



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Alocação das reservas internacionais com base em
previsão de fatores econômicos e financeiros
utilizando redes neurais recorrentes e modelo
Nelson-Siegel Dinâmico**

Ezer Fernandes da Silva

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientador
Prof. Dr. Gladston Luiz da Silva

Brasília
2019

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

Fa Fernandes da Silva, Ezer
 Alocação das reservas internacionais com base em previsão
 de fatores econômicos e financeiros utilizando redes neurais
 recorrentes e modelo Nelson-Siegel Dinâmico / Ezer Fernandes
 da Silva; orientador Gladston Luiz da Silva. -- Brasília,
 2019.
 69 p.

 Dissertação (Mestrado - Mestrado Profissional em
 Computação Aplicada) -- Universidade de Brasília, 2019.

 1. otimização de portfólio. 2. Banco Central do Brasil. 3.
 redes neurais. 4. curva de juros. 5. modelo Nelson-Siegel.
 I. Luiz da Silva, Gladston, orient. II. Título.

Dedicatória

A Deus, autor da Vida e consumidor da Salvação, que no meio das muitas batalhas me trouxe para mais perto d'Ele.

À minha esposa Míryam, que me incentivou a fazer este mestrado, mesmo sabendo que isso significaria renúncia de preciosos momentos juntos; que me deu suporte com amor nos momentos de estresse e de noites acordadas, animando-me o espírito; esta vitória é sua tanto quanto é minha.

À minha filha Eliza, nascida junto com o mestrado, sem cuja simpatia e energia esta dissertação teria sido terminada em metade do tempo, e cuja existência me traz muitas alegrias.

Aos meus pais, Marli e Almir, que me deram as condições de realizar o meu potencial.

Agradecimentos

Meus agradecimentos ao meu orientador Prof. Dr. Gladston Luiz da Silva pelo apoio e dedicação na correção de rumo do projeto: sua experiência e seus conselhos se provaram inestimáveis à feitura deste trabalho.

Meus agradecimentos aos colegas mestrandos, cuja organização facilitou a troca efetiva de informação e cujo exemplo de persistência me incentivou a tentar com mais afinco. Sob risco de omissão, gostaria de agradecer especialmente aos colegas Urias, Sérgio Leite, José Eduardo, Frank Ned, Renan e Yago pela companhia e apoio.

Meus agradecimentos aos professores do Mestrado em Computação Aplicada pela ministração de aulas que permitiram a aquisição de conhecimentos, teóricos e práticos, relevantes à realização da pesquisa; em especial, ao Prof. Dr. Guilherme Novaes Ramos por instilar o hábito da leitura de artigos científicos e ao Prof. Dr. Donald Matthew Pianto por instigar o interesse em aprofundar o conhecimento estatístico.

Meus agradecimentos à Juliana Fernandes, minha irmã, por revisar o texto desta dissertação com o cuidado e o olho clínico de quem tem talento e vocação para o magistério.

Meus agradecimentos ao Prof. Dr. Daniel Cajueiro pelas ponderações durante a qualificação, que me levaram a identificar pontos relevantes de melhoria no projeto.

Meus agradecimentos ao colega de trabalho Paulo Mergulhão, por corrigir o pré-projeto antes de sua submissão ao Programa de Pós-Graduação em Computação Aplicada.

Por fim, agradeço ao Banco Central do Brasil pelo apoio institucional e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) pelo fornecimento de acesso ao Portal de Periódicos.

Resumo

O problema da seleção de portfólio é central para a gestão de investimentos. O presente trabalho propõe um modelo que utiliza previsões sobre a curva de juros, com base em previsões do estado futuro da economia, para sugestão de mudanças na alocação de um portfólio de renda fixa. Para realizar tais previsões, em um primeiro passo, foram utilizados indicadores de antecipação para a previsão de variáveis econômicas relevantes; posteriormente, o modelo Nelson-Siegel Dinâmico foi utilizado para decompor a curva de juros em séries de fatores, que foram então previstos com base nas previsões sobre variáveis macroeconômicas, combinadas às previsões dos analistas por meio do modelo Black-Litterman. As previsões foram feitas por redes neurais recorrentes, cujo desempenho foi comparado aos de outros modelos. O modelo proposto foi aplicado às reservas internacionais do Brasil e seu desempenho foi avaliado em comparação a uma carteira de referência. As redes neurais recorrentes, do tipo LSTM e *Minimal* RNN, não tiveram desempenho superior a um modelo autorregressivo na previsão das variáveis macroeconômicas. Já na previsão dos fatores do modelo Nelson-Siegel Dinâmico, a rede neural do tipo LSTM teve desempenho superior ao modelo autorregressivo, porém o retorno das carteiras sugeridas pela estratégia foi similar ao retorno das carteira de referência.

Palavras-chave: otimização de portfólio, Banco Central do Brasil, redes neurais, fatores macroeconômicos, curva de taxa de juros, modelo Nelson-Siegel Dinâmico

Abstract

The portfolio selection problem is central for investment management processes. The present work proposes a model that uses forecasts about the yield curve, based on forecasts of the future state of the economy, to suggest changes in the allocation of a fixed-income portfolio. For such prediction, at first we used leading indicators to predict relevant economic variables; afterwards, we used the Dynamic Nelson-Siegel model to decompose the yield curve into factor series and then we predicted it using the forecasted economic variables as inputs, combined to the analysts' views using the Black-Litterman model. We used recurrent neural networks for forecasting, which had its performances compared to other models. The proposed model was finally applied to the Brazilian international reserves and its performance was evaluated relative to a benchmark portfolio. The recurrent neural networks, with LSTM and *Minimal* RNN layers, did not have a superior performance at predicting macroeconomic variables when compared to an autoregressive model. However, the LSTM neural network had a performance superior to the performance of an autoregressive model at predicting the factors of the Dynamic Nelson-Siegel model. The returns of the suggested portfolios were similar to the returns of the benchmark portfolio.

Keywords: portfolio optimization, Central Bank of Brazil, neural networks, macroeconomic factors, yield curve, Dynamic Nelson-Siegel model

Sumário

1	Introdução	1
1.1	Problema de Pesquisa	1
1.2	Justificativa	3
1.3	Objetivo Geral	3
1.4	Objetivos Específicos	3
1.5	Estrutura da Dissertação	4
2	Revisão da Literatura	6
2.1	Fundamentação Teórica	6
2.1.1	Seleção de Portfólios	6
2.1.2	Modelagem em Fatores	8
2.1.3	Precificação de Ativos	9
2.1.4	Modelo <i>Black-Litterman</i>	10
2.1.5	Modelagem da Curva de Juros	13
2.1.6	Aprendizado de Máquina em Finanças	17
2.1.7	Aprendizado de Máquina <i>Online</i>	21
2.1.8	Redes Neurais Artificiais	22
2.1.9	Redes Neurais Recorrentes	23
2.1.10	Florestas Aleatórias	29
2.2	Trabalhos Correlatos	30
3	Metodologia de Pesquisa	31
3.1	Entendimento do Negócio	31
3.2	Compreensão dos Dados	33
3.3	Preparação dos Dados	35
3.3.1	Interpolação e <i>Resampling</i>	35
3.3.2	Diferenciação Fracional	35
3.4	Modelagem	36
3.4.1	Previsão das Variáveis Macroeconômicas	36

3.4.2	Previsão da Curva de Juros	39
3.4.3	Sugestão de Alocação do Portfólio	39
3.5	Avaliação	40
3.5.1	Validação Cruzada	41
3.5.2	Métricas de Avaliação	42
3.6	Implantação	43
4	Resultados	44
4.1	Previsão das Variáveis Macroeconômicas	44
4.1.1	Seleção das Variáveis e Redução de Dimensionalidade	44
4.1.2	Ajuste dos Hiperparâmetros	47
4.1.3	Comparação do Desempenho dos Modelos	53
4.2	Previsão dos Fatores da Curva de Taxas de Juros	53
4.2.1	Seleção das Variáveis de Entrada	54
4.2.2	Fatores do Modelo Nelson-Siegel Dinâmico	54
4.3	Implementação da Estratégia de Alocação Sugerida	55
5	Conclusões e Trabalhos Futuros	58
5.1	Conclusões	58
5.2	Trabalhos futuros	59
	Glossário	61
	Referências	65

Lista de Figuras

2.1 Exemplo de fronteira eficiente.	7
2.2 Fluxo do modelo Black-Litterman. Adaptado de Mishra <i>et al</i> [1].	11
2.3 Exemplo de uma curva de juros.	14
2.4 Ilustração da equivalência entre uma rede neural recorrente e uma rede neural multicamadas em que cada camada é uma cópia da rede recorrente e passa mensagem para a camada seguinte. Obtido de Olah (2015) [2].	24
2.5 Visão geral da arquitetura da célula de memória LSTM. Adaptado de Olah (2015) [2].	24
2.6 Diagrama da célula de uma rede LSTM. Adaptado de Goodfellow <i>et al</i> [3].	26
2.7 Diagrama da célula de uma rede <i>Minimal</i> RNN. Obtido de Chen (2017) [4].	27
3.1 Alocação das reservas por classe de ativo (em 08/05/2019)	32
3.2 Alocação das reservas por moeda (em 08/05/2019)	32
3.3 Comparação entre os métodos de interpolação spline e PCHIP.	35
4.1 Ilustração de uma das redes LSTM treinadas e avaliadas no ajuste dos hiperparâmetros. Esta rede específica tem 3 neurônios na camada recorrente e valor de <i>dropout</i> de 0.4.	48
4.2 Mapas de calor dos RMSEs obtidos na previsão das variáveis macroeconômicas utilizando método de busca em grade com redes LSTM.	50
4.3 Fatores do modelo Nelson-Siegel Dinâmico para a curva de juros norte-americana entre 31/01/1978 e 31/05/2019.	54
4.4 Previsão do fatores da curva de juros com a rede LSTM para os próximos 12 meses.	56

Lista de Tabelas

3.1 Índices utilizados para modelar a curva de juros americana	33
3.2 Variáveis macroeconômicas selecionadas	33
3.3 Indicadores pré-selecionados para previsão das variáveis econômicas	34
4.1 Variáveis selecionadas como relevantes para a previsão da variável Inflação .	44
4.2 Variáveis selecionadas como relevantes para a previsão da variável Desemprego	45
4.3 Variáveis selecionadas como relevantes para a previsão da variável Produção	
Industrial	45
4.4 Variáveis selecionadas como relevantes para a previsão da variável Lucros	
Corporativos	46
4.5 Número de componentes principais selecionados como variáveis de entrada	
e a variância explicada por eles para cada variável macroeconômica.	47
4.6 <i>Rankings</i> das 5 redes LSTM com melhor desempenho por fator econômico .	49
4.7 <i>Rankings</i> das 5 redes <i>Minimal</i> RNN com melhor desempenho por fator	
econômico	52
4.8 Desempenho dos modelos de redes neurais recorrentes e de modelo autorre-	
gressivo.	53
4.9 Desempenho (RMSEs) da rede LSTM e do modelo autorregressivo para os	
últimos 12 meses dos fatores da curva de juros.	55

Lista de Abreviaturas e Siglas

APT Teoria da Precificação por Arbitragem, do inglês *Arbitrage Pricing Theory*.

BCB Banco Central do Brasil.

CAPM Modelo de Precificação de Ativos Financeiros, do inglês *Capital Asset Pricing Model*.

CRISP-DM Cross-Industry Standard Process for Data Mining.

Depin Departamento das Reservas Internacionais.

Deris Departamento de Riscos Corporativos e Referências Operacionais.

GRC Comitê de Governança, Riscos e Controle.

GRU *Gated Recurrent Unit*.

IndRNN *Independently Recurrent Neural Network*.

LSTM *Long Short-Term Memory*.

NSD Modelo Nelson-Siegel Dinâmico.

NSDR Modelo Nelson-Siegel Dinâmico Rotacionado.

PCA Análise de Componentes Principais, do inglês *Principal Component Analysis*.

PCHIP Piecewise Cubic Hermite Interpolating Polynomial.

RMSE Raíz do Erro Quadrático Médio.

RNN Rede Neural Recorrente, do inglês *Recurrent Neural Network*.

spline Piecewise Polynomial.

SRU *Simple Recurrent Unit.*

VaR *Value at Risk.*

Capítulo 1

Introdução

1.1 Problema de Pesquisa

As reservas internacionais de um país consistem de ativos financeiros denominados em moeda estrangeira, negociáveis no mercado financeiro internacional, que estão sob a administração de um banco central ou de outra autoridade monetária. A gestão de reservas internacionais é um processo que busca garantir a pronta disponibilidade dos ativos estrangeiros à autoridade monetária para a execução de objetivos definidos por um país ou união, sendo esta, em geral, responsável pelo gerenciamento das reservas e dos riscos associados [5].

No contexto brasileiro, compete ao Banco Central do Brasil (BCB) ser o depositário das reservas internacionais, de ouro e de direitos especiais de saque [6]. No âmbito do Banco Central, compete à Diretoria Colegiada, em reunião do Comitê de Governança, Riscos e Controle (GRC), definir as diretrizes e parâmetros (carteira de referência) para a administração das reservas internacionais. Compete ao Departamento de Riscos Corporativos e Referências Operacionais (Deris) apoiar o GRC nas discussões técnicas de definição da carteira de referência e de limites para os desvios em relação à carteira. Por fim, compete ao Departamento das Reservas Internacionais (Depin) administrar as reservas internacionais do país, por meio de operações com essas reservas, observados os parâmetros estabelecidos pelo GRC [7].

As reservas internacionais brasileiras totalizam cerca de US\$ 384,165 bilhões de dólares em valor de mercado¹, alocadas na maior parte em instrumentos de renda fixa, especialmente em títulos governamentais soberanos², mas também em depósitos em bancos

¹Reservas Internacionais - Total - mensal, posição em março de 2019. Disponível em: <https://www3.bcb.gov.br/sgspub/>

²Títulos soberanos são emitidos pelos governos centrais para financiar seus balanços fiscais [8].

centrais e em organismos supranacionais³, títulos de agências governamentais de diferentes países, títulos de organismos supranacionais e em depósitos bancários a prazo fixo [9]. A gestão das reservas internacionais do Brasil está ancorada em sistema de controle e aferição diários de resultados e de acompanhamento dos investimentos, baseado em três pilares: o portfólio de referência, que define o perfil de risco e retorno esperados; os limites operacionais para os desvios autorizados em relação a esse portfólio de referência; e os critérios de mensuração de resultados [9].

A política de investimento é definida pela Diretoria Colegiada com a preocupação de atender aos compromissos externos e de se ter uma alocação com características anticíclicas e de redução da exposição a oscilações cambiais. Uma vez considerada a política de cobertura cambial da dívida externa bruta e de alocação com caráter anticíclico, o investimento das reservas internacionais é realizado com o auxílio de técnicas de otimização risco-retorno de carteira, observando-se critérios de segurança, liquidez e rentabilidade, nessa ordem de prioridade [9].

Em dezembro de 2018, a carteira de referência definia a alocação por moeda da seguinte forma: 89,93% em dólar norte-americano, 5,13% em euro, 1,92% em libra esterlina, 1,49% em iene, 0,75% em ouro, 0,47% em dólar canadense e 0,30% em dólar australiano, refletindo a composição da dívida externa soberana. Além disso, a carteira de referência definia a alocação de 93,18% das reservas em títulos governamentais [9].

No investimento das reservas internacionais, o BCB monitora os riscos de mercado⁴, de liquidez⁵, de crédito⁶ e operacional⁷. Para avaliar o risco de mercado das reservas internacionais, o BCB utiliza o Value at Risk (VaR)⁸ como principal indicador da perda esperada de um portfólio para um nível de confiança e um horizonte de tempo dados. Para mitigar o risco de liquidez, são adotados limites operacionais no que se refere aos montantes e aos prazos de investimento. O risco de crédito é reduzido por grande parte das reservas estar alocada em títulos soberanos e, majoritariamente, em títulos do Tesouro Americano [9].

³Supranacionais são organismos multilaterais que promovem o desenvolvimento econômico de seus países membros. Estes podem emitir títulos e agir como intermediários financeiros. Exemplos de supranacionais são o Banco Mundial e o Banco Interamericano de Desenvolvimento [8].

⁴Risco de mercado é o risco de perdas decorrentes de variações nos preços dos ativos [10].

⁵Risco de liquidez corresponde à facilidade com que um ativo pode ser vendido por seu valor real [10].

⁶Risco de crédito se refere ao risco de que o emissor de um ativo não seja capaz de fazer pagamentos tempestivos do valor principal ou dos juros deste ativo [10].

⁷Risco operacional pode ser definido como a estimativa de perdas resultantes de falha, deficiência ou inadequação de processos internos, pessoas e sistemas, ou de eventos externos [9].

⁸O BCB calcula o VaR diário das reservas internacionais com nível de confiança de 95% [9].

1.2 Justificativa

Para definir estratégias ativas de investimento que tragam retornos diferenciais positivos em relação à carteira de referência, o Depin realiza análises econômico-financeiras independentes e análises quantitativas com dados financeiros históricos. Nesse contexto, o Depin pode se beneficiar de modelos que o auxiliem a definir uma alocação das reservas internacionais que reflita suas visões sobre condições econômicas futuras, buscando a maximização do retorno, respeitadas as restrições estabelecidas e a tolerância a risco da alta administração.

Para melhorar o aproveitamento dos estudos realizados pelos analistas do Depin, torna-se imprescindível a utilização de modelos de alocação de ativos que permitam incorporar as previsões para diversas variáveis da economia. Dado o nível de incerteza inerente às previsões econômicas, é essencial que níveis de confiança estejam atrelados a elas, combinando-as também com informações disponíveis *a priori* para alcançar uma sugestão de alocação ancorada à realidade.

1.3 Objetivo Geral

O objetivo deste projeto é desenvolver uma ferramenta que incorpore previsões econômicas diretamente ao processo de otimização da alocação da carteira de renda fixa das reservas internacionais, por meio de previsões com redes neurais combinadas com as previsões dos analistas e utilizando essas previsões das variáveis macroeconômicas para prever a curva de juros⁹ de títulos governamentais americanos e fornecer sugestões de estratégias de investimento das reservas. O resultado deste trabalho deverá ser incorporado ao processo de gestão ativa das reservas através da implementação em sistema.

Dado que os títulos de renda fixa precificados em dólar compõem a maior parte das reservas internacionais, o modelo proposto terá como escopo a previsão da curva das taxas de juros norte-americana e a utilização desta previsão para sugerir mudanças na alocação das reservas internacionais.

1.4 Objetivos Específicos

Para atingir o objetivo geral, foram definidos os seguintes objetivos específicos:

1. Selecionar variáveis macroeconômicas relevantes, de frequência mensal, para utilização em modelo de fatores macroeconômicos.

⁹A curva de juros demonstra a relação entre taxas de juros e a maturidade de títulos de renda fixa, sendo estimada dos preços dos títulos [11].

2. Selecionar indicadores econômicos e financeiros relevantes e realizar as transformações necessárias para utilizá-los como *features* no desenvolvimento de modelos de previsão das variáveis macroeconômicas selecionadas previamente.
3. Treinar redes neurais recorrentes para realizar a previsão do próximo valor das variáveis macroeconômicas.
4. Modificar o modelo Black-Litterman¹⁰ para permitir combinar as previsões dadas pelas redes neurais com as previsões dos analistas para as variáveis econômicas.
5. Utilizar o modelo Nelson-Siegel Dinâmico para decompor as curvas de taxas de juros de títulos governamentais americanos em fatores e utilizar as variáveis macroeconômicas para previsão desses fatores.
6. Gerar a curva de juros prevista, utilizando a previsão para os fatores do modelo Nelson-Siegel dinâmico, e utilizar a curva de juros atual e a curva de juros prevista para sugerir estratégias de investimento que levem a uma melhor alocação das reservas.
7. Avaliar o desempenho do modelo proposto por meio de: avaliação do desempenho na previsão dos fatores macroeconômicos; avaliação do desempenho na previsão das curvas de taxas de juros; e comparação do desempenho da carteira de referência com desempenho de portfólio sugerido pelo modelo.

