devem ser usados modelos ARIMA.

Seguindo a metodologia de Box-Jenkins, o ajuste dos modelos ARIMA à série temporal deve ser tal que a diferença entre os valores gerados pelos modelos e os valores observados resulte em séries de resíduos de comportamento aleatório em torno de zero (Shumway & Stoffer, 2011).

A implementação da metodologia de Box-Jenkins na obtenção de previsões para os fatores latentes da função de NSS obedeceu aos seguintes passos:

1. Análise gráfica das séries de parâmetros e percepção do comportamento das mesmas, tentando analisar a presença de sazonalidade ou tendência.
2. Análise dos gráficos ACF e PACF para verificar se a série é estacionária na média ou se algum dos *lags* se desvia das barreiras. Entenda-se por *lag* os níveis de desfasamento considerados.
3. Caso a série não seja considerada estacionária, tornar a série estacionária na média por diferenciação.
4. Apesar da análise gráfica ser um passo importante, é necessário confirmar a necessidade de transformar a série através de testes de raízes unitárias. Existem diversos testes para este efeito, os mais utilizados e que foram feitos nesta dissertação são: Phillips-Perron (PP) e Augmented Dickey-Fuller (ADF). O primeiro teste utiliza correções não-paramétricas baseadas nas variações a longo prazo da série temporal o que significa que não é obrigatório determinar o nível de correlação da série. O segundo utiliza uma estrutura paramétrica autorregressiva para capturar a correlação da série. O *p-value* escolhido para a realização destes testes foi de 10%.
5. Caso depois das diferenciações se continue a verificar volatilidade na série, verificar se é necessário usar a transformação Box-Cox para estabilizar a variância.
6. Utilização da função *auto.arima* para confirmar qual o modelo que se adequa melhor à série temporal. A utilização desta função reduz significativamente o tempo computacional na seleção dos melhores modelos.
7. Seleção do modelo com base em:
 - Análise gráfica das auto-correlações, auto-correlações parciais;
 - Critério de AIC, este critério é baseado na função de log-verosimilhança, onde dado um conjunto de modelos candidatos para explicar os dados, aquele que tem menor AIC é preferível aos outros.

- Análise das medidas estatísticas: erro médio (ME), erro médio absoluto (MAE) e raiz quadrada do erro médio (RMSE).
8. Verificar se o modelo encontrado é adequado aos dados e conforme com os pressupostos através de uma análise de resíduos. Para confirmar a seleção dos modelos que melhor se ajustam aos dados foi feita uma análise de diagnóstico do modelo obtido através do teste de Box-Ljung. Este teste é usado para analisar a variância dos resíduos, verificando assim se o modelo se ajusta à série temporal e é aplicado aos resíduos da série temporal depois de prevermos os parâmetros desta. O teste examina a autocorrelação dos resíduos e se as correlações forem baixas podemos concluir que o modelo é válido.
 9. Prever os valores dos parâmetros.
 10. Analisar os resultados da previsão.

Para o desenvolvimento dos modelos univariados ARIMA (p,d,q) foi utilizado o *package forecast* do software R.⁴

A título ilustrativo, apresenta-se de seguida a análise feita para o parâmetro β_0 , considerando os valores das observações mensais referentes ao período inicial, ou seja, de junho 2009 a dezembro 2013. As análises para os restantes parâmetros são semelhantes estando disponíveis para consulta nos anexos 8 a 13.

Tal como foi referido na metodologia, o primeiro passo deste método consiste em verificar se a série é estacionária, através da observação gráfica e dos respetivos testes estatísticos, *Phillips-Perron* e *Augmented-Dickey-Fuller*.

Na figura 5 abaixo representam-se os gráficos ACF e PACF do parâmetro β_0 entre o período de junho 2009 e dezembro 2013 considerando os valores das observações mensais.

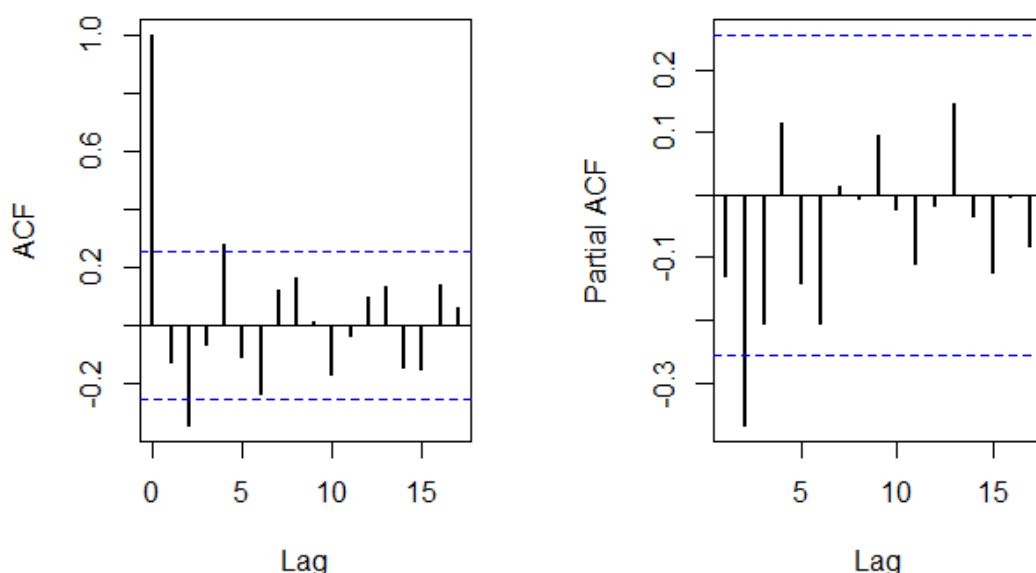


Figura 5: Gráficos ACF e PACF do parâmetro β_0

⁴ Para mais informações consultar o manual do package forecast (Hyndman, 2015)

Os resultados do teste *Phillips-Perron* evidenciam a rejeição da hipótese nula da presença de raiz unitária, ou seja, não foi necessário diferenciar nenhuma das séries observadas. Tal como se pode verificar nos resultados apresentados na seguinte tabela:

Tabela 7 : Resumo dos resultados dos testes de raízes unitárias

Parâmetro	<i>Phillips-Perron</i>	<i>P-value</i>
β_0	-3.4809	0.0516
β_1	-3.8347	0.02293
β_2	-3.7587	0.02742
β_3	-10.933	0.01

A tabela 8 sintetiza os principais indicadores resultantes da aplicação da metodologia de Box-Jenkins na identificação do modelo ótimo para cada parâmetro da função de NSS considerando os valores das observações mensais referentes ao período de observação inicial considerado no procedimento *recursive expanding window forecast* (junho 2009 - dezembro 2013).

Tabela 8: Estatísticas resumo do modelo univariado ótimo selecionado para cada parâmetro

Parâmetro	AIC	ME	MAE	RMSE	ARIMA (p,d,q)
β_0	139.21	-0.0168	0.5475	0.7302	ARIMA (1,0,0)
β_1	138.21	0.0041	0.5388	0.7256	ARIMA (1,0,0)
β_2	347.31	0.3525	1.6075	4.0376	ARIMA (1,0,1)
β_3	351.88	0.7958	1.4125	4.2249	ARIMA (2,0,0)

Nota: Os valores foram obtidos com observações mensais referentes ao período de junho 2009 a dezembro 2013 e os modelos selecionados foram os que apresentavam erros mais baixos.

Para confirmar a hipótese nula de que os resíduos dos modelos selecionados como sendo aqueles que melhor se ajustam aos dados são aleatórios e independentes ao longo do tempo, i.e., as observações não apresentam sinais de autocorrelação temporal, foi usada a estatística de *Ljung-Box*. Como é sabido, na presença de autocorrelação a precisão de um modelo na previsão é reduzida e pode levar a uma interpretação incorreta dos dados. A tabela 9 sintetiza os resultados da aplicação do teste às quatro séries temporais de parâmetros da função NSS relativas ao período compreendido entre junho de 2009 e dezembro de 2013.

Como se observa no quadro, a análise dos valores da estatística de teste e respectivos valores de prova não nos coloca, em todos os casos analisados, em condições de rejeitar a hipótese nula de ausência de correlação entre os resíduos dos modelos ARIMA.

Tabela 9: Resultados do Teste Ljung-Box

Parâmetro	Estatística	Lag	<i>p-value</i>
β_0	2.74	2	0.2541
β_1	3.1404	2	0.208
β_2	0.52231	2	0.7702
β_3	1.1776	2	0.555

O gráfico seguinte representa, de forma mais generalizada, as estatísticas de diagnóstico dos resíduos. A leitura do gráfico parece indicar que os resíduos são aleatórios ou *white noise*. O gráfico de ACF dos resíduos não apresenta nenhum valor fora dos limites e o teste de Ljung-Box apresenta quase todos os valores com *p-values* superiores ao limite. Podemos assim concluir que o modelo respeita os pressupostos de base.

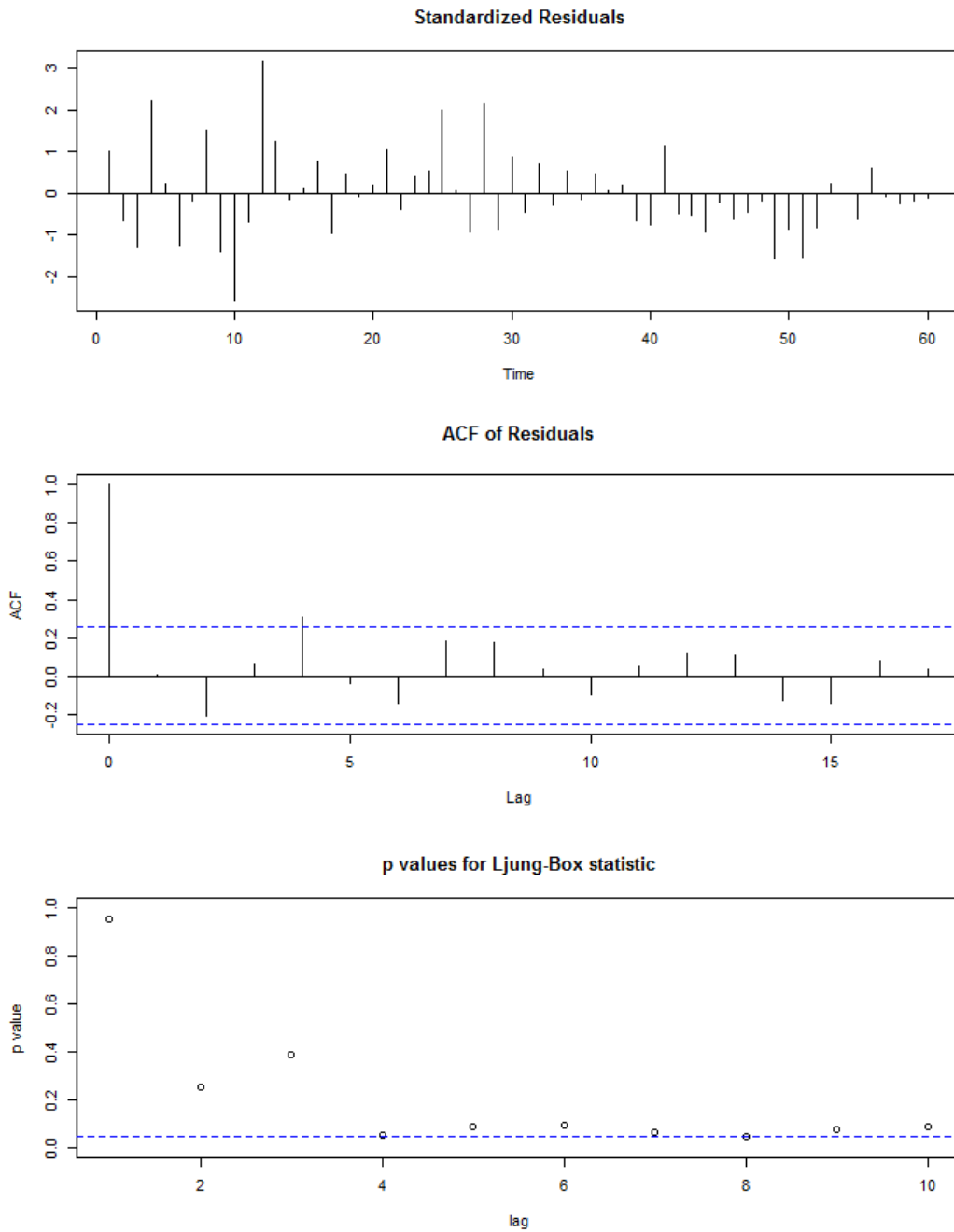


Figura 6: Análise dos resíduos pelo Ljung-Box Test do parâmetro β_0

Finalmente, foram obtidas as equações e a variância do erro para cada parâmetro:

Tabela 10: Parâmetros obtidos com os modelos univariados ARIMA(p,d,q)

Parâmetro	Equação de regressão	Variância do erro
β_0	$\hat{\beta}_{0,t} = 2.4201 + 0.6976 \hat{\beta}_{0,t-1}$ (0.3006) (0.0906)	$\sigma_\varepsilon^2 = 0.5332$
β_1	$\hat{\beta}_{1,t} = -2.0501 + 0.5891 \hat{\beta}_{1,t-1}$ (0.2227) (0.1019)	$\sigma_\varepsilon^2 = 0.5265$
β_2	$\hat{\beta}_{2,t} = -16.0598 + 0.9519 \hat{\beta}_{2,t-1} - 0.4116 \varepsilon_{2,t-1}$ (4.9827) (0.0413) (0.1326)	$\sigma_\varepsilon^2 = 16.3$
β_3	$\hat{\beta}_{3,t} = 0.6668 \hat{\beta}_{3,t-1} + 0.3038 \hat{\beta}_{3,t-2}$ (0.1213) (0.1225)	$\sigma_\varepsilon^2 = 17.85$

Nota: Os valores do desvio padrão da estimativa dos parâmetros encontram-se entre parêntesis.

A análise dos valores de prova permite-nos concluir que todos os parâmetros são significativos ao nível de significância de 5%.

3.2.2. Modelos VAR

Tal como já foi referido anteriormente, o modelo VAR trata todas as relações lineares existentes entre as variáveis endógenas e os valores passados das mesmas, sem impor restrições quanto à sua independência e dependência, permitindo ainda a inclusão de variáveis exógenas no modelo. É um modelo mais flexível do que os modelos uni-variados e permite testar a direção de causalidade determinando previsões de melhor qualidade relativamente aos modelos estruturais tradicionais (Moreira, 2011).

No que respeita ao processo de estimação do VAR, o processo pode ser sintetizado nas seguintes etapas (Pfaff, 2008):

1. Análise preliminar dos dados da série, análise essa que é composta pelas estatísticas descritivas da série (mínimo, máximo, média e mediana) e pelos respetivos gráficos.
2. Testes de raízes unitárias, nomeadamente o teste ADF.
O teste rejeitou a hipótese de diferenciar para todos os parâmetros.
3. Determinar o número ótimo de *lags* a considerar, o qual é escolhido a partir de critérios de informação/seleção:
 - AIC;
 - HQ (Hannan-Quinn Criterion);
 - SC (Schwarz Criterion);
 - FPE (Final Prediction Error Criterion).
4. Estimar os parâmetros do modelo.

5. Verificar a adequabilidade do modelo, analisando os erros do modelo através dos seguintes testes estatísticos:
- Portmanteau-Test;
 - Teste à normalidade.

Para a implementação desta metodologia e elaboração das previsões foi construída uma rotina no *software R* fazendo uso do *package vars*⁵.

Na tabela 11 apresentamos os valores dos critérios AIC, HQ, SC e FPE para diferentes níveis de desfasamento das variáveis do modelo VAR para uma amostra de observações relativa ao período de junho 2009 a dezembro 2013.

Tabela 11: *Lags* testados para os modelos VAR

CRITÉRIO	LAGS TESTADOS					
	1	2	3	4	5	6
AIC(N)	-1.85607	-1.82729	-1.64042	-1.82441	-1.75481	-1.75557
HQ(N)	-1.65054	-1.41623	-1.02384	-1.00229	-0.72717	-0.5224
SC(N)	-1.33802	-0.79118	-0.08626	0.247806	0.835454	1.352749
FPE(N)	0.156367	0.161525	0.196688	0.166953	0.185187	0.195063

Como se observa pela leitura dos resultados, os quatro critérios são coincidentes e sugerem que o *lag* ótimo a considerar no modelo VAR é 1, pelo que o consideraremos na análise subsequente dos resultados. A tabela seguinte sintetiza os resultados da estimação do modelo VAR considerando a amostra relativa ao período junho 2009 a dezembro 2013.

⁵ Para mais detalhe consultar manual de *package vars* (Pfaff & Stigler, 2013).

Tabela 12: Resultados da Estimação do modelo VAR (período junho 2009-dezembro2013)

Resultados da estimação da equação para $\beta_{0,t}$

	Estimativa	SE	t-value	p-value
$\beta_{0,t-1}$	1.10042	0.32771	3.358	0.00127**
$\beta_{1,t-1}$	0.29702	0.38172	0.778	0.43913
$\beta_{2,t-1}$	0.07085	0.02749	2.577	0.01208*
$\beta_{3,t-1}$	0.06743	0.03116	2.164	0.03388*

Resultados da estimação da equação para $\beta_{1,t}$

	Estimativa	SE	t-value	p-value
$\beta_{0,t-1}$	-0.34748	0.33525	-1.036	0.3036
$\beta_{1,t-1}$	0.41678	0.39051	1.067	0.2895
$\beta_{2,t-1}$	-0.06396	0.02813	-2.274	0.0260*
$\beta_{3,t-1}$	-0.05800	0.03188	-1.820	0.0731

Resultados da estimação da equação para $\beta_{2,t}$

	Estimativa	SE	t-value	p-value
$\beta_{0,t-1}$	-4.1112	2.2686	-1.812	0.0742
$\beta_{1,t-1}$	-5.3780	2.6426	-2.035	0.0456*
$\beta_{2,t-1}$	1.1934	0.1903	6.270	2.59e-09***
$\beta_{3,t-1}$	0.3366	0.2157	1.560	0.1232

Resultados da estimação da equação para $\beta_{3,t}$

	Estimativa	SE	t-value	p-value
$\beta_{0,t-1}$	4.3301	2.1395	2.024	0.0468*
$\beta_{1,t-1}$	5.3246	2.4922	2.137	0.0361*
$\beta_{2,t-1}$	-0.3618	0.1795	-2.015	0.0477*
$\beta_{3,t-1}$	0.5029	0.2034	2.472	0.0159*

Notas: Códigos de significâncias 0 '***'; 0.001 '**'; 0.01 '*'; 0.1 '.'. As estimativas apresentadas dizem respeito ao período de observações inicial (junho de 2009 a dezembro de 2013).

Tal como mencionado anteriormente, a última etapa para previsão através dos modelos VAR é a verificação da adequabilidade do modelo. O objetivo principal é verificar se os modelos propostos capturam toda a estrutura de dependência das variáveis a ser modeladas, ou seja, se os modelos capturam toda a informação contida nos dados.

Se o modelo for adequado, espera-se que os resíduos não apresentem nenhum padrão bem definido, i.e., possam ser classificados como ruído branco. Caso contrário, se existir estrutura de dependência nos resíduos e este puder ser modelado, não podemos classificá-lo como ruído branco e, conseqüentemente, o modelo deverá ser modificado para incluir esta explicação adicional. Para classificar o resíduo como sendo ruído branco, deve-se investigar se os resíduos são independentes e identicamente distribuídos (i.i.d.) de média zero.

Na tabela 13 podemos verificar, através dos valores do *p-value* dos testes efetuados, que rejeitamos a hipótese nula de os resíduos serem i.i.d, o que nos permite concluir que o modelo obtido é o adequado para os dados em análise.

Tabela 13:Quadro Resumo das estatísticas de teste dos resíduos dos modelos VAR

Teste	Chi-squared	Df	<i>p-value</i>
Portmanteau Test	214.45	240	0.881
Kurtosis	1709.3	4	<2.2e-16
JB-Test	1734.6	8	<2.2e-16
Skewness	25.323	4	4.331e-05

Nota: Os valores apresentados dizem respeito as observações mensais do período inicial da amostra, ou seja, de junho 2009 a dezembro 2013.

3.2.3. Modelos de Aprendizagem Automática

Neste subcapítulo serão descritas sucintamente os procedimentos de estimação e os resultados da aplicação dos modelos de aprendizagem automática: (i) considerando a previsão dos parâmetros da especificação de NSS individualmente; (ii) considerando a modelação simultânea de todos os fatores latentes da especificação de NSS, tal como no estudo de Vela (2013).

Para o desenvolvimento dos modelos de RNA foi utilizado o *package* neuralnet⁶ do *software* R. A melhor parametrização para a realização das previsões foi encontrada através da experimentação das diferentes configurações possíveis e que gerassem os melhores resultados, ou seja, que minimize o erro.

3.2.3.1. Desenvolvimentos de Modelos de Redes Neurais

De acordo com diferentes autores, como por exemplo, Kaastra & Boyd (1996) ou Coakley & Brown (2000), a metodologia a ser seguida para a construção das redes neurais como modelos de previsão em dados de séries temporais deverá englobar os seguintes passos:

A. Seleção das variáveis

A seleção das variáveis é dada pelo número de *lags* que irá determinar o comportamento da rede. Considerando que foram feitas duas abordagens, as previsões poderão ser explicadas ou pelos valores observados do fator ou pelos valores observados dos fatores complementares da expressão de NSS.