1.5 Estrutura da Dissertação

Esta dissertação é organizada em cinco capítulos, contando com o presente capítulo 1, cuja finalidade é introduzir o leitor ao tema da gestão das reservas internacionais do Brasil e apresentar as principais motivações e objetivos do trabalho realizado.

No Capítulo 2, é feita uma revisão dos conceitos e modelos financeiros e de aprendizagem de máquina necessários para os próximos capítulos, bem como são apresentados outros trabalhos relacionados com o tema deste projeto.

No Capítulo 3, a metodologia de pesquisa adotada para a pesquisa é descrita, estruturada de acordo com as fases de um projeto de mineração de dados proposta pelo *Cross-Industry Standard Process for Data Mining* (CRISP-DM): Entendimento do Negócio (3.1), Compreensão dos Dados (3.2), Preparação dos Dados (3.3), Modelagem (3.4),

¹⁰A ideia central do modelo definido por Black e Litterman [12] é tratar as previsões dos analistas como observações sobre eventos futuros, com incerteza e ruído associados, utilizáveis para realizar inferência estatística dos parâmetros de um modelo do retorno de ativos, sendo, portanto, um modelo *bayesiano* [13].

Avaliação (3.5) e Implantação (3.6). Esta estrutura induziu a pesquisa a focar em problemas relevantes aos objetivos do processo de gestão ativa das reservas internacionais.

O Capítulo 4 apresenta os resultados da aplicação da metodologia, descrita no capítulo que o antecedeu, utilizando-se de dados reais. Os resultados foram comparados a outros modelos-padrão da área de previsão de séries temporais e aos resultados esperados do portfólio de referência.

Por fim, o Capítulo 5 trata das conclusões obtidas nesta dissertação, bem como de possíveis trabalhos futuros.

Capítulo 2

Revisão da Literatura

2.1 Fundamentação Teórica

2.1.1 Seleção de Portfólios

A seleção de portfólios é um dos problemas mais importantes na gestão de investimentos [14]. O problema da alocação de portfólio consiste na alocação de recursos em um número definido de ativos. Na ausência de um ativo livre de risco¹, o problema da alocação ótima resume-se à escolha de um vetor de pesos que represente a alocação relativa nos ativos e que maximize o alcance dos objetivos do investidor, utilizando como entrada do processo informações sobre o perfil do investidor e sobre o mercado, disponíveis no momento da tomada de decisão [15][16].

O processo de seleção de portfólios pode ser dividido em dois estágios: o primeiro inicia-se com os dados históricos e a experiência do gestor e termina com crenças sobre os desempenhos futuros de ativos; o segundo inicia-se com as crenças sobre os desempenhos futuros e tem por saída a escolha do portfólio de ativos. A teoria moderna de portfólio se preocupa com o segundo estágio, de seleção de um portfólio de ativos dadas estimativas de retornos futuros, não tratando do processo de estimativa de retornos futuros [17].

A teoria moderna de portfólio, também conhecida como análise de média-variância, é um modelo matemático para a definição de portfólios de ativos cujo retorno esperado é maximizado para cada nível de risco dado. Os portfólios de maior retorno para cada nível de risco definem uma fronteira eficiente de portfólios [17], conforme ilustrado pela Figura 2.1. Markowitz, em artigos fundacionais da teoria moderna de portfólio, formalizou analiticamente a relação entre risco e retorno na seleção de portfólios ótimos, demonstrando a necessidade de se levar em conta o retorno e o risco do portfólio como um todo, e não

¹Um ativo livre de risco tem um retorno futuro certo. Os títulos emitidos pelo Departamento do Tesouro dos Estados Unidos (*treasuries*) são considerados pelos participantes dos mercados como ativos livres de risco, embora, tecnicamente, todos os ativos financeiros carreguem um certo grau de risco [8].

apenas dos ativos individualmente. Por definição, nenhum portfólio pode ficar acima da fronteira eficiente. Na prática, no caso de as expectativas de retorno utilizadas na otimização se mostrarem incorretas, algum portfólio pode ter retorno maior que o portfólio de mesmo risco na fronteira eficiente [17][18].

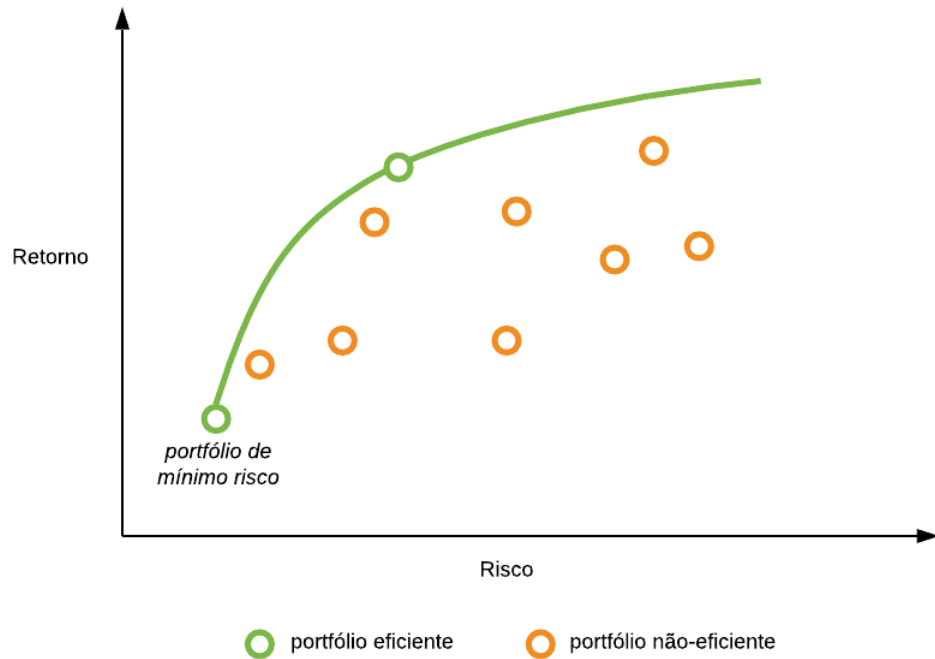


Figura 2.1: Exemplo de fronteira eficiente.

De acordo com a teoria moderna de portfólio [17][18], o retorno R_p de um portfólio p é dado pela fórmula:

$$E(R_p) = \sum_i w_i E(R_i) \quad (2.1)$$

Onde R_i e w_i são o retorno de um ativo i e o peso de um ativo i no portfólio p , respectivamente.

A variância do retorno do portfólio é definida por:

$$\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij} \quad (2.2)$$

Onde σ_i é o desvio padrão dos retornos de um ativo i e ρ_{ij} é o coeficiente de correlação entre os retornos dos ativos i e j .

A fronteira eficiente é calculada por minimizar a seguinte expressão:

$$w^T \Sigma w - q * R^T w \quad (2.3)$$

Onde:

- w é um vetor de pesos de um portfólio em que $\sum_i |w_i| = 1$;
- Σ é a matriz de covariância para os retornos dos ativos no portfólio;
- q é um parâmetro de tolerância a risco, onde $q \geq 0$ e 0 resulta no portfólio com menor risco e ∞ resulta em um portfólio infinitamente afastado na fronteira;
- R é um vetor de retornos esperados;
- $w^T \Sigma w$ é a variância do retorno do portfólio;
- $R^T w$ é o retorno esperado do portfólio.

Esse procedimento deve ser realizado para diversos valores do parâmetro q .

2.1.2 Modelagem em Fatores

Diversificação é um dos conceitos fundamentais na gestão de risco de investimentos. A ideia é que, ao combinar categorias de investimento, pode-se limitar a volatilidade de preço do portfólio sem afetar os retornos esperados. Entretanto, desde a crise financeira de 2007-2009, quando quase todas as categorias de investimento declinaram massivamente, o senso comum quanto à diversificação tem sido criticado, levantando a questão sobre a possibilidade de se diversificar efetivamente um portfólio. Investimento utilizando fatores é uma das respostas a essa questão [19].

Ativos financeiros oferecem prêmios de retorno para compensar investidores por exposição a perdas. Embora seja possível definir fator como a causa dos prêmios de risco dos ativos [20], este não é o caso para alguns fatores, em que o prêmio pode até estar associado a um risco mais baixo, como, por exemplo, o fator “baixa volatilidade” [19]. Por esse motivo, pode-se melhor definir fator como um atributo subjacente que explica parte do retorno de diversos ativos [19]. Sob essa ótica, o retorno de um ativo é a combinação linear do retorno de fatores de risco, mais um termo de ruído [20].

Existem dois tipos básicos de fatores de risco: fatores macroeconômicos, tais como crescimento econômico, inflação, volatilidade e risco demográfico; e fatores de estilo de investimento, tais como momento, valor e tamanho das companhias. Fatores macroeconômicos tendem a afetar de maneira pervasiva todos os investidores e classes de ativos, com variação no nível de exposição de cada investidor e cada ativo aos fatores de risco [20].

2.1.3 Precificação de Ativos

A precificação de ativos refere-se a um desenvolvimento formal de dois princípios: preços são determinados pela precificação de mercado via oferta e demanda, e preços de derivativos são calculados de maneira que eles são livres de arbitragem com respeito aos preços dos ativos subjacentes [21].

2.1.3.1 Modelo de Precificação de Ativos Financeiros (CAPM)

Utilizando a Teoria Moderna de Portfólio como base, Sharpe [22] e Lintner [23] construíram o Modelo de Precificação de Ativos Financeiros (CAPM). Este modelo é utilizado para a precificação de ativos e descreve o relacionamento entre risco sistêmico e retorno esperado de um ativo como a combinação do retorno de um ativo livre de risco com um prêmio de risco que leva em conta o retorno esperado do mercado e a sensibilidade do ativo ao risco não-diversificável (também conhecida como beta) [21]. O modelo CAPM é definido da seguinte forma:

$$E(r_i) = r_f + \beta_i(E(r_m) - r_f) \quad (2.4)$$

Onde:

- $E(r_i)$ é o retorno esperado do ativo
- r_f é a taxa de retorno do ativo livre de risco
- β_i é a sensibilidade dos retornos do ativo i aos retornos esperados do mercado
- $E(r_m)$ é o retorno esperado do mercado
- $E(r_m) - r_f$ é também conhecido como prêmio de mercado
- $E(r_i) - r_f$ é também conhecido como prêmio de risco

O beta (β) é calculado da seguinte forma:

$$\beta_i = \frac{Cov(r_i, r_m)}{Var(r_m)} = \rho_{i,m} \frac{\delta_i}{\delta_m} \quad (2.5)$$

Onde $\rho_{i,m}$ é o coeficiente de correlação entre o ativo e o mercado.

No CAPM, o prêmio de risco do ativo é igual a $E(r_i) - r_f = \beta_i(E(r_m) - r_f)$, ou seja, é igual ao prêmio de mercado multiplicado pelo β do ativo.

O modelo CAPM é um dos modelos mais utilizados na precificação de ativos e na avaliação de portfólios. Entretanto, o modelo tem falhado em inúmeros testes empíricos, o que implica que a maioria de suas aplicações é inválida, apesar de manter sua alta popularidade devido à sua simplicidade e ao seu apelo intuitivo [24].

2.1.3.2 Teoria da Precificação por Arbitragem (APT)

A Teoria da Precificação por Arbitragem (APT) é uma teoria geral de precificação de ativos que define o retorno de um ativo como um modelo linear de fatores, em que a sensibilidade a cada fator é representada por um coeficiente beta [21]. O retorno dos ativos pode ser definido da seguinte forma:

$$r_i = a_i + \sum_{n=1}^N (b_{i,n} F_n) + \epsilon_i \quad (2.6)$$

Onde:

- a_i é uma constante para o ativo i
- F_n é um fator de risco
- N é o número de fatores
- $b_{i,n}$ é a sensibilidade do ativo i ao fator n
- ϵ_i é uma variável aleatória com valor esperado zero

A intuição por trás do modelo APT é a de que os componentes idiossincráticos dos retornos dos ativos não carregam preços de risco, já que investidores podem mitigar o risco através de suficiente diversificação em um portfólio. Dessa maneira, as contribuições de retorno esperadas de um ativo deveriam ser relacionadas à covariância dos ativos com os fatores comuns [21].

2.1.4 Modelo *Black-Litterman*

Um dos problemas de modelos quantitativos que utilizam a teoria moderna de portfólio para otimização da alocação é que, sem imposição de restrições sobre as alocações, o resultado tende a ter posições vendidas muito grandes e, com restrição de se ter apenas posições compradas, o resultado tende a ser uma solução de canto, com peso zero em muitos ativos e peso grande em poucos ativos, muitas vezes com baixa capitalização [25]. Esses resultados injustificados tendem a ser originados pelo fato de que retornos esperados são difíceis de se estimar e de que modelos de alocação de ativos tendem a ser muito sensíveis às suposições de retorno utilizadas.

Para tentar mitigar os problemas descritos acima, Black e Litterman propuseram um modelo matemático que combina retornos esperados para os ativos obtidos por otimização reversa a partir dos pesos de um portfólio que reflita a capitalização dos ativos, utilizando o modelo de precificação de ativos financeiros (CAPM) e as *views* dos investidores sobre retornos esperados como forma de desviar do retorno esperado do portfólio global [25].

Do ponto de vista estatístico, a ideia central do modelo Black-Litterman é tratar as *views* do gestor como observações imprecisas do futuro, úteis para realizar inferência estatística sobre os parâmetros de um modelo subjacente para os retornos dos ativos. Pode-se dizer que este modelo é um caso especial de um modelo bayesiano em que o retorno dos ativos é uma distribuição normal multivariada com média no vetor de retornos esperados, há uma função de verossimilhança normal para a regressão das *views* do investidor, e *a priori* é a distribuição normal que é ótima em relação ao portfólio de mercado [13].

De fato, o modelo pode ser considerado uma rede bayesiana em que os vetores aleatórios de retornos esperados e de *views* são condicionalmente independentes dado o vetor de parâmetros do modelo [13]. A Figura 2.2 sintetiza os resultados dos passos principais do modelo Black-Litterman. É importante notar que a obtenção dos pesos revisados do portfólio não faz parte do escopo do modelo Black-Litterman, mas é obtido a partir da aplicação de um método de otimização de portfólio, como a análise de média-variância proposta por Markowitz, utilizando-se das expectativas de retorno revisadas que são o produto final do modelo Black-Litterman.

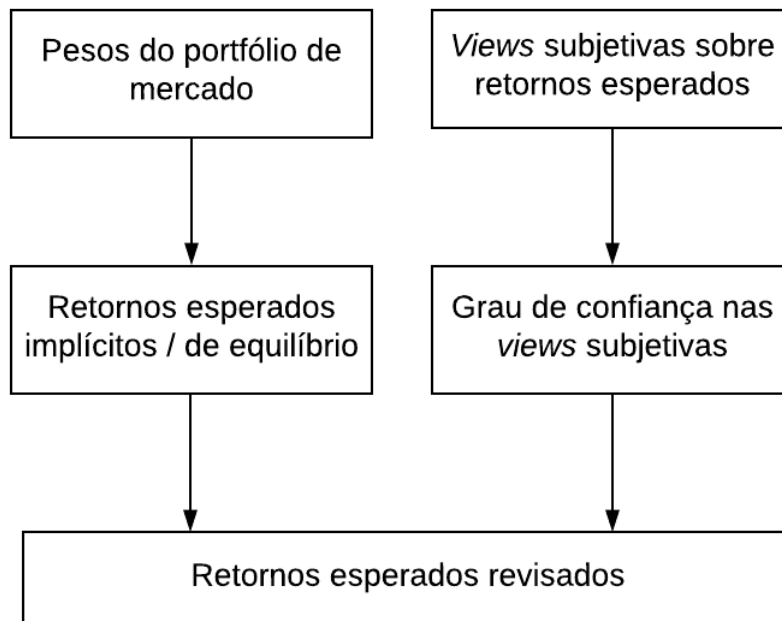


Figura 2.2: Fluxo do modelo Black-Litterman. Adaptado de Mishra *et al* [1].

Outro aspecto-chave do modelo Black-Litterman é que o gestor também deve especificar um nível de incerteza para cada *view*, informação utilizada como fonte independente de ruído [13]. Com uma previsão de retornos para os ativos que reflita o retorno esperado de um portfólio global em equilíbrio e o entendimento dos investidores, espera-se

que o processo de seleção de portfólio tenha como resultado um portfólio ou conjunto de portfólios com melhor perfil de risco-retorno.

Dadas as entradas [25]:

- k é o número de *views*
- n é o número de ativos utilizados
- w são os pesos dos ativos proporcionais à capitalização de mercado
- Σ é a matriz de covariância entre os ativos
- r_f é a taxa livre de risco
- δ é o coeficiente de aversão a risco do portfólio de mercado
- τ é uma medida de incerteza das estimativas *a priori* dos retornos

O modelo Black-Litterman é definido matematicamente da seguinte forma [25]:

1. Primeiramente, utiliza-se procedimento de otimização reversa para calcular o vetor de retornos de equilíbrio a partir do portfólio de mercado:

$$\Pi = \delta \sum w \quad (2.7)$$

2. O investidor também necessita quantificar a sua incerteza em relação à *priori* por meio da definição de um valor para τ .
3. Em seguida, o investidor deve formular suas *views*, especificando P , Ω e Q . Dadas k *views* em n ativos, P é uma matriz $k \times n$, em que a soma de cada linha totaliza 1 (um) no caso de uma *view* absoluta, e 0 (zero) no caso de uma *view* relativa. Q é um vetor $k \times 1$ com os retornos em excesso para cada *view*. Ω é a matriz diagonal $k \times k$ da variância da média estimada das *views*. Por exemplo, para três ativos e duas *views*, um investidor poderia fornecer como entrada a matriz P , Q e Ω com os valores da Equação 2.8, Equação 2.9 e Equação 2.10, respectivamente. Nesse exemplo, o investidor estabeleceu duas *views*: uma de que o ativo 1 terá um retorno de 3%, com variância de 1,6; e outra de que o retorno do ativo 2 será 2% superior ao retorno do ativo 3, com variância de 0,4. Pode-se ver, pela variância definida, que o investidor hipotético está mais confiante na segunda *view* que na primeira.

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \quad (2.8)$$

$$Q = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad (2.9)$$

$$\Omega = \begin{bmatrix} 1,6 & 0 \\ 0 & 0,4 \end{bmatrix} \quad (2.10)$$

4. Deve-se, então, aplicar a equação principal do modelo Black-Litterman para calcular a estimativa posterior dos retornos:

$$\hat{\Pi} = \Pi + \tau \sum P^T [(P\tau \sum P^T) + \Omega]^{-1} |Q - P\Pi| \quad (2.11)$$

5. Após isso, deve-se computar a variância posterior da média estimada sobre a média desconhecida utilizando a seguinte equação:

$$M = \tau \sum -\tau \sum P^T (P\tau \sum P^T + \Omega)^{-1} P\tau \sum \quad (2.12)$$

6. Assumindo que a incerteza nas estimativas é independente da covariância conhecida dos retornos sobre a média desconhecida, pode-se adicionar as duas para obter a covariância dos retornos sobre as médias estimadas:

$$\sum_p = \sum + M \quad (2.13)$$

Com as informações de médias estimadas e covariância dos retornos sobre as médias estimadas, obtidas na Equação 2.11 e na Equação 2.13, respectivamente, pode-se utilizar um modelo de otimização de portfólio para obter uma alocação que reflita esses resultados, sendo comumente utilizada a otimização de média-variância.

2.1.5 Modelagem da Curva de Juros

A curva de juros mostra diversas taxas de juros associadas com diferentes durações de contratos para um instrumento de débito específico. Ela sumariza o relacionamento entre o termo (tempo para maturidade do instrumento) do débito e a taxa de juros a ele associada [26]. A Figura 2.3 é um exemplo hipotético de uma curva de juros.