B. Recolha dos dados

Este passo foi previamente explicado na secção 3.1 da dissertação.

⁶ Para mais detalhes sobre o algoritmo consultar manual *package* neuralnet (Riedmiller, 1994).

C. Pré-Processamento de dados

Este passo tem como objetivo fazer uma análise e transformação das variáveis de *input* para garantir que a previsão final terá o menor ruído possível. Como tal, pretende-se detetar tendências, correlações e analisar a distribuição da série. Também este passo foi previamente descrito na secção 3.1 da dissertação.

No entanto, para a análise das RNA os dados foram normalizados com o intuito de estabelecer um intervalo de valores entre 0 e 1, o que irá tornar o treino da rede mais rápido, pois não irão existir valores muito elevados nem muito baixos que tornem mais difícil a convergência do modelo para um ótimo global.

D. Treino, testes e validação

Neste passo foi necessário definir a divisão dos dados para assegurar a aprendizagem da rede. O mais usual segundo a literatura é formar três conjuntos distintos:

- Treino: Conjunto que contém um maior número de dados pois será a partir deste que a rede irá aprender a tendência atual com os dados;
- Validação: Conjunto utilizado para avaliar a capacidade de generalização da rede e determinar o fim do treino;
- Teste: Conjunto usado para a avaliação final da rede com dados que não foram usados no processo de aprendizagem. Ao definir a dimensão deste grupo é necessário garantir que existem dados suficientes para testar a rede e que as restantes observações são as necessárias para treinar e validar a rede.

Neste estudo fizeram-se diversas experiências de forma a conferir a melhor divisão entre os grupos, tendo-se concluído que seriam feitos apenas dois conjuntos de dados uma vez que a amostra recolhida não tinha dimensão significativa para ser dividida nos três conjuntos habituais. Assim, foi feita a seguinte divisão: treino (com 70% das observações) e teste (com os restantes 30%).

E. Definição da estrutura da rede

Relativamente à estrutura da rede neuronal criada, foi necessário definir o número de neurónios em cada uma das camadas:

- *Input*: o número de neurónios na camada de *input* é fácil de determinar, uma vez que cada variável independente do modelo será um neurónio na camada de *input*. O número de neurónios corresponde ao número de *lags* que define as observações futuras.

Neste caso serão considerados 12 neurónios de *input* que correspondem a 12 *lags* – primeiro objeto desde o primeiro valor da série até ao décimo-segundo.

- *Hidden*: são as camadas responsáveis pelo processo de comunicação entre a camada de *input* e a camada de *output*. É esta a camada que fornece à rede a capacidade de generalizar. Segundo Kaastra (1996), uma camada *hidden* com o número suficiente de neurónios é o suficiente para realizar uma aproximação fiável a qualquer função contínua. Uma rede com mais do que uma ou duas camadas *hidden* tem maior tempo de computação da rede, o que aumenta o risco de ser encontrado um ótimo local e não global e diminui a capacidade de generalização da rede. A rede apresentada tem apenas uma camada *hidden*.

O número de neurónios na camada *hidden* foi selecionado através de experimentação e foi selecionado o número de neurónios que gerava a melhor rede. Ao testar diferentes números de neurónios todos os outros parâmetros de definição da rede se mantiveram inalterados. Foram testadas redes com 5, 10, 15 neurónios.

- *Output*: camada responsável pela obtenção das previsões finais da rede para um dado objeto. Estará extremamente relacionada com o horizonte temporal da previsão. Neste caso o *output* será um parâmetro, no caso da rede individual, e todos os parâmetros no caso da RNA que prevê todos os parâmetros em simultâneo.

F. Definição e aplicação do critério de avaliação da rede

Posteriormente, é necessário escolher o critério de avaliação da rede para determinar a fiabilidade e consequentemente qual a RNA que se adequa melhor à série temporal.

Neste caso, foi escolhido o modelo que gerava um erro médio menor. A análise foi feita para todas as observações do conjunto de teste.

G. Treino da rede

Depois de estar definida a estrutura da RNA, passamos para o processo de aprendizagem da rede que tem como objetivo otimizar os pesos de cada neurónio de forma a minimizar o erro da função.

Foi utilizado o algoritmo RPROP+ (*Resilient BackPropagation*) para treinar a rede, sendo este um treino supervisionado. Este algoritmo é definido como uma adaptação do algoritmo de aprendizagem de *backpropagation* que tem como objetivo produzir um *output* com o menor erro possível, recorrendo, para tal, à combinação dos pesos utilizados (Rojas, 1996).

Com o algoritmo RPROP+, os pesos são ajustados da camada de *output* para a camada de *input*. Esta técnica pretende perceber quais os pesos que mais contribuíram para o erro gerado pela rede e, assim, ajustá-los de forma a obter melhores resultados.

H. Implementação

Com o objetivo de escolher a configuração que gera melhores resultados, foram definidas as seguintes condições para a implementação da rede:

- Função logística que descreve a relação não linear entre a camada de *input* e a camada *hidden* e uma função logística que relaciona a camada *hidden* com a camada de *output*;
- Algoritmo de aprendizagem *Resilient BackPropagation* para o processo de aprendizagem;
- A taxa de aprendizagem é uma constante que auxilia a atualização dos pesos.

Definir uma taxa de aprendizagem correta não é trivial. Se a taxa for demasiado pequena, o algoritmo irá demorar demasiado tempo a convergir para um mínimo global ou ficar preso num mínimo local. Por outro lado, se a taxa for demasiado elevada, o algoritmo poderá divergir (Cilimkovic, 2010).

Foram testadas taxas de aprendizagem de 0.01, 0.02, 0.03 e 0.05.

- A inicialização dos pesos é feita de modo aleatório, o que significa que mesmo utilizando os restantes parâmetros e o mesmo conjunto de treino cada vez que se inicia um novo treino os resultados obtidos serão diferentes. Para reproduzir os mesmos resultados é necessário usar a mesma *seed* para geração dos pesos iniciais da rede.
- Critérios de paragem:
 - Número de iterações da rede

Ao atingir este número, a rede irá terminar. Foram testadas redes com 1000, 10000, 30000, 50000 e 100000 iterações, para ter uma perceção melhor do número de iterações necessárias para gerar um bom modelo. Uma vez que a rede não apresenta muitos registos as diferenças entre as redes não são significativas e optou-se por apenas realizar os restantes testes com redes com 10000 iterações.

- *Threshold*

Enquanto a rede está a correr, se o resultado obtido for superior a este valor, a rede irá terminar. Caso não seja, irá continuar até atingir este valor.

Quanto menor for este valor maior será a probabilidade de todas as iterações serem finalizadas na fase de treino, portanto o mínimo testado para este parâmetro foi de 0.01.

3.2.3.2. Rede para previsão simultânea de todos os parâmetros de NSS

Na construção da rede com a previsão de β_0 β_1 β_2 β_3 em simultâneo, foi necessário definir o número de neurónios na *hidden layer*, a taxa de aprendizagem e os critérios de paragem da rede. Para tal, elaborou-se uma análise prévia com diversas parametrizações, com o intuito de escolher a rede que geraria melhores resultados. Assim, foi configurada uma rede com 10 neurónios na *hidden layer* e com uma taxa de aprendizagem de 0.03, que irá terminar ao atingir um valor limiar de 0.01 (*threshold*) ou as 1000 iterações, esta escolha resultou num erro médio de 0.13956.

Segue uma tabela resumo com as principais análises de sensibilidade efetuadas, para conjunto de treino, até chegar à conclusão supra mencionada:

Tabela 14: Configurações testadas à RNA conjunta

Configuração/Resultados						
Neurónios <i>hidden</i>	5	5	5	5	5	5
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.20866	0.20195	0.25037	0.3174	0.29798	0.30608
Neurónios <i>hidden</i>	10	10	10	10	10	10
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.13956	0.17782	0.18546	0.28044	0.27435	0.27764
Neurónios <i>hidden</i>	15	15	15	15	15	15
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.2085	0.19672	0.18359	0.24213	0.2421	0.23633

Nota: Os resultados foram obtidos com os dados do conjunto de treino, ou seja, com as observações mensais de junho 2009 até dezembro 2013. O valor assinalado a negrito representa o menor valor registado.

Assim, foi obtida a rede neuronal que se encontra em anexo na figura 14.

3.2.3.3. Redes individuais para previsão dos parâmetros de NSS

Na construção de redes para a previsão dos parâmetros individualmente, foram testadas diversas configurações para o conjunto de treino dos parâmetros, tal como tinha sido efetuado na abordagem anteriormente referida.

Nesta modelização o objetivo é a otimização de cada parâmetro independentemente, assim, depois de diversas experimentações, que podem ser consultadas nas tabelas 17-20 presentes em anexo, foram escolhidas as seguintes configurações para cada β :

Tabela 15: Configurações final da RNA individual

Configuração/Resultados				
Parâmetros	B0	B1	B2	B3
Neurónios <i>hidden</i>	5	15	5	15
Nº máximo de iterações	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.01
Taxa de Aprendizagem	0.03	0.02	0.02	0.01
Erro Médio	0.17032	0.15464	0.0627	0.35532

Nota: Os resultados foram obtidos com os dados do conjunto de treino, ou seja, com as observações mensais de junho 2009 até dezembro 2013.

De seguida é apresentada a representação da rede neuronal obtida para o parâmetro β_0 . As restantes ilustrações podem ser consultadas nos anexos 15-17.

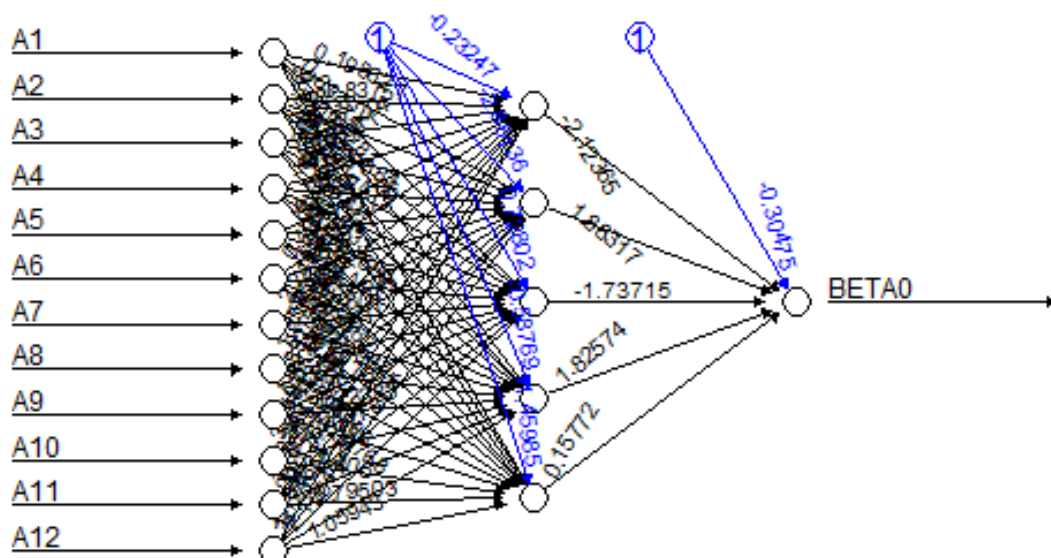


Figura 7: Rede Neuronal β_0

3.2.4. *Random walk*

Segundo a hipótese de que as taxas de juro para as diferentes maturidades seguem um passeio aleatório sem tendência, a previsão do valor de uma determinada taxa de juros no horizonte temporal h corresponde ao último valor observado. Dito de outro modo, a previsão do modelo é a de que, para cada maturidade, não existem variações face à última observação (Diebold & Li, 2006). Em termos formais:

$$\hat{y}_{t+h}(\tau) = \hat{y}_t(\tau) \quad [13]$$

onde $\hat{y}_{t+h}(\tau)$ denota a previsão da taxa para a maturidade τ no momento t e h denota o horizonte temporal da previsão.

Assim, os resultados obtidos por ambas as aproximações (métodos paramétricos e aprendizagem automática) foram comparados com os resultados obtidos por um simples *Random walk* para aferir a capacidade de previsão de eventos futuros não usados no desenvolvimento dos modelos.

3.3. CRITÉRIOS DE DESEMPENHO DOS MODELOS

Os resultados obtidos nos diferentes modelos efetuados foram comparados usando o indicador estatístico RMSE. Este indicador mede a amplitude média do erro do modelo e é definido na seguinte fórmula:

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (\hat{x}_t - x_t)^2}{T}} \quad [14]$$

onde T denota o número de previsões efetuadas, e \hat{x}_t e x_t representam, respetivamente, os valores estimado e observado da série.

Esta foi a medida estatística escolhida uma vez que tem sido a mais usada na literatura e em estudos similares (veja-se, por exemplo, Vela (2013) e Arantes et al., (2014)).

4. RESULTADOS E DISCUSSÃO

Neste capítulo é apresentada e analisada a capacidade preditiva dos diferentes modelos na obtenção de previsões de taxas de juro para a Zona Euro.

A título recapitulativo, recordamos que na modelização da ETTJ foi considerada a forma funcional de NSS atualmente utilizada pelo BCE, tendo sido testadas quatro metodologias alternativas de previsão da *yield curve*: (i) Modelos ARIMA considerando cada um dos fatores da especificação de NSS individualmente; (ii) Modelos VAR, modelo que prevê todos os fatores tendo em conta a sua interação; (iii) RNA com a previsão de cada fator individualmente e (iv) RNA com a previsão de todos os fatores em conjunto. Com o intuito de verificar a performance dos modelos elaborados, estes foram ainda comparados com os resultados obtidos pelo modelo de passeio aleatório sem tendência (*Random Walk*).

Depois de analisar os modelos que melhor se ajustavam aos dados, procedeu-se à previsão dos parâmetros e à introdução desses parâmetros na fórmula de NSS para obter a *yield curve*. A partir destes valores, foi calculado o RMSE para verificar a qualidade das previsões feitas.

A análise do poder preditivo dos modelos é feita tendo em conta quatro dimensões distintas:

1. Considerando os indicadores (RMSE) obtidos pelos diferentes modelos testados;
2. Relativamente aos três horizontes de previsão considerados (1 mês, 2 meses e 3 meses);
3. Relativamente às distintas maturidades avaliadas no estudo (de 3 meses a 30 anos);
4. Tendo em conta os resultados obtidos em estudos similares realizados no âmbito da previsão da ETTJ.

A tabela 16 apresenta a comparação entre os valores de RMSE dos cinco modelos de previsão elaborados para 1, 2 e 3 meses de previsão. Apresentam-se assinalados a verde os valores de RMSE mais baixos para cada uma das maturidades testadas, ou seja, o modelo que prevê melhor tendo em conta as diferentes maturidades.

Tabela 16: Resultados da previsão para diferentes horizontes temporais

A: Horizonte Temporal de 1 mês

<i>MATURIDADE</i>	<i>MODELOS UNIVARIADOS ARIMA(P,D,Q)</i>	<i>VAR(1)</i>	<i>RNA (IND)</i>	<i>RNA (CONJ)</i>	<i>RW</i>
3 MESES	1.0857	1.0284	0.7344	0.4032	0.0415
6 MESES	1.4022	1.3171	0.7674	0.4343	0.0467
1 ANO	1.1331	1.0433	0.8015	0.4594	0.0440
2 ANOS	0.1602	0.1008	0.7911	0.4389	0.0619
3 ANOS	0.4913	0.5199	0.6901	0.3689	0.0872
5 ANOS	0.9413	0.9535	0.4601	0.3195	0.1298
7 ANOS	1.0115	1.0188	0.3870	0.582	0.1582
10 ANOS	0.9321	0.9372	0.7508	1.0585	0.1828
15 ANOS	0.7623	0.7624	1.3134	1.601	0.2041
30 ANOS	0.5467	0.5223	1.6949	1.8774	0.1900

B: Horizonte Temporal de 2 meses

<i>MATURIDADE</i>	<i>MODELOS UNIVARIADOS ARIMA(P,D,Q)</i>	<i>VAR(1)</i>	<i>RNA (IND)</i>	<i>RNA (CONJ)</i>	<i>RW</i>
3 MESES	1.09381	0.9841	0.7311	0.4486	0.04668
6 MESES	1.42232	1.2643	0.7645	0.4777	0.04775
1 ANO	1.17608	1.0097	0.8001	0.499	0.04959
2 ANOS	0.24444	0.1311	0.7935	0.4775	0.06721
3 ANOS	0.43566	0.4756	0.7137	0.4118	0.09183
5 ANOS	0.87188	0.886	0.4741	0.3577	0.1499
7 ANOS	0.94666	0.9483	0.4117	0.594	0.20152
10 ANOS	0.88189	0.8719	0.7701	1.0564	0.24816
15 ANOS	0.74132	0.7081	1.3271	1.5925	0.26893
30 ANOS	0.59556	0.4864	1.7047	1.8665	0.23457

C: Horizonte Temporal de 3 meses

<i>MATURIDADE</i>	<i>MODELOS UNIVARIADOS ARIMA(P,D,Q)</i>	<i>VAR(1)</i>	<i>RNA (IND)</i>	<i>RNA (CONJ)</i>	<i>RW</i>
3 MESES	1.10581	0.9497	0.7316	0.4900	0.07295
6 MESES	1.44292	1.2224	0.7643	0.51916	0.07231
1 ANO	1.21628	0.9822	0.8003	0.54235	0.06882
2 ANOS	0.33185	0.127	0.7936	0.52633	0.09051
3 ANOS	0.41812	0.4444	0.7141	0.46709	0.12887
5 ANOS	0.83233	0.8359	0.4743	0.40795	0.21305
7 ANOS	0.91396	0.8957	0.4103	0.61196	0.28474
10 ANOS	0.86805	0.8234	0.7705	1.0537	0.3568
15 ANOS	0.76053	0.6689	1.3293	1.58006	0.4052
30 ANOS	0.6746	0.5223	1.7092	1.85112	0.35313

Nota: Os valores indicados nas tabelas acima correspondem ao RMSE obtido entre as taxas spot observadas e as previsões da taxa obtidas a partir da especificação de Nelson-Siegel & Svensson. Ao todo, foram considerados 12 períodos para realizar as previsões, desde dezembro 2013 a janeiro 2014, com observações mensais.

A estrutura das análises elaboradas é a seguinte:

- Modelos paramétricos:
 - Modelos ARIMA;
 - Modelos VAR(1);
 - Comparação dos modelos paramétricos (ARIMA vs VAR).
- Redes Neurais Artificiais:
 - RNA Individuais;
 - RNA Conjuntas;
 - Comparação entre as duas redes neuronais elaboradas.
- Comparação entre os 5 modelos elaborados.
- Comparação dos resultados da dissertação com 4 estudos distintos.

Analisando os valores de RMSE obtidos pelos modelos univariados ARIMA (p,d,q) no período em estudo, podemos observar que estes modelos se comportam melhor nas maturidades de médio-prazo (2 e 3 anos) e de muito longo-prazo (30 anos). Estes resultados verificam-se nos três horizontes de previsão testados. Ainda assim, o erro de previsão para o horizonte de 1 mês é inferior, em todas as maturidades, comparativamente com os valores de erro obtidos nas previsões a 2 e 3 meses.

Nos modelos VAR(1) verificamos um comportamento similar ao observado nos modelos ARIMA, na medida em que também estes apresentam melhores resultados para as previsões nas maturidades de médio-prazo (2 e 3 anos) e de muito longo-prazo (30 anos) nos três horizontes de previsão, apresentando contudo uma performance ligeiramente melhor nas previsões a 3 meses do que nas previsões a 1 e 2 meses.

Comparando as duas abordagens paramétricas, verificamos que os modelos univariados preveem melhor para horizontes temporais mais curtos (1 mês), ao contrário dos modelos multivariados que são capazes de produzir previsões mais fiáveis para horizontes temporais mais longos (3 meses).

O RMSE das RNA individuais evidencia que este método prevê melhor nas maturidades de 5 e 7 anos, nos três horizontes de previsão testados. No entanto, constata-se que nas previsões a 1 mês os valores do RMSE são sensivelmente mais baixos. É ainda possível concluir que estas redes produzem previsões pouco fiáveis para horizontes temporais de muito longo prazo (15 e 30 anos).

Para as RNA com a previsão simultânea de todos os parâmetros, podemos concluir que as maturidades que apresentam um erro menor são as de 3 e 5 anos, qualquer que seja o horizonte de previsão, apesar de as redes apresentarem valores semelhantes entre as maturidades de 3 meses a 7 anos. No entanto, para as maturidades de longo prazo (15 e 30 anos) o erro é, uma vez mais, consideravelmente maior.

A comparação entre os valores do indicador RMSE obtidos pelas duas variantes das redes neuronais artificiais permite concluir que a RNA Conjunta apresenta resultados melhores nas previsões de curto e médio prazo relativamente às RNA Individuais, enquanto que nas maturidades longas a performance das RNA Individuais é ligeiramente superior, pese embora ambas apresentarem resultados insatisfatórios em termos preditivos. Em termos globais, os resultados obtidos neste estudo sugerem que a RNA Conjunta apresenta previsões mais fidedignas.