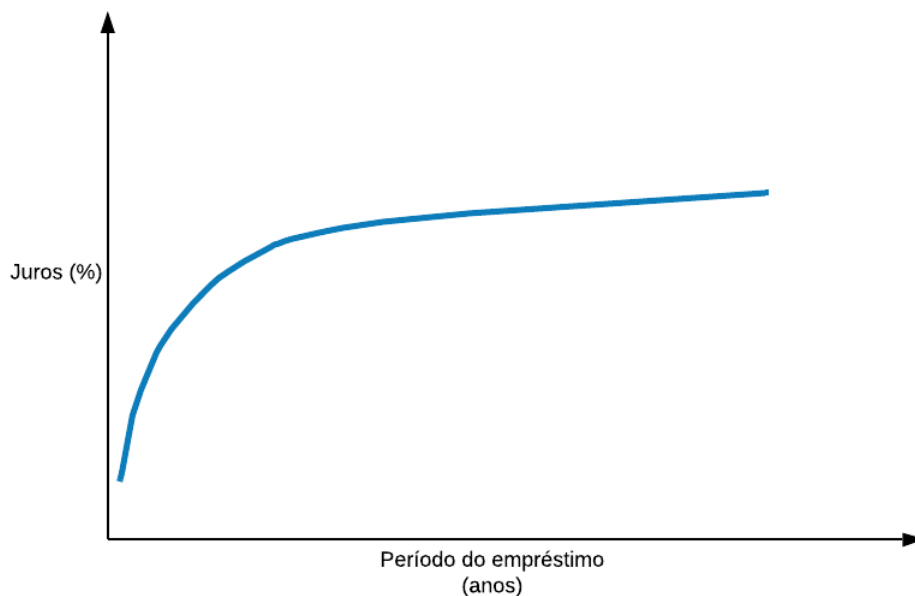


Figura 2.3: Exemplo de uma curva de juros.

Em condições normais, conforme o tempo para maturidade de um instrumento de débito aumenta, também cresce a taxa de juros associada. A razão para isto é que débito emitido para termos mais longos carrega mais risco pela maior probabilidade de inflação inesperada ou de *default* (não pagamento do débito) e pela maior probabilidade de ser necessário vender o ativo antes do termo e assumir perdas financeiras [26].

Embora a curva de juros seja mostrada como contínua, dados para todas as possíveis datas de maturidade de um dado instrumento de débito não costumam estar disponíveis, e muitos pontos sobre a curva são calculados por interpolação com datas para maturidades conhecidas [26].

Entender a dinâmica de evolução da curva de juros é importante para precificar ativos financeiros, alocar portfólios, conduzir a política monetária e avaliar bens de capitais, dentre outras tarefas. Para investigar a dinâmica da curva de juros, pesquisadores tem produzido literatura acadêmica com uma ampla variedade de modelos, embora a maioria dos modelos sejam teoricamente rigorosos ou empiricamente bem-sucedidos, mas não ambos. Em contraste, o modelo Nelson-Siegel Dinâmico (NSD), uma versão dinâmica do modelo clássico proposto por Nelson e Siegel [27], tem sido altamente bem-sucedido empiricamente sem perder o rigor teórico [11].

2.1.5.1 Modelo Nelson-Siegel Dinâmico

O modelo Nelson-Siegel Dinâmico (NSD) decompõe a curva da taxa de juros em três fatores que evoluem no tempo, que podem ser interpretados como o nível da curva de juros, a inclinação da curva de juros e a curvatura da curva de juros, respectivamente [28]. O modelo NSD tornou-se padrão para modelagem de curva de juros orientada a políticas em bancos centrais [11].

Define-se $y_t(\tau_i)$ como uma única observação de taxa de juros feita na data t para uma maturidade τ_i e $y_t(\tau_i) \forall i$ como a curva de juros no tempo t . O conjunto de curvas de juros para as datas $t = 1, \dots, T$ mostra a evolução da curva de juros no tempo. Uma representação da dinâmica dos fatores da curva de juros pode ser escrita como [28][29]:

$$y_t(\tau) = H \cdot \beta_t + \epsilon_t \quad (2.14)$$

$$\beta_t = \mu + F \cdot \beta_{t-1} + v_t \quad (2.15)$$

Onde β_t é o vetor com os valores dos fatores no tempo t , H é a matriz de carregamento dos fatores e ϵ_t é um termo de erro. A Equação 2.14 provê um relacionamento transversal entre taxas de juros observadas e a representação na forma dos fatores escolhidos, $\beta = [\text{Nível} \quad -\text{Inclinação} \quad \text{Curvatura}]$ [29]. A parametrização da matriz de carregamento de fator é dada por [29]:

$$H = \left[1 \quad \frac{1 - \exp(-\lambda_t \cdot \tau)}{\lambda_t \cdot \tau} \quad \frac{1 - \exp(-\lambda_t \cdot \tau)}{\lambda_t \cdot \tau} - \exp(-\lambda_t \cdot \tau) \right] \quad (2.16)$$

Onde a variável λ_t é um parâmetro de decaimento que determina a forma e velocidade de decaimento dos carregamentos da inclinação e da curvatura [29].

A evolução dinâmica dos fatores da curva de juros é especificada pela Equação 2.15, em que a matriz F coleta os parâmetros auto-regressivos e v_t é um termo de erro [29].

2.1.5.2 Modelo Nelson-Siegel Dinâmico Rotacionado

Para fins de análise, pode ser que os fatores tenham interpretação econômica clara. Para transformar os fatores do modelo NSD de $[\text{Nível} \quad -\text{Inclinação} \quad \text{Curvatura}]$ para $[\text{Taxa curta} \quad \text{Inclinação} \quad \text{Curvatura}]$, o modelo deve ser modificado para incluir uma matriz de rotação, denotada por A , escolhida de tal forma que a desejada interpretação dos fatores emerja e de maneira que $I = A^{-1} \cdot A$, onde I é a matriz identidade. Dessa forma, o modelo aumentado, chamado Nelson-Siegel Dinâmico Rotacionado (NSDR), tem as mesmas propriedades estatísticas do modelo NSD clássico [29].

O modelo NSDR é então sintetizado na seguinte forma:

$$y_t(\tau) = H \cdot A^{-1} \cdot A \cdot \beta_t + \epsilon_t \quad (2.17)$$

$$A \cdot \beta_t = A \cdot \mu + F \cdot A \cdot \beta_{t-1} + A \cdot v_t \quad (2.18)$$

Para simplificar a notação da Equação 2.17 e da Equação 2.18:

$$G = H \cdot A^{-1} \quad (2.19)$$

$$\gamma_t = A \cdot \beta_t \quad (2.20)$$

$$m = A \cdot \mu \quad (2.21)$$

$$z_t = A \cdot v_t \quad (2.22)$$

Tendo como forma padrão proposta do modelo MSDR, as seguintes equações [29]:

$$y_t(\tau) = G \cdot \gamma_t + \epsilon_t \quad (2.23)$$

$$\gamma_t = m + F \cdot \gamma_{t-1} + z_t \quad (2.24)$$

O último elemento do modelo NSDR, a matriz de rotação A , tem que ser parametrizado. Dado que uma interpretação econômica dos fatores é desejável e que, conforme Equação 2.20, cada linha de A pode ser vista como uma combinação linear dos fatores do modelo NSD original, que fornecem a interpretação desejada dos fatores, tem-se [29]:

$$\begin{bmatrix} \text{Taxa curta} \\ \text{Inclinação} \\ \text{Curvatura} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \cdot \begin{bmatrix} \text{Nível} \\ -\text{Inclinação} \\ \text{Curvatura} \end{bmatrix} \quad (2.25)$$

Se a taxa curta fosse definida para maturidade $\tau = 0$, a matriz de rotação seria [29]:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.26)$$

Entretanto, por motivos institucionais e com o objetivo de integrar variáveis macroeconômicas ao modelo, pode ser preferível trabalhar com uma taxa curta no segmento de maturidades de 1 ou 3 meses. Nesse caso, a matriz de rotação A deve ser modificada para refletir o carregamento atual dos fatores na maturidade escolhida como taxa curta e, naturalmente, diferirá da matriz apresentada na Equação 2.26.

Para obter uma expressão geral para A , pode-se pensar na taxa curta como a ponta longa da curva menos a inclinação, adicionada de uma pequena correção por $t_s > 0$. No modelo clássico, o sinal do fator inclinação é invertido e, portanto, um valor negativo representa uma curva com inclinação positiva. Por conseguinte, a primeira linha da matriz de rotação geral é dada por [29]:

$$A_{(1,1:3),\tau_s} = \left[1 \quad \frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} \quad \frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} - \exp(-\lambda \cdot \tau_s) \right] \quad (2.27)$$

A segunda linha de A expressa a combinação linear dos fatores do modelo NSD clássico que formam a inclinação. Esta pode ser empiricamente calculada como a diferença entre a ponta longa e a ponta curta da curva, sendo portanto:

$$A_{(2,1:3),\tau_s} = \left[1 \quad 0 \quad 0 \right] - \left[1 \quad \frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} \quad \frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} - \exp(-\lambda \cdot \tau_s) \right] \quad (2.28)$$

$$A_{(2,1:3),\tau_s} = \left[0 \quad - \left(\frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} \right) \quad - \left(\frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} - \exp(-\lambda \cdot \tau_s) \right) \right] \quad (2.29)$$

Usando a definição empírica do fator “curvatura” como duas vezes o juros de maturidade média menos o juros de termo longo [28], a terceira e última linha de A é representada como:

$$A_{(3,1:3),\tau_s} = \left[0 \quad 1 - \frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} \quad 1 - \left(\frac{1 - \exp(-\lambda_t \cdot \tau_s)}{\lambda_t \cdot \tau_s} - \exp(-\lambda \cdot \tau_s) \right) \right] \quad (2.30)$$

Além disso, Nyholm [29] demonstrou que a matriz A é não-singular, ou seja, $I = A^{-1} \cdot A$, para valores economicamente significativos de λ .

2.1.6 Aprendizado de Máquina em Finanças

Muitos investidores não percebem a complexidade da aplicação de técnicas de aprendizado de máquina à gestão de investimentos. A porcentagem de modelos quantitativos que não conseguem entregar resultados superiores ao mercado é alta, especialmente quando utilizando técnicas de aprendizado de máquina [30].

Diversas peculiaridades relacionadas a problemas em economia e finanças levam a esse alto índice de falhas. Dentre elas, pode-se citar:

- a utilização do conjunto de testes múltiplas vezes durante o desenvolvimento dos modelos, levando à superestimação da significância dos resultados do modelo final; esse procedimento tende a levar ao sobreajuste ao *backtest*² [30];
- observações nesses domínios tendem a não serem obtidas de um processo independente e identicamente distribuído, invalidando o uso de técnicas tradicionais como, por exemplo, a validação cruzada *k-fold* [30];
- a utilização do universo atual de investimentos para treinamento do modelo, ignorando que algumas companhias faliram e títulos venceram no período [32];
- a elaboração de uma estória *ex post* para justificar algum padrão aleatório nos dados [32];
- a não consideração dos custos de transação no treinamento e na seleção dos modelos [32];
- o foco apenas em retorno enquanto se ignora outras métricas [30];
- a não consideração dos custos de empréstimo nas operações vendidas [32];
- a confusão entre correlação e causalidade [30]; e
- a seleção de um período não representativo para treinamento do modelo [30].

Dentre os problemas enumerados, o sobreajuste ao *backtest* é um dos mais difíceis de evitar, mesmo para profissionais experientes. Sobreajuste ao *backtest* é o viés de seleção após múltiplos *backtests* e tende a levar à detecção de padrões históricos aleatórios. O propósito dessa ferramenta é descartar modelos ruins, e não ajustar o modelo até atingir um resultado satisfatório [30].

Enquanto não há uma saída fácil para o problema de sobreajuste ao *backtest*, algumas medidas podem reduzir a sua influência, tais como [30]:

- desenvolver modelos para classes de ativos financeiros ou para o universo de investimento, em vez de desenvolvê-los para um ativo específico;
- aplicar *bagging* como uma forma de prevenir sobreajuste e de reduzir a variância do erro de previsão;
- não realizar *backtest* até que a pesquisa esteja completa;
- registrar cada *backtest* conduzido de maneira que seja possível estimar o sobreajuste ao *backtest* no resultado final; e
- simular cenários em vez de utilizar dados históricos.

²*Backtest* é uma simulação histórica de uma estratégia algorítmica de investimento [31].

2.1.6.1 Métodos de Validação em Finanças

De maneira geral, o desempenho de um modelo sobre dados utilizados para seu treinamento não é um bom indicador do desempenho deste modelo em dados não-observados, devido ao problema do sobreajuste do modelo. Se os dados são independentes e identicamente distribuídos e numerosos, pode-se manter parte dos dados fora do conjunto de treinamento, num conjunto de dados de validação [33].

Um método de validação frequentemente utilizado em aprendizado de máquina em finanças é a divisão dos dados temporalmente em dois grupos, treinar o modelo no primeiro grupo e testar *out-of-sample* no segundo grupo. Como vantagem, este método evita o risco de sobreajuste advindo da captura de tendências que surgiram em dados de testes anteriores aos dados de treinamento utilizados. Como potenciais problemas, os modelos são selecionados com viés para o período mais recente e somente uma fração dos dados é utilizada para avaliar o desempenho *out-of-sample* [30]. Se a quantidade de dados é limitada, é importante, para se obter um bom modelo, a utilização do máximo possível de dados no treinamento do modelo.

Uma solução para esse problema é o uso de métodos de validação cruzada [33], cujo propósito é determinar o erro de generalização de um modelo de aprendizado de máquina, de maneira a prevenir o sobreajuste e a produzir previsões sobre instâncias futuras e não observadas previamente [30].

O algoritmo mais popular para validação cruzada é o método de validação *k-fold*. Neste método, os dados são divididos em k partes iguais. O modelo, então, é treinado em $k - 1$ partes e usado para prever a parte k dos dados. Este procedimento é realizado k vezes, com cada subconjunto dos dados sendo usado como dados de teste, e os erros de predição são combinados [34].

O algoritmo *k-fold* não funciona em finanças porque não se pode assumir que instâncias sejam independentes em uma série temporal, ou seja, pode existir auto-correlação, podendo levar ao vazamento de informação do conjunto de treinamento para o conjunto de testes e ao sobreajuste do modelo selecionado [30].

A forma mais comum de validação cruzada utilizada para avaliar o desempenho histórico de uma estratégia de investimento é conhecida como *walk-forward*. Este método consiste em avaliar o desempenho no passado, em diversos momentos de decisão do passado, com base em dados que precederam cada decisão específica. O método *walk-forward* tem a vantagem de ter uma interpretação histórica clássica, mas pode apresentar as seguintes desvantagens [30]:

- apenas o cenário histórico é testado, podendo levar a sobreajuste;

- a sequência específica de pontos de dados do passado não é necessariamente representativa do desempenho futuro; e
- o uso do método *walk-forward* leva a que decisões iniciais sejam feitas com pequenas porções dos dados, usualmente os dados menos recentes.

O método de validação cruzada *k-fold* combinatório purgado foi proposto por De Prado (2018) [30] e busca prevenir os problemas do método *walk-forward* e os associados ao uso de validação cruzada em finanças de maneira geral. O método proposto envolve:

1. realizar N de partições nos dados, cada partição com tamanho T/N , onde T é o número total de ponto de dados;
2. selecionar k partições a cada divisão para serem os dados de teste, gerando $\binom{N}{N-k} = \frac{\prod_{i=0}^{k-1} (N-i)}{k!}$ caminhos possíveis (combinações);
3. para cada par de dados (y_i, y_j) em que y_i pertence ao dados de treinamento e y_j pertence aos dados de teste, purgar y_i se este está dentro do período utilizado para determinar y_j ;
4. aplicar uma janela de embargo em dados após cada partição de treinamento de maneira que não haja correlação entre os dados de treinamento e os dados de teste;
5. treinar os modelos sobre os $\binom{N}{N-k}$ conjuntos de treinamento e produzir previsões sobre os $\binom{N}{N-k}$ respectivos conjuntos de teste; e
6. calcular as métricas para cada caminho, gerando uma distribuição empírica da métrica utilizada na avaliação (e extraindo estatísticas desta).

2.1.6.2 Redução de Dimensionalidade

A Análise de Componentes Principais (PCA) é uma técnica amplamente utilizada para redução de dimensionalidade e extração de variáveis [35]. O PCA pode ser definido como uma projeção ortogonal dos dados em um espaço linear de menor dimensionalidade ou como a projeção linear que minimiza a média da distância quadrática entre os pontos de dados e as suas projeções [33].

Outra definição mais intuitiva [36] do algoritmo envolve considerar um conjunto de n séries estacionárias X_i e a combinação linear dessas séries na forma de um portfólio P com pesos w_p e variância σ_p^2 e, por fim, considerar um portfólio normalizado, cuja soma dos quadrados dos pesos seja um, com a maior variância possível. Assumindo-se que a média dos retornos seja zero e que as séries sejam sequências independentes e identicamente

distribuídas, com distribuição normal conjunta e matriz de covariância Σ , o retorno de cada um dos portfólios será normalmente distribuído com variância dada por:

$$\sigma_p^T = w_P \Sigma w_P \quad (2.31)$$

Sendo assim, o portfólio de máxima variância será dado pela seguinte função de maximização:

$$\operatorname{argmax}_{\|w\|=1} \{w_p^T \Sigma w_P\} \quad (2.32)$$

Em que a solução é o autovetor w_1 associado ao maior autovalor λ_1 da matriz de covariâncias Σ .

Considerando-se o conjunto de todos os portfólios normalizados ortogonais a w_1 e repetindo-se o processo anterior n vezes, é possível chegar a n portfólios P_i de máxima variância. Os pesos desses portfólios são os autovetores ortonormais da matriz de covariância Σ . A cada um desses portfólios corresponde uma série temporal dada pela combinação linear das séries originais X_i . Por construção, as séries dos portfólios assim obtidos (componentes principais) são mutuamente não-correlacionadas. Consequentemente, para recuperar o retorno das séries originais, é necessária apenas a combinação linear dos componentes principais:

$$X_j = \sum_i \alpha_{j,i} P_i \quad (2.33)$$

Se as séries originais apresentam alto grau de correlação, um reduzido número de componentes principais possuirão variância estatisticamente diferente de zero. Dessa maneira, pode-se reduzir a dimensionalidade do problema se utilizando apenas dos componentes principais de maior variância.

Em resumo, para se implementar a análise de componentes principais (PCA), deve-se calcular os autovalores e autovetores da matriz de covariância e escolher os maiores autovalores [33] [36]. Os autovetores correspondentes equivalem aos pesos dos portfólios que formam os componentes principais.

2.1.7 Aprendizado de Máquina *Online*

Desde os primórdios do aprendizado de máquina, duas visões de como abordar o problema de aprendizado automático coexistem: aprendizado em lote e aprendizado *online*. O aprendizado em lote assume que todo o conjunto de dados de treinamento pode ser acessado de maneira a ajustar o modelo, que não há restrições de tempo e que o processo subjacente à geração dos dados não muda.

Se as premissas supracitadas são assumidas como verdadeiras, o escopo de aplicação das técnicas de aprendizado de máquina se tornaria estreito em demasia. Por outro lado, se algumas dessas premissas são relaxadas, é possível abarcar alguns desafios enfrentados na prática [37]:

- situação em que os dados são continuamente entregues, havendo a necessidade de que o modelo esteja atualizado a cada instante, pelo fato de o conjunto de dados nunca estar completo nem imediatamente acessível;
- situação em que o conjunto de dados é de tamanho volume que torna impossível a centralização dos dados, levando à necessidade de paralelização do processo de treinamento; e
- situação em que o processo subjacente à geração de dados muda e o conjunto de treinamento inicial perde sua validade com o passar do tempo, tornando necessário mecanismo de atualização dos modelos.

Não há consenso na definição do significado do termo “aprendizado *online*”. Alguns autores o definem apenas como a primeira das situações supramencionadas, enquanto outras fontes também incluem as duas últimas situações. Tal indefinição dá-se, em parte, pelo fato de os problemas apresentados poderem ser solucionados de maneira similar [37].

2.1.8 Redes Neurais Artificiais

O modelo de rede neural artificial básico pode ser descrito como uma série de transformações funcionais, em que a entrada de cada função é o somatório das entradas da rede neural ou das saídas das funções da camada anterior, multiplicadas por pesos, mais um parâmetro de viés. A esse resultado, conhecido como ativação, aplica-se uma função de ativação não-linear, que é a saída da função, comumente com valor entre zero e um. Essa saída será utilizada por uma próxima camada oculta de funções ou será saída da rede neural [33]. O valor de cada neurônio, portanto, pode ser definido da seguinte maneira [38]:

$$v_{t,i} = a_{t,i} \left(\sum_{j=0}^{k_{t-1}} w_{t,i,j} v_{t-1,j} \right) \quad (2.34)$$

Onde:

- $v_{t,i}$ é o valor de saída do neurônio i da camada t
- $v_{t-1,j}$ é o valor de saída do neurônio j da camada $t - 1$

- $a_{t,i}$ é a função de ativação do neurônio i da camada t
- k_{t-1} é o número de neurônios na camada $t - 1$, anterior à camada t
- $w_{t,i,j}$ é o peso aplicado ao valor de saída do neurônio j da camada $t - 1$ ao ser utilizado como entrada do neurônio i da camada t
- $w_{t,i,0}$ é multiplicado por um para representar o parâmetro de viés

Redes neurais artificiais tem sido utilizadas com sucesso para a previsão de diversas séries temporais financeiras e econômicas, tais como preços no mercado de ações, ciclos de negócios, taxas de juros, desemprego, produto interno bruto, dentre outras [25]. No campo de análise financeira, modelos de redes neurais podem melhorar significativamente a acurácia da previsão do lucro por ação de empresas, especialmente se variáveis fundamentalistas forem utilizadas. Redes neurais foram utilizadas com bom desempenho no apreçamento de opções, sem sofrer dos vieses sistemáticos resultantes de suposições simplificadas, tais como as do modelo *Black-Scholes*. Redes neurais foram também utilizadas no mercado de câmbio para prever a magnitude e mudanças de direção nas taxas com resultados promissores. Na previsão de inflação, as evidências sugerem que redes neurais são competitivas e tem alto grau de robustez, em comparação a outros modelos [39].

2.1.9 Redes Neurais Recorrentes

Redes neurais recorrentes são modelos com a capacidade de seletivamente passar informação através de passos de uma sequência, durante o processamento de dados sequenciais elemento a elemento. Dessa maneira, uma rede neural recorrente é capaz de modelar entrada e/ou saída consistindo de sequências de elementos que não são independentes entre si. Além disso, redes neurais recorrentes são capazes de modelar dependências sequenciais e temporais em múltiplas escalas, não estando presas a uma janela de contexto fixa [38].

Para ir do modelo de redes neurais multicamadas para um modelo de redes neurais recorrentes, é necessário incorporar o compartilhamento de parâmetros entre diferentes partes do modelo. Esse compartilhamento de estado (h^{t-1}), passado para a rede do momento $t - 1$ para t e ao qual se incorpora o vetor de entrada x^t , equivale a uma rede neural padrão com múltiplas camadas, cada uma passando uma mensagem para sua sucessora [3], conforme exemplificado pela Figura 2.4.

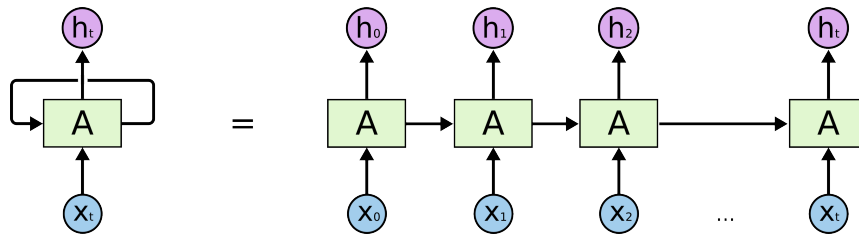


Figura 2.4: Ilustração da equivalência entre uma rede neural recorrente e uma rede neural multicamadas em que cada camada é uma cópia da rede recorrente e passa mensagem para a camada seguinte. Obtido de Olah (2015) [2].

2.1.9.1 Camada *Long Short-Term Memory* (LSTM)

O modelo LSTM é, em parte, similar a uma rede neural recorrente com uma camada oculta, mas cada nó da camada oculta é substituído por uma célula de memória. Uma célula de memória é uma unidade composta, projetada a partir de nós mais simples com um padrão específico de conectividade [38]. A Figura 2.5 apresenta uma visão geral da arquitetura de uma célula LSTM.

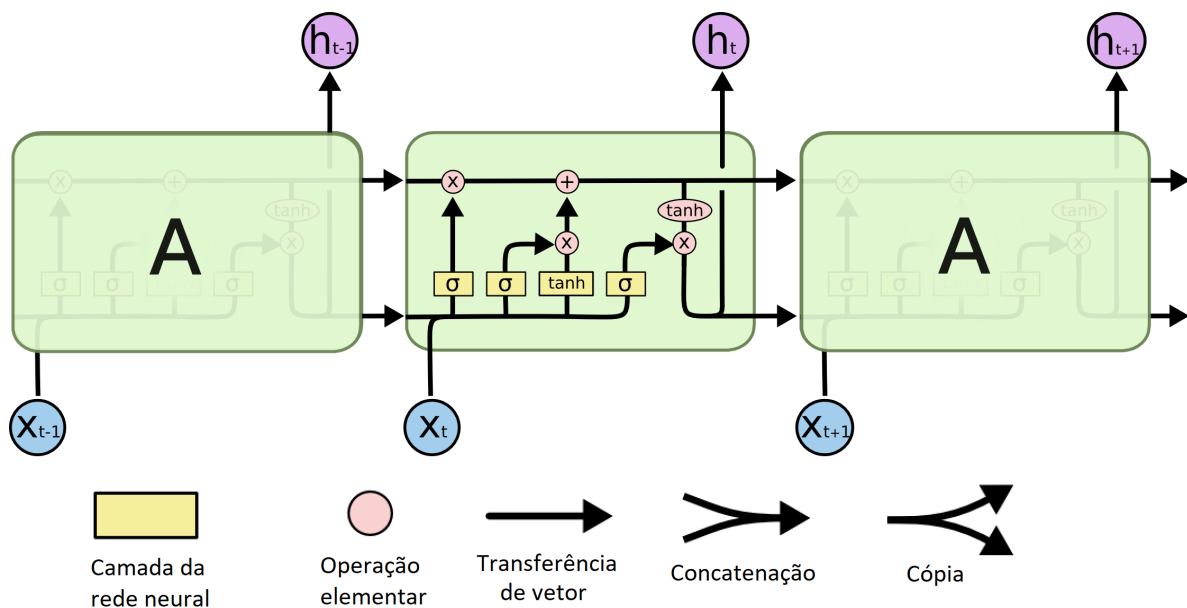


Figura 2.5: Visão geral da arquitetura da célula de memória LSTM. Adaptado de Olah (2015) [2].

Um dos elementos distintivos da camada LSTM (e outras camadas recorrentes baseadas nela) é a porta. Uma porta é uma unidade sigmóide cujo valor é utilizado para multiplicar o valor de outro nó. É uma porta no sentido que, se seu valor é zero, o fluxo de dados do nó relacionado é eliminado; se seu valor é um, todo o fluxo é passado adiante,

e valores intermediários controlam quantos dos dados gerados pelo nó relacionado são passados adiante [38].

Uma camada do tipo LSTM apresenta os elementos listados abaixo. Foi utilizado o c subscrito para valores de uma célula de memória específica e a notação vetorial sem o c subscrito para valores resultantes de uma camada de células inteira [38] [3]:

- *Nó de entrada* (g_c): Este nó realiza a ativação de maneira padrão da camada de entrada $x^{(t)}$ do instante atual e da camadas oculta no instante anterior $h^{(t-1)}$. Tipicamente, a soma ponderada da entrada é fornecida como entrada para a função de ativação tangente hiperbólica (\tanh), mas pode utilizar qualquer função de “esmagamento”, que é uma função que transforma um valor para um dos limites de um intervalo;
- *Porta de entrada* (i): o valor desta porta é utilizado para multiplicar o valor do nó de entrada;
- *Estado interno* (s_c): o estado interno é um nó com ativação linear que tem uma aresta recorrente auto-conectada com peso fixo de 1, permitindo que a informação anterior flua livremente entre vários instantes. A atualização do estado interno se dá, em notação vetorial, por $s^{(t)} = g^{(t)} \odot i^{(t)} + s^{(t-1)}$;
- *Porta de esquecimento* (f_t): a porta de esquecimento provê um forma pela qual a rede pode aprender a eliminar o conteúdo do seu estado interno. Incluindo-se a porta de esquecimento, a equação para calcular o estado interno é $s^{(t)} = g^{(t)} \odot i^{(t)} + f^{(t)} \odot s^{(t-1)}$; e
- *Porta de saída* (o_c): a porta de saída multiplica o valor de s_c para gerar o valor de saída da célula de memória. É comum que, antes dessa multiplicação, o valor do estado interno s_c seja transformado por uma função de ativação tangente hiperbólica (\tanh).

Todas as portas da célula de memória apresentam não-linearidade sigmoideal [3]. A estrutura da célula da rede LSTM pode ser observada na Figura 2.6.

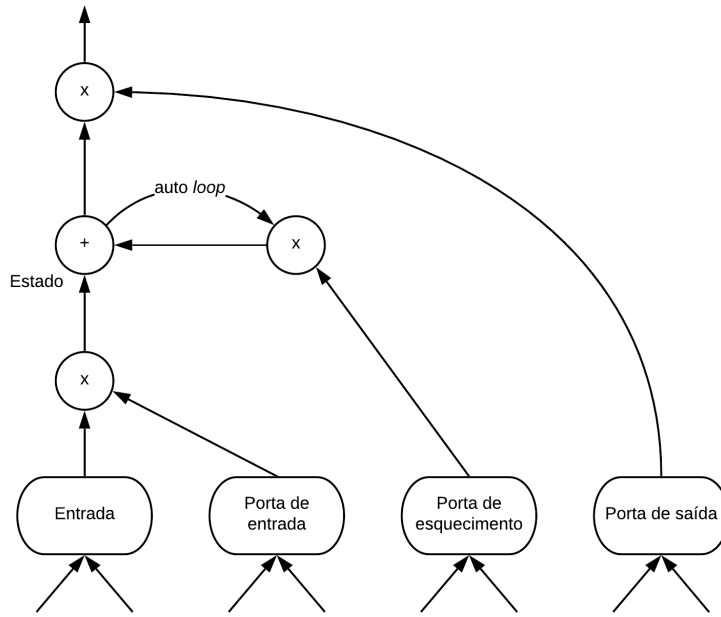


Figura 2.6: Diagrama da célula de uma rede LSTM. Adaptado de Goodfellow *et al* [3].

Antes da introdução da rede neural LSTM, tentativas de treinar redes neurais recorrentes para cobrir períodos longos não obtinham resultados melhores que uma escolha aleatória [40].

2.1.9.2 Camada *Gated Recurrent Unit* (GRU)

A camada do tipo GRU foi introduzida em 2014 e é similar à camada LSTM, porém sem incluir a porta de saída [41]. Um modelo GRU teve desempenho similar a um modelo LSTM na tarefa de modelagem de música polifônica e de modelagem de sinal de linguagem falada, embora ambos os modelos sejam superiores a um modelo RNN mais simples [41] [42].

2.1.9.3 Camada *Minimal Recurrent Neural Network* (Minimal RNN)

Como o próprio nome sugere, a rede *Minimal RNN* emprega atualizações mínimas de sua memória, com o objetivo de alcançar aprendizado eficiente e melhor interpretabilidade do modelo aprendido.

A cada passo t , o modelo mapeia o vetor de entrada x_t para um espaço latente através de

$$z_t = \Phi(x_t) \quad (2.35)$$

Onde Φ pode ser qualquer função, mesmo outra rede neural. Dada a representação latente da entrada z_t , a atualização do estado é feita com

$$h_t = u_t \odot h_{t-1} + (1 - u_t) \odot z_t \quad (2.36)$$

Onde u_t é a porta de atualização, dada por

$$u_t = \sigma(U_h h_{t-1} + U_z z_t + b_u) \quad (2.37)$$

Dessa forma, a rede *Minimal RNN* define um espaço latente e passa a se mover apenas nesse espaço, atualizando seu estado (h_t) com base na média ponderada entre seu estado anterior (h_{t-1}) e a entrada codificada (z_t), utilizando, para isso, o valor da porta de atualização (u_t). Esse mecanismo é, comparado ao de outras redes neurais recursivas, como LSTM e GRU, muito mais simples, embora retenha uma porta, o que é crucial para manter as dependências de longo prazo. A Figura 2.7 exemplifica o funcionamento de sua célula [4].

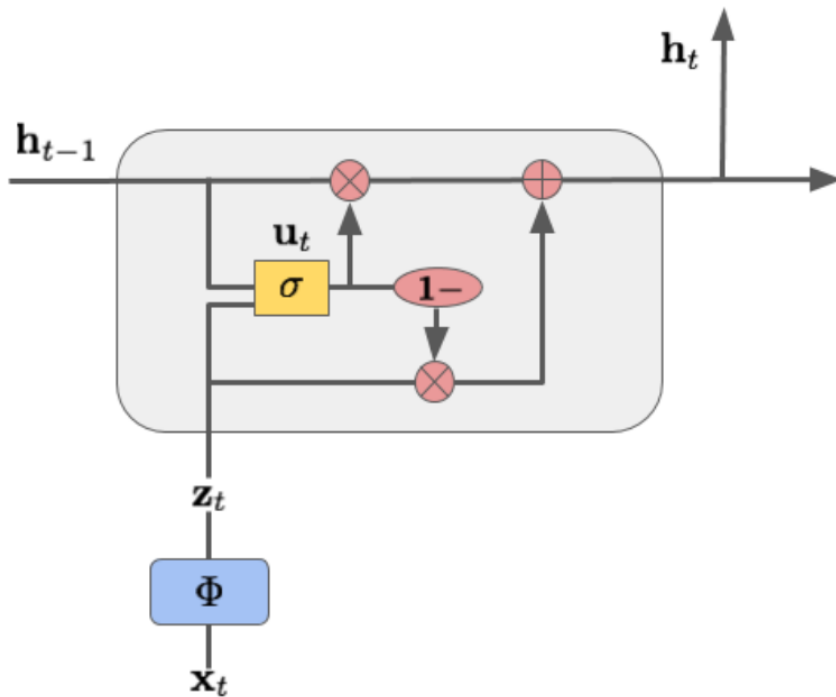


Figura 2.7: Diagrama da célula de uma rede *Minimal RNN*. Obtido de Chen (2017) [4].

2.1.9.4 Camada *Simple Recurrent Unit* (SRU)

A camada do tipo SRU foi projetada para prover capacidade expressiva e uma implementação altamente paralelizada, com o intuito de resolver o problema de falta de escalabilidade

das redes neurais recursivas [43].

Uma única camada SRU envolve as seguintes computações [43][44]:

$$f_t = \sigma(W_f x_t + v_f \odot c_{t-1} + b_f) \quad (2.38)$$

$$c_t = f_t \odot c_{t-1} + (1 - f_t) \odot (W x_t) \quad (2.39)$$

$$r_t = \sigma(W_r x_t + v_r \odot c_{t-1} + b_r) \quad (2.40)$$

$$h_t = r_t \odot c_t + (1 - r_t) \odot x_t \quad (2.41)$$

Onde W , W_f e W_r são matrizes de parâmetros e v_f , v_r , b_f e b_r são vetores de parâmetros a serem aprendidos durante o treinamento. A Equação 2.38 e a Equação 2.39 formam o chamado “subcomponente de recorrência leve”, que lê sucessivamente o vetor de entrada x_t e computa a sequência de estados c_t que captura informação sequencial. A computação deste componente é similar a de outras unidades recorrentes, tais como LSTM e GRU, com algumas modificações. Em especial, a substituição da multiplicação de matrizes pela multiplicação elemento a elemento torna independente cada dimensão dos vetores e, por conseguinte, permite a sua paralelização. A Equação 2.40 e a Equação 2.41 formam o subcomponente de “*highway*” da rede, que facilita o treinamento de redes neurais profundas baseado em gradiente [43][44].

Em experimentos com múltiplas tarefas de processamento de linguagem natural, a rede SRU foi treinada em tempo de 5 a 9 vezes menor que uma rede LSTM otimizada para uso da GPU, fornecendo resultados melhores que os modelos LSTM e convolucional [43].

2.1.9.5 Camada *Independently Recurrent Neural Network* (IndRNN)

A camada do tipo IndRNN foi desenvolvida com foco na construção de redes neurais recorrentes profundas de maneira eficiente. Como características distintas de outras redes neurais recorrentes, a rede IndRNN tem neurônios da mesma camada independentes entre si e conectados entre camadas; pode ser treinada com funções de ativação não-saturadas tal como unidade linear retificada (relu); e permite a construção eficiente de redes neurais mais profundas [45].

Cada neurônio da camada IndRNN pode ser descrito como [46]:

$$h_t = \sigma(W x_t + u \odot h_{t-1} + b) \quad (2.42)$$

Onde u é um vetor que deve ser aprendido no treinamento. Para o neurônio n , o estado oculto pode ser obtido por [46]:

$$h_{n,t} = \sigma(w_n x_t + u_n h_{n,t-1} + b_n) \quad (2.43)$$

Onde w_n e u_n são a matriz de pesos de entrada e o vetor de pesos recorrente, respectivamente. Cada neurônio só recebe informação da entrada e de seu estado interno no passo anterior.

Resultados experimentais mostraram que a rede IndRNN é capaz de processar sequências muito longas e construir redes profundas de maneira robusta [45].

2.1.10 Florestas Aleatórias

O modelo de Florestas Aleatórias (*Random Forests*, em inglês) é uma modificação do método de *bagging*³ que treina uma grande coleção de árvores de decisão descorrelacionadas e então combina suas previsões [34].

De maneira geral, o algoritmo das Florestas Aleatórias pode ser descrito da seguinte forma [34]:

1. Para b de 1 a B :
 - (a) Obtenha amostra Z_b de tamanho N dos dados de treinamento;
 - (b) Treine uma árvore de decisão T_b com os dados Z_b por repetir os seguintes passos recursivamente para cada nó terminal da rede, até que o tamanho mínimo de nó n_{min} seja alcançado:
 - i. Selecione m variáveis aleatoriamente das p variáveis existentes;
 - ii. Selecione a melhor variável para realizar a divisão dos dados entre as m variáveis selecionadas; e
 - iii. Separe o nó em dois nós-filhos.
2. Retorne a coleção de árvores $\{T_b\}_1^B$.

Onde B é o número de árvores de decisão a treinar e p é o número de variáveis.

Para realizar uma previsão utiliza-se para [34]:

- regressão: $R = \frac{1}{B} \sum_{b=1}^B T_b(x)$; e
- classificação: $C = \text{voto majoritário}\{C_b(x)\}_1^B$.

³Técnica utilizada para reduzir a variância de uma função de predição, que funciona especialmente bem para modelos com alta variância e baixo viés, como árvores de decisão.

2.2 Trabalhos Correlatos

No uso de redes neurais para a previsão de fatores macroeconômicos, Nakamura (2005) [47] utilizou redes neurais para previsão da inflação americana com resultados superiores a modelos autoregressivos para horizontes de um e dois trimestres, com dados da inflação entre 1960 e 2003.

Sambasivan e Das (2017) [48] utilizaram processos gaussianos em conjunto com uma estratégia de modelagem dinâmica para modelar a curva de juros e compararam o modelo proposto a um modelo de séries temporais multivariadas, com acurácia superior na região de termo médio e longo e acurácia inferior na região de termo curto. Os resultados do estudo são interessantes pois a natureza da curva de juros é tal que existem mais dados para as regiões de termo curto que para as regiões de termo longo, tornando a previsão de longo prazo mais desafiadora.