No que se refere à comparação da qualidade de previsão entre todos os modelos testados, apesar de não ser possível identificar apenas um modelo com uma performance melhor em todos os horizontes temporais e para todas as maturidades, conclui-se que:

- RNA conjuntas são melhores a prever as taxas de juro com maturidades de curto-prazo (entre os 3 meses e 1 ano) e para as maturidades de medio-longo prazo (3 a 5 anos), nos três horizontes de previsão testados.
- RNA individuais apresentam o melhor desempenho preditivo nas maturidades entre 7 e 10 anos), nos três horizontes de previsão testados.
- Os modelos VAR(1) destacam-se na obtenção de previsões de taxa de juro para maturidades de 2 anos e para maturidades longo prazo e muito-longo prazo (15 e 30 anos respetivamente).
- Os modelos ARIMA apenas apresentam a melhor capacidade de previsão para o horizonte-temporal de 1 mês e para maturidade de 15 anos, em todos os outros horizontes temporais de previsão não estes modelos não se destacam.

No entanto, é de salientar que no período amostral em estudo nenhum dos modelos conseguiu superar os resultados obtidos pelo passeio aleatório sem tendência, modelo que apresenta, para todos os horizontes temporais de previsão, os valores mais baixos para o indicador RMSE.

Antes de comparar os resultados obtidos neste estudo com os estudos anteriores, é importante referir que as capacidades de previsão dos modelos desenvolvidos nesta dissertação são testados dentro de um período em que mudanças substanciais na *yield curve* tiveram lugar, contrastando assim com exercícios similares realizados em contextos de maior estabilidade nos mercados de dívida, o que poderá ter originado resultados díspares dos expectáveis.

Começando por comparar com os resultados obtidos por Vela (2013), onde se analisa o poder preditivo da ETTJ em quatro países da América Latina (Colômbia, México, Peru e Chile), verificamos que a qualidade das previsões é bastante díspar dependendo do país em análise e do período de observação. Vela concluiu numa análise geral que as redes neuronais artificiais preveem melhor para o horizonte temporal de um mês quando comparadas com os modelos paramétricos e o autor aponta ainda que, nas previsões para 1 mês, as RNA se distanciam bastante dos resultados obtidos pelo RW. O estudo menciona também que para os restantes horizontes temporais de previsão estudados (de 2 a 6 meses) os resultados não foram conclusivos pois em alguns países as redes neuronais funcionavam melhor que os modelos paramétricos mas noutros casos não conseguiam superar as previsões feitas através do RW.

Assim, em termos genéricos, conclui-se que o resultado das previsões está intrinsecamente relacionado com a estrutura da *yield curve* de cada país e com a sua situação económica. Podemos ainda depreender que os resultados dependem do número de observações consideradas na amostra de dados recolhida, quanto maior for a amostra, melhores serão as previsões para os horizontes temporais mais longos.

Analisando agora o estudo feito por Arantes et al. (2014), cujo âmbito era a previsão da ETTJ no Brasil, o autor indica que em média os modelos estimados através de RNA conseguiram superar as previsões realizadas pelo RW em todos os horizontes de previsão, sendo que a vantagem (diferença) em termos de qualidade preditiva é estimada em torno de 2% e 5% para horizontes de previsão de 1

dia e 1 semana, e de 12% e 7% em horizontes de previsão mais longos, de 1 mês e 3 meses. Para o estudo da curva brasileira foram escolhidos horizontes de previsão de curto prazo (1 dia até 3 meses) e os resultados foram bastante mais animadores relativamente à capacidade preditiva das RNA quando comparados com as previsões do RW e dos modelos tradicionalmente utilizados (Diebold e Li, o modelo de Svensson, o autorregressivo, o vetor-autorregressivo). O que também se verifica nesta dissertação, as RNA apresentam previsões de melhor qualidade para os horizontes temporais mais curtos.

Comparando os resultados desta dissertação com o estudo de Diebold (2006) cuja temática era modelar e prever a *yield curve* usando diferentes variações do modelo Nelson-Siegel. Embora os autores tenham usado diversas metodologias, apenas iremos comparar os resultados obtidos com os modelos AR (1) e VAR (1). O estudo indica que a previsão dos fatores latentes da *yield curve* não produz previsões melhores do que o *Random walk* para o horizonte de previsão de 1 mês. No entanto, para períodos de previsão de 1 ano, os modelos desenvolvidos apresentam resultados bastante superiores. Assim, através do estudo de Diebold e Li podemos verificar, que à semelhança do que se conclui neste estudo, os modelos paramétricos apresentam resultados melhores para horizontes de previsão mais longos.

É ainda importante comparar os resultados obtidos com o estudo elaborado por Tabak et al (2012) que tinha também como objetivo prever a *yield curve* para a Zona Euro. Este estudo compara as capacidades de previsão do modelo *Functional Signal plus Noise* (FSN), do *Dynamic Nelson-Siegel* (DL) e do *Random walk*. Os autores utilizaram dados diários desde 1 de janeiro de 1999 até 29 de abril de 2011, considerando 20 maturidades diferentes, desde 6 meses até 10 anos. Como resultados importantes a reter, o estudo indica que os modelos FSN e DL apresentam uma capacidade preditiva relativamente superior comparando com o RW para *yield curve* da Zona Euro, apesar de em alguns casos essa diferença ser muito ténue. E os autores referem ainda que comparando os modelos DL e os modelos FSN, estes últimos apresentam melhores previsões para horizonte temporais de curto prazo.

5. CONCLUSÕES

No presente estudo investigou-se a capacidade preditiva da estrutura temporal das taxas de juro na Zona Euro, considerando diferentes métodos de previsão para o período entre 2009 e 2015. Em termos mais específicos, foi analisada a performance preditiva do modelo de Nelson-Siegel & Svensson, assumindo que os seus parâmetros podem ser modelizados através de vários métodos distintos: univariados (modelos ARIMA, *Random walk*), multivariados (modelos VAR) e Redes Neurais Artificiais (RNA) individuais e conjuntas.

É importante lembrar que obter previsões fiáveis da ETTJ para toda a Zona Euro é uma tarefa de extrema relevância, tendo em conta não só a crescente globalização dos mercados financeiros mas também o facto de os países da Zona Euro partilharem uma moeda única.

Este estudo destaca-se por dois motivos: em primeiro lugar, porque compara o desempenho em termos preditivos de um conjunto de métodos alternativos de previsão da *yield curve* da Zona Euro e em segundo lugar, porque avalia o desempenho preditivo dos modelos num contexto e num período temporal marcados por uma extrema volatilidade nos mercados de taxa de juro da Zona Euro, na sequência da crise financeira de 2008, da crise económica global que se lhe sucedeu, da crise da dívida pública e da implementação de planos de resgate financeiro em vários países da Zona Euro (Portugal, Irlanda, Grécia).

Após a análise dos resultados empíricos obtidos no estudo verifica-se que não existem evidências suficientes para afirmar que existe um modelo que apresente sistematicamente uma performance preditiva melhor em todas as maturidades e para todos os horizontes temporais testados. Ainda assim, e de acordo com os resultados obtidos, conclui-se que:

- O modelo que apresenta valores de erro mais baixos para as maturidades de curto e médio prazo (3 meses a 5 anos) é o modelo RNA com a previsão conjunta dos parâmetros;
- As RNA individuais são sempre melhores para prever as taxas de juro para as maturidade de 7 e 10 anos;
- Para as maturidades de longo e muito longo prazo (15 e 30 anos respetivamente), o modelo com melhor capacidade preditiva é o VAR(1).

No entanto, assinala-se que no período estudado nenhum dos modelos mais complexos testados apresenta valores de erro inferiores aos erros obtidos com o modelo *Random walk*. Este resultado pode dever-se ao facto de no período em análise a *yield curve* ter tomado valores atípicos, dificultando a previsão e originando um modelo que não tem em conta as variações.

Em suma, apesar de o período em análise apresentar *yield curves* bastante voláteis, o que causa dificuldades em obter previsões fidedignas, os resultados demonstram que, para prever a ETTJ na Zona Euro, as Redes Neurais Artificiais são uma boa alternativa para as maturidades de curto e médio prazo, enquanto que para prever as maturidades de longo prazo, os modelos VAR(1) apresentam melhor desempenho.

5.1. LIMITAÇÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

A principal limitação desta dissertação resulta de uma amostra de dados reduzida, o que diminui a robustez dos resultados em análise. Caso a amostra recolhida contivesse mais dados, seria possível fazer estudos para diferentes horizontes temporais utilizando, por exemplo, dados semanais ou mesmo com dados diários. No entanto, dada a volatilidade causada pela crise internacional desde 2008 e com repercussões no sistema financeiro até ao presente, não foi possível utilizar todos os dados disponibilizados pelo BCE.

Como análise futura, propõe-se um estudo que adicione às redes neuronais outras variáveis relacionadas com a *yield curve* e que ajudem a descrever o seu comportamento permitindo uma melhor performance dos modelos, como por exemplo, variáveis relacionadas com o risco ou o preço das ações, tal como é feito por Ang e Piazzesi (2003). A inclusão de outras variáveis explicativas poderia complementar a análise do presente estudo.

Posteriormente, seria também interessante realizar um estudo comparativo entre os diferentes países da Zona Euro onde se analisasse se as taxas de juro de cada país dependem dos seus fatores internos e esses fatores são influenciados pelos fatores globais da Zona Euro.

6. BIBLIOGRAFIA

- Aljinović, Z., & Poklepović, T. (2013). Neuronal networks and vetor autoregressive model in forecasting yield curve.
- Allende, Hector. Moraga, Claudio. Salas, R. (2002). *An artificial neural network in timeseries forecasting: A compartive analysis* (Vol. 88).
- Alpaydin, E. (2004). *Introduction to Machine Learning*. Massachusetts: Massachusetts Institute of Technology.
- Ang, A., & Piazzesi, M. (2003). A no-arbitrage vetor autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, 50(4), 745–787.
- Basheer, I. a., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1), 3–31.
- Bech, M. L., & Lengwiler, Y. (2012). *The financial crisis and the changing dynamics of the yield curve*.
- Berk, J. M. (1998). The information content of the yield curve for monetary policy: a survey. *De Economist*, (2), 303–320.
- Berk, J. M., & Bergeijk, P. a. G. Van. (2000). *Is the yield curve a useful information variable for the Eurosystem?*
- BIS. (2005). Zero-Coupon yield curves: technical documentation. *BIS Papers*, No.25.
- Bjork, T., & Christensen, B. J. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance*, 9(4), 1–38.
- Bodyanskiy, Y., & Popov, S. (2006). Neural network approach to forecasting of quasiperiodic financial time series. *European Journal of Operational Research*, 175(3), 1357–1366.
- Bolder, D. (2006). *Modelling Term-Structure Dynamics for Risk Management : A Practitioner ' s Perspective*. Bank of Canada Working Paper No. 2006–48.
- Bolder, D., & Stréliški, D. (1999). *Yield curve modelling at the Bank of Canada*. Bank of Canada.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, 31, 307–327.
- Bowsher, C.G., Meeks, R., 2008. The dynamics of economic functions: modeling and forecasting the yield curve. *Journal of the American Statistical Association* 103 (484), 1419–1437
- Box, G. E. P., & Jenkins, G. M. (1994). *Time Series Analysis: Forecasting and Control*. *Journal of Time Series Analysis* (Vol. Third).
- Bravo, J. (2001). *Modelos de risco de taxa de juro: Estratégias de cobertura e imunização*.
- Caiado, J. (2002). *Modelos VAR, Taxas de Juro e Inflação*.
- Campbell, J. Y. (1995). *Some Lessons from the Yield Curve*. *Journal of Economic Perspectives* (Vol. 9).
- Cieslak, A., & Povala, P. (2013). Information in the term structure of yield curve volatility.