Nyholm (2015) [29] aplicou ao modelo Nelson-Siegel Dinâmico um esquema de rotação de fatores para permitir a incorporação de fatores macroeconômicos na previsão da taxa de juros curta através de aplicação da Regra de Taylor, com a execução de experimentos que resultaram nos fatores macroeconômicos impactando a evolução da taxa de juros curta entre 1990 e 2002 e tornando-se estatisticamente insignificantes a partir de 2002.

Caldeira, Moura e Santos (2016) [49] examinaram a acurácia estatística e o valor econômico de modelar a curva de taxas de juros utilizando combinações de previsões de diversos modelos autoregressivos incluindo modelos dinâmicos de fatores, modelos de equilíbrio e modelos de regressão, demonstrando que, em uma grande quantidade de contratos futuros de taxas de juros brasileiros com diferentes maturidades, as previsões combinadas consistentemente tiveram melhor desempenho que modelos individuais, especialmente quando critérios econômicos foram levados em consideração.

Capítulo 3

Metodologia de Pesquisa

Este capítulo apresenta a metodologia utilizada neste projeto de pesquisa. Seguindo a estrutura proposta pelo padrão *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [50], as atividades da pesquisa foram compartimentalizadas nas seguintes fases: entendimento do negócio, compreensão dos dados, preparação dos dados, modelagem, avaliação e implantação.

Procedeu-se a pesquisa bibliográfica, buscando sintetizar o estado-da-arte dos conhecimentos necessários para a conseqüente pesquisa primária, o desenvolvimento do modelo proposto e a execução das etapas exigidas à sua validação.

A pesquisa utilizou o período de tempo - ou simplesmente janela, conforme jargão do mercado financeiro - de um mês para previsão das variáveis e para realocação dos portfólios e avaliação do desempenho da estratégia utilizada para avaliar o método.

O projeto foi desenvolvido utilizando a linguagem Python 3. Para modelagem das redes neurais, foi utilizado o pacote Keras. Outros pacotes utilizados foram o Pandas, Numpy e scikit-learn, dentre outros.

3.1 Entendimento do Negócio

As reservas internacionais brasileiras estão alocadas em diversas classes de ativos e diversas moedas. A fim de se buscar um modelo com impacto significativo sobre a alocação das reservas, analisou-se quais classes de ativos e quais moedas recebem maior fatia das reservas internacionais.

As Figuras 3.1 e 3.2 mostram que as reservas internacionais estão alocadas majoritariamente em títulos governamentais (75,6%) e em ativos precificados em dólar (87,5%), respectivamente. De fato, títulos precificados em dólar correspondem a 36,9% das reservas internacionais como um todo.

Figura 3.1: Alocação das reservas por classe de ativo (em 08/05/2019)

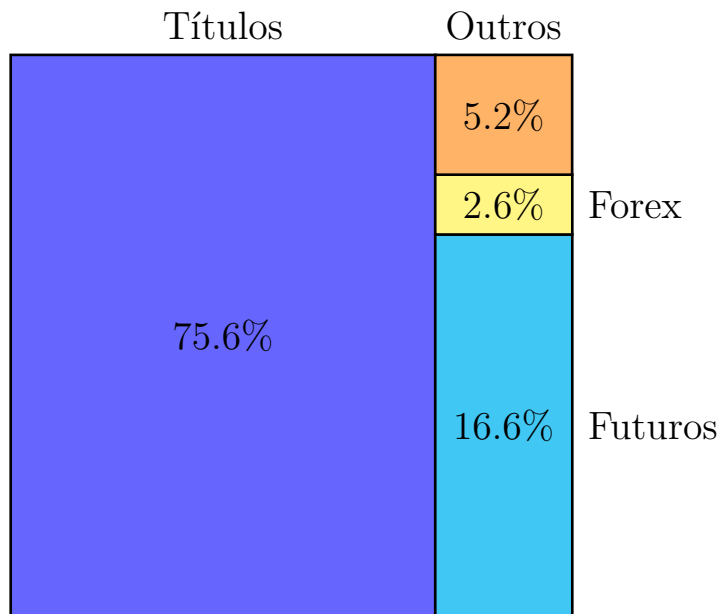
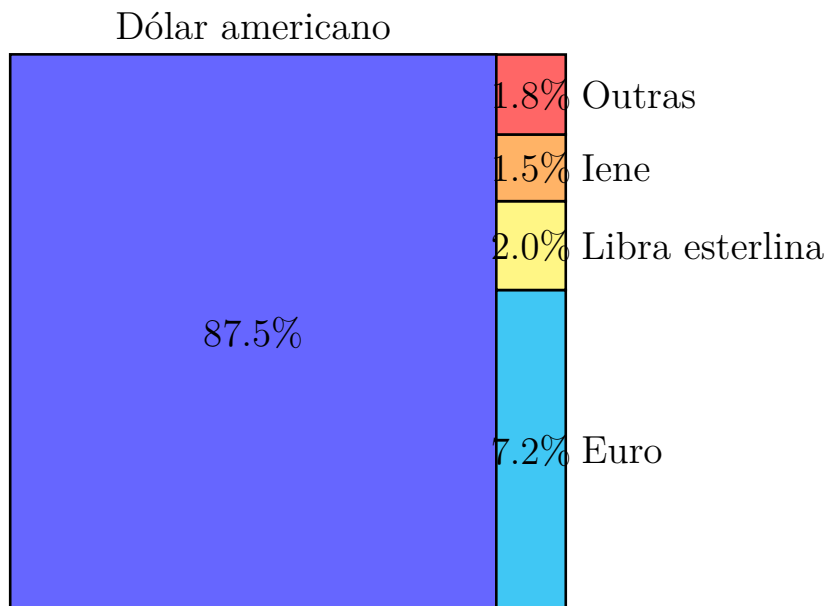


Figura 3.2: Alocação das reservas por moeda (em 08/05/2019)



Dadas as informações expostas, o foco desta pesquisa foi a otimização da alocação da carteira de títulos precificados em dólar, o que não preclui a aplicação futura do método desenvolvido para outros instrumentos de renda fixa e para outras moedas.

3.2 Compreensão dos Dados

As séries temporais foram obtidas no mínimo de 31/03/1978 até 31/03/2019, embora algumas séries apresentem dados iniciando em datas anteriores a 31/03/1978.

Para modelar a curva da taxa de juros americana, foram obtidos dados mensais, de 31/03/1978 a 31/03/2019, das taxas de juros governamentais genéricas para diversas maturidades, conforme mostra a Tabela 3.1.

Tabela 3.1: Índices utilizados para modelar a curva de juros americana

<i>Bloomberg Ticker</i>	Nome
USGG3M Index	US Generic Govt 3 Month Yield
USGG6M Index	US Generic Govt 6 Month Yield
USGG12M Index	US Generic Govt 12 Month Yield
USGG2Y Index	US Generic Govt 2 Year Yield
USGG3Y Index	US Generic Govt 3 Year Yield
USGG5YR Index	US Generic Govt 5 Year Yield
USGG7YR Index	US Generic Govt 7 Year Yield
USGG10YR Index	US Generic Govt 10 Year Yield
USGG30YR Index	US Generic Govt 30 Year Yield

Foram selecionadas variáveis macroeconômicas que representam o estado da economia americana e que tenham baixa correlação entre si. As variáveis selecionadas estão descritas na Tabela 3.2.

Tabela 3.2: Variáveis macroeconômicas selecionadas

<i>Bloomberg Ticker</i>	Nome	Fator representado
CPI CHNG Index	US CPI Urban Consumers MoM SA	Inflação
USURTOT Index	U-3 US Unemployment Rate Total	Desemprego
CPFTQOQ Index	US Corporate Profits with IVA	Lucros Corporativos
IP CHNG Index	US Industrial Production MOM SA	Produção Industrial

Para prever o valor futuro das variáveis econômicas, foram selecionados indicadores de antecipação relevantes para essa previsão. Indicador de antecipação é qualquer fator econômico que muda antes do resto da economia começar a ir em uma direção particular e pode ajudar na previsão de mudanças significativas na economia. Os 40 indicadores descritos na Tabela 3.3 foram inicialmente selecionados considerando sua relevância e a existência de dados no período mínimo definido. Desse conjunto inicial, foram obtidas novas séries através de transformações, e um subconjunto foi selecionado para a previsão de cada variável macroeconômica apresentada na Tabela 3.2.

Tabela 3.3: Indicadores pré-selecionados para previsão das variáveis econômicas

<i>Bloomberg Ticker</i>	Nome
CSRPAISB Index	Credit Suisse RavenPack Artificial Intelligence Sentiment Balanced
CBKIMS5U Index	Commerzbank Media Sentiment Lo
CSRPAIS Index	Credit Suisse RavenPack Artificial Intelligence Sentiment
US0003M Index	ICE LIBOR USD 3 Month
FDTR Index	Federal Funds Target Rate - Up
INJCJC Index	US Initial Jobless Claims SA
INJCJCNS Index	US Initial Jobless Claims NSA
INJCSP Index	US Continuing Jobless Claims S
JPMEQGS Index	J.P. Morgan Global Equity Sent
DOEASCRD Index	DOE Total Change in Crude Oil
BAKETOT Index	Baker Hughes United States Oil
INJCJC4 Index	US Initial Jobless Claims 4 We
SPX Index	S&P 500 INDEX
NHSLSUPP Index	US New One Family Houses Month
NHSLCHNG Index	US New One Family Houses Sold
MTIB Index	US Manufacturing & Trade Inventory
MTIBCHNG Index	US Manufacturing & Trade Inventory Change
RSTAMOM Index	Adjusted Retail & Food Service
NHSPSTOT Index	US New Priv. Owned Housing Units Started Total SAAR
CONSENT Index	University of Michigan Consumer Sentiment Index
SNTEUSGX Index	sentix Economic Indices US Aggregate
ETSLMOM Index	US Existing Homes Sales MoM SA
USRINDEX Index	U.S. Recession Indicator
FDDSGDP Index	US Treasury Federal Budget Deficit
NFP TCH Index	US Employees on Nonfarm Payrolls Total MoM Net Change SA
CHPMINDX Index	Chicago Purchasing Managers Index
DGNOCHNG Index	US Durable Goods New Orders Industries MoM SA
PPI CHNG Index	US PPI Finished Goods SA MoM%
TIC NLST Index	Net Long Term Securities Trans
TMNOCHNG Index	US Manufacturers New Orders Total MoM SA
CNSTTOTA Index	Census Bureau US Construction
CONCCONF Index	Conference Board Consumer Confidence SA
PITLCHNG Index	US Personal Income MoM SA
PCE CRCH Index	US Personal Consumption Expenditures Nom. Dollars MoM SA
GDPCTOT%	GDP US Personal Consumption Change
GPDITOC%	US Gross Private Domestic Investments
GDPTEXP%	GDP US Exports
CONSEXP	The University of Michigan Consumer Expectations Index
GDPTIMP%	GDP US Imports
EHBBUS	US Budget Balance (% GDP)

3.3 Preparação dos Dados

As séries temporais obtidas foram transformadas antes do treinamento dos modelos por meio de:

- interpolação e *resampling*, para que as séries tenham a mesma frequência (mensal);
- uma diferenciação da série, quando necessário, para gerar séries estacionárias; e
- diferenciação fracional das séries, quando possível, para gerar séries estacionárias com mais memória.

3.3.1 Interpolação e *Resampling*

Para séries com frequência maior que mensal (semanal, diária, intradiária), a frequência foi transformada para mensal se obtendo o último valor do período ou a média do período, de acordo com a série.

Para séries com frequência trimestral, valores mensais foram obtidos utilizando o método de interpolação *Piecewise Cubic Hermite Interpolating Polynomial* (PCHIP). A vantagem deste método em relação ao método *Piecewise Polynomial* (spline) é que, enquanto ambos ajustam uma curva aos pontos para interpolar, o método PCHIP não introduz novos extremos e gera uma curva em que os pontos gerados por interpolação nunca excederão o *range* da série original, conforme pode ser visto na Figura 3.3.

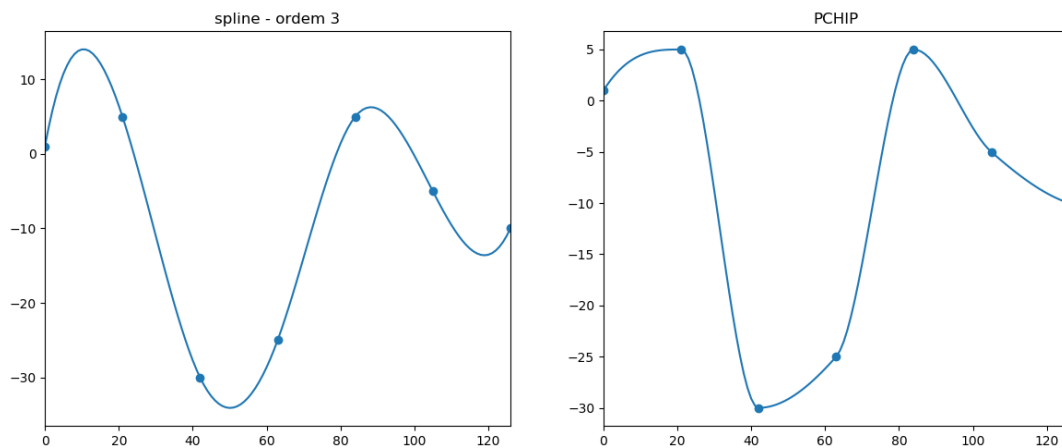


Figura 3.3: Comparação entre os métodos de interpolação spline e PCHIP.

3.3.2 Diferenciação Fracional

É comum, em finanças, encontrar séries temporais não-estacionárias. O que torna estas séries não-estacionárias é a presença de memória, ou seja, uma história de níveis prévios

que alteram a média ao longo do tempo. Para realizar análise inferencial, é desejável trabalhar com processos invariantes, tais como retornos sobre preços, mudanças em volatilidade, dentre outros [30]. Estas transformações tornam uma série estacionária às custas de eliminar toda memória presente na série original [51]. Embora estacionariedade seja uma propriedade necessária para propósitos de inferência, a eliminação de toda memória pode reduzir o poder preditivo do modelo [11].

Considere o operador de defasagem B aplicado a uma matriz de variáveis de valores reais $\{X_t\}$, onde $B^k X_t = X_{t-k}$ para qualquer inteiro $k \geq 0$. Note que $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$ para n positivo. Para um número real d , $(1 + x)^d = \sum_{k=0}^{\infty} \binom{d}{k} x^k$. Em um modelo fracional, permite-se que o expoente d seja um número real, com a seguinte expansão formal da série binomial [30]:

$$(1 - B)^d = 1 - dB + \frac{d(d-1)}{2!} B^2 - \frac{d(d-1)(d-2)}{3!} B^3 + \dots \quad (3.1)$$

Uma das formas de se realizar a diferenciação fracional é utilizando uma janela de tamanho fixo, isto é, descartando os pesos após o seu módulo cair abaixo de um dado limite (τ). Isto é o equivalente a achar os primeiros l tal que $|w_l| \geq \tau$ e $|w_{l+1}| \leq \tau$ [30].

Após a transformação de cada série com um determinado valor de d , a série resultante foi testada para avaliar se esta apresenta estacionariedade. O teste utilizado foi o Teste de Dickey-Fuller aumentado, e a série foi considerada estacionária se o valor- p resultante foi abaixo de 5%. Após essa avaliação inicial, as séries consideradas não-estacionárias foram fracionalmente diferenciadas com o menor valor de d para o qual a série foi considerada estacionária, desde que a série resultante apresentasse dados para o período definido para a pesquisa.

3.4 Modelagem

O modelo definido para otimização do portfólio de títulos de renda fixa se dividiu em três fases principais: previsão de variáveis macroeconômicas, com base em indicadores de antecipação; previsão da curva de juros, utilizando-se da previsão das variáveis macroeconômicas combinada com as *views* dos analistas para a economia; e otimização do portfólio, com base na curva de juros prevista.

3.4.1 Previsão das Variáveis Macroeconômicas

3.4.1.1 Seleção das Variáveis e Redução de Dimensionalidade

A fase de seleção das variáveis independentes recebeu como entrada as séries estacionárias originais, as séries estacionárias geradas pelas transformações aplicadas nas séries originais

não-estacionárias (interpolação, *resampling*, diferenciação inteira e diferenciação fracional) e a própria variável dependente.

Para a seleção das variáveis relevantes, foi utilizado o algoritmo Boruta, definido por Kursa e Witold (2010) [52], para encontrar todas as variáveis relevantes e não apenas as variáveis não-redundantes. Esse método tem a vantagem de educar o usuário sobre quais variáveis tem um impacto na previsão de cada variável macroeconômica.

O algoritmo Boruta utiliza o algoritmo de Florestas Aleatórias, obtendo a medida de importância de cada variável independente para cada árvore, calculando a média e o desvio-padrão das importâncias das variáveis independentes para todas as árvores e calculando o *z-score* de cada variável ao dividir um pelo outro. Como o *z-score* não pode ser utilizado diretamente para decidir se uma variável é significativa, o algoritmo cria atributos “sombra”, por definição aleatórios, para utilizar como referência.

O algoritmo Boruta consiste dos seguintes passos [52]:

1. Para cada atributo, criar um atributo “sombra” correspondente, cujos valores são obtidos por embaralhar os valores do atributo original.
2. Rodar um modelo de Florestas Aleatórias e capturar os *z-scores* computados.
3. Achar o máximo *z-score* entre atributos “sombra” (MZSA) e adicionar um sucesso para cada atributo que teve um resultado melhor que MZSA.
4. Para cada atributo com importância indeterminada, realizar um teste de igualdade com MZSA.
5. Considerar atributos que pontuaram significativamente abaixo de MZSA como “não-importantes” e os remover permanentemente da busca.
6. Considerar atributos que pontuaram significativamente acima de MZSA como “importantes”.
7. Remover todos os atributos “sombra”.
8. Repetir o procedimento até que a importância esteja definida para todos os atributos ou até que o algoritmo alcance limite previamente definido do número de execuções de Florestas Aleatórias.

A importância dos atributos “sombra” só pode ser não-zero devido a flutuações aleatórias e, portanto, suas importâncias são utilizadas como uma referência para decidir quais atributos são realmente importantes. A medida de importância também varia devido à natureza estocástica do modelo de Florestas Aleatórias. Além disto, é fornecida uma semente diferente para gerar aleatoriamente os resultados em cada execução do algoritmo da

Floresta Aleatória, de maneira que os valores de importância das variáveis sejam também diferentes entre execuções.

Para confirmar ou rejeitar uma variável original, o algoritmo Boruta compara o número de sucessos a uma distribuição binomial a cada iteração. Se o número de sucessos de uma variável tem p-valor igual ou maior ao limite estabelecido, esta é rejeitada e classificada como não-importante. Se, por outro lado, o número de sucessos ultrapassa de maneira significativa o valor esperado pela distribuição binomial, ou seja, é menor que o limite estabelecido, a variável é considerada importante.

O algoritmo Boruta foi aplicado apenas sobre as observações da primeira metade das séries temporais, que foram posteriormente utilizadas apenas para treinamento e não para avaliação dos modelos. Este cuidado seguiu a orientação de James *et al* (2014) [53] para que o conjunto de dados utilizado na seleção de variáveis não seja utilizado na validação dos modelos, pois tal fato levaria a que os erros de validação fossem estimativas incorretas dos erros do modelo com dados desconhecidos.

Para este projeto, o p-valor para determinar a importância das variáveis independentes foi definido como 0,005, a profundidade máxima das árvores foi definida como 5, o número de árvores de decisão foi definido como 10 e o número máximo de variáveis a considerar quando buscando a melhor divisão foi definido como um terço do número total de variáveis a selecionar.

Após a seleção das variáveis importantes, foi aplicada Análise de Componentes Principais sobre estas e foi selecionado o menor conjunto dos componentes principais cuja variância explicada fosse maior ou igual a 75% para serem utilizados como entrada no treinamento dos modelos de previsão.

3.4.1.2 Treinamento das Redes Neurais

Foram treinadas redes neurais com uma camada do tipo LSTM e *Minimal* RNN, todas redes do tipo recorrente.