- Cilimkovic, M. (2010). *Neural Networks and Back Propagation Algorithm. Institute of Technology Blanchardstown.*
- Coakley, J. R., & Brown, C. E. (2000). Artificial Neural Networks in Accounting and Finance: Modeling Issues. *International Journal of Intelligent Systems in Accounting Finance & Management*, 9(july 1999), 119–144.
- Cox, J. C., Ingersoll Jonathan E., J., & Ross, S. a. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*.
- Cui, G., Wong, M. L., & Lui, H.-K. (2006). Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming. *Management Science*, 52(4), 597–612.
- Czirák, D. (2007). Forecasting the Yield Curve with S-Plus. *Wilmott Magazine*, 86–93.
- De Pooter, M. (2007). Examining the Nelson-Siegel class of term structure models: In-sample fit versus out-of-sample forecasting performance. *Econometric Institute and Tinbergen Institute Erasmus University Rotterdam.*
- De Rezende, R. B., & Ferreira, M. S. (2011). Modeling and forecasting the yield curve by an extended Nelson-Siegel class of models: A quantile autoregression approach. *Journal of Forecasting*, 32, 111–123.
- Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130, 337–364.
- Dolan, C. P. (1999). Forecasting the Yield Curve Shape: Evidence in Global Markets. *The Journal of Fixed Income*, 92-99
- Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6, 379–406.
- Estrella, A., & Trubin, M. R. (2006). The Yield Curve as a Leading Indicator : Some Practical Issues. *Finance*, 12, 1–7.
- Fabozzi, F. J., Martellini, L., & Priaulet, P. (2005). Predictability in the Shape of the Term Structure of Interest Rates. *The Journal of Fixed Income*, 15(1), 40–53.
- Félix, R. M., & Nunes, L. (2003). *Forecasting euro area aggregates with Bayesian VAR and VECM models.*
- Gilli, M., Große, S., & Schumann, E. (2010). Calibrating the Nelson – Siegel – Svensson model. *Comisef Working Papers Series*, 1–22.
- Hamilton, J. D. (1994). *Time Series Analysis. Book.*
- Haykin, S. (1999). Neural networks: a comprehensive foundation. *The Knowledge Engineering Review.*
- Hill, T., O'Connor, M., & Remus, W. (1996). Neural Network Models for Time Series Forecasts. *Management Science*, 42(7), 1082–1092.
- Hull, J., & White, A. (1990). Pricing Interest-Rate-Derivative Securities.

- Hyndman, R. J. (2015). *Forecasting Functions for Time Series and Linear Models*.
- Ioannides, M. (2003). A comparison of yield curve estimation techniques using UK data. *Journal of Banking and Finance*.
- Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), 215–236.
- Kohonen, T. (1987). Adaptive, associative, and self-organizing functions in neural computing. *Applied Optics*, 26, 4910–4918.
- Luo, X., Han, H., & Zhang, J. E. (2012). Forecasting the term structure of Chinese Treasury yields. *Pacific Basin Finance Journal*, 20(5), 639–659.
- Makridakis, S., & Hibon, M. (1997). ARMA Models and the Box–Jenkins Methodology, 16(3), 147–163.
- Martellini, L. e Priaulet, P. (2001), *Fixed-Income Securities: Dynamic Methods for Interest Rate Risk Pricing and Hedging*, Chichester, John Wiley & Sons.
- McNelis, P. D. (2005). *Neural Networks in Finance: Gaining Predictive Edge in the Market*.
- Monteiro, O. (2005). *A hipótese das expectativas e a estrutura temporal de taxas de juro: um resumo da literatura e aplicação empírica ao caso português*.
- Moreira, B. (2011). *Modelização de empréstimos bancários de empresas não financeiras na Zona Euro: uma abordagem VAR/VECM*. Faculdade de Ciências, Instituto Superior das Ciências do Trabalho e da Empresa.
- Moshiri, S., & Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19(3), 201–217.
- Nelson C, & Siegel A. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*.
- Oliveira Arantes, B., Fróis de Caldeira, J., & Savino de Portugal, M. (2012). *Previsão da Estrutura a Termo da Taxa de Juros Brasileira Usando redes neuronais artificiais*.
- Pfaff, B. (2008). VAR, SVAR and SVEC Modelos: Implementation Within R Package vars. *Jornal of Statistical Software*, 27(4).
- Pfaff, B., & Stigler, M. (2013). Package “vars”.
- Ramos, F. (2011). *Cointegração, Modelos VAR E BVAR*.
- Riedmiller, M. (1994). Rprop-description and implementation details. *Report*, 5–6.
- Rojas, R. (1996). *Neural Networks - A Systematic Introduction*. Berlin: Springer-Verlag (Vol. 7).
- Shiller, R. J., & Huston McCulloch, J. (1990). The term structure of interest rates. *Handbook of Monetary Economics*, 1, 627–722.
- Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications With R Examples*.
- Sims, C. (1980). Macroeconomics and Reality.pdf. *Econometrica*, 1–48.

- Stimpson, A. J., & Cummings, M. L. (2014). *Assessing Intervention Timing in Computer-Based Education Using Machine Learning Algorithms*.
- Svensson, L. (1994). Estimating and interpreting forward interest rates: sweden 1992-1994. *NBER Working Paper Series*.
- Vicente, J., Tabak, B., 2008. Forecasting bond yields in the Brazilian fixed income market. *International Journal of Forecasting* 24, 490–497.
- Tabak, B. M., Sollaci, A. B., Gomes, G. M., & Cajueiro, D. O. (2012). Forecasting the yield curve for the Euro region. *Economics Letters*, 117(2), 513–516.
- Täppinen, J. (1998). *Interest rate forecasting with neuronal networks*.
- Tissot, H. C., Camargo, L. C., & Pozo, A. T. R. (n.d.). Treinamento de Redes Neurais Feedforward : comparativo dos algoritmos Backpropagation e Differential Evolution.
- Vasicek, O. (1977). An Equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177–188.
- Vela, D. (2013). Forecasting Latin-american yield curves: An artificial neural network approach.
- Vereda, L., Lopes, H., Kubrusly, J., & Pizzinga, A. (2014). Yield Curve Forecasts and the Predictive Power of Macro Variables in a VAR Framework. *Journal of Reviews on Global Economics*, 3, 377–393.
- Warner, Brad. Misra, M. (1996). Understanding Neural Networks as Statistical Tools. *The American Statistician*.
- Winandy, C.-E., Borges Filho, E., & Bento, L. V. (2007). Algoritmos para Aprendizagem Supervisionada.
- Yu, W. C., & Zivot, E. (2011). Forecasting the term structures of Treasury and corporate yields using dynamic Nelson-Siegel models. *International Journal of Forecasting*, 27(2), 579–591.
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175.
- Zhang, G. P., Patuwo, E. B., & Michael Y., H. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35–62.