Para selecionar a melhor configuração de parâmetros para as redes neurais recorrentes, foi realizado ajuste dos hiperparâmetros de cada tipo de rede neural e selecionada a configuração com melhor resultado em relação às 4 variáveis macroeconômicas, conforme descrito na Equação 3.6.

Após a seleção da arquitetura para cada tipo de rede, o desempenho foi avaliado utilizando o método *walk-forward* padrão e avaliado com os dados dos últimos 12 meses. Os RMSEs dos modelos foram comparados entre si e comparados com um modelo autorregressivo, para cada variável.

3.4.1.3 Incorporação das *views* dos estrategistas

A fim de permitir que os estrategistas incorporem suas previsões sobre as variáveis macroeconômicas às previsões das redes neurais, o modelo Black-Litterman, descrito na Subseção 2.1.4, foi modificado, com a eliminação do procedimento de otimização reversa sobre um portfólio global de ativos financeiros e com a inclusão das previsões das redes neurais em seu lugar.

Dessa forma, em vez de obter-se o vetor de retornos esperados implícitos para ativos financeiros, o modelo Black-Litterman modificado recebe diretamente o vetor de mudanças nas variáveis macroeconômicas e, portanto, as *views* dos analistas e os desvios-padrões associados a elas são relativos à variação das variáveis macroeconômicas em vez de serem associados a retornos de ativos financeiros, que é o uso clássico do modelo Black-Litterman. Com essa alteração, a Equação 2.7 passou a não ser utilizada no modelo proposto.

Não se objetivou avaliar o desempenho do modelo proposto com a incorporação das *views* dos analistas, sendo esta etapa do modelo implementada na ferramenta a ser entregue aos estrategistas do Banco Central do Brasil.

3.4.2 Previsão da Curva de Juros

Na etapa de previsão dos fatores do modelo Nelson-Siegel Dinâmico, o método de seleção de variáveis descrito na Subseção 3.4.1.1 foi aplicado na seleção das variáveis macroeconômicas mais relevantes para prever cada um dos fatores do modelo NSD. O método de treinamento descrito na Subseção 3.4.1.2 e o método de validação cruzada descrito na Subseção 3.5.1 também foram utilizados para previsão dos fatores do modelo NSD e avaliação dos modelos treinados, respectivamente.

Com a previsão obtida para os fatores do modelo Nelson-Siegel Dinâmico, foram calculadas as curvas de taxas de juros previstas.

3.4.3 Sugestão de Alocação do Portfólio

O desempenho do modelo sugerido foi comparado ao desempenho mensal, nos últimos 12 meses, da carteira de títulos precificados em dólar do portfólio de referência das reservas internacionais. Dado que as informações sobre alocação e retorno das reservas internacionais estão disponíveis oficialmente apenas em nível agregado, não foi possível fornecer informações no nível de detalhe de retornos e alocações mensais das carteiras e títulos, preservando-se o mesmo nível de agrupamento das informações publicamente disponibilizadas.

Para sugerir mudanças de alocação do portfólio de títulos de renda fixa, as curvas de juros previstas foram comparadas às curvas de juros reais. Dado que as informações sobre

as curvas de juros são conhecidas pelos gestores das reservas internacionais no momento da tomada de decisão sobre alocação, considerou-se que esse dado é levado em consideração para sugerir tal alocação, em conjunto com outras informações relevantes. Portanto, também considerou-se que mudanças nas curvas de juros impactam as alocações dos ativos em carteira de maneira proporcional à mudança, ou seja, as mudanças nas taxas de juros levam os gestores a modificar as alocações de seus portfólios conforme títulos chegam a termo. Portanto, para sugerir mudança na alocação, considerou-se a diferença entre a curva de juros prevista para o tempo $t + 1$ e a curva de juros real no tempo t , conforme descrito a seguir.

Tendo os dados da curva de juros atual $C^{(t)}$ e da curva de juros prevista $\hat{C}^{(t+1)}$, vetores compostos de taxas de juros $j_m^{(t)}$ e $\hat{j}_m^{(t+1)}$, respectivamente, onde m é a maturidade obtida do conjunto de maturidades M , gerou-se o seguinte vetor de proporções:

$$P^{(t+1)} = \hat{C}^{(t+1)} \oslash C^{(t)} \quad (3.2)$$

Onde \oslash é símbolo para divisão elemento a elemento.

Para cada tipo de instrumento de renda fixa, obteve-se o vetor de alocação $A^{(t)}$, com as taxas de juros para as maturidades em M e gerou-se nova alocação sugerida por meio de:

$$\hat{A}^{(t+1)} = \frac{A^{(t)} \odot P^{(t+1)}}{\text{soma}\{A^{(t)} \odot P^{(t+1)}\}} \quad (3.3)$$

garantindo desta forma que a alocação sugerida se mantenha com soma igual a 1. Esse procedimento é repetido para cada tipo de instrumento de renda fixa e modifica a alocação entre títulos de mesmo tipo mas com diferentes maturidades mas não aumenta ou diminui o montante total destinado àqueles papéis como um todo.

Tendo a alocação sugerida $\hat{A}^{(t+1)}$ mês a mês e um vetor de retornos mensais R , foi feita a comparação do retorno obtido se 1 dólar fosse aplicado no início do período até a atualidade, tanto na carteira de referência quanto na alocação sugerida pelo modelo.

3.5 Avaliação

O modelo proposto teve diversos de seus produtos intermediários avaliados, além de seu resultado final. Avaliações intermediárias envolveram a comparação das predições das redes neurais com predições de modelos de autoregressão. A avaliação do resultado final envolveu a comparação do resultado obtido com a estratégia de alocação sugerida pelo modelo com os resultados obtidos pela carteira de referência.

Para a avaliação do desempenho das redes neurais, foi utilizado método descrito na Subseção 3.4.1.2. Para avaliação do resultado final, foi utilizado método de validação *walk-forward* em que os retornos foram calculados mês a mês e reinvestidos.

3.5.1 Validação Cruzada

Foi adotado como método de validação, uma versão híbrida do método *walk-forward* e do método de divisão treinamento/teste. Dado que as séries tem T pontos de dados, ordenados temporalmente:

1. As redes neurais foram treinadas na primeira metade dos dados das séries (de 1 a $T/2$);
2. Sobre a segunda metade dos dados (de $T/2 + 1$ a T), foi utilizado método de aprendizado de máquina *online* em que os pontos de dados das séries no tempo t foram utilizados para atualizar o treinamento das redes neurais, que foram então utilizadas para prever o valor da variável macroeconômica em $t + 1$; e
3. A série de previsões e a série de cada variável macroeconômica, de $T/2 + 1$ a T , foram utilizadas para calcular a Raíz do Erro Médio Quadrático.

Esse procedimento foi adotado com o objetivo de se alcançar as seguintes vantagens:

- utilizar a primeira metade dos dados para seleção das variáveis de entrada para previsão de cada fator macroeconômico, com menor risco de sobreajuste, uma vez que estes dados foram utilizados apenas para o treinamento do modelo, e não para sua avaliação;
- dar enfoque maior nos dados mais recentes;
- ter modelos que possam ser rapidamente atualizados com novos dados, através do aprendizado de máquina *online*; e
- manter a característica de simular a forma como um investidor agiria e de como o modelo será utilizado na prática, isto é, observando os dados passados e buscando prever o próximo ponto de dado de cada série no futuro.

Optou-se por não utilizar o método *k-fold* combinatório purgado por este ser recente e não ter sido identificada a sua utilização em nenhum trabalho acadêmico até o presente momento.

3.5.2 Métricas de Avaliação

3.5.2.1 Avaliação da previsão de séries temporais

Para avaliar o desempenho das redes neurais na predição dos valores futuros das variáveis macroeconômicas e dos fatores da curva de juros calculados pelo modelo Nelson-Siegel Dinâmico, foi utilizada como métrica a Raíz do Erro Quadrático Médio (RMSE), calculada da seguinte forma:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}} \quad (3.4)$$

Onde y_t é o valor da variável no tempo t , \hat{y}_t é a previsão do modelo para a variável no tempo t e T é o número de pontos de dados. Para fins de comparação entre séries temporais diversas, o RMSE será calculado após a transformação das séries para z -score, da seguinte forma:

$$z = \frac{x - \mu}{\sigma} \quad (3.5)$$

Onde μ é a média dos valores na série temporal e σ é o desvio padrão da série.

Como os modelos foram treinados sobre diversas variáveis econômicas, para selecionar uma única arquitetura de cada tipo de rede neural para todas as variáveis, buscou-se uma métrica a maximizar que fosse inversamente proporcional à média e ao desvio-padrão dos RMSEs (μ_{RMSE} e σ_{RMSE} , respectivamente) das variáveis macroeconômicas, para selecionar uma única arquitetura para todas as variáveis. O objetivo de utilizar também o desvio-padrão, e não apenas a média, é buscar uma arquitetura com um desempenho mais consistente e, assim, reduzir o sobreajuste. A fórmula utilizada foi:

$$\frac{1}{\mu_{\text{RMSE}} \sigma_{\text{RMSE}}} \quad (3.6)$$

Onde $\mu_{\text{RMSE}} = \frac{\sum_{v=1}^V \text{RMSE}_v}{V}$ e $\sigma_{\text{RMSE}} = \sqrt{\frac{\sum_{v=1}^V (\text{RMSE}_v - \mu_{\text{RMSE}})^2}{V}}$.

3.5.2.2 Avaliação dos resultados da estratégia

Para avaliar o resultado da implementação da estratégia sugerida pelo modelo proposto, sem a incorporação das *views* dos analistas, na alocação das reservas internacionais, a métrica Índice de Sharpe foi utilizada, calculada da seguinte forma:

$$\text{Índice de Sharpe} = \frac{R_p - R_f}{\sigma_p} \quad (3.7)$$

Onde R_p é o retorno do portfólio, R_f é o retorno de um ativo livre de risco e σ_p é o desvio padrão do retorno do portfólio.

3.6 Implantação

O modelo decorrente da aplicação do método descrito será disponibilizado como um modelo quantitativo, através de interface gráfica, para uso dos analistas da divisão de estratégia em tempo real. Os resultados serão exibidos na forma de tabelas e gráficos. As matrizes de pesos das redes neurais serão atualizadas mensalmente, após treinamento das redes, conforme os dados de entrada estejam disponíveis.

Capítulo 4

Resultados

4.1 Previsão das Variáveis Macroeconômicas

4.1.1 Seleção das Variáveis e Redução de Dimensionalidade

Utilizando o método Boruta, descrito na Subseção 3.4.1.1, os indicadores de antecipação descritos nas Tabelas 4.1, 4.2, 4.3 e 4.4 foram selecionados dentre os apresentados na Tabela 3.3, por serem classificados pelo algoritmo como relevantes para a previsão das variáveis Inflação, Desemprego, Produção Industrial e Lucros Corporativos, respectivamente.

Tabela 4.1: Variáveis selecionadas como relevantes para a previsão da variável Inflação

<i>Bloomberg Ticker</i>	Nome
GDPCTOT%	GDP US Personal Consumption Change
GPDITOC%	US Gross Private Domestic Investments
GDPTEXP%	GDP US Exports
GDPTIMP%	GDP US Imports
EHBBUS	US Budget Balance (% GDP)
MTIBCHNG	US Manufacturing & Trade Inventory Change
PPI CHNG	US PPI Finished Goods SA MoM%
CONSEXP	The University of Michigan Consumer Expectations Index
PITLCHNG	US Personal Income MoM SA
SPX_frac0.72	S&P 500 Index (diferenciado fracionalmente, $d=0.72$)
CPI CHNG	Inflação

Tabela 4.2: Variáveis selecionadas como relevantes para a previsão da variável Desemprego

<i>Bloomberg Ticker</i>	Nome
GDPCTOT%	GDP US Personal Consumption Change
USRINDEX	US Recession Indicator
NFP TCH	US Employees on Nonfarm Payrolls Total MoM Net Change SA
CHPMINDX	Chicago Purchasing Managers Index
CONSEXP	The University of Michigan Consumer Expectations Index
TMNOCHNG	US Manufacturers New Orders Total MoM SA
USURTOT	Desemprego

Tabela 4.3: Variáveis selecionadas como relevantes para a previsão da variável Produção Industrial

<i>Bloomberg Ticker</i>	Nome
INJCJC	US Initial Jobless Claims SA
INJCJC4	US Initial Jobless Claims 4 Week Moving Average SA
GDPCTOT%	GDP US Personal Consumption Change
GPGSTOC%	US GDP Govt Purchases & Investment
GDPTEXP%	GDP US Exports
GDPTIMP%	GDP US Imports
EHUPUS	US Unemployment Rate (%)
USRINDEX	US Recession Indicator
NFP TCH	US Employees on Nonfarm Payrolls Total MoM Net Change SA
CHPMINDX	Chicago Purchasing Managers Index
IP CHNG	Produção Industrial
CONSEXP	The University of Michigan Consumer Expectations Index
TMNOCHNG	US Manufacturers New Orders Total MoM SA
PCE CRCH	US Personal Consumption Expenditures Nom. Dollars MoM SA
CNSTTOTA_frac0.86	Census Bureau US Construction (frac. dif d=0.86)
SPX_diff1	S&P 500 Index (diferenciado)

Tabela 4.4: Variáveis selecionadas como relevantes para a previsão da variável Lucros Corporativos

<i>Bloomberg Ticker</i>	Nome
INJCJC	US Initial Jobless Claims SA
INJCSP	US Continuing Jobless Claims SA
INJCJC4	US Initial Jobless Claims 4 Week Moving Average SA
GDPCTOT%	GDP US Personal Consumption Change
GPDITOC%	US Gross Private Domestic Investments
GDPTEXP%	GDP US Exports
GDPTIMP%	GDP US Imports
EHUPUS	US Unemployment Rate (%)
EHBBUS	US Budget Balance (% GDP)
NFP TCH	US Employees on Nonfarm Payrolls Total MoM Net Change SA
CHPMINDX	Chicago Purchasing Managers Index
IP CHNG	Produção Industrial
CONSEXP	The University of Michigan Consumer Expectations Index
MTIB_frac0.72	US Manufacturing & Trade Inventory (frac. dif d=0.72)
FDTR_diff1	Federal Funds Target Rate - Up (diferenciado)
CNSTTOTA_diff1	Census Bureau US Construction (diferenciado)
CPFTQOQ	Lucros Corporativos

Após a seleção das variáveis relevantes para cada variável macroeconômica, as séries selecionadas foram transformadas com a aplicação de Análise de Componentes Principais e foram selecionados os componentes principais que explicavam 75% ou mais da variância da respectiva variável-alvo. Dessa forma, foram obtidas as variáveis independentes a serem utilizadas no treinamento das redes neurais. O número de variáveis independentes resultantes para previsão de cada variável econômica e a variância explicada por essas variáveis independentes estão apresentados na Tabela 4.5.

Tabela 4.5: Número de componentes principais selecionados como variáveis de entrada e a variância explicada por eles para cada variável macroeconômica.

Variável econômica	Nº de componentes principais	Variância explicada (%)
Inflação	5	78,1%
Desemprego	3	78,3%
Produção Industrial	6	76,9%
Lucros Corporativos	6	76,5%

4.1.2 Ajuste dos Hiperparâmetros

Para a seleção do modelo a ser utilizado e definição dos hiperparâmetros de cada modelo, foi utilizado método de busca em grade. Cada modelo foi treinado e avaliado utilizando a metade mais recente de cada variável macroeconômica, após estas serem convertidas para *z-score*.

Após a seleção dos hiperparâmetros, os modelos foram comparados utilizando como métrica a RMSE, descrita na Equação 3.4. O modelo cuja média do desempenho foi melhor para todas as variáveis macroeconômicas foi então utilizado para a previsão de todas as variáveis e também utilizado na previsão dos fatores da curva de juros.

4.1.2.1 Rede LSTM

Para treinamento das redes LSTM, foi utilizada a camada CuDNNLSTM do pacote Keras, uma implementação da camada LSTM para execução sobre a GPU, de mais rápido treinamento que a camada LSTM padrão.

Os parâmetros listados abaixo, com os respectivos valores assumidos, foram utilizados na busca pelo melhor modelo:

- **Número de neurônios na camada recorrente** (neurônios): número de neurônios da camada LSTM. Número inteiro positivo. Valores utilizados: 1, 3, 5, 8 e 13; e

- **Dropout dos valores de saída da camada recorrente** (dropout): proporção de unidades, escolhidas aleatoriamente, a serem eliminadas, ou seja, definidas com o valor 0, a cada atualização durante o treinamento, para ajudar a prevenir o sobreajuste. Quando o valor é zero, considera-se que não há unidade de *dropout*. Número real entre 0 e 1. Valores utilizados: 0.0, 0.2, 0.4, 0.6 e 0.8.

A combinação dos hiperparâmetros descritos acima resultou em 25 redes neurais LSTM distintas sendo treinadas para cada uma das 4 variáveis macroeconômicas e comparadas por meio do cálculo da Raiz do Erro Médio Quadrático (RMSE) sobre a segunda metade das séries macroeconômicas, utilizando método *walk-forward*. A Figura 4.1 ilustra a arquitetura de uma das redes LSTM que foram treinadas e avaliadas.

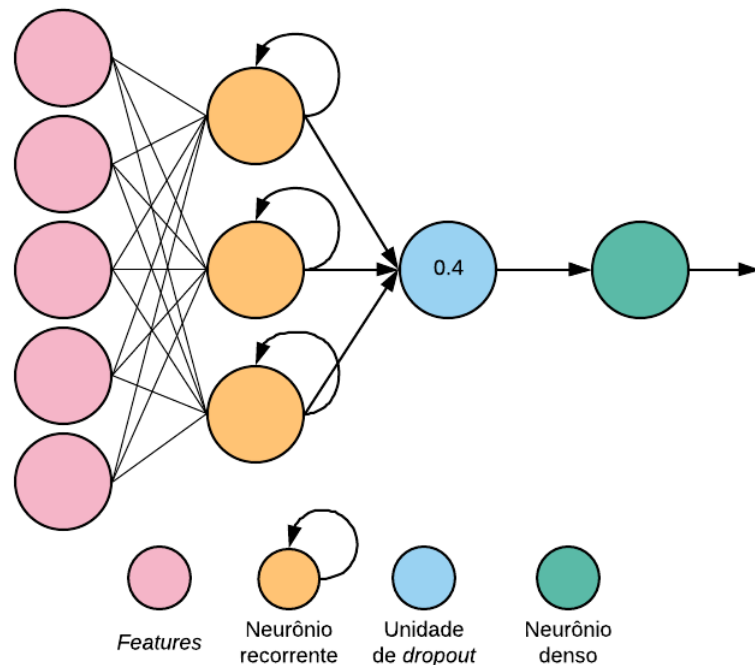


Figura 4.1: Ilustração de uma das redes LSTM treinadas e avaliadas no ajuste dos hiperparâmetros. Esta rede específica tem 3 neurônios na camada recorrente e valor de *dropout* de 0.4.

A Tabela 4.6 apresenta as 5 redes com melhor RMSE para cada variável macroeconômica. Já os mapas de calor na Figura 4.2 sintetizam a informação sobre a avaliação de todas as redes LSTM treinadas.

Tabela 4.6: *Rankings* das 5 redes LSTM com melhor desempenho por fator econômico

Fator	Posição	Hiperparâmetros		<i>RMSE</i>
		neuronios	dropout	
Inflação	1	1	0.8	0.961
	2	3	0.8	0.980
	3	1	0.6	1.009
	4	8	0.8	1.023
	5	5	0.8	1.037
Desemprego	1	1	0.6	1.137
	2	1	0.4	1.143
	3	3	0.8	1.143
	4	1	0.8	1.145
	5	5	0.8	1.152
Produção Industrial	1	1	0.4	1.011
	2	1	0.8	1.031
	3	5	0.8	1.035
	4	3	0.8	1.036
	5	1	0.6	1.041
Lucros Corporativos	1	13	0.2	0.775
	2	3	0.4	0.777
	3	5	0.8	0.783
	4	1	0.4	0.791
	5	3	0.8	0.796

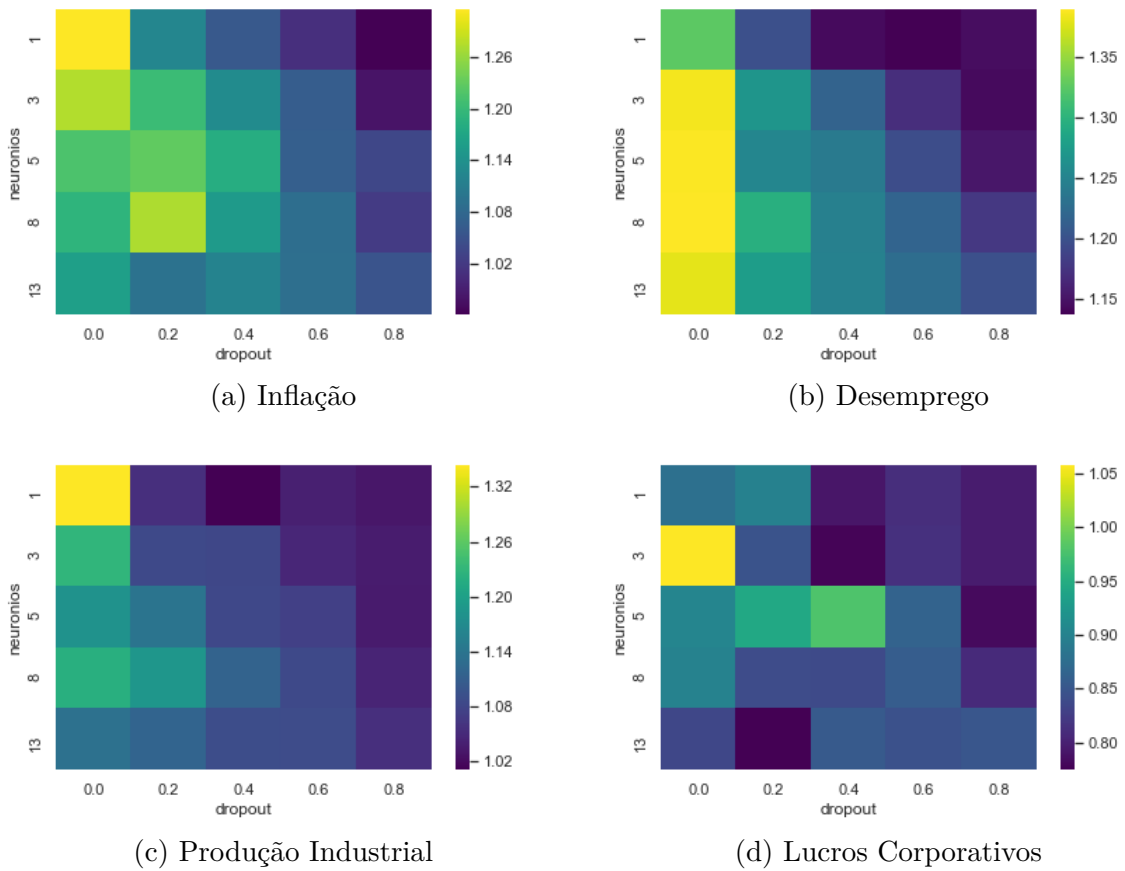


Figura 4.2: Mapas de calor dos RMSEs obtidos na previsão das variáveis macroeconômicas utilizando método de busca em grade com redes LSTM.

A arquitetura LSTM selecionada para previsão das variáveis macroeconômicas e dos fatores da curva de juros, utilizando a Equação 3.6, tem **5** neurônios na camada LSTM e *dropout* de **0.4**, e teve RMSE médio de 1,1.

4.1.2.2 Rede *Minimal* RNN

À rede *Minimal* RNN, criada por Chen (2017) [4], foram aplicados os mesmos procedimentos adotados para a rede LSTM de ajuste dos hiperparâmetros e seleção da arquitetura.

Os parâmetros listados abaixo, com os respectivos valores assumidos, foram utilizados na busca pelo melhor modelo:

- **Número de neurônios** (`num_neur`): número de neurônios da camada *Minimal* RNN. Número inteiro positivo. Valores utilizados: 1, 5, 10 e 15;
- **Dropout** (`drop`): proporção das unidades a deixar de fora da transformação linear das entradas. Número real entre 0 e 1. Valores utilizados: 0.0, 0.3 e 0.6;

- **Dropout recorrente** (`drop_rec`): proporção das unidades a deixar de fora da transformação linear do estado recorrente. Número real entre 0 e 1. Valores utilizados: 0.0 e 0.3;

A Tabela 4.7 apresenta as 5 melhores redes para cada variável econômica, por ordem ascendente dos RMSEs calculados.

A combinação dos hiperparâmetros descritos acima resultou em 24 redes neurais *Minimal RNN* distintas sendo treinadas para cada uma das 4 variáveis macroeconômicas e comparadas por meio do cálculo da Raiz do Erro Médio Quadrático (RMSE) sobre a segunda metade das séries macroeconômicas, utilizando método *walk-forward*.

A arquitetura *Minimal RNN* selecionada para previsão das variáveis macroeconômicas e dos fatores da curva de juros, utilizando a Equação 3.6, tem **1** neurônio na camada *Minimal RNN*, *dropout* de **0.6** e *dropout* recorrente de **0**, e teve RMSE médio de 0,991.

Tabela 4.7: *Rankings* das 5 redes *Minimal RNN* com melhor desempenho por fator econômico

Fator	Posição	Hiperparâmetros			Score
		neuronios	dropout	recur_dropout	
Inflação	1	1	0.6	0.0	1.0291
	2	1	0.6	0.3	1.0306
	3	1	0.3	0.0	1.0837
	4	5	0.3	0.0	1.0842
	5	5	0.3	0.3	1.0898
Desemprego	1	1	0.6	0.3	1.0912
	2	1	0.6	0.0	1.0958
	3	1	0.3	0.3	1.1702
	4	1	0.3	0.0	1.1780
	5	5	0.6	0.0	1.2113
Produção Industrial	1	1	0.6	0.3	0.9731
	2	1	0.6	0.0	0.9737
	3	1	0.3	0.3	1.0292
	4	1	0.3	0.0	1.0326
	5	15	0.3	0.3	1.0397
Lucros Corporativos	1	10	0.3	0.3	0.7245
	2	15	0.3	0.0	0.7351
	3	15	0.3	0.0	0.7367
	4	15	0.6	0.3	0.7471
	5	5	0.3	0.0	0.7600

4.1.3 Comparação do Desempenho dos Modelos

Com as arquiteturas selecionadas para as redes do tipo LSTM e *Minimal* RNN, as redes foram treinadas com método *walk-forward* e treinamento em lote, seus desempenhos avaliados para os últimos 12 meses e comparados ao desempenho de um modelo autorregressivo univariado com *lag* de 18 observações.

Os desempenhos das redes neurais e do modelo autorregressivo estão apresentados na Tabela 4.8. Os resultados mostram que a rede neural recorrente teve desempenho similar ao do modelo autorregressivo.

Variável econômica	LSTM	Minimal RNN	Autorregressivo
Inflação	0.396	0.464	0.420
Desemprego	1.227	1.249	1.242
Produção Industrial	0.807	0.743	0.712
Lucros Corporativos	0.305	0.306	0.360
<i>RMSE médio</i>	0.684	0.691	0.684

Tabela 4.8: Desempenho dos modelos de redes neurais recorrentes e de modelo autorregressivo.

4.2 Previsão dos Fatores da Curva de Taxas de Juros

Os fatores nos quais o modelo Nelson-Siegel Dinâmico, descrito na Subseção 2.1.5.1, decompõe a curva de juros podem ser interpretados como o nível da curva de juros, a inclinação da curva de juros e a curvatura da curva de juros. Esses fatores foram previstos utilizando como entrada as previsões das variáveis macroeconômicas selecionadas e seu próprio valor anterior.

Primeiramente, o modelo Nelson-Siegel Dinâmico foi implementado e executado para obter os fatores que melhor explicam a curva de juros americana no período da pesquisa, utilizando-se os índices listados na Tabela 3.1. Os fatores podem ser visualizados na Figura 4.3.

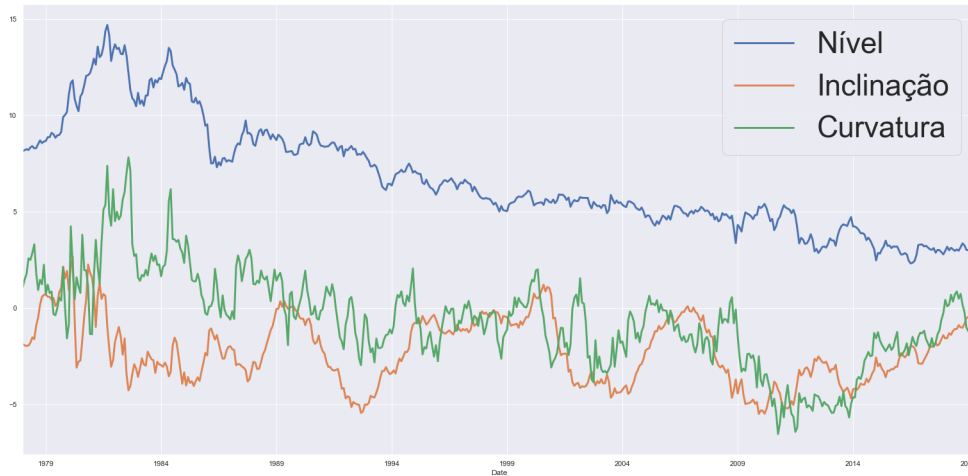


Figura 4.3: Fatores do modelo Nelson-Siegel Dinâmico para a curva de juros norte-americana entre 31/01/1978 e 31/05/2019.

4.2.1 Seleção das Variáveis de Entrada

Para selecionar quais variáveis macroeconômicas foram posteriormente utilizadas como variáveis de entrada, foi aplicado o método Boruta, descrito na Subseção 3.4.1.1. A única distinção é que as séries não foram transformadas utilizando PCA, uma vez que as variáveis macroeconômicas apresentam baixa correlação entre si.

Os resultados da aplicação do método Boruta foi a seleção das seguintes variáveis como relevantes para a previsão de cada um dos fatores do modelo NSD:

- **Nível:** Inflação, Produção Industrial e Lucros Corporativos;
- **Inclinação:** Inflação e Lucros Corporativos;
- **Curvatura:** Nenhuma variável.

Às variáveis selecionadas, foi acrescentada a série com o valor imediatamente anterior do próprio fator.

4.2.2 Fatores do Modelo Nelson-Siegel Dinâmico

Para prever os fatores do modelo Nelson-Siegel Dinâmico, utilizou-se procedimento similar ao descrito na Subseção 3.5.1, com a diferença de que a primeira parte dos dados abrangue todos os dados, menos os últimos 12 meses, e sobre a última parte dos dados, foi utilizado o método *walk-forward* com aprendizado *online*.

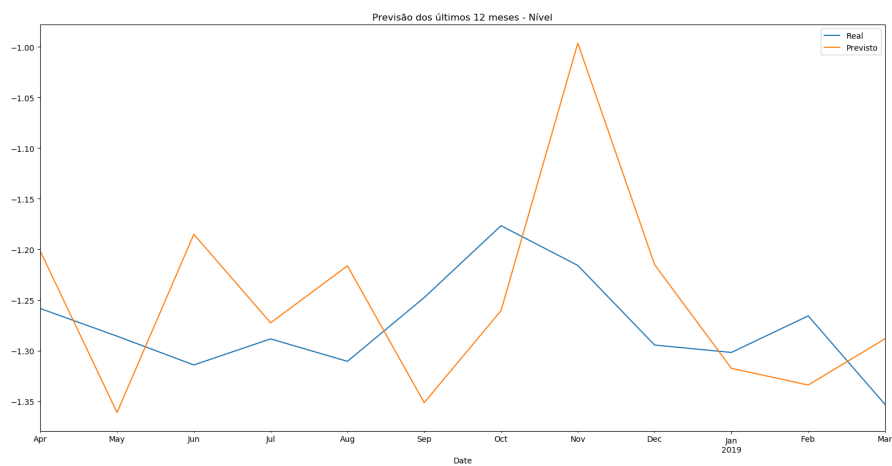
A rede neural LSTM teve o seu desempenho comparado a modelo autorregressivo univariado. Os resultados constantes da Tabela 4.9 mostram que a rede LSTM foi, em média, 6,21 vezes superior ao modelo autorregressivo na previsão dos fatores da curva de juros.

Variável econômica	LSTM	Auto-regressivo
Nível	0.085	1.345
Inclinação	0.110	0.750
Curvatura	0.200	0.369
<i>RMSE médio</i>	0.132	0.821

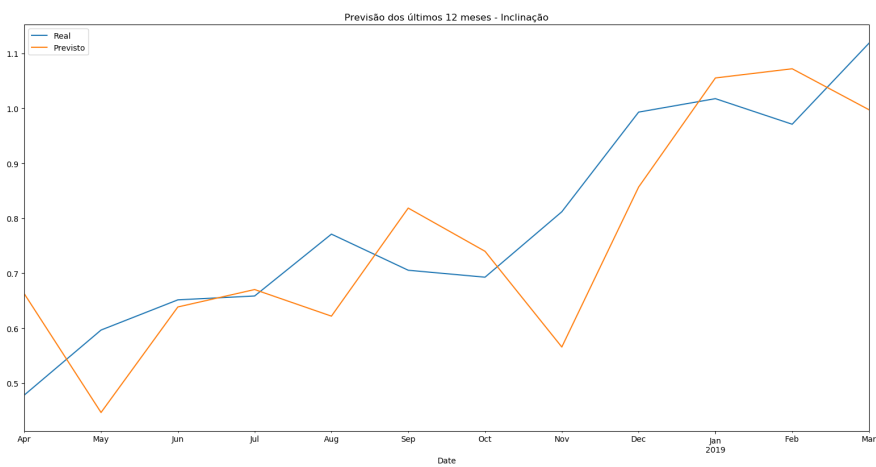
Tabela 4.9: Desempenho (RMSEs) da rede LSTM e do modelo autorregressivo para os últimos 12 meses dos fatores da curva de juros.

4.3 Implementação da Estratégia de Alocação Sugerida

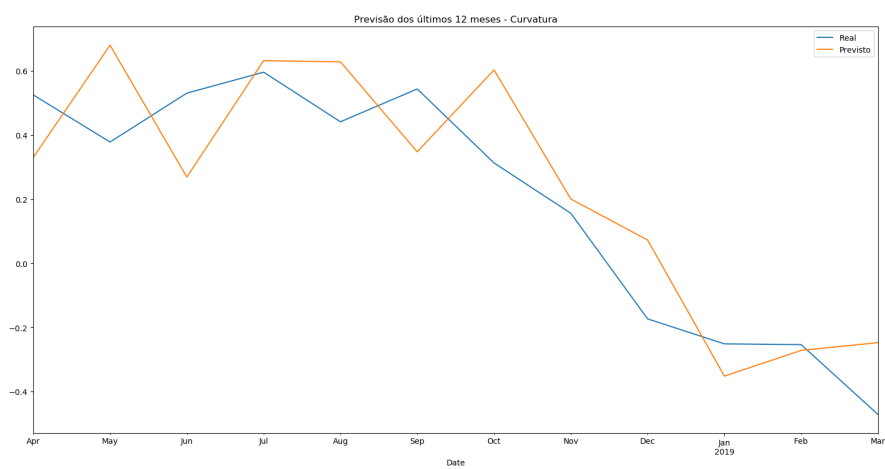
A fim de avaliar a estratégia de investimento sugerida pelo modelo proposto, a rede foi treinada, em lote, com os dados até o último ano e foi treinada e avaliada com método *walk-forward* para os últimos 12 meses das séries. A Figura 4.4 mostra as previsões das redes LSTM para os fatores da curva de juros.



(a) Nível



(b) Inclinação



(c) Curvatura

Figura 4.4: Previsão do fatores da curva de juros com a rede LSTM para os próximos 12 meses.

A aplicação da estratégia sugerida na Subseção 3.4.3 à carteira de títulos precificados em dólar, para os últimos 12 meses, trouxe retornos correspondentes a 100,9% dos retornos da carteira do portfólio de referência e, utilizando a métrica Índice de Sharpe, trouxe retornos ajustados ao risco correspondentes a 102,8% dos retornos ajustados ao risco da carteira do portfólio de referência, para o mesmo período, insuficientes para afirmar a efetividade da estratégia implementada em relação à carteira do portfólio de referência. Dadas as restrições expostas na Subseção 3.4.3, quanto ao sigilo das informações sobre as estratégias de investimento, não foi possível detalhar os ganhos individuais do portfólio de referência.

Capítulo 5

Conclusões e Trabalhos Futuros

Este capítulo apresenta as conclusões e os resultados obtidos no presente trabalho. Apresenta ainda os trabalhos futuros que se vislumbram a partir deste estudo.

5.1 Conclusões

A responsabilidade pela gestão ativa das reservas internacionais é do Departamento das Reservas Internacionais (Depin), e as estratégias de desvio do portfólio de referência são decididas em conjunto com o Diretor de Política Monetária, sendo, portanto, decisão técnica não passível de automação, tomada com base em informações de mercado, em expectativas sobre a economia e sobre os retornos dos ativos e em resultados de modelos financeiros. Tendo em vista tais peculiaridades, não se buscou, neste trabalho, substituir o processo decisório de alocação das reservas internacionais por um processo automatizado, mas acrescentar, aos modelos já existentes, cujos resultados são entrada para o processo decisório, novo modelo cujo diferencial se dá por permitir a incorporação direta das expectativas econômicas e do conhecimento técnico da equipe de analistas de estratégias na sugestão de nova alocação do portfólio, reduzindo-se assim a subjetividade na tradução desse conhecimento em parâmetros para o modelo.

Ao substituir o procedimento de otimização reversa do modelo Black-Litterman clássico pela previsão das variáveis macroeconômicas por redes neurais, o modelo adquiriu outra característica vantajosa, que é a de não depender de suposições teóricas (como a suposição de um mercado eficiente) ou meramente expedientes (como assumir um valor específico para um parâmetro), garantindo uma visão diversa da visão dos modelos clássicos já utilizados pela equipe de estrategistas, o que enriquece a análise por eles realizada.

Embora houvesse a expectativa de que, com a seleção dos indicadores de antecipação relevantes e o treinamento utilizando-se de métodos de ajuste dos hiperparâmetros, isso não se confirmou, já que os resultados obtidos com a previsão das variáveis ma-

croeconômicas pelas redes neurais recorrentes não foram melhores que os de um modelo autorregressivo univariado.

Ocorre que, apesar de o resultado apresentado ter ficado aquém do esperado, o modelo proposto pode trazer benefícios para o Banco Central do Brasil, uma vez que as previsões das redes neurais substituem o vetor de retornos esperados do modelo Black-Litterman e são combinadas com as previsões dos estrategistas, fornecendo uma previsão que combina padrões históricos capturados pelos modelos e o conhecimento econômico e financeiro dos analistas. Além disso, por ser modular, ele é capaz de evoluir conforme aperfeiçoamentos em seus componentes.

5.2 Trabalhos futuros

A partir do presente trabalho vislumbram-se diversos caminhos a serem percorridos no futuro. Por ter envolvido diversas etapas, com uso de modelos financeiros distintos e treinamento de vários modelos de aprendizado de máquina, é possível identificar futuros aperfeiçoamentos e modificações do ponto de vista do objetivo final da pesquisa, que é a otimização da alocação das reservas internacionais, ou pode-se aplicar as técnicas e modelos aqui desenvolvidos a problemas distintos, porém de valia para o Banco Central do Brasil e do ponto de vista acadêmico.

Fica como aprimoramento futuro a aplicação do mesmo modelo aqui desenvolvido às carteiras de títulos de renda fixa denominados em outras moedas que não o dólar. Para tal fim, será necessário levantar indicadores e variáveis macroeconômicas relevantes para cada mercado em questão. Outra possibilidade simples de expansão do modelo desenvolvido é a incorporação de novas variáveis macroeconômicas ao modelo e avaliação de seu poder preditivo, sem a necessidade de modificar a metodologia desenvolvida e aplicada.