7. ANEXOS

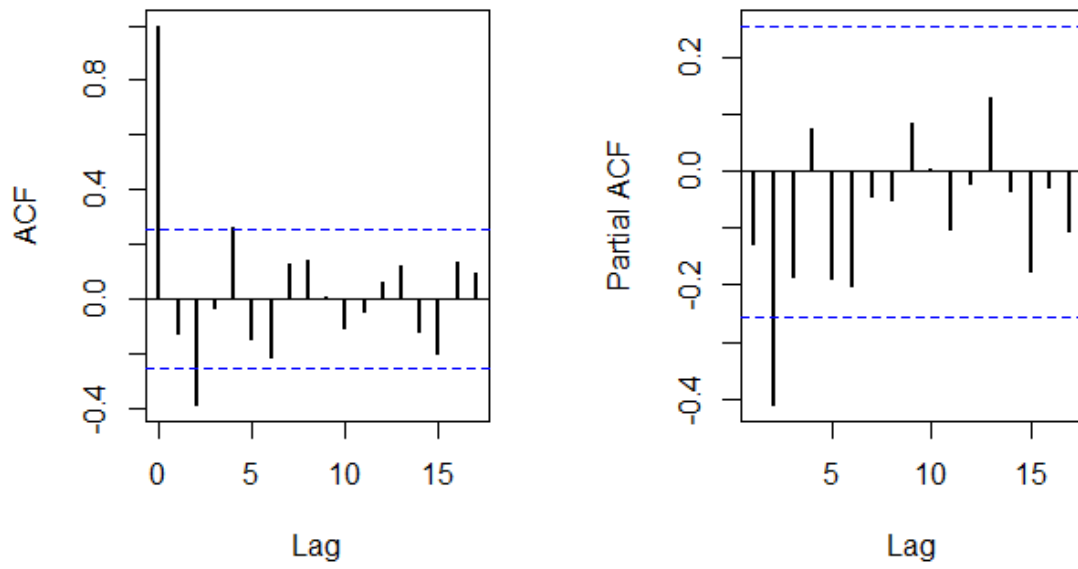


Figura 8: Gráficos ACF e PACF do parâmetro β_1

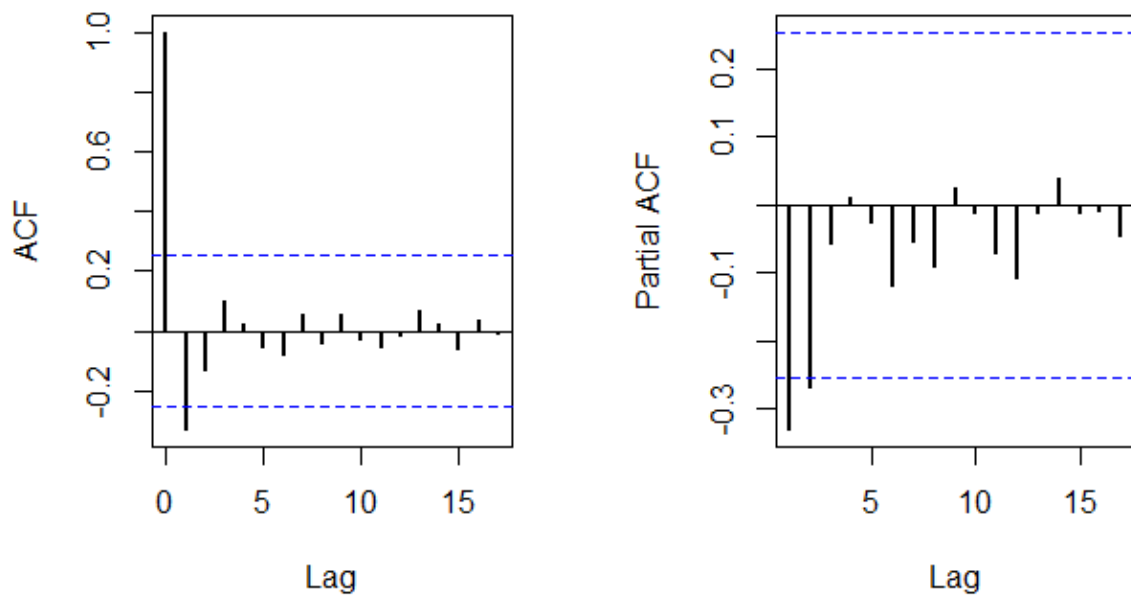


Figura 9: Gráficos ACF e PACF do parâmetro β_2

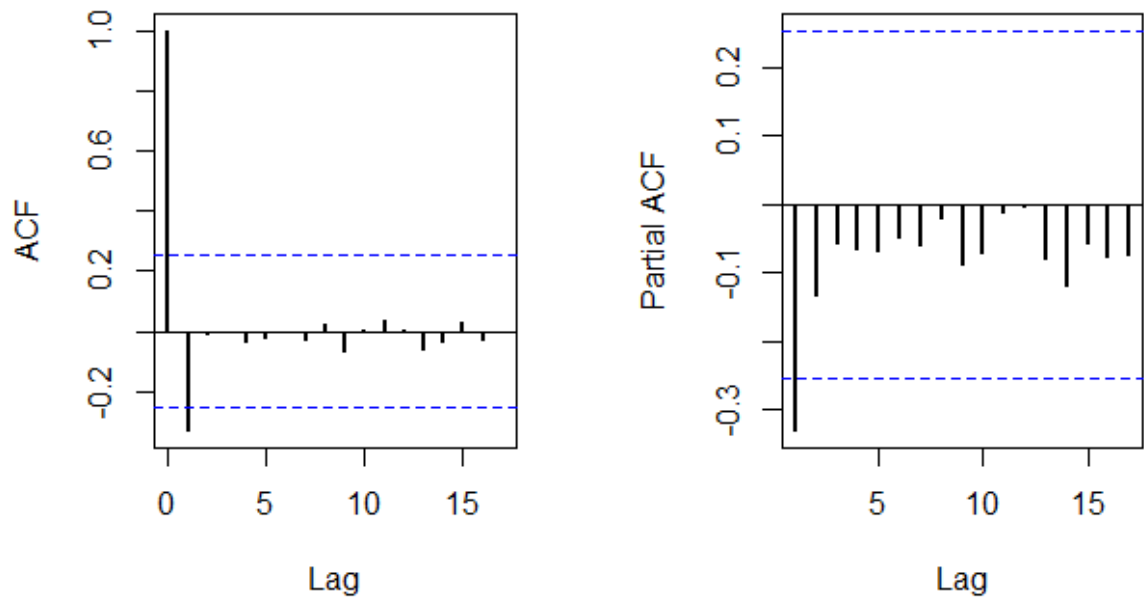


Figura 10: Gráficos ACF e PACF do parâmetro β_3

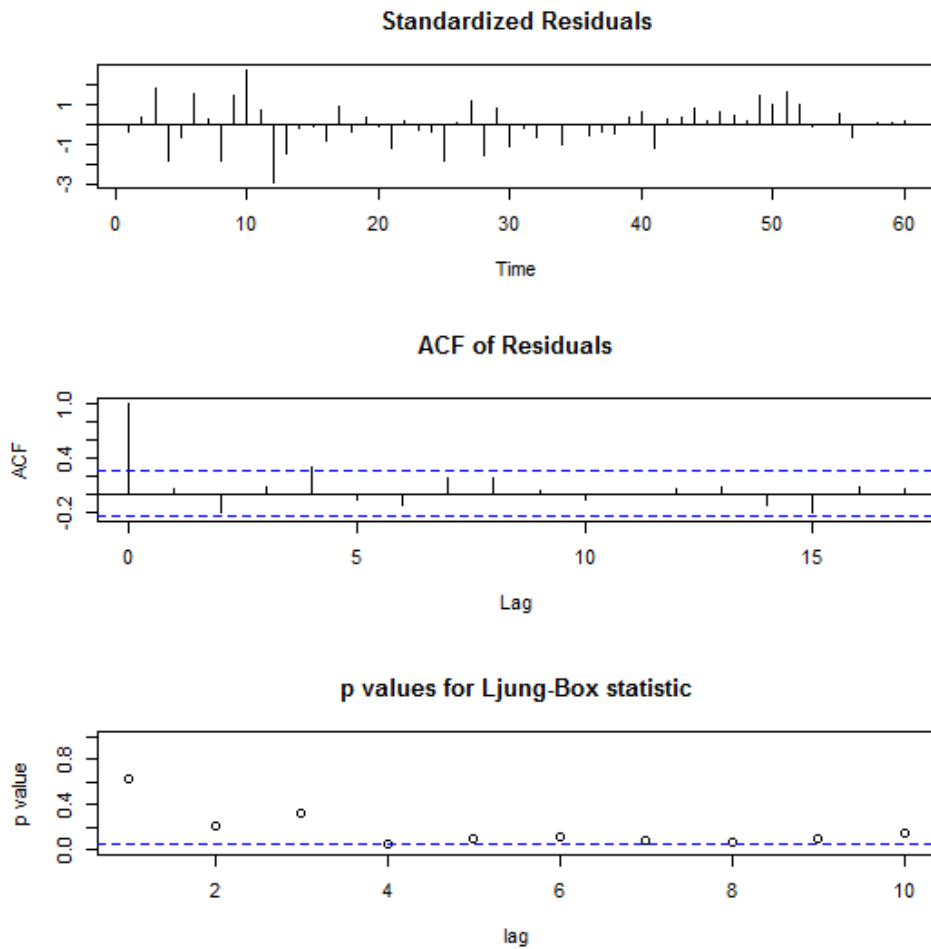


Figura 11: Análise dos resíduos do parâmetro β_1

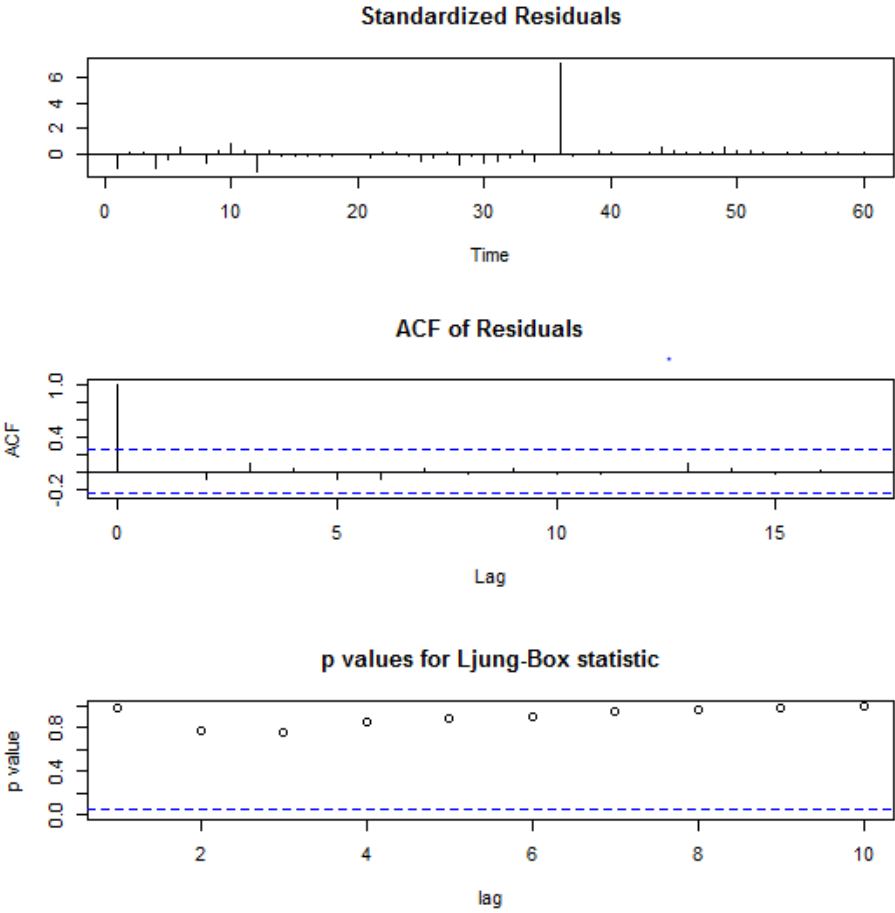


Figura 12: Análise dos resíduos do parâmetro β_2

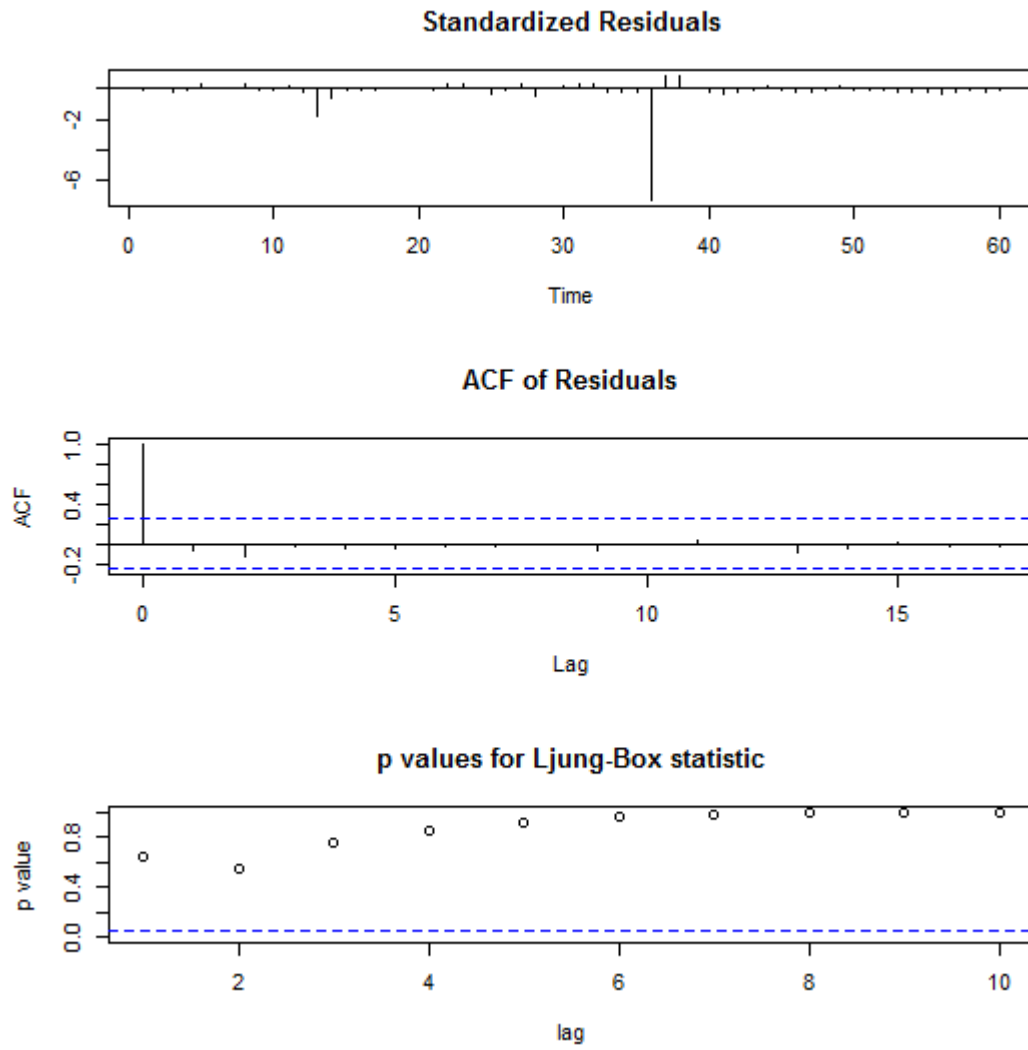


Figura 13: Análise dos resíduos do parâmetro β_3

Tabela 17: Configurações testadas à RNA para o parâmetro β_0

Configuração/Resultados						
Neurónios <i>hidden</i>	5	5	5	5	5	5
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.17038	0.19688	0.19831	0.20968	0.2327	0.23418
Neurónios <i>hidden</i>	10	10	10	10	10	10
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.2394	0.23644	0.23697	0.26039	0.2692	0.26579
Neurónios <i>hidden</i>	15	15	15	15	15	15
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.23654	0.21206	0.23157	0.24292	0.27055	0.26441

Tabela 18: Configurações testadas à RNA para o parâmetro β_1

Configuração/Resultados						
Neurónios <i>hidden</i>	5	5	5	5	5	5
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.17038	0.19688	0.19831	0.20968	0.2327	0.23418
Neurónios <i>hidden</i>	10	10	10	10	10	10
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.15843	0.15545	0.16386	0.22224	0.20748	0.21381
Neurónios <i>hidden</i>	15	15	15	15	15	15