Uma possibilidade de modificação é substituir o procedimento de sugestão de alocação de portfólio utilizado e descrito na Subseção 3.4.3 pela utilização do método de média-variância, descrito na Subseção 2.1.3, e avaliar o impacto desta modificação na alocação sugerida.

Um melhoramento que se pode buscar no futuro é o uso de outros modelos de redes neurais recorrentes, como os delineados na Subseção 2.1.9.

Outra possibilidade é a utilização do modelo Nelson-Siegel Dinâmico Rotacionado, descrito na Subseção 2.1.5.2, no lugar do modelo Nelson-Siegel clássico, descrito em 2.1.5.1. Essa mudança traz como vantagem uma maior interpretabilidade, do ponto de vista econômico, dos fatores do modelo descritor da curva de juros. Em conjunto com essa mudança, faria sentido também utilizar o modelo Black-Litterman para incorporar as previsões dos analistas para os fatores da curva de juros.

Também, como forma de aumentar ainda mais o escopo de alcance da pesquisa, um trabalho futuro é a implementação de um modelo genérico de fatores macroeconômicos e com poder explicativo do retorno de ativos de renda variável, como futuros, ETFs, dentre outros. Esse modelo extenderia a utilidade das previsões das variáveis macroeconômicas pelos modelos combinadas com as previsões pelos analistas para outros tipos de ativos financeiros que não apenas os títulos de renda fixa. A vantagem desse desenvolvimento, além da óbvia vantagem de um modelo adicional para renda variável, é poder utilizar o mesmo conjunto de previsões dos analistas para este modelo, preservando estas previsões como o ponto focal de discussão da equipe sobre o estado futuro da economia.

Do ponto de vista da avaliação, do ajuste dos hiperparâmetros e da seleção dos modelos, fica como trabalho futuro a aplicação do método de validação cruzada combinatória purgada, descrito na Subseção 2.1.6.1, no lugar do método de validação *walk-forward*, tradicional em finanças e utilizado neste trabalho. Por ser um método recente e com poucas publicações relacionadas, optou-se por não o utilizar neste trabalho. Todavia, o método é promissor e pode trazer ganhos, como o melhor aproveitamento dos dados no treinamento e maior probabilidade de se evitar o sobreajuste dos modelos.

Outro caminho de pesquisa interessante, do ponto de vista teórico e prático, é a aplicação das técnicas de preparação dos dados, seleção de variáveis e validação e dos modelos de redes neurais recorrentes avaliados na previsão dos retornos de classes de ativos. Os ativos financeiros tem séries de retornos de muito maior frequência que as séries de variáveis econômicas, que costumam ser mensais ou trimestrais. Esse volume de dados pode permitir a construção de redes neurais com mais camadas e mais neurônios e mais capaz de capturar padrões complexos de difícil identificação com o baixo volume de dados das variáveis macroeconômicas.

Por fim, também é interessante avaliar o desempenho de outros modelos de previsão de séries temporais e, mais especificamente, de outros modelos de redes neurais recorrentes, na previsão das variáveis macroeconômicas.

Glossário

Ativo financeiro Ativo líquido que deriva seu valor de um direito contratual ou reivindicação de propriedade. Seu valor reflete fatores de oferta e demanda nos mercados em que são negociados e o grau de risco que carregam.

Autoridade monetária Entidade que controla a quantidade de moeda em circulação em uma dada economia, frequentemente com o objetivo de controlar inflação e/ou a taxas de juros.

Bloomberg Companhia global provedora de informações e notícias financeiras, incluindo dados em tempo real e dados históricos.

Busca em grade Do inglês “*grid search*”. A busca em grade realiza uma busca exaustiva pela combinação de conjuntos de valores finitos para os hiperparâmetros de um modelo, retornando a combinação que resultou no melhor erro de validação.

Capitalização de mercado Valor de mercado total de um ativo, obtido pela multiplicação do preço de cada ativo pela quantidade de ativos em circulação.

Curva de juros Linha que mostra as taxas de juros, em um certo momento no tempo, de títulos com mesma qualidade de crédito mas diferentes datas de maturidade.

Desemprego Proporção de pessoas capazes buscando um trabalho que não conseguem emprego.

ETF Do inglês “*Exchange-Traded Fund*”. É um fundo de investimento, normalmente negociado em bolsa de valores, com portfólio de ativos que persegue replicar um determinado índice, como o S&P 500 Index.

Federal Funds Target Rate Taxa definida como meta pelo *Federal Open Market Committee*, órgão do *Federal Reserve System* (Fed) responsável pela política monetária americana, para a taxa de juros diária que bancos cobram para emprestar dinheiro a outros bancos. Para alcançar a taxa-meta, o Fed ajusta a oferta monetária.

Forex Do inglês “*Foreign exchange*”. É a negociação de uma moeda pela outra. Também usado para se referir ao mercado no qual as moedas são negociadas.

Futuros Contratos financeiros derivativos que obrigam as partes a transacionar um ativo a um preço e em uma data predeterminados. O comprador deve comprar ou o vendedor deve vender o ativo subjacente a um preço definido, independente do preço de mercado na data de vencimento.

Indicador de antecipação Do inglês “*leading indicator*”. Qualquer fator econômico que mude antes do resto da economia começar a ir numa direção específica. Indicadores de antecipação auxiliam observadores do mercado a prever mudanças significativas na economia.

Índice de Sharpe Métrica muito utilizada no mercado financeiro para compreender o retorno de um investimento ajustado ao seu risco. É a razão entre o retorno em excesso médio por unidade de volatilidade.

Inflação Aumento generalizado nos preços e queda no poder de compra da moeda.

LIBOR *London Interbank Offered Rate*. Taxa de juros de *benchmark* que indica os custos de empréstimos de curto prazo entre bancos no mercado interbancário internacional.

Lucros corporativos Indicador econômico, calculado trimestralmente pelo *Bureau of Economic Analysis* (BEA) americano, que sumariza o lucro líquido das corporações.

Modelo bayesiano Modelo que, em essência, faz inferência baseado no teorema de Bayes para obter uma distribuição *a posteriori* para uma quantidade ou quantidades de interesse, baseado em alguma distribuição *a priori*.

Modelo Black-Scholes Modelo de variação de preço, no tempo, de instrumentos financeiros tais como ações que pode, dentre outras coisas, ser utilizado para determinar o preço de opções.

Modelo autoregressivo Modelo que especifica que a variável de saída depende linearmente de seus próprios valores anteriores e de um termo estocástico.

MoM Mês a mês (do inglês “*Month over Month*”). Uma série mês a mês compara o resultado de cada mês com o mês anterior.

Otimização reversa Procedimento desenvolvido por William Sharpe e descrito em 1974 em seu artigo “Imputing Expected Security Returns from Portfolio Composition”.

Resumidamente, dados pesos de ativos de um portfólio, volatilidades, correlações e uma expectativa de retorno para o portfólio, o método infere os retornos esperados ótimos para os ativos que fazem o vetor de pesos sob o modelo CAPM e de acordo com certas suposições consideradas razoáveis pelo autor. Este procedimento é utilizado no modelo Black-Litterman clássico.

Portfólio Agrupamento de ativos financeiros tais como ações, títulos, *commodities* e moedas, dentre outros.

Portfólio de referência Um portfólio de ativos específicos a ser utilizado para comparar com o desempenho de um portfólio real.

Processo gaussiano Conjunto de variáveis aleatórias indexadas por tempo em que cada subconjunto finito destas variáveis tem distribuição normal multivariada.

Produção industrial Medida da saída real do setor industrial da economia.

Prêmio de risco Retorno acima do retorno livre de risco que é esperado de um dado investimento.

Retorno em excesso Retornos de investimento em um ativo ou portfólio que excedem a taxa de retorno de um ativo percebido como sendo livre de risco.

S&P 500 Índice das 500 maiores companhias americanas negociadas publicamente, com o peso proporcional à sua capitalização de mercado. É considerado o melhor indicador para ações americanas de elevada capitalização.

SA Ajustado Sazonalmente (do inglês “*Seasonally Adjusted*”).

SAAR Taxa Anual Ajustada Sazonalmente (do inglês “*Seasonally Adjusted Annual Rate*”).

Teste de Dickey-Fuller Aumentado é um teste de raiz unitária em séries temporais utilizado para rejeitar a hipótese de que uma série não é estacionária.

Ticker Abreviação utilizada para identificar unicamente uma ação, título, mercado ou índice específico.

Validação *out-of-sample* Validação de modelo que utiliza dados da amostra, reservados e não utilizados para o treinamento, para fazer previsões, de maneira a avaliar o modelo e determinar se as estatísticas dos seus erros são similares a de outro(s) modelo(s) avaliados com os mesmos dados.

Valor- p Probabilidade de, quando a hipótese nula é verdadeira, se obter uma estatística de teste igual ou mais extrema que aquela observada em uma amostra.

Variável macroeconômica Variável que representa um fenômeno que afeta a economia como um todo e todos os membros de uma sociedade.

Referências

- [1] Mishra, Alok, Subramanyam Pisipati e Iti Vyas: *An equilibrium approach for tactical asset allocation: Assessing black-litterman model to indian stock market*. Journal of Economics and International Finance, Vol. 3:553–563, setembro 2011. x, 11
- [2] Olah, Christopher: *Understanding lstms*. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>, acesso em 2015-08-27. x, 24
- [3] Goodfellow, Ian, Yoshua Bengio e Aaron Courville: *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. x, 23, 25, 26
- [4] Chen, Minmin: *Minimalrnn: Toward more interpretable and trainable recurrent neural networks*, 2017. x, 27, 50
- [5] INTERNATIONAL MONETARY FUND: *Guidelines for Foreign Exchange Reserve Management*, 2001. <http://www.imf.org/external/np/mae/ferm/eng/>, Atualizado em: 30 de mai. 2018. 1
- [6] BRASIL: *Lei nº 4.595*, 31 de dezembro de 1964. Dispõe sobre a Política e as Instituições Monetárias, Bancárias e Creditícias, cria o Conselho Monetário Nacional e dá outras providências. 2 Diário Oficial da União, p. 12081, 31 de dez. 1964. Suplemento. 1
- [7] BANCO CENTRAL DO BRASIL: *Regimento Interno do Banco Central do Brasil*. <http://www.bcb.gov.br/Adm/RegimentoInterno/RegimentoInterno.pdf>, Atualizado em: 24 de abr. 2019. 1
- [8] Fabozzi, Frank J. (editor): *The handbook of fixed income securities*. Wiley, 2a ed edição, 2007, ISBN 978-0-47-005221-1. 1, 2, 6
- [9] BANCO CENTRAL DO BRASIL: *Relatório de Gestão das Reservas Internacionais*, Março de 2019. <http://www.bcb.gov.br/?id=GESTAORESERVAS&ano=2016>, Acesso em: 24 abr. 2019. 2
- [10] Fabozzi, Frank J. e Steven V. Mann (editores): *The handbook of fixed income securities*. McGraw-Hill, New York, 8a ed edição, 2012, ISBN 978-0-07-176846-7. 2
- [11] Diebold, Francis X. e Glenn D. Rudebusch: *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. The Econometric and Tinbergen Institutes Lectures. Princeton University Press, 1a ed edição, 2013, ISBN 978-0691146805. 3, 14, 15, 36

- [12] Black, Fischer e Robert B Litterman: *Asset Allocation: Combining Investor Views with Market Equilibrium*. The Journal of Fixed Income, 1(2):7–18, setembro 1991, ISSN 1059-8596, 2168-8648. <http://jfi.iiijournals.com/lookup/doi/10.3905/jfi.1991.408013>, acesso em 2018-06-16TZ. 4
- [13] Kolm, Petter e Gordon Ritter: *On the Bayesian interpretation of Black–Litterman*. European Journal of Operational Research, 258(2):564–572, abril 2017, ISSN 03772217. <http://linkinghub.elsevier.com/retrieve/pii/S037722171630861X>, acesso em 2018-06-16TZ. 4, 11
- [14] Avramov, Doron e Guofu Zhou: *Bayesian Portfolio Analysis*. Annual Review of Financial Economics, 2(1):25–47, dezembro 2010, ISSN 1941-1367, 1941-1375. <http://www.annualreviews.org/doi/10.1146/annurev-financial-120209-133947>, acesso em 2018-06-16. 6
- [15] Brandt, Michael W: *Portfolio Choice Problems*. página 68. 6
- [16] Meucci, Attilio: *Risk and asset allocation*. Springer finance. Springer, Berlin ; New York, 2009, ISBN 978-3-642-00964-8 978-3-540-22213-2. OCLC: ocn752143921. 6
- [17] Markowitz, Harry: *Portfolio Selection*. The Journal of Finance, 7(1):77, março 1952, ISSN 00221082. <https://www.jstor.org/stable/2975974?origin=crossref>, acesso em 2018-06-16TZ. 6, 7
- [18] Markowitz, Harry: *Mean-variance analysis in portfolio choice and capital markets*. Frank J. Fabozzi Assoc, New Hope, Pa, rev. reissue edição, 2000, ISBN 978-1-883249-75-5. OCLC: 248342992. 7
- [19] Koedijk, Kees G., Alfred M.H. Slager e Philip A. Stork: *Investing in Systematic Factor Premiums: Investing in Systematic Factor Premiums*. European Financial Management, 22(2):193–234, março 2016, ISSN 13547798. <http://doi.wiley.com/10.1111/eufm.12081>, acesso em 2018-06-17TZ. 8
- [20] Ang, Andrew: *Asset management: a systematic approach to factor investing*. Financial Management Association survey and synthesis series. Oxford University Press, Oxford, 2014, ISBN 978-0-19-995932-7. 8
- [21] Cochrane, John H.: *Asset pricing*. Princeton University Press, 2001, ISBN 978-0-69-107498-6. 9, 10
- [22] Sharpe, William F.: *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*. The Journal of Finance, 19(3):425, setembro 1964, ISSN 00221082. <https://www.jstor.org/stable/2977928?origin=crossref>, acesso em 2018-06-16TZ. 9
- [23] Lintner, John: *The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets*. The Review of Economics and Statistics, 47(1):13, fevereiro 1965, ISSN 00346535. <https://www.jstor.org/stable/1924119?origin=crossref>, acesso em 2018-06-16TZ. 9

- [24] Fama, Eugene F e Kenneth R French: *The Capital Asset Pricing Model: Theory and Evidence*. página 37. 9
- [25] Black, Fischer e Robert Litterman: *Global Portfolio Optimization*. Financial Analysts Journal, 48(5):28–43, setembro 1992, ISSN 0015-198X. <http://www.cfapubs.org/doi/full/10.2469/faj.v48.n5.28>, acesso em 2018-06-16TZ. 10, 12
- [26] *Encyclopædia Britannica*. Britannica Academic, 2019. Yield curve. 13, 14
- [27] Nelson, Charles R. e Andrew F. Siegel: *Parsimonious modeling of yield curves*. The Journal of Business, 60(4):473–489, 1987, ISSN 00219398, 15375374. <http://www.jstor.org/stable/2352957>. 14
- [28] Diebold, Francis X. e Canlin Li: *Forecasting the term structure of government bond yields*. Journal of Econometrics, 130(2):337–364, fevereiro 2006. <https://doi.org/10.1016/j.jeconom.2005.03.005>. 15, 17
- [29] Nyholm, Ken: *A rotated dynamic nelson-siegel model with macro-financial applications*. Working paper 1851, European Central Bank, setembro 2015. 15, 16, 17, 30
- [30] De Prado, Marcos L.: *Advances in Financial Machine Learning*. Wiley, 1a ed edição, 2018, ISBN 978-1119482086. 17, 18, 19, 20, 36
- [31] *Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance*. Notices of the American Mathematical Society, 61(5), 2014. 18
- [32] Luo, Y., M. Alvarez, S. Wang e G. Rohal: *Seven sins of quantitative investing*. White paper, Deutsche Bank Markets Research, setembro 2014. 18
- [33] Bishop, Christopher M.: *Pattern recognition and machine learning*. Information science and statistics. Springer, New York, 2006, ISBN 978-0-387-31073-2. 19, 20, 21, 22
- [34] Hastie, Trevor, Robert Tibshirani e Jerome Friedman: *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, 2a ed edição, 2009. 19, 29
- [35] Jolliffe, I. T.: *Principal Component Analysis*. Spring, 2a ed edição, 2002. 20
- [36] 20, 21
- [37] Igel'nik, Boris e Jacek M. Zurada (editores): *Efficiency and Scalability Methods for Computational Intellect*. IGI Global, 2013. <https://doi.org/10.4018/978-1-4666-3942-3>. 22
- [38] Lipton, Zachary C., John Berkowitz e Charles Elkan: *A Critical Review of Recurrent Neural Networks for Sequence Learning*. arXiv:1506.00019 [cs], maio 2015. <http://arxiv.org/abs/1506.00019>, acesso em 2018-06-09TZ, arXiv: 1506.00019. 22, 23, 24, 25

- [39] Tkáč, Michal e Robert Verner: *Artificial neural networks in business: Two decades of research*. Applied Soft Computing, 38:788–804, janeiro 2016, ISSN 15684946. <http://linkinghub.elsevier.com/retrieve/pii/S1568494615006122>, acesso em 2018-06-16TZ. 23
- [40] Hochreiter, Sepp e Jürgen Schmidhuber: *Bridging long time lags by weight guessing and "long short term memory"*. Em *SPATIOTEMPORAL MODELS IN BIOLOGICAL AND ARTIFICIAL SYSTEMS*, páginas 65–72. IOS Press, 1996. 26
- [41] Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk e Yoshua Bengio: *Learning phrase representations using rnn encoder-decoder for statistical machine translation*, 2014. 26
- [42] Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho e Yoshua Bengio: *Empirical evaluation of gated recurrent neural networks on sequence modeling*, 2014. 26
- [43] Lei, Tao, Yu Zhang, Sida I. Wang, Hui Dai e Yoav Artzi: *Simple recurrent units for highly parallelizable recurrence*. Em *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, páginas 4470–4481, Brussels, Belgium, 10-11 2018. Association for Computational Linguistics. <https://www.aclweb.org/anthology/D18-1477>. 28
- [44] Lei, Tao, Yu Zhang, Sida I. Wang, Hui Dai e Yoav Artzi: *Simple recurrent units for highly parallelizable recurrence*, 2017. 28
- [45] Li, Shuai, Wanqing Li, Chris Cook, Ce Zhu e Yanbo Gao: *Independently recurrent neural network (indrnn): Building a longer and deeper rnn*. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 28, 29
- [46] Li, Shuai, Wanqing Li, Chris Cook, Ce Zhu e Yanbo Gao: *Independently recurrent neural network (indrnn): Building a longer and deeper rnn*, 2018. 28
- [47] Nakamura, Emi: *Inflation forecasting using a neural network*. Economics Letters, 86(3):373–378, março 2005. <https://doi.org/10.1016/j.econlet.2004.09.003>. 30
- [48] Sambasivan, Rajiv e Sourish Das: *A statistical machine learning approach to yield curve forecasting*, 2017. 30
- [49] Caldeira, João F., Guilherme V. Moura e André A.P. Santos: *Predicting the yield curve using forecast combinations*. Computational Statistics & Data Analysis, 100:79–98, agosto 2016. <https://doi.org/10.1016/j.csda.2014.05.008>. 30
- [50] Shearer, C: *The crisp-dm model: the new blueprint for data mining*. J Data Warehouse, 5:13–22, janeiro 2000. 31
- [51] Alexander, C.: *Market Models*. John Wiley Sons, 1a ed edição. 36
- [52] Kursá, Miron B. e Witold R. Rudnicki: *Feature selection with the Boruta Package*. Journal of Statistical Software, 36(11), 2010. <https://doi.org/10.18637/jss.v036.i11>. 37

- [53] James, Gareth, Daniela Witten, Trevor Hastie e Robert Tibshirani: *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014, ISBN 1461471370, 9781461471370. 38