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.16013	0.15464	0.15688	0.16635	0.18038	0.17665

Tabela 19: Configurações testadas à RNA para o parâmetro β_2

Configuração/Resultados						
Neurónios <i>hidden</i>	5	5	5	5	5	5
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.28513	0.0627	0.27037	0.29018	0.2992	0.29636
Neurónios <i>hidden</i>	10	10	10	10	10	10
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.30365	0.28841	0.3133	0.36887	0.35757	0.351
Neurónios <i>hidden</i>	15	15	15	15	15	15
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.28091	0.28433	0.27614	0.30827	0.31157	0.30985

Tabela 20: Configurações testadas à RNA para o parâmetro β_3

Configuração/Resultados						
Neurónios <i>hidden</i>	5	5	5	5	5	5
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.36385	0.36653	0.35889	0.38107	0.37589	0.37541
Neurónios <i>hidden</i>	10	10	10	10	10	10
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.3625	0.37139	0.36666	0.38042	0.3858	0.38659
Neurónios <i>hidden</i>	15	15	15	15	15	15
Nº máximo de iterações	10000	10000	10000	10000	10000	10000
<i>Threshold</i>	0.01	0.01	0.01	0.02	0.02	0.02
Taxa de Aprendizagem	0.03	0.02	0.01	0.03	0.02	0.01
Erro Médio	0.35956	0.36074	0.35532	0.36632	0.37106	0.37184

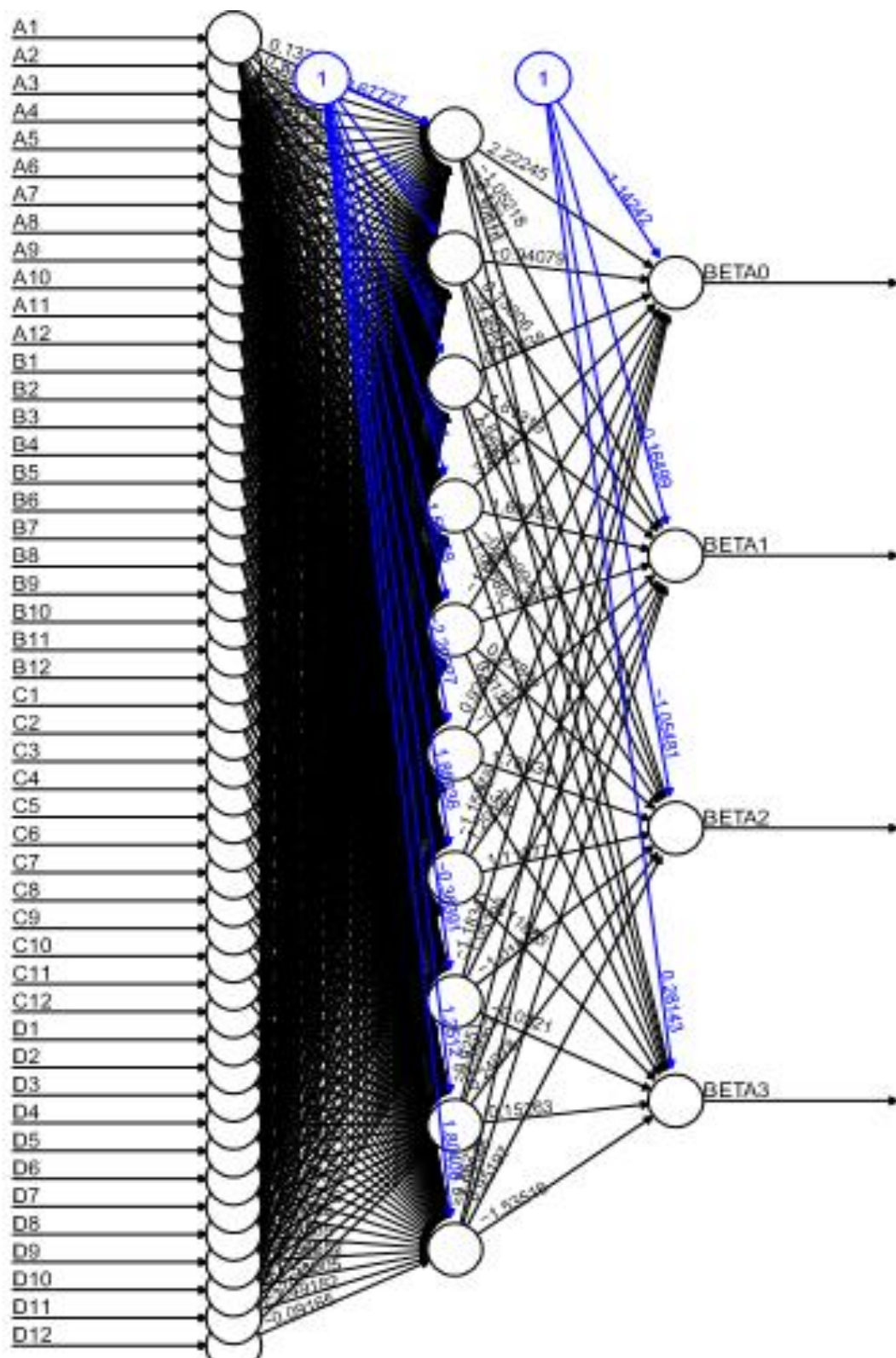


Figura 14: Rede Neuronal Conjunta

Figura 15: Rede Neuronal β_1

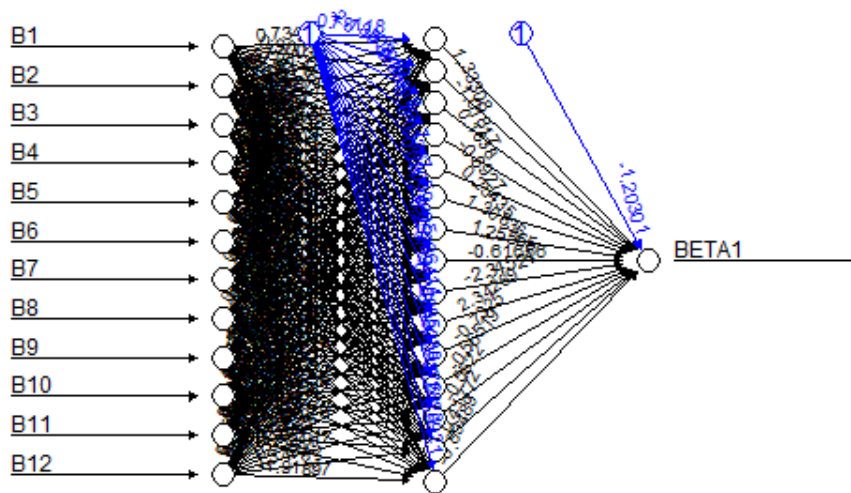


Figura 16: Rede Neuronal β_2

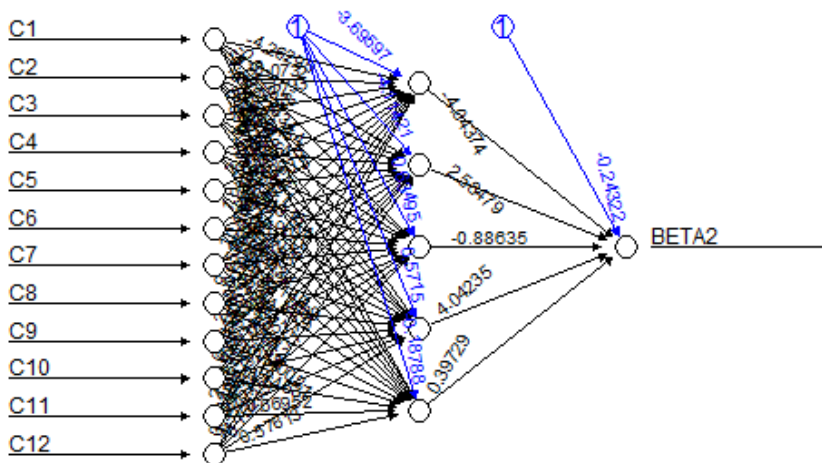


Figura 17: Rede Neuronal β_3